



Faculté des Sciences

Louvain Institute of Biomolecular Science and Technology

Engineering *N*-glycosylation profile of pharmaceutical glycoproteins produced in *Nicotiana tabacum* BY-2 cells

Thèse présentée par Xavier Herman en vue de l'obtention du grade de
docteur en Sciences

Promoteur.rice.s :

Prof. François CHAUMONT

Dr. Catherine NAVARRE

Président du jury :

Prof. Bernard HALLET

Membres du jury :

Mme Marianne DEWERCHIN

Prof. Muriel BARDOR

Prof. Richard STRASSER

Prof. Henri BATOKO

Janvier 2023

Acknowledgements

Cette thèse est le fruit de cinq années de travail, au cours desquelles j'ai eu la chance de rencontrer, collaborer, être soutenu par, partager des moments forts professionnels et extra-professionnels avec de nombreuses personnes. Je tiens à tou.te.s les remercier, cette thèse est aussi la leur.

Catherine et François, merci de m'avoir fait confiance et accueilli dans le laboratoire de Molecular Farming. Merci Catherine du temps que tu as toujours su me consacrer, ainsi qu'au reste de ton équipe qui ne cesse de s'agrandir, le signe indéniable d'une dynamique d'entraide, de fun et d'excellence. Ton savoir scientifique et technique mais bien sûr aussi ton amitié au quotidien vont me manquer. Merci François pour la supervision bienveillante, les échanges francs et constructifs ainsi que les barbecues chez toi et les verres afterwork en terrasse (et désolé de t'avoir battu aux 24h vélo).

Ma gratitude va également à Mme Marianne Dewerchin, Prof. Bernard Hallet, Prof. Muriel Bardor, Prof. Henri Batoko and Prof. Richard Strasser, membres du comité d'accompagnement et/ou du jury de thèse. Merci du temps que vous m'avez octroyé, ainsi que pour la discussion que nous avons eue lors de la défense privée.

Merci aux deux formidables mémorant.e.s, Marie et Laurent, que j'ai eu la chance d'encadrer puis d'accueillir dans l'équipe et qui ont contribué directement à cette thèse. Marie, ta motivation, ton organisation et ta rigueur scientifique m'ont impressionné dès ton mémoire et au cours de ta thèse, dont je suis impatient d'entendre les conclusions dans quelques années. Laurent, ton dévouement, ton enthousiasme et ton imagination sont des atouts qui te mèneront où tu le veux dans le futur (un post-doc sur les crossing-over, peut-être ?). Promis, tu peux me claquer la bise maintenant, on se voit à notre prochaine séance d'UVs.

Merci Adeline pour ton aide précieuse et les bons moments passés ensemble ; merci Jo pour tes coups de main, ta disponibilité, ton savoir et ton grand cœur ; merci à Marc pour les encouragements, les conseils scientifiques et la relecture du manuscrit ; merci à Marie-Eve pour le soutien technique et la bonne humeur permanente.

Merci aussi à tou.te.s les autres membres du labo, passé.e.s et actuel.le.s, pour l'incroyable ambiance qui a régné et règne grâce à vous: Sébastien, Nicolas, Mathilde, Valentin, Camille, Rik, Marie G., Antoine, Christophe, Romane.

Merci à mes voisins de bureau Alice, Louis, Nicolas G., Max V. et les autres pré-cités, partager ces 8 m² avec vous a été un honneur !

Qu'aurais-je fait de mes midis et de nombreuses soirées sans la Team Grailée, Pain et salade, Zanzi, karaoké, cercle et autres ? Estelle, Maxime L., Quentin, Adrian, Manon et tou.te.s les autres que j'ai déjà cité.e.s, promis je repasserai à midi mais aussi et surtout à minuit. Laurie, Seb, Tim et Jahed, merci pour l'accueil en FYMO et les instants de retrouvailles mémorables encore bien après votre départ. Merci Cyril pour ton hospitalité aux soirées raclettes, barbecue et foot, quand tu veux pour la prochaine ! Vous avez mis la barre haute, mes futur.e.s collègues ont du pain sur la planche pour vous arriver à la cheville.

Merci à tou.te.s les membres de FYMO, désolé je ne peux pas citer chacun.e.s d'entre vous sinon je risque d'en oublier un.e et ça ferait tache. Ce groupe est superbe et j'en garde des souvenirs fantastiques. J'étais venu pour faire de la Science mais au final, ma plus belle découverte c'est vous.

Merci à tou.te.s les LIBSTers rencontré.e.s dans les couloirs, aux 24h vélo, vélectro, cheese and wine, quizz, barbecues, welcome drink, halloween drink, baraki night et autres festivités.

Merci aux collaborateurs de l'ULiège et particulièrement à Johann pour le travail accompli malgré les difficultés. Merci aussi aux partenaires de GSK pour les discussions et le suivi du projet.

Merci à la Mowhaker d'être une team d'amis unique et merveilleuse, vous êtes les rois de lantwuf.

Merci à mes parents et à ma sœur qui se sont toujours montré.e.s intéressé.e.s malgré la complexité et le temps que je pouvais passer à parler des glycanes.

Enfin, merci à toi, Isa, d'avoir été à mes côtés tout au long de cette thèse. Ta patience et ton enthousiasme sont pour moi une source d'inspiration permanente. Merci de me supporter et soutenir depuis 9 ans, j'ai hâte de voir ce que l'avenir nous réserve.

Table of contents

Table of contents	v
Summary.....	ix
Abbreviations	xiii
Introduction.....	1
0. Contextualization	3
1. Heterologous expression systems for glycoprotein production	4
2. <i>N</i> -glycosylation	11
3. Glyco-engineering in plants.....	17
4. CRISPR/Cas9 genome editing tool	38
5. Generation of marker-free transgenic plants.....	44
6. Human cytomegalovirus	58
Objectives	65
Results and Discussion	69
1. Inactivation of <i>N</i> -acetylglucosaminyltransferase I and α 1,3-fucosyltransferase genes in <i>N. tabacum</i> BY-2 cells results in glycoproteins with highly homogeneous, high-mannose <i>N</i> -glycans	71
2. <i>In vivo</i> deglycosylation of recombinant glycoproteins in <i>Nicotiana tabacum</i> BY-2 cells.....	90
3. Excision of the transgenes.....	109
General Discussion and Perspectives.....	125
Supplementary Figures and Tables	133
Materials and Methods.....	163
Annex.....	179
Inactivation of β 1,2-xylosyltransferase and GDP-mannose 4,6-dehydratase genes to obtain a BY-2 cell line producing <i>N</i> -glycans without β 1,2-xylose and depleted for α 1,3-fucose and α 1,4-fucose	181
References	185

Summary

Nicotiana tabacum BY-2 cells are among the most used plant cell lines for producing biopharmaceutical glycoproteins. However, *N*-glycosylation, which increases heterogeneity and raises safety concerns due to the presence of non-human residues, is poorly controlled. *N*-glycan composition is a concern for therapeutic or vaccine glycoproteins because it can alter their activity and lead to batch-to-batch variability. To tackle these issues, we are developing a set of glyco-engineered BY-2 cell lines through knock out and ectopic expression of genes involved in the *N*-glycosylation pathway. Generation of glyco-engineered cell lines is a powerful tool to produce recombinant therapeutics with fine-tuned biological activities but also to broaden our understanding of the impact of *N*-glycans on their properties.

We first generated a cell line producing glycoproteins devoid of complex *N*-glycans. It was obtained using CRISPR/Cas9-mediated edition of *N*-acetylglucosaminyltransferase I (*GnTI*) genes, whose activity is a prerequisite for the formation of all complex *N*-glycans, as well as *α 1,3-fucosyltransferase* genes. Absence of complex *N*-glycans on secreted glycoproteins was confirmed by Western blotting and mass spectrometry.

Next, we generated cell lines producing *in vivo* deglycosylated proteins. It was obtained by expressing the fungal endoglycosidase T (EndoT) in *GnTI/FucT* knockout cell lines. EndoT cleaves high mannose *N*-glycans to generate single, asparagine-linked, *N*-acetylglucosamine residues.

Finally, an immunoglobulin G and the glycoprotein B of human cytomegalovirus were expressed in the *GnTI* knockout background expressing or not EndoT. Glycoproteomic mass spectrometry analysis of the secreted antibody and *in vitro* enzymatic mobility shift assay of glycoprotein B revealed that each cell type generates glycovariants with the anticipated *N*-glycan structures.

Abbreviations

Ac/Ds	Activator/Dissociation
ALG3	Asparagine-linked glycosylation 3
Asn	Asparagine
BY-2	Bright Yellow 2
Cas	CRISPR-associated protein
CaMV	Cauliflower mosaic virus
Cgl1	Complex glycan-less 1
CHO	Chinese hamster ovary
CM	Culture medium
CMAS	CMP-Neu5Ac synthetase
CMP	Cytidine monophosphate
CodA	Cytosine deaminase A
Cre	Cyclization recombination
CRISPR	Clustered regularly interspaced short palindromic repeats
CST	CMP-sialic acid transferase
CTS	Cytoplasmic tail, transmembrane domain and stem
dCas9	Dead Cas9
Dol-PP	Dolichol pyrophosphate
DSB	Double strand break
dsDNA	Double stranded DNA
Endo H/T	Endo- β -N-acetylglucosaminidase H/T
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum-associated protein degradation
eVLP	Enveloped virus-like particle
Fc	Fragment crystallizable
Fuc	Fucose
FucT	Fucosyltransferase
Gal	Galactose
GalT	Galactosyltransferase
gB	Glycoprotein B
GCS	glucosidase
Glc	Glucose
GlcNAc	N-acetylglucosamine
GDP	Guanosine diphosphate
GM	Golgi α -mannosidase
GMD	GDP-d-mannose 4,6-dehydratase
GNE	UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase
GnTI/II/III/IV/V	N-acetylglucosaminyltransferase I/II/III/IV/V
GOI	Gene of interest

GSK	GlaxoSmithKline
GT	Glycosyltransferase
HCMV	Human cytomegalovirus
HDR	Homology directed repair
HEK	Human embryonic kidney
Hexo	β - <i>N</i> -acetylhexosaminidases
Hgl1	Hybrid glycosylation 1
HR	Homologous recombination
Ig	Immunoglobulin
Indel	Insertion/deletion
LB	Left Border
Le ^a	Lewis A
<i>LoxP</i>	Locus of X-over of P1
MolSys-ULiege	Laboratory of mass spectrometry, University of Liege
mAb	Monoclonal antibody
Man	Mannose
ManNAc	<i>N</i> -acetylmannosamine
Mgat1	α -1,3-mannosyl-glycoprotein 2-beta- <i>N</i> -acetylglucosaminyltransferase
MNS	α -mannosidase
MS	Mass spectrometry/Murashige and Skoog
NANP	Neu5Ac-9-phosphatase
NANS	CMP-Neu5Ac synthetase
nCas9	Nickase Cas9
Neu5Ac	<i>N</i> -acetylneuraminic acid or sialic acid
NHEJ	Non homologous end joining
PAM	Protospacer adjacent motif
PEG	Polyethylene glycol
PTM	Post-translational modification
PNGase	Peptide- <i>N</i> -glycosidase
Pre-crRNA	Pre-CRISPR RNA
PTG	Polycistronic tRNA-gRNA
RB	Right border
RBD	Receptor binding domain
RNAi	RNA interference
RNP	Ribonucleoprotein
sgRNA	Single guide RNA
ST	α 2,6-sialyltransferase
TALEN	Tale effector-like nucleases
T-DNA	Transfer DNA

TracrRNA	Trans-activating crRNAs
tRNA	Transfer RNA
TSCP	Total soluble cellular proteins
UDP	Uridine diphosphate
UGGT	UDP-glucose:glycoprotein glucosyltransferase
WT	Wild-type
Xyl	β 1,2-xylose
XylT	β 1,2-xylosyltransferase
ZFN	Zinc finger nucleases
5-FC	5-fluorocytosine
5-FU	5-fluorouracile

Introduction

0. Contextualization

Pharmaceutical recombinant proteins are a powerful, cutting-edge, tool to treat or prevent many diseases and infections. As a result, the global market demand for such proteins is continuously growing. However, despite a steady decrease in costs, most recombinant proteins are still extremely expensive. These prohibiting costs, as well as the variety of recombinant protein candidates, justify the need for the development of various heterologous production systems. Mammalian cells, yeast and bacteria are currently the leading production platforms but other systems such as insect cells, plants or plant cells are emerging as valuable alternatives. Our laboratory aims at further developing a tobacco (*Nicotiana tabacum*) suspension cell platform for the production of therapeutic or vaccine glycoproteins. One of the focus is the improvement of the efficiency of these products by modulating their post-translational modifications (PTMs). In this introduction, the main recombinant protein expression systems will be briefly described, with an emphasis on plant systems. Next, the plant *N*-glycosylation pathway will be thoroughly described, and an exhaustive list of the *N*-glyco-engineering attempts will be presented. The most recent studies, including this thesis, rely on the use of the CRISPR/Cas9 system to inactivate genes. Therefore, this technology will be described. Once the gene inactivation is effective, the CRISPR/Cas9 transgenes are not desirable anymore. The main techniques to generate a transgene-free genome of edited plants and plant cells will be discussed in the penultimate chapter. Finally, the glycoprotein B (gB) of human cytomegalovirus (HCMV), which is used as a model viral envelope recombinant protein in this thesis, will be quickly described.

1. Heterologous expression systems for glycoprotein production

1.1. Main expression systems

The production systems of conventional pharmaceuticals use bacteria, yeast, insect, or mammalian cell cultures. Each platform has its intrinsic advantages and drawbacks. The most suitable expression system to produce a given pharmaceutical must be determined regarding different features such as protein folding machinery, presence of relevant PTMs, yields and production costs. Other criteria such as the available genetic tools, patent situation and regulatory aspects, which can vary according to the production host, must also be considered.

Among prokaryotic expression systems, *Escherichia coli* is an attractive platform because of its short generation time, the ease of handling, the scientific knowledge acquired and its capacity to accumulate high concentrations of foreign proteins. However, these proteins sometimes aggregate in inclusion bodies rather than being secreted into the extracellular medium, which is preferred because it allows simpler product purification. As a result, *E. coli* secretion systems reaching high product titers of several grams per liter have been developed (Schmidt, 2004). Despite this, prokaryotic expression systems are not suitable for the production of many pharmaceuticals due to their lack of some eukaryotic features. Eukaryotic protein folding systems can be mandatory to obtain particular proteins in the proper conformation. Furthermore, prokaryotic PTMs such as glycosylation are limited and highly different from eukaryotes. This is an important issue because more than half of human proteins are glycosylated, which can affect both protein folding and activity (Brooks, 2009). As a result, prokaryotic expression systems are mainly used to produce proteins that are not glycosylated or pharmacologically active without glycosylation. To produce complex, glycosylated proteins, the use of eukaryotic expression systems is often mandatory.

Similar to *E. coli*, yeast expression systems are well characterized and display a short generation time allowing to quickly reach high cell density. *Saccharomyces cerevisiae* and *Pichia pastoris* are the most frequently used species for heterologous protein expression. *P. pastoris* secretion system is highly efficient, reaching more than 10 g/l for recombinant proteins such as albumin, collagen, tetanus toxin, hydroxynitrile lyase, manganese peroxidase, lysostaphin and α -carbonic anhydrase (Schmidt, 2004; Pan *et al.*, 2022). Furthermore, PTMs in yeast show similarities with those of higher eukaryotes. Yeast and higher eukaryotes share the same N-glycan precursor. However, the processing of this precursor in the Golgi apparatus leads to the

formation of different structures in yeast and higher eukaryotes. This is a concern for glycoproteins requiring specific *N*-glycans for their pharmacological activity as well as for the safety of the final product.

Baculovirus-mediated expression in insect cells is attractive because it is fast and offers high levels of protein expression with PTMs approaching those of mammalian cells despite the presence of some non-human residues. However, the infection with actively replicating viruses results in the presence of large numbers of contaminants along with cellular proteins and debris from lysed cells in the culture medium. Costly purification steps are then necessary to meet pharmaceutical-grade purity (Yee *et al.*, 2018). Also, insect cells are susceptible to contamination by some human pathogens.

The use of mammalian cell cultures to produce recombinant proteins confers characteristics closer to endogenous human proteins. However, their cultivation requires a more complex and expensive growth medium and generally gives lower yields. Mammalian cell lines are the most adapted to achieve complex folding and human-like PTMs. Chinese hamster ovary (CHO) cells and human embryonic kidney (HEK) cells are the most commonly used. These cell lines can produce high titers (1-5 g/L) of complex glycoproteins (Santos *et al.*, 2016). However, the development of a new mammalian cell line to produce a glycoprotein is time-consuming and requires heavy investments. Production costs are also higher because of the complexity of the growth medium and the risk for contamination with human pathogens. Indeed, most mammalian cell culture media require mammalian-derived components that are capable of transmitting mammalian viruses or prions. Moreover, the complexity of culture media leads to greater variability between batches of media. Changes in culture conditions can strongly affect *N*-glycosylation patterns, which in turn can impact glycoprotein activity (Tekoah *et al.*, 2015). As a result, high costs are associated with upstream and downstream quality control processes ensuring health security and batch-to-batch consistency.

1.2. Molecular farming

The expression of pharmaceutical proteins in plants or plant cells is called molecular farming. It has emerged as an alternative to animal cells to produce complex recombinant glycoproteins. Rice, tobacco, maize, soybean, potato, carrot and other plant species have been engineered to stably or transiently produce recombinant glycoproteins such as antibodies and vaccines, either as a whole plant or as a tissue/cell culture (Zhu *et al.*, 2022). Plant systems are attractive because they require

simple and inexpensive growth media, provide low-cost scalability and cannot be contaminated by human or zoonotic pathogens (Moustafa *et al.*, 2016, Zhu *et al.*, 2022). Furthermore, they allow efficient protein folding and PTMs close to those in humans, while being much simpler. As a result, they have more homogenous glycan moieties, with significantly fewer glycoforms (glycoproteins having different glycans attached) and less batch-to-batch variability. The main disadvantages are the low production levels and the high rate of proteolysis.

Transient expression is obtained by the introduction and expression of transgenes without integration in the genome. *Nicotiana benthamiana* plants are the dominant plant species for transient expression of glycoproteins. The main technique used is leaf infiltration with the bacterium *Agrobacterium tumefaciens*, or agroinfiltration. Small-scale agroinfiltration relies on the use of an *Agrobacterium*-filled syringe to inject the bacteria into the plant leaves through their stomata. Large-scale agroinfiltration relies on vacuum infiltration (Figure 1). Briefly, the whole plant is submerged into an *Agrobacterium*-containing medium and then placed in a vacuum chamber. When vacuum is applied, the air is drawn out of the interstitial spaces. As the vacuum is released, the interstitial spaces are filled by the *Agrobacterium* suspension (Simmons *et al.*, 2009). This leads to the transfection of a high number of cells and quick production of recombinant glycoproteins. Plant viruses can also be used for transient expression (Santos *et al.*, 2016).

Transient expression is characterized by a quick development cycle but requires more handling and hence are less scalable than stable transgenic plants. The ability to rapidly produce a new recombinant protein is of great value during new disease outbreaks. Recently, transient expression in *N. benthamiana* has been used to produce prophylactic antibodies against Ebola virus (Santos *et al.*, 2016) as well as a virus-like particle-based vaccine against Influenza A virus and SARS-CoV-2, the latter being authorized by Health Canada (Medicago Inc., Quebec, Canada). However, transient expression in *N. benthamiana* leaves faces different issues: approval by the regulatory authorities, the presence of endotoxins resulting from the large number of bacteria introduced and the reluctance of an industry specialized in bioreactor-based production (Hundleby *et al.*, 2022).



Figure 1. *N. benthamiana* vacuum infiltration unit (Fraunhofer USA). When the *N. benthamiana* plants are about 12 cm tall, the rack is placed upside down into a container containing an *Agrobacterium* culture. A vacuum is applied and all the air in the plant is pulled out. When the vacuum is turned off, the interstitial spaces are filled with *Agrobacterium*, which will transform the cells.

Stable expression in plants relies on the genomic integration of a transgene, usually via *A. tumefaciens* or *Agrobacterium rhizogenes*-mediated transformation or biolistics. *Agrobacterium*-mediated transformation is the dominant technology. Disarmed *Agrobacterium* strains have been engineered to remove the oncogenes from the Transfer-DNA (T-DNA). These strains are used to transform plant cells by transferring a T-DNA containing the desired transgenes. This T-DNA is placed on a binary plasmid (Hwang *et al.*, 2017). It is transferred to the plant genome thanks to the *vir* genes expressed by the bacterium in trans. Biolistics, on the other hand, relies on the coating of gold microparticles with the DNA to be transferred and their shooting onto plant tissues or cells to integrate the genome. When plant tissues are used, transformed cells develop into calli which give rise to cell lines, or which can be regenerated into transgenic plants. Transformed suspension cells also lead to calli, which can be grown as cell lines. Moreover, stable expression can be ubiquitous or targeted to specific organs such as seeds or trichomes, according to the transcription promoters used (Piron *et al.*, 2015; Pottier *et al.*, 2020). While transient expression is the most rapid expression system, the use of stably transformed plants or plant cell lines decreases the production and quality control costs by suppressing the agroinfiltration step. Also, stable transformation techniques have been implemented

in many plant species, while highly efficient transient transformation is mainly restricted to *N. benthamiana*.

1.3. The *Nicotiana tabacum* BY-2 suspension cell platform

Plant suspension cells were first used in basic research but are gaining importance in molecular farming because they combine the advantages of whole plants with those of microorganisms. Plant suspension cells can be cultured in bioreactor systems similar to those used for mammalian cell cultures and fall under the same regulations. Moreover, recombinant glycoproteins can be secreted and purified directly from the extracellular medium. While human pathogens pose a huge threat for animal-based cell lines, plant suspension cells are not susceptible. Furthermore, plant viruses cannot propagate well in plant cell cultures as they use plasmodesmata to spread infection cell-to-cell. These factors are not only important for safety, but also significantly reduce operational costs (Tekoah *et al.*, 2015).

Suspension cell lines have been obtained from several plant species. Numerous pharmaceutical proteins have been expressed in cell lines from different plant species, but very few reached the level of industrial production or even clinical trials. Currently, the main cell lines for molecular farming are derived from tobacco (*N. tabacum*), rice (*Oryza sativa*), carrot (*Daucus carota*) and alfalfa or lucerne (*Medicago sativa*) (Santos *et al.*, 2016).

The first licensed biotherapeutic obtained by molecular farming is taliglucerase alfa (commercialized as Elelyso). It is a recombinant human glucocerebrosidase, an enzyme whose deficiency causes Gaucher disease. It is produced by Protalix Biotherapeutics in carrot suspension cells. Plant-based production of this glycoprotein gained attention after a shortage due to virus contamination of CHO bioreactors producing glucocerebrosidase (Tekoah *et al.*, 2016).

Other Protalix therapeutic proteins in phase II and III clinical trials are produced in *N. tabacum* Bright Yellow 2 (BY-2) cells (Figure 2). It is the most widely used plant cell line. It was obtained in 1968 from an induced callus of a tobacco cultivar BY-2 seedling in the Central Research Institute of the Japan Tobacco and Salt Public Corporation (now the Tobacco Science Research Laboratory, Japan Tobacco, Inc.) (Nagata *et al.*, 1992). *N. tabacum* is a natural allotetraploid derived by interspecific hybridization and chromosomal doubling from ancestral diploid progenitors related to modern *Nicotiana sylvestris* and *Nicotiana tomentosiformis* (Kovarík *et al.*, 2012; Sierro

et al., 2013). As a result, the number of isoforms of most *N. tabacum* genes is doubled compared to its ancestors.

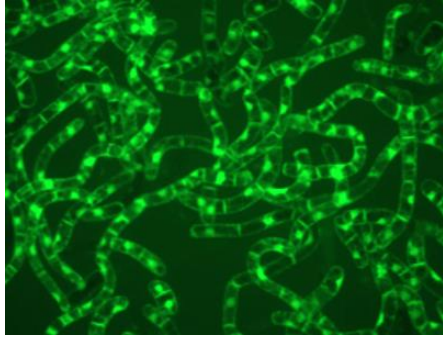


Figure 2. *N. tabacum* BY-2 suspension cells treated with fluorescein diacetate to stain living cells. BY-2 cells form small chains in liquid culture. Scale bar = 100 μ m

The initial project involving BY-2 cells aimed at using them as a raw material for cigarettes. The project did not succeed but showed that the long-term continuous culture of BY-2 cells on an industrial level is possible. Later on, Nagata *et al.*, established the synchrony of this cell line using aphidicolin and a peak mitotic index of 70-80%, which is the ratio between the number of cells in mitosis and the total number of cells, was measured. At that time, other cell lines reached 10-20% mitotic index at most. This feature highlighted BY-2 cells as an important tool to study gene expression during the cell cycle of higher plants (Nagata *et al.*, 1992).

Over time, the intrinsic features of BY-2 cells established them as a powerful tool not only for basic research, but also for recombinant protein production. Its multiplication rate, reaching up to 100-fold per week of subculture, is a valuable asset (Srba *et al.*, 2016, Santos *et al.*, 2016). Unlike algae and mosses, BY-2 cells do not require light in the presence of a carbon source. Indeed, they lack chloroplasts but possess undifferentiated heterotrophic plastids. This is an advantage because it is difficult to maintain a homogenous illumination in a bioreactor and there may be a risk of overheating (Moustafa *et al.*, 2015; Takahashi *et al.*, 2015). BY-2 cells also have a very efficient secretory pathway, allowing purifying recombinant proteins directly from the culture medium (Mercx *et al.*, 2016, Yao *et al.*, 2015). They are also easily amenable for *N*-glycosylation engineering (detailed in chapter 3). Finally, *Agrobacterium*-mediated transformation of BY-2 cells is highly efficient and different glycoproteins have been successfully produced using this system, including human erythropoietin, interleukins, several immunoglobulins (Ig), the house dust mite Derp1 allergen and human DNase I (Santos *et al.*, 2016; Hanania *et al.*, 2017). In our

laboratory, IgGs (De Muynck *et al.*, 2009; Magy *et al.*, 2014; Mercx *et al.*, 2017; Navarre *et al.*, 2017), human DNase I (unpublished data) and vaccine proteins against Lyme disease (Navarre *et al.*, 2006), and HCMV (Smargiasso *et al.*, 2019) have been produced in BY-2 cells.

This heterologous expression platform can be further improved. First, the legal framework associated with this production system is still limited but should become stronger with the arrival of more and more plant-produced proteins in clinical phases II and III. Next, the recombinant protein yields are still much lower than the conventional platforms. Moreover, recombinant proteins secreted from plant suspension cells are susceptible to degradation by host proteases secreted during growth (Navarre *et al.*, 2012). Finally, generation of a panel of glyco-engineered BY-2 cell lines would allow the production of different glycoforms of the recombinant glycoproteins. This should lead to the identification of the most valuable *N*-glycan structure for a -or a class of- pharmaceutical glycoprotein.

2. *N*-glycosylation

2.1. Plant *N*-glycosylation pathway

Hundreds of different PTMs have been detected on proteins. The most frequently reported PTMs are phosphorylation, acetylation, ubiquitination, succinylation, methylation, malonylation as well as *N*- and *O*-glycosylations, according to the dbPTM databank (Ramazi and Zahiri, 2021). Glycosylation is one of the most complex PTMs, consisting of the attachment of oligosaccharides to proteins and their maturation throughout the secretory pathway. There are two main types of glycosylation: *O*- and *N*-glycosylation. *O*-glycosylation consists of the sequential addition of sugars to hydroxyl residues of amino acids present in a fully formed protein in the Golgi apparatus. They are typically found on serine, threonine or hydroxylated amino acids such as hydroxyproline or hydroxylysine (Strasser, 2016). *O*-glycosylation differs greatly between mammals, plants and yeast. *N*-glycosylation, on the contrary, is a partly conserved process among eukaryotes. It consists of the addition of an oligosaccharide precursor to the nitrogen of an asparagine (Asn) amide residue present on a nascent protein in the endoplasmic reticulum (ER). The *N*-glycan is subsequently trimmed and modified throughout the secretory pathway. Here, we will only focus on *N*-glycosylation.

The *N*-glycosylation pathway starts in the ER with the transfer of the lipid-linked oligosaccharide precursor Glc3Man9GlcNAc2 to a nascent protein (Figure 3). This precursor is common to most eukaryotes, with few exceptions such as some parasitic species and microalgae (Mathieu-Rivet *et al.*, 2020). The site of *N*-glycosylation is defined by the consensus sequence N-X-S/T (in which X is any amino acid but a proline). The synthesis of this precursor starts at the cytosolic side of the ER membrane by the assembly of Man5GlcNAc2 on the lipid carrier dolichol pyrophosphate (Dol-PP). N-acetylglucosamine (GlcNAc) and mannose (Man) residues are transferred stepwise on Dol-PP by glycosyltransferases (GT) until Man5GlcNAc2-Dol-PP is obtained. Then, it is flipped to the luminal side of the ER membrane. Four Man and three glucoses (Glc) residues are further transferred to obtain the Glc3Man9GlcNAc2-PP-Dol precursor. The oligosaccharide is transferred from Dol-PP to a consensus Asn by the oligosaccharyltransferase complex (Lannoo and Van Damme, 2015).

Before leaving the ER, the newly synthesized glycoprotein has to pass a quality control process. Once α -glucosidases (GCS) I and II have trimmed the two outermost Glc, GlcMan9GlcNAc2 is recognized by the lectin site of the chaperones calnexin and calreticulin (Figure 3). This interaction promotes folding of

glycoproteins and prevents aggregation of folding intermediates (Strasser, 2016). GCSII then trims the last Glc residue and the glycoprotein is released. If it is incorrectly or incompletely folded, the uridine diphosphate (UDP)-glucose:glycoprotein glucosyltransferase (UGGT) unfolded glycoprotein sensor transfers a Glc residue to the oligosaccharide to re-form a GlcMan9GlcNAc2, which will be detected again by the chaperones. UGGT and ER chaperones then successively interact with the glycoprotein until it is correctly folded or eliminated. The α -mannosidase 3 (MNS3) removes a Man residue from the middle branch of correctly folded glycoproteins, which can then continue their journey into the Golgi apparatus. Clearance of misfolded glycoproteins is carried out through the action of MNS4 and 5, which trim two Man residues from Man9GlcNAc2. This leaves a terminal α 1.6-Man, which is a signal for ER-associated degradation (ERAD) (Lannoo and Van Damme, 2015).

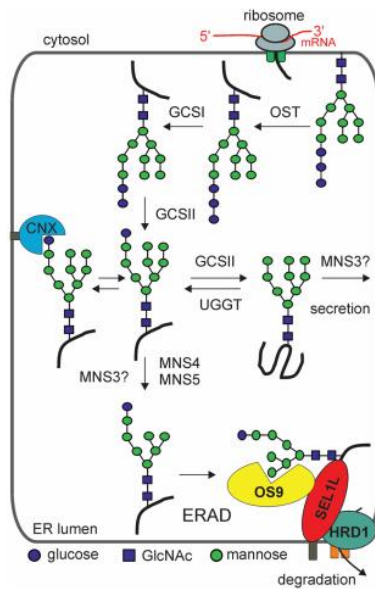


Figure 3 *N*-Glycan processing and *N*-glycan-mediated quality control in the ER of plants (from Strasser 2016). OST, oligosaccharyltransferase; GCS, α -glucosidase; UGGT, UDP-glucose:glycoprotein glucosyltransferase; MNS, class I α -mannosidase; ERAD, Endoplasmic reticulum-associated degradation; OS9, SEL1L, HRD1: proteins involved in the sorting out of terminally misfolded polypeptides for ERAD.

Whether and how an *N*-glycan moiety is modified in the Golgi is determined by the accessibility of the glycan site on the folded protein, but also by its journey through

the Golgi. Indeed, the glycosylation reactions are carried out sequentially and one reaction can be a prerequisite for or inhibit another. The final processed glycoforms are thus the outcome of a very complex spatiotemporal glycosylation network. As a result, the compartmental localization of GTs impacts the final outcome of the *N*-glycans. All known Golgi-resident GTs are type II transmembrane proteins. They consist of a cytosolic N-terminus, a transmembrane domain, a stem and the catalytic domain. The cytosolic tail, transmembrane domain and stem form the CTS region, which is responsible for the sub-Golgi localization (Figure 4). The first Golgi GTs, acting on high mannose or early complex *N*-glycan substrates display a cis/medial localization, while the GTs recognizing more processed substrates are localized in the medial/trans-Golgi (Loos *et al.*, 2014). Man5-8GlcNAc2 is referred to as high mannose, while the further processed forms are hybrid or complex if one or two antennas are further processed (Figure 5). Paucimannosidic *N*-glycans (Man3 $\pm$ XF) are considered complex because they result from the trimming of complex *N*-glycans by β -*N*-acetylhexosaminidases (HEXOs) (Schoberer and Strasser, 2018).

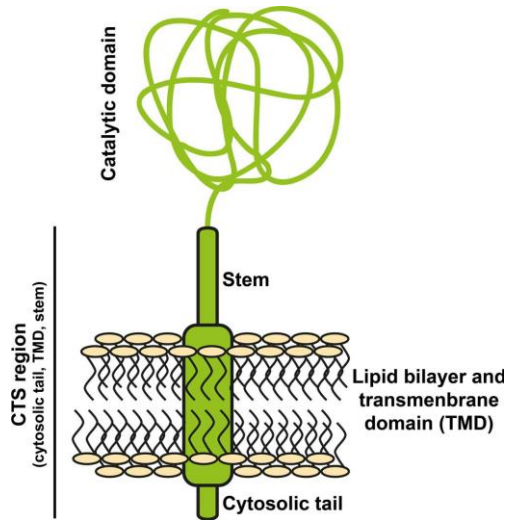


Figure 4. Structure of *N*-glycosyltransferases. Golgi-resident glycosyltransferases are type II transmembrane proteins. Their localization within the Golgi depends on the N-terminal CTS region, consisting of the cytosolic tail, the transmembrane domain and the stem. The C-terminal catalytic domain is located in the Golgi lumen (from Loos *et al.*, 2014).

In the cis-Golgi, the Man8GlcNAc2 glycan is further trimmed and/or complexified. MNS1 and MNS2 can trim up to three Man residues, resulting in Man5GlcNAc2

(Figure 5) (Strasser, 2016). The first step that initiates the formation of hybrid or complex *N*-glycans is catalyzed by the *N*-acetylglucosaminyltransferase I (GnTI) in the cis/medial-Golgi. GnTI catalyzes the transfer of a GlcNAc on the α 1,3-Man branch of Man5GlcNAc2, thus generating a hybrid *N*-glycan. This step is thought to be required for the action of all downstream GTs (Strasser *et al.*, 1999).

Three different enzymes can directly use GlcNAcMan5GlcNAc2 as a substrate in the medial- and trans-Golgi (Figure 5). β 1,2-xylosyltransferase (XylT) and α 1,3-fucosyltransferase (α 1,3-FucT) add a β 1,2-xylose on the first Man or a fucose on the first GlcNAc, respectively. In parallel, Golgi α -mannosidase II (GMII) can cleave two Man residues, which results in the hybrid structure (Fuc/Xyl)-GlcNAcMan3GlcNAc2. Then, the medial-Golgi-resident GT GnTII catalyzes the formation of complex *N*-glycans by adding a GlcNAc on the antenna previously trimmed by GMII to form (Fuc/Xyl)-GlcNAc2Man3GlcNAc2 (Strasser, 2016).

In the trans-Golgi, Lewis A (Le^a) structures can be generated on terminal GlcNAc. First, β 1,3-galactosyltransferase (β 1,3-GalT) catalyzes the addition of a galactose residue to a terminal GlcNAc and then α 1,4-fucosyltransferase (α 1,4-FucT) transfers a fucose residue on the same GlcNAc (Figure 5).

Alternatively, *N*-glycans can leave the Golgi without Le^a processing. They can undergo one last modification by HEXOs. Two active HEXOs have been identified in plants. If targeted to the vacuole, *N*-glycans with one or two terminal GlcNAc can be processed by the soluble HEXO1 (Figure 5). HEXO1 trims terminal GlcNAc to form (semi-) paucimannosidic *N*-glycans. If the glycoprotein is instead secreted to the apoplast, *N*-glycans with terminal GlcNAc can be trimmed by the plasma membrane-located, HEXO3 in the same manner as HEXO1 (Liebminger *et al.*, 2011).

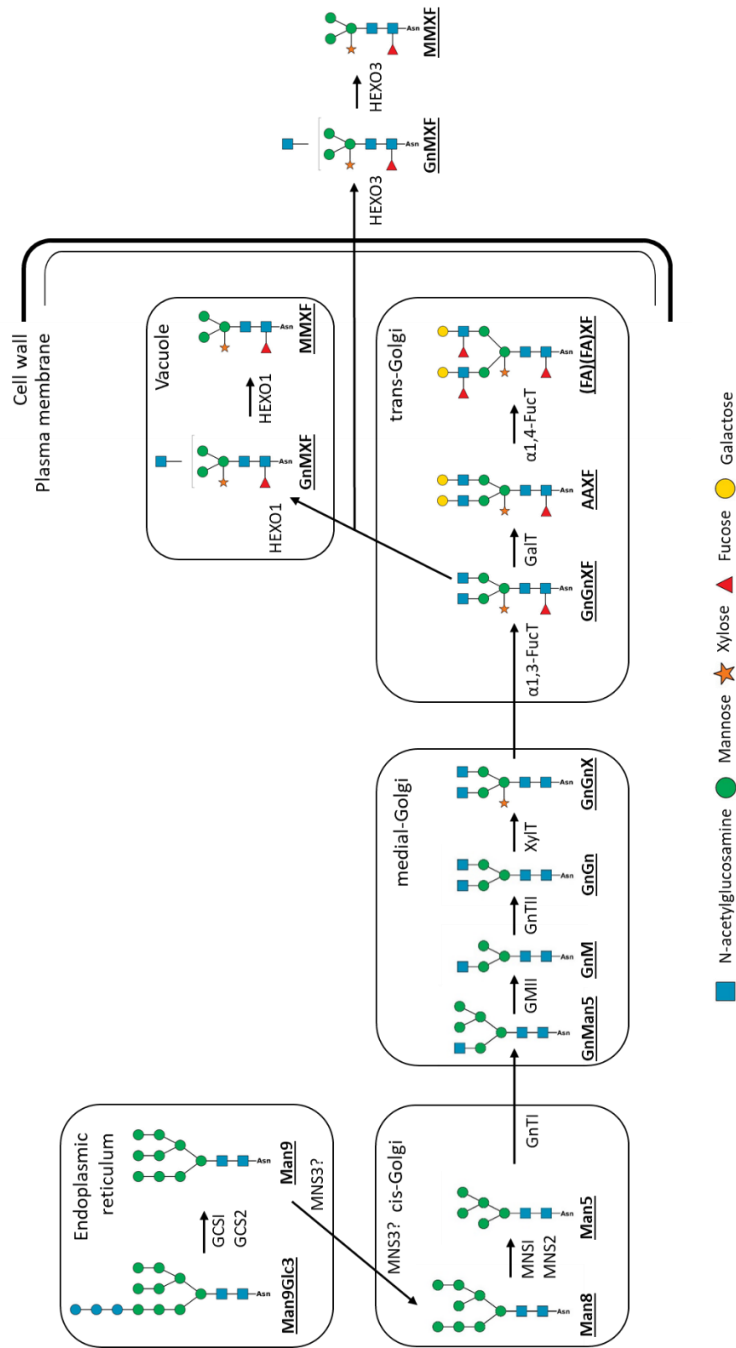


Figure 5: Schematic plant N-glycosylation pathway. N-glycan abbreviations are used according to the ProGlycan short nomenclature (www.proglycan.com). GCS, α -glucosidase; MNS, class I α -mannosidase; GnT, N-acetylglucosaminyltransferase; GMn, Golgi- α -mannosidase II; XylT, β 1,2-xylosyltransferase; FucT, fucosyltransferase; GalT, galactosyltransferase; HEXO, hexosaminidase. Adapted from Strasser, 2016.

2.2. Differences between plant and human *N*-glycans

N-glycans are more diverse in humans than in plants. While the steps preceding the formation of hybrid and complex *N*-glycans are well conserved, downstream processes diverge and lead to a higher number of structures in humans.

First, human *N*-glycans do not contain β 1,2-xylose or α 1,3-fucose. Instead, the basal GlcNAc can be modified with α 1,6-fucose (Figure 6). While in plants terminal complex GlcNAc can be modified with Le^a structure (β 1,3-galactose and α 1,4-fucose), human complex *N*-glycans are mainly decorated with GlcNAc-linked β 1,4-galactose that can be further elongated with sialic acids. However, rarely, *N*-glycan-linked Le^a can be expressed in humans in specific cell types, such as in polymorphonuclear leucocytes (Brazil *et al.*, 2017) or in cancerous colon cells (Zhang *et al.*, 1994). Finally, plant *N*-glycans are biantennary whereas mammalian *N*-glycans can be bi, tri or tetraantennary due to the presence of GnTIII, GnTIV and GnTV. As each antenna is modified individually, this increases greatly the heterogeneity of *N*-glycans. Finally, plant *N*-glycans can be shortened to form paucimannosidic structures by HEXOs (Liebminger *et al.*, 2011). Paucimannoses are common *N*-glycoepitopes in invertebrates and plants. They have also been detected in vertebrates but only in specific cell types (Dahmen *et al.*, 2015).

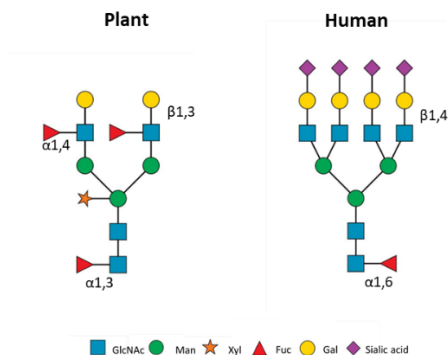


Figure 6. Example of highly processed complex *N*-glycans produced in plants and humans.

The diversity of glycan structures that are synthesized at each *N*-glycosylation site of a single glycoprotein can be therefore very high in mammalian cells and is referred to as microheterogeneity (Rösenlocher *et al.*, 2016). This can lead to a huge diversity of glycoforms for the same recombinant glycoprotein. The consequence for molecular farming is that it complicates the quality control. In addition, this diversity prevents identification of the most appropriate glycoform.

3. Glyco-engineering in plants

Producing a recombinant glycoprotein in plant cells is a simple way to limit the number of distinct glycoforms compared to mammalian cells. However, the presence of non-human residues and the absence of some human residues may be a concern, depending on the glycoprotein produced. Indeed, *N*-glycosylation is a key post-translational modification, which deeply influences the properties of glycoproteins, e.g., making them more soluble, protecting them from proteolysis, covering antigenic sites and affecting their three-dimensional configuration (Ramazi and Zahiri, 2021). In this section, we will go through the glyco-engineering strategies developed for glycoprotein production in plants, with a special emphasis on BY-2 cells. There are three main approaches, which can be combined: depletion of non-human GTs, addition of human GTs and simplification of the *N*-glycan repertoire (Figure 7).

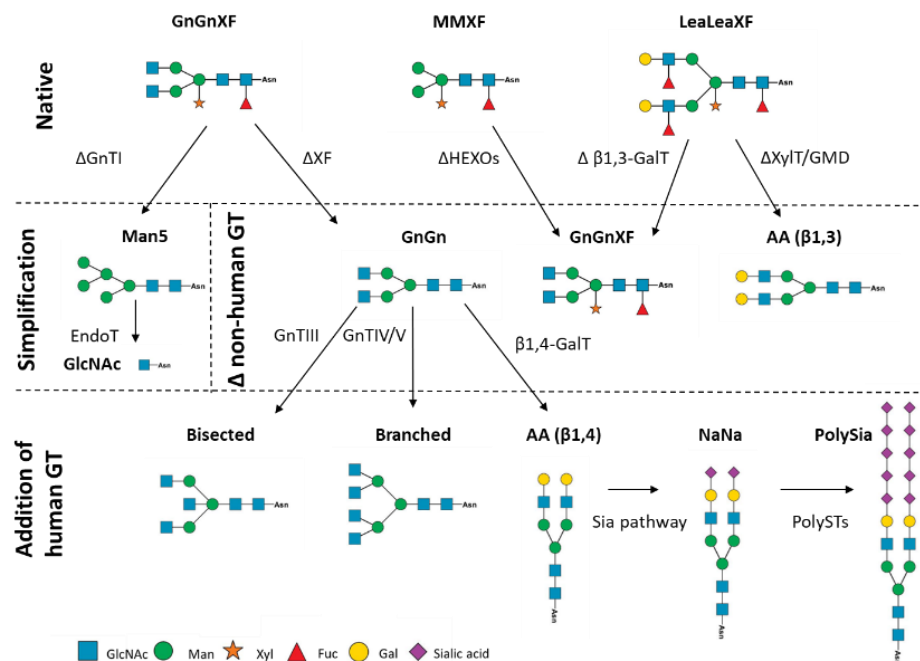


Figure 7. Schematic representation of plant-based *N*-glycan engineering. Native: The three major complex glycoforms frequently detected in wild-type plants are GnGnXF, MMXF paucimannosidic structure, and Le^a structures. Simplification: Glycan simplification and homogenization by gene knockout and/or ectopic gene expression approaches: Man5 structure generated by GnTI knockout/down; single GlcNAc (GlycoDelete). Suppression (Δ) of non-human *N*-glycans: GnGn (common eukaryotic core); GnGnXF; AA, GnGn with terminal β1,3-galactose residues. Addition of human GT: Augmentation of plant *N*-glycans by stable/transient overexpression of foreign GTs: Bisected structure; Branched ([GnGn][GnGn]) structure; AA, GnGn with terminal β1,4-galactose residues; NaNa, α2,6-sialylated glycoforms; PolySia, polysialylated *N*-glycans. ΔXF: knockout/down of α1,3-FucT and XylT; ΔHEXO: knockout of β-N-acetylhexosaminidases HEXO1 and HEXO3; Sia Pathway: recombinant expression of the sialic acid pathway; PolySTs: α2,8-polysialyltransferase II, IV; GT: glycosyltransferase (Adapted from Montero-Morales and Steinkellner, 2018). *N*-glycans are labeled following ProGlycan nomenclature.

3.1. Depletion of non-human *N*-glycan structures

3.1.1. Depletion of α 1,3-fucoses and β 1,2-xyloses

One limitation of plant-made glycoproteins is the presence of potentially allergenic α 1,3-fucose and β 1,2-xylose epitopes linked to complex structures, and this is particularly of concern for vaccine glycoproteins (Bosch *et al.*, 2013). Despite the absence of adverse effects detected upon the administration of plant-produced proteins to patients, the elimination of plant glycans is still considered beneficial to anticipate regulatory issues (Fischer *et al.* 2018).

As a result, knockdowns and knockouts of *a1,3-FucT* and *XylT* genes have been obtained for most plant species that are valuable for recombinant glycoprotein production. The first glyco-engineered plants were mainly knockdowns obtained by RNA interference (RNAi) or knockouts originating from chemical/radiation/transposon mutagenesis libraries. Plant-specific residues in RNAi hosts were suppressed or significantly reduced, without any noticeable phenotype change in *Lemna minor* (duckweed) (Cox *et al.*, 2006), *M. sativa* (Sourrouille *et al.*, 2008), *N. benthamiana* plants (Strasser *et al.*, 2008), *O. sativa* suspension cells (Shin *et al.*, 2011) and BY-2 cells (Yin *et al.*, 2011). In *Physcomitrella patens* (spreading earth moss), *XylT/a1,3-FucT* double knockout lines were obtained by gene targeting via homologous recombination (HR) -highly efficient in this organism- and did not show any growth or morphological abnormalities (Koprivova *et al.*, 2004). In an *A. thaliana* *a1,3-FucT* and *XylT* T-DNA insertion mutant, the *N*-glycan content was completely free of β 1,2-xylose and α 1,3-fucose, without any altered phenotype in normal growth conditions (Strasser *et al.*, 2004b). In contrast, *Lotus japonicus* *FucT* T-DNA insertion mutants showed a strong reduction of α 1,3-fucose, associated with significantly reduced growth and altered seed production (Pedersen *et al.*, 2017). *O. sativa* *XylT* knockout plants obtained via radiation-induced mutagenesis also showed developmental defects, although mainly in abiotic stress conditions (Takano *et al.*, 2015). Moreover, *O. sativa* *FucT* knockout plants, generated by T-DNA insertion, displayed developmental abnormalities associated with hormone signaling (Harmoko *et al.*, 2016, Sim *et al.*, 2018). Thanks to the advent of CRISPR/Cas9 technology, double knockouts were obtained in BY-2 cells (Hanania *et al.*, 2017; Mercx *et al.*, 2017), *O. sativa* cells (Jung *et al.*, 2021) as well as *N. benthamiana* (Jansing *et al.*, 2018) and *N. tabacum* plants (Göritz *et al.*, 2022). They were totally devoid of β 1,2-xylose and α 1,3-fucose and did not show any altered phenotype under normal growth conditions. However, regeneration of rice plants from the mutant calli could not be obtained. The authors suggest that the plant regeneration failed due to altered hormone function.

An alternative strategy was developed to simultaneously reduce β 1,2-xylose and α 1,3-fucose as well as α 1,4-fucose residues present in Le^a structures in *N. benthamiana* plants (Matsuo *et al.*, 2011, Matsuo *et al.*, 2014). Fucose residues are transferred to *N*-glycans from the sugar nucleotide guanosine diphosphate (GDP)-l-fucose. Ninety-five percent of GDP-l-fucose is synthesized from GDP-Man by the *de novo* pathway, the remaining 5% being produced by the salvage pathway from cytosolic fucose (Matsuo *et al.*, 2011; Ebert *et al.*, 2017). Depletion of fucose residues in *N*-glycans is thus possible by repression of GDP-l-fucose *de novo* biosynthesis. Modification of GDP-d-mannose into GDP-l-fucose requires the action of GDP-d-mannose 4,6-dehydratase (GMD) (Figure 8). Inactivation of *GMD* thus makes it possible to eliminate together most α 1,3-fucose and α 1,4-fucose residues. Matsuo *et al.*, (2011) achieved repression of *GMD* by using virus-induced gene silencing and RNAi. An increase of 95% in the proportion of fucose-free *N*-glycans was observed in the total soluble cellular proteins (TSCP) of *N. benthamiana* plants following RNAi. RNAi was then used to generate plants in which *XyIT* genes are repressed. Finally, these two lines were crossed and a strong reduction of β 1,2-xylose and α 1,3-fucose-substituted *N*-glycans was observed: from 87% in wild-type (WT) plants to 10% in the double-repressed mutant (Matsuo *et al.*, 2014).

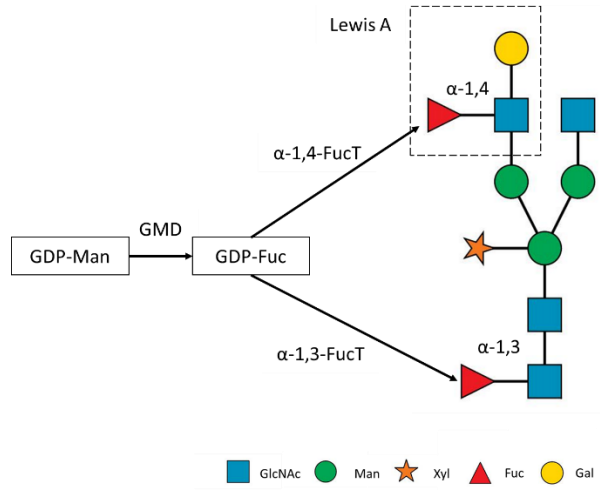


Figure 8. *De novo* biosynthesis pathway of fucose epitopes on plant *N*-glycans. GDP: Guanosine diphosphate; GMD: GDP-d-mannose 4,6-dehydratase.

3.1.2. Depletion of Lewis A epitopes

The Le^a epitope, consisting of terminal GlcNAc-linked β 1,3-galactose and α 1,4-fucose (Figure 5), can be detected on plants glycoproteins, mainly localized at the surface of plant cells, suggesting that they might play a role in cell-to-cell communication in plants (Fitchette-Lainé *et al.*, 1997). Important disparities can be observed between species. Indeed, N-glycans from rice calli intra- and extracellular proteins having β 1,3-galactose and/or α 1,4-fucose account for 36% and 24% of all N-glycans, respectively (Shin *et al.*, 2011; Harmoko *et al.*, 2016). In contrast, BY-2 cells extracellular proteins contained less than 2% of N-glycans with these residues (Mercx *et al.*, 2017). Similarly, low levels of Le^a epitopes were observed in *N. tabacum* plants, *P. patens* and in some organs of *A. thaliana* (Elbers *et al.*, 2001; Lhernould *et al.*, 2008; Strasser *et al.*, 2007; Parsons *et al.*, 2012). Nevertheless, the presence of these particular residues, even in small amounts, can raise the same safety concerns as β 1,2-xylose and α 1,3-fucose residues.

P. patens and *A. thaliana* β 1,3-GalT knockouts/knockdowns were obtained by HR and T-DNA insertion or RNAi, respectively (Parsons *et al.*, 2012; Strasser *et al.*, 2007). No β 1,3-galactose could be detected in these plants. No α 1,4-fucose was identified either. This is due to the strict acceptor substrate specificity of α 1,4-FucT reported for Gal β 1-3GlcNAc. Accordingly, β 1,3-galactose was still detected in *a*1,4-FucT knockouts *P. patens* (Parsons *et al.*, 2012). No altered phenotype was observed for *P. patens* mutants under normal growth conditions. Root hair elongation was slightly impaired in *A. thaliana* Le^a-deficient plants (Frank *et al.*, 2021).

The absence of modified phenotype in β 1,3-GalT and *a*1,4-FucT-knockout *A. thaliana* and *P. patens* cannot rule out any important role in other plant species, particularly those with high amounts of Le^a N-glycans such as rice.

3.1.3. Depletion of paucimannoses

Paucimannosidic N-glycans are generated by the truncation of complex N-glycans catalyzed by HEXOs. These truncated forms are not commonly observed in humans but are frequent in invertebrates and plants (Tjondro *et al.*, 2019). Again, there is concern that this may negatively affect the biological activity of recombinant therapeutic glycoproteins produced by plants.

A. thaliana double *hexo1/3* mutant plants obtained by T-DNA insertions do not show any particular phenotype. A dramatic reduction of paucimannosidic N-glycans as compared to WT plants was detected by mass spectrometry (MS) analysis of total N-glycans of the double mutant. MMXF N-glycans could not be detected, while it is

the predominant *N*-glycan species found in WT plants (Figure 9). GnMXF/MGnXF could still be detected, although in lower levels. Additional MS experiments demonstrated the complete absence of the GnMXF isomer. While GnMXF requires the action of HEXOs, MGnXF can be obtained following *N*-glycan processing of Man5GnXF by GMII (Figure 5). This suggests that the remaining MGnXF is not obtained through HEXO activity and hence HEXO1 and HEXO3 are the only HEXOs involved in paucimannosidic *N*-glycans generation (Liebminger *et al.*, 2011).

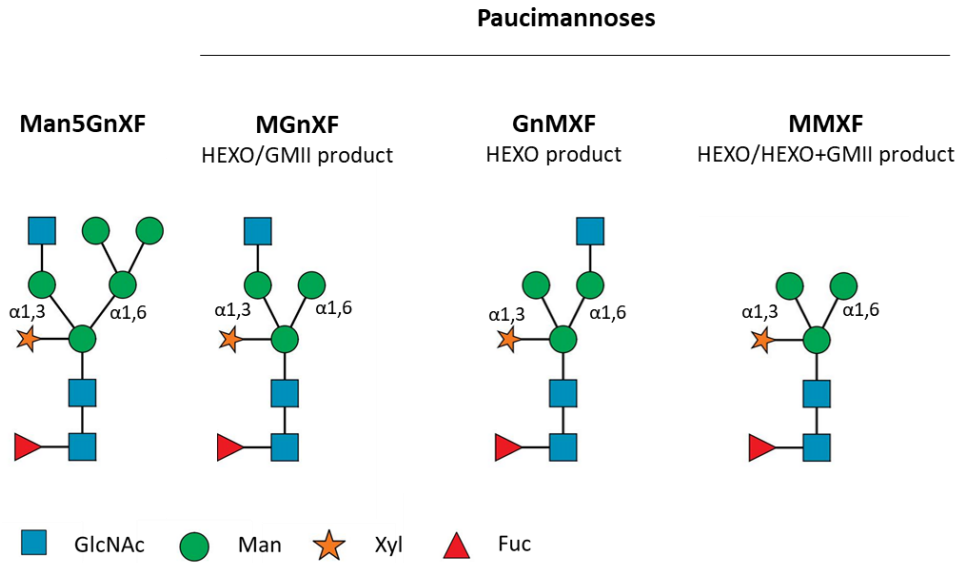


Figure 9. Origin of paucimannose *N*-glycan structures. *N*-glycans are labeled following ProGlycan nomenclature.

Later, an *hexo3* knockdown was obtained via RNAi in *N. benthamiana*. The *N*-glycans present on the recombinant secreted human α 1-antitrypsin were analyzed. A reduction of paucimannosidic structures and an increase in complex *N*-glycans were measured. Paucimannosidic MMXF, GnMXF and MGnXF *N*-glycans were still detected. The authors suggest that either their transient silencing approach was not very efficient or that some HEXO3 proteins made before the silencing establishment were still active in the apoplast (Shin *et al.*, 2017).

3.2. Addition of human *N*-glycan structures

Monoclonal antibodies (mAbs), mainly IgGs, are by far the most produced recombinant glycoproteins. As a result, most glyco-engineering applications have been performed to generate the most valuable cell lines for IgG production. IgGs share a conserved *N*-glycosylation site on Asn 297 of the two heavy chain constant domain 2 (CH2) of the crystallizable fragment (Fc). The type of *N*-glycan structure on this *N*-site strongly affects Fc effector functions (Kiyoshi *et al.*, 2017).

3.2.1. Addition of β 1,4-galactoses

Most *N*-glycans present on the conserved Asn 297 of human serum IgGs are β 1,4-galactosylated, which can influence their biological activity (Strasser *et al.*, 2009). Moreover, the presence of β 1,4-galactose residue is a prerequisite for sialic acid elongation, which enhances antibody serum half-life. Hence, human β 1,4-GalT was one of the first exogenous GTs expressed in plants for molecular farming (Palacpac *et al.*, 1999).

Different variants have been expressed in BY-2 cells (Palacpac *et al.*, 1999; Navarre *et al.*, 2017), *N. tabacum* plants (Bakker *et al.*, 2001; Bakker *et al.*, 2006; Hesselink *et al.*, 2014; Kittur *et al.*, 2020), *N. benthamiana* plants (Vézina *et al.*, 2009) and alfalfa plants (Sourrouille *et al.*, 2008). There are two natural isoforms of hGalT, resulting from an alternative transcription initiation (Hesselink *et al.*, 2014). The shorter isoform lacks thirteen amino acids in the cytosolic tail. Since the composition of the CTS determines hGalT sub-Golgi localization, different *N*-glycan repertoires were obtained when expressing the two natural isoforms or when fusing hGalT to other CTS. Expression of hGalT with its natural short CTS or swapped with the CTS from NtGnTI or AtXylT resulted in hybrid monogalactosylated *N*-glycans as well as low β 1,2-xylose and α 1,3-fucose content (Palacpac *et al.*, 1999; Bakker *et al.*, 2001; Misaki *et al.*, 2003; Bakker *et al.*, 2006; Fujiyama *et al.*, 2007; Sourrouille *et al.*, 2008; Vézina *et al.*, 2009; Hesselink *et al.*, 2014; Navarre *et al.*, 2017). These versions of hGalT are targeted to the cis/medial-Golgi and can thus modify *N*-glycans as soon as the first terminal GlcNAc is added by GnTI (Figure 5). Early β 1,4-galactosylation then inhibits the transfer of β 1,2-xylose and α 1,3-fucose residues as well as the formation of the second complex antenna, resulting in hybrid, unxylosylated, unfucosylated *N*-glycans. On the other hand, expression of the long natural isoform or fusion to the CTS from rat α 2,6-sialyltransferase (ST) resulted in complex, mono- and digalactosylated *N*-glycans containing the usual epitopes β 1,2-xylose and α 1,3-fucose (Hesselink *et al.*, 2014; Navarre *et al.*, 2017). These hGalT versions are targeted to the medial/trans-Golgi and will thus act on more processed *N*-glycans. As a result,

galactosylation occurs on structures that are already xylosylated, fucosylated and biantennary complex. This highlights the importance of targeting exogenous GTs to the adequate sub-Golgi compartment.

hGalT was also expressed in *P. patens* and targeted to the trans-Golgi via fusion to the moss α 1,4-FucT CTS (Bohlender *et al.*, 2022). While the expected galactosylation was observed, α -linked arabinoses were detected as well. This was already reported in *N. tabacum* (Kittur *et al.*, 2020). These residues are absent in humans and hence constitute an issue for full glycan humanization in plants.

A complementary strategy to increase β 1,4-galactose (or β 1,3-galactose) was carried out in *N. benthamiana* by silencing β -galactosidase 1 gene, responsible for the trimming of terminal galactoses (Kriechbaum *et al.*, 2020). Both genome edition and RNAi reduced galactosidase activity and increase fully galactosylated *N*-glycans.

Anti-HIV antibodies were transiently expressed in *N. benthamiana* *XyII/FucT* knockout plants expressing hGalT fused to rat ST CTS. The mAbs were fully galactosylated and their virus neutralization potency increased (Strasser *et al.*, 2009).

3.2.2. Addition of fucoses

Transient expression of human *α 1,6-fucosyltransferase* was successfully carried out in *N. benthamiana* *XyII/FucT* knockouts to replace the core α 1,3-fucose by the typical mammalian α 1,6-fucose (Forthal *et al.*, 2010, Castilho *et al.*, 2011). Expression of *α 1,6-fucosyltransferase* in the rice endosperm resulted in the addition of the mammalian epitope, in association with a decrease in α 1,3-fucose (Wang *et al.*, 2017).

Recently, an association between specific ABO blood groups and COVID-19-linked hospitalization was established (Wu *et al.*, 2020). Since glycosylation of viral envelope proteins is susceptible to affect virus infectivity and immune evasion, it was suggested that the presence of distinct ABO carbohydrate structures could improve immunological reactions. Glyco-engineering was carried out in *N. benthamiana* by co-expressing human *GalT* or *A. thaliana GalT*, human *α 1,2-FucTs*, human *α 1,3-GalNAc-transferase* (ABO A) or human *α 1,3-galactosyltransferase* (ABO B) and the receptor binding domain (RBD) from SARS-CoV-1 and SARS-CoV-2 spike proteins (König-Beihammer *et al.*, 2022). Binding of these viral antigen glycovariants to the human ACE2 receptor was confirmed. Moreover, binding of IgGs originating from sera of non-infected blood group O and B individuals to the RBD “ABO A” glycovariant showed that anti-A antibodies could be protective against viral infection.

Other structures resulting from the modification of terminal GlcNAcs, namely Lewis X (β 1,4-Gal and α 1,3-Fuc linked to a GlcNAc), LDN (LacdiNAc; β 1-4GalNAc linked to a GlcNAc) and LDN-F (β 1-4GalNAc and α 1,3-Fuc linked to a GlcNAc) antigens have been engineered in *N. benthamiana* plants (Wilbers *et al.*, 2017). The aim was to produce helminth parasite omega-1 and kappa-5 immunomodulatory glycoproteins with typical helminth *N*-glycosylations for the treatment of immune-related diseases. Indeed, the type of glycan structures present on these glycoproteins affects their immunomodulatory properties. Transient co-expression of these glycoproteins with specific glycosyltransferases yielded *N*-glycans bearing the expected terminal Le^x, LDN, LDN-F along with the usual xylose and core α 1,3-fucose, that are also present in helminths.

3.2.3. Addition of a bisecting GlcNAc

Another mammalian feature that is absent in plants is the production of tri and tetra-antennary *N*-glycans. Multiantennary *N*-glycans can indirectly increase the serum half-life of therapeutic glycoproteins. Indeed, when combined with the expression of *GalT* and *ST*, the multiplication of antennas increases the overall sialylation of the *N*-glycan. The degree of sialylation slows down the clearance through the liver receptors for non-sialylated glycoproteins (Nagels *et al.*, 2011).

Trans-Golgi-resident GnTIII adds a bisecting GlcNAc into *N*-glycans in mammalian cells, which has a strong impact over other modifications (Figure 10). Indeed, the presence of a bisecting GlcNAc prevents the action of GMII, GnTII, α 1,6-FucT, GnTIV and GnTV (Karg *et al.*, 2010; Kizuka *et al.*, 2016). The expression of human *GnTIII* in tobacco plants resulted in most complex-type *N*-glycans having a bisecting GlcNAc (Rouwendal *et al.*, 2007). In the same study, it was shown *in vitro* that a bisecting GlcNAc efficiently inhibited the activity of both At α 1,3-FucT and AtXylT. However, despite high levels of bisecting GlcNAc on *N*-glycans, xylose levels were not affected and core-fucosylation was only slightly reduced. It was suggested that hGnTIII was also localized in the plant trans-Golgi and its activity was thus downstream to that of XylT and α 1,3-FucT. Later on, GnTIII was fused to CTS from GTs localized in the cis- (GMII), medial- (α 1,3-FucT and XylT) and trans-Golgi (ST) and transiently expressed in *N. benthamiana* (Castilho *et al.*, 2011). Similar to what was observed for hGalT localization, targeting to later stages of the *N*-glycosylation pathway was more effective in generating bisected *N*-glycans. In BY-2 cells, GnTIII fused to the CTS of AtMNSII increased the amount of *N*-glycans devoid of α 1,3-fucose and β 1,2-xylose (Karg *et al.*, 2010).

Surprisingly, stable expression of hGnTIII with its natural CTS in *N. benthamiana* resulted in a low level of bisected glycans (Kittur *et al.*, 2020), possibly due to improper localization of GnTIII. Swapping the endogenous CTS with that of XylT or ST should improve the level of bisecting GlcNAc.

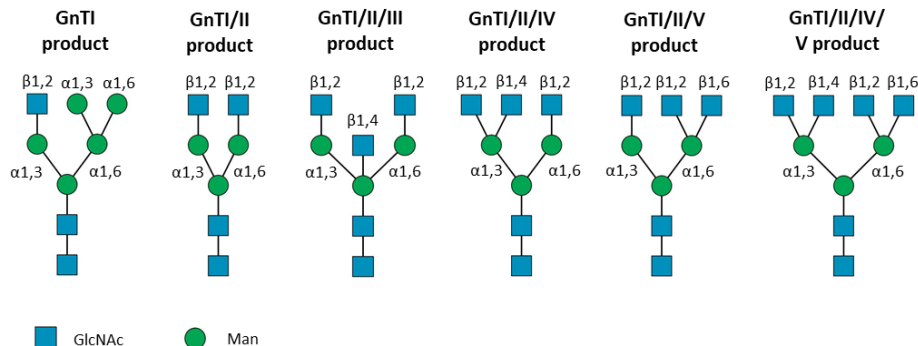


Figure 10. *N*-glycan structures related to GnTI, GnTII, GnTIII, GnTIV and GnTV activity.

3.2.4. Addition of third and fourth antennae

GnTIV and GnTV are responsible for the formation of non-bisecting tri and tetra-antennary *N*-glycans in mammals (Figure 10). GnTIV adds a second GlcNAc on the α 1,3-Man branch, while GnTV modifies the α 1,6-Man. GnTIV and GnTV are not mutually exclusive and can thus generate tetraantennary structures. WT and *XylT/a1,3-FucT*-silenced *N. benthamiana* plants were used to transiently or stably express *GnTIV* and *GnTV*. To target these GTs to the medial/trans-Golgi, they were fused to the CTS of XylT and α 1,3-FucT. GnTIV and GnTV activity was higher (more triantennary structures) in *XylT/FucT*-silenced plants. This is either due to competition for the same substrate or a less efficient conversion rate of β 1,2-xylose/ α 1,3-fucose-modified *N*-glycans by GnTIV/V. The plant presenting maximal *N*-glycan branching had 10% tetraantennary, 25% triantennary, and 35% biantennary structures. Thus, at least 70% of the *N*-glycans of this plant could be galactosylated and sialylated to generate low clearance therapeutic antibodies. Most plants did not show any abnormal phenotype (Nagels *et al.*, 2011). Next, Nagels *et al.*, (2012) also expressed *GnTIV* and *GnTV* separately in *XylT/FucT*-KO *A. thaliana*, which yielded triantennary *N*-glycans without any noticeable phenotype change.

GnTIV and *GnTV* were also transiently co-expressed with *GnTIII* in *XylT/FucT*-silenced *N. benthamiana*, which resulted in a significant amount of complex tetra-antennary glycans also carrying a bisecting GlcNAc (Castilho *et al.*, 2011).

3.2.5. Addition of sialic acid residues

Obtaining sialic acid (or *N*-acetylneuraminic acid (Neu5Ac)) residues on *N*-glycans in plants is a more complicated process than the different modifications reported so far in this chapter. Indeed, while the substrates for GalT, GnTIII, IV and V (UDP-galactose and UDP-GlcNAc) are present in plants, the substrate for ST (cytidine monophosphate (CMP)-sialic acid, also called CMP-Neu5Ac) is not. Thus, plants lack the machinery for biosynthesis of the nucleotide sugar CMP-Neu5Ac and its transport from the cytosol to the Golgi apparatus, where it serves as the donor substrate for different ST (Kallolimath *et al.*, 2016).

In mammals, sialic acid biosynthesis pathway starts in the cytosol with the conversion of UDP-GlcNAc to *N*-acetylmannosamine (ManNAc) and the phosphorylation of ManNAc to ManNAc-6-phosphate, catalyzed by the UDP-*N*-acetylglucosamine 2-epimerase/*N*-acetylmannosamine kinase (GNE) (Figure 11). Next, Neu5Ac phosphate synthetase (NANS) condenses ManNAc-6-phosphate and phosphoenolpyruvate, resulting in Neu5Ac-9-phosphate. Subsequently, Neu5Ac-9-P phosphatase (NANP) dephosphorylates Neu5Ac-9-P into Neu5Ac. CMP-Neu5Ac synthetase (CMAS) activates the resulting primary sialic acid in the nucleus to CMP-Neu5Ac (Castilho *et al.*, 2008). Then, CMP-sialic acid transferase (CST) transfers CMP-Neu5Ac into the Golgi. Finally, a ST transfers Neu5Ac to the acceptor substrate: a β 1,4-galactosylated, biantennary *N*-glycan.

The sialic acid pathway was engineered in plants. In a first step, microbial Neu5Ac synthase was successfully expressed in the cytosol of BY-2 cells and alfalfa plants. Correct enzymatic activity was confirmed by *in vitro* activity assay (Paccalet *et al.*, 2006). In parallel, the human *CMAS*, *CST* and *ST* were expressed in BY-2 cells and the confirmation of their respective biological activities *in vitro* paved a route for the introduction of the sialic acid biosynthetic pathway (Misaki *et al.*, 2006; Kajiura *et al.*, 2011).

Synthesis of CMP-sialic acid was first obtained in *A. thaliana* by co-expressing GNE, NANS and CMAS (Figure 11) (Castilho *et al.*, 2008). Then, Castilho *et al.* (2010) combined the transient expression of GNE, NANS, *CMAS* with that of *CST*, *ST* and *bGalT* into a *XylT/a1,3-FucT* knocked-down *N. benthamiana* plant. Introduction of NANP was not mandatory as NANP-like activity was detected in plants. A mAb

produced in this plant showed a high level of sialylation on *N*-glycan at Asn 297. Later on, to obtain tri- and tetra-branched sialylated *N*-glycans, GnTIV and GnTV were co-expressed with the first six enzymes (Catilho *et al.*, 2013).

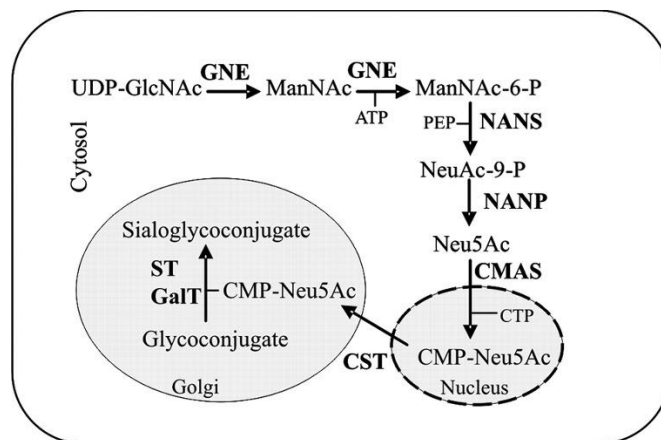


Figure 11. Schematic representation of the mammalian pathway for sialylation of glycoconjugates. GNE: UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase; NANS: Neu5Ac phosphate synthetase; NANP: Neu5Ac-9-P phosphatase, CMAS: CMP-Neu5Ac synthetase; CST: CMP-sialic acid transferase; GalT: β 1,4-galactosyltransferase; ST: α 2,6-sialyltransferase; UDP-GlcNAc: Uridine diphosphate-N-acetylglucosamine; ManNAc: N-acetylmannosamine; ManNAc-6-P: ManNAc-6-phosphate; NeuAc-9P: N-acetyl neuraminic acid-9-phosphate; CMP-Neu5Ac: cytidine monophosphate N-acetylneuraminic acid; PEP: phosphoenolpyruvate (Castilho *et al.*, 2010).

An approach combining stable and transient expression was also designed. Indeed, the transient expression methodology is less suitable for large-scale production as it is more prone to batch-to-batch inconsistencies and necessitates more manipulations than stable expression. The mammalian sialylation pathway used in the aforementioned study was stably expressed in *N. benthamiana* *XylT/FucT*-knockdown, resulting in sialylated *N*-glycans without β 1,2-xylose/ α 1,3-fucose plant-specific residues. Next, two human α 2,8-polysialyltransferases (ST8Sia-II and ST8Sia-IV) were transiently co-expressed, which resulted in *N*-glycan structures with a degree of polymerization greater than forty. Polysialylation has a substantial impact on pharmacokinetic properties of recombinant glycoprotein therapeutics (Kallolimath *et al.*, 2016).

Much effort was also put to implement the sialic acid biosynthesis pathway in *P. patens* *XylT/FucT* knockouts. To this end, the endogenous GalT was knocked out and seven mammalian enzymes were stably expressed: GNE, NANS, NANP, CMAS, CST, ST, hGalT (Bohlender *et al.*, 2020). Sialylated *N*-glycans were detected by MS, although in low amounts (around 6.3%).

3.3. Simplification of the *N*-glycan repertoire

3.3.1. Modification of the *N*-glycan precursor

Simplification of the *N*-glycan repertoire can be attempted by altering the formation of the *N*-glycan precursor, Glc3Man9GlcNAc2. Asparagine-linked glycosylation 3 (ALG3) mediates the addition of an α 1,3-mannose on the Man5-Dol-PP to form Man6-Dol-PP. *A. thaliana* *ALG3* knockout plants were isolated from the collections of T-DNA insertion lines (Kajiura *et al.*, 2010). In these mutants, an immature lipid-linked oligosaccharide structure, Man5^{ER}, was synthesized and transferred to nascent glycoproteins. It was then processed by α 1,2-mannosidases, mainly into Man3 structures. A cell culture was obtained from *alg3* plants and was used to produce human acid- α -GCS with predominantly Man3 structures (Sariyatun *et al.*, 2021). Acid- α -GCS is used for enzyme replacement therapy of Pompe disease and benefits from the presence of *N*-glycans with terminal mannose to allow cellular delivery through mannose receptors.

3.3.2. Inactivation of mannosidases

A. thaliana GMII knockouts were identified from a T-DNA insertion collection and their *N*-glycosylation profile was determined (Strasser *et al.*, 2006). Mainly high mannose and hybrid *N*-glycans were identified. Absence of complex *N*-glycans was due to the presence of mannose residues on the α 1,6-arm, which are normally cleaved by GMII. These *hybrid glycosylation 1* (*hgl1*) mutants suffered from enhanced salt sensitivity and both core fucose and xylose residues were found to be masked from full immunogenic recognition by untrimmed mannoses on the α 1,6 arm (Kaulfurst-Soboll *et al.*, 2011).

The latter feature was used to reduce the immunogenic potential of tomato fruits, in the context of food allergies (Kaulfurst-Soboll *et al.*, 2021). GMII-RNAi lines were generated. However, these plants suffered from reproductive defects and morphological abnormalities.

Lotus MNSI knockouts (*Ljman 1*) were identified within the LORE1 T-DNA insertion collection. Their *N*-glycan content was almost exclusively oligomannose structures, which is expected given the very upstream function of MNSI. Again, strong phenotypes such as reduced growth and seed production were observed for this mutant compared with the WT (Pedersen *et al.*, 2017).

3.3.3. *GnTI* knockdowns/outs

Humanization of plant *N*-glycans via the addition of human GTs improves several features such as clearance and complement-dependent cytotoxicity when applied to mAb production. However, it also ends up generating a mix of more or less processed complex *N*-glycans. *N*-glycan heterogeneity on therapeutic glycoproteins causes drawbacks for protein purification and process reproducibility and can lead to variable therapeutic efficacy (Meuris *et al.*, 2014). Furthermore, the impact of the glycoprofile of vaccine glycoproteins on their immunogenicity is not known. Therefore, the development of a recombinant glycoprotein production platform with a simplified and homogenized *N*-glycan repertoire could be a key to designing new highly effective vaccines and therapeutic glycoproteins. Glycoproteins presenting such homogenized *N*-glycan structures can also be modified *in vitro* by chemoenzymatic transglycosylation to generate different glycoforms with highly homogeneous and well-defined *N*-glycans (Liu *et al.*, 2018; Wada *et al.*, 2019).

Suppression of *N*-glycosylation sites by mutagenesis is not expected to give a functional deglycosylated protein. Indeed, the *N*-glycan precursor is added on the nascent polypeptide chain and hence affects the glycoprotein folding process. Likewise, the inactivation of ER-resident enzymes might have an impact on the folding catalyzed by ER chaperones. Rather, impairing the synthesis of complex *N*-glycans, responsible for most *N*-glycan diversity, is a powerful strategy to reduce the *N*-glycan repertoire. GnTI is necessary to generate all complex *N*-glycans, both in animals and plants. As a result, it has been targeted in several mammalian and plant species, using different approaches. Another approach to produce recombinant glycoproteins with unprocessed high-mannose *N*-glycans is to prevent their trafficking through the Golgi by the addition of a C-terminal H/KDEL ER retention sequence (Fujiyama *et al.*, 2009).

3.3.3.1. Animals

More than forty years ago, CHO mutated cells were isolated by selection for resistance to cytotoxic plant lectins. Indeed, thanks to their ability to bind to specific carbohydrates, lectins can induce cell aggregation and cross-linking of surface

glycoproteins, resulting in mitogenic and death-inducing effects (Parshenkov and Hennet, 2022). Hence, glycosylation-deficient mutants can become resistant to cytotoxic lectins through the suppression of certain *N*-glycan structures. These cell lines contained a wide variety of alterations in protein and lipid glycosylation (Stanley, 1983). Since then, this mutant CHO library was used to characterize glycosylation pathways and their components. Leuco-phytohemagglutinin (L-PHA) resistant (*Lec1*) CHO mutants were isolated as early as in 1974 (Gottlieb *et al.*, 1974; Stanley *et al.*, 1975). Genetic complementation allowed identifying the mutated gene responsible for the resistance to L-PHA and other plant lectins. This gene is *α1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase* (*Mgat1*), the mammalian homologue of *GnTI*. Its inactivation in CHO was associated with the absence of hybrid and complex *N*-glycans. Three point mutations in *Mgat1* were detected in *Lec1* CHO, all resulting in an amino acid substitution (Puthalakath *et al.*, 1996). Only one of them (C123R) affected enzyme activity and resulted in complete loss of activity.

Later, HEK 293S cells were engineered to produce only oligomannosidic *N*-glycans by random mutagenesis and lectin resistance selection (Reeves *et al.*, 2002). The HEK293S *Mgat1*⁻ cell line is currently the most popular cell line for transient and stable expression in structural biology. However, this system is not suitable for clinical and large-scale production due to genetic instability and the inability to grow for sustained periods at high cell densities (Byrne *et al.*, 2018). High-mannose type glycosylation was also obtained by cultivation with the α -mannosidase I inhibitor kifunensine (Büssow, K. 2015).

Inactivation of *Mgat1* in mouse was carried out by disrupting the gene via HR (Ioffe *et al.*, 1994; Metzler *et al.*, 1994). However, while *Mgat1* was not required for the cultured cell viability, mouse *Mgat1* knockouts were embryonic lethal. Mutations of genes involved in the *N*-glycosylation pathway acting upstream *Mgat1* were also embryonic lethal, revealing an essential role for hybrid and complex *N*-glycans in mouse development (Stanley, P. 2016).

GnTI knockout in other animal models such as the worm *Caenorhabditis elegans* and the housefly *Drosophila melanogaster* also reduces viability (Schachter, H. 2010).

3.3.3.2. Plants

Inactivation of *GnTI* in plants has been extensively carried out. Besides the fundamental interest, it allows for *N*-glycan simplification of recombinant glycoproteins but also the suppression of all non-human residues since they are found on complex *N*-glycans. Here is an exhaustive description of all *GnTI*-deficient plant species generated.

3.3.3.2.1. *A. thaliana*

The first plant *GnTI* mutants were isolated in *A. thaliana* from a pool of random mutagenized seeds (von Schaewen *et al.*, 1993). The *complex glycan-less1 allele 1 (cgl1-1)* mutant lacks *GnTI* activity and cDNA sequencing identified the point mutation D144N, which introduced an extra *N*-glycosylation site impairing *GnTI* folding (Strasser *et al.*, 2005; Frank *et al.*, 2008). Early reports of the *cgl1-1* mutant did not detect any complex *N*-glycans on glycoproteins, with the predominant type of glycans being Man5 (75%) (Strasser *et al.*, 2005). However, the *cgl1-1* mutant enzyme recovers its activity when the wrong *N*-glycosylation site is skipped (e.g., under stress, at certain developmental stages, or in the *stt3a* background, for the latter see Frank *et al.*, 2008). *A. thaliana cgl1-1* plant growth and morphology were not strongly affected under normal conditions but were more impacted by salt stress than WT plants (Kang *et al.*, 2008). A decrease in photosynthetic capacity of the *cgl1-1* mutant was also reported (Jiao *et al.*, 2020). The *cgl1-2* mutant (translational frameshift) also lacked *GnTI* activity (von Schaewen *et al.*, 1993; Frank *et al.*, 2008). Root hair length was increased in *Cgl1-2* plants and this phenotype was linked to altered hormone homeostasis (Frank *et al.*, 2021). Analysis of the activation associated secretory protein 1 produced in *cgl1-2* seeds showed only high-mannose glycans (Piron *et al.*, 2015), suggesting a total loss of *GnTI* activity. Many enzymes used as therapeutics for lysosomal storage diseases require phosphorylated high-mannose *N*-glycans for their targeting to the lysosomes (Grubb *et al.*, 2010). Hence, the human lysosomal enzymes glucocerebrosidase and α -L-iduronidase were produced in *cgl1-1* seeds and showed a large proportion of high-mannose-type *N*-glycans (85–94%), predominantly Man5. Nonetheless, significant amounts of MMXF (4–9%), Man5F (0.4–3%), GnMXF (0–3.1%) or MMX (0–1.3%) on specific *N*-sites were also produced, suggesting the presence of a residual *GnTI* activity in the seeds of *cgl1-1* plants (He *et al.*, 2012 and 2013; Pierce *et al.*, 2017).

3.3.3.2.2. *O. sativa*

The rice *gnt1* mutant line (SAI3G12) contains a T-DNA insertion in the sixth exon/intron junction of *GnTI* leading to the absence of native *GnTI* transcripts (Fanata *et al.*, 2013). As a result, only high-mannose *N*-glycans were identified. In contrast to *A. thaliana cgl1* mutants, rice *gnt1* plants displayed a severe growth phenotype, resulting in early lethality under normal culture conditions. These plants had a reduced cell wall thickness and cellulose content and were insensitive to cytokinin signaling. A rice suspension line was derived from a *gnt1* plant and used to produce the recombinant human acid α - or β -GCS. Only mannosylated *N*-glycans were detected on the two recombinant proteins secreted in the culture medium, with

a high proportion of Man5 (58% for α -GCS and 80% for β -GCS) (Jung *et al.*, 2017, 2019).

3.3.3.2.3. *L. japonicus*

Several mutants in key enzymes of the *N*-glycosylation pathway were selected from the *L. japonicus* retrotransposon 1 insertion population (Pedersen *et al.*, 2017). The *Ljgnt I* mutated plants had a severely altered growth phenotype, as they predominantly died before the flowering stage. Minimizing the biotic and abiotic stress increased the number of *Ljgnt I* mutants entering the reproduction stage. These results confirmed the differences in tolerance to *GnTI* inactivation between plant species and highlighted the implication of complex *N*-glycans in stress resistance.

3.3.3.2.4. *P. patens*

GnTI was disrupted in the moss *P. patens* by HR. However, no alteration in the *N*-glycan repertoire could be identified in the *GnTI* knockout compared to WT moss (Koprivova *et al.*, 2003). This surprising result was probably associated with the presence of a second *GnTI* allele. Indeed, later on, *GnTI* was successfully knocked out in *P. patens* protoplasts via PEG-mediated transformation with a knockout construct (Shen *et al.*, 2016). *P. patens* *GnTI*-knockout plants did not synthesize hybrid and complex *N*-glycans and were used to produce recombinant α -galactosidase A against Pompe disease with mainly Man5 *N*-glycans (Shen *et al.*, 2016; Hintze *et al.*, 2020).

3.3.3.2.5. *Solanum tuberosum* (potato)

Reduction of *GnTI* activity was obtained in the Solanaceae *S. tuberosum* by gene silencing using an antisense technology. Transgenic plants showed a substantial reduction of complex *N*-glycans and normal growth phenotype (Wenderoth and von Schaewen, 2000).

3.3.3.2.6. *Solanum lycopersicum* (tomato)

Originally created to investigate the ability of β 1,2-xylose and α 1,3-fucose to elicit a food allergy, tomato plants depleted for these residues were generated using RNAi targeting *GnTI* (Kaulfürst-Soboll *et al.*, 2011). *GnTI* knockdown plants had a severe phenotype consisting of early fruit drop, incomplete ripening and necrotic fruit-attached stalks, highlighting a role for complex *N*-glycans in fruit development (Kaulfürst-Soboll *et al.*, 2021).

3.3.3.2.7. *N. benthamiana*

GnTI knockdown in *N. benthamiana* plants were generated in two different studies by RNAi-mediated silencing. In the first one, *GnTI* activity was below the detection limit (3%) in one line but sufficient for the synthesis of complex *N*-glycans in

amounts comparable to WT plants as no impact on the *N*-glycan repertoire was observed. Complex *N*-glycans still accounted for 80% of total *N*-glycans (Strasser *et al.*, 2004a). It was hypothesized that low *GnTI* expression and activity were sufficient to maintain complex *N*-glycans synthesis. On the other hand, transgenic plants showing a strong reduction in complex *N*-glycans abundance (with only 9% of all *N*-glycan structures as compared to 89% in WT plants) were obtained in the second study (Limkul *et al.*, 2016). No altered growth phenotype was observed in those plants when compared with WT plants. Glycoproteomics analysis of the purified glucocerebrosidase expressed in this mutant background showed a large proportion of high-mannose-type *N*-glycans (73–85%), predominantly Man5, on the four *N*-sites, but also the presence of significant amounts of paucimannose *N*-glycans with β 1,2-xylose and α 1,3-fucose residues, suggesting a residual *GnTI* activity (Limkul *et al.*, 2015, 2016). These results clearly indicate that it is possible to deplete complex *N*-glycans by knocking down *GnTI* in *N. benthamiana*. It would be interesting to determine if *N. benthamiana* *GnTI* knockout plants are viable. Later on, root culture from *GnTI*-down regulated *N. benthamiana* plants were generated and used to produce β -glucocerebrosidase with mainly oligomannose *N*-glycans (Uthailak *et al.*, 2022).

3.3.3.2.8. *N. tabacum*

NtGnTI was first identified by Strasser *et al.* (1999) and the presence of two isoforms in both *N. tabacum* and *N. benthamiana*, due to the amphidiploid nature of these two genomes was confirmed. Reduction of *NtGnTI* activity in *N. tabacum* was obtained by gene silencing using the antisense technology and evaluated by immunoblot analysis using a complex glycan antiserum on leaf extracts. Transformants showed a substantial reduction in complex glycans compared with untransformed controls but this reduction varied according to the leaf age and developmental stage (Wenderoth and von Schaewen, 2000). Normal growth phenotype was observed but, again, a knockout *N. tabacum* line should be obtained to draw definitive conclusions on the importance of complex *N*-glycans in higher plants development.

In conclusion, *GnTI* knockdowns/outs are depleted for complex *N*-glycans, which results either in strongly affected growth and development or in defects only observed in stress conditions, depending on the plant species. This is surprising because the *N*-glycosylation pathway is highly similar between most plant species. However, *N*-glycosylation defects are generally pleiotropic and affect numerous glycoproteins (Strasser, F. 2022). Consequently, the observed phenotype in *GnTI* mutants could arise from several different glycoproteins that are dysfunctional in the absence of complex *N*-glycans. On the other hand, since Le^a structures –which are far more abundant in rice- play a role in cell-to-cell recognition in higher plants, their

suppression via inactivation of *GnTI* could explain the strong phenotype observed. Complex *N*-glycans have also been linked to plant immunity and control of trafficking through the Golgi (Strasser, 2022).

3.3.4. *GnTII* knockouts

Arabidopsis *Gnt2* T-DNA insertion mutants do not produce any complex *N*-glycans, due to the absence of elongation of the α 1,6-mannose arm with a GlcNAc (Yoo *et al.*, 2021). The growth and development of these mutants are strongly affected, but only under stress conditions.

3.3.5. *In vivo* deglycosylation of *N*-glycans

While glycosylation is a critical feature for some recombinant glycoproteins, the presence of *N*-glycans might be dispensable for others. The proteins whose properties are not significantly affected by the presence of *N*-glycans could benefit from a glycosylation shortening strategy. Indeed, removal of *N*-glycans suppresses the presence of non-human residues, increases homogeneity and simplifies downstream processing (Van Landuyt *et al.*, 2019).

In addition, the absence of *N*-glycan structures can confer useful properties to glycoproteins, even if their native form is glycosylated. Besides the gain in homogeneity, deglycosylated IgGs can be desirable for applications that do not require Fc effector functions because these can cause undesired inflammation and cytotoxicity (Ju and Jung, 2014). Moreover, deglycosylated IgGs are a promising class of therapeutics for autoimmune diseases such as arthritis because they display a dominant suppressive effect on immune complex-mediated inflammation (Nandakumar *et al.*, 2013). Non-glycosylated IgGs can also be useful in diagnostic procedures such as *in vitro* immunoassays since Fc glycosylation might interfere with the detection of aberrant glycoforms of a disease biomarker (Lee *et al.*, 2020). Vaccine antigens can also benefit from the suppression of large *N*-glycans. Pathogens, especially viruses, use an immune evasion strategy called glycan shielding which consists of masking conserved surface antigen epitopes with glycans to avoid detection by the host immune system (Schön *et al.*, 2021).

Mice immunization studies evaluating the potential of viral antigens with shortened (or no) *N*-glycosylation showed promising results. *In vitro* deglycosylated influenza hemagglutinin elicits higher cross-strain immune responses, stronger neutralization and higher protective effect than the fully glycosylated form produced in HEK 293 (Wang *et al.*, 2009; Chen *et al.*, 2014a). Partially deglycosylated SARS-CoV spike protein RBD elicits stronger specific antibody responses and a higher level of

neutralizing antibodies (Chen *et al.*, 2014b). Deglycosylation of SARS-CoV2 spike protein also results in stronger immune responses as well as better and broader protection against variants of concern (Huang *et al.*, 2022; Wu *et al.*, 2022).

In vitro enzymatic deglycosylation is a relevant strategy to produce deglycosylated proteins for small-scale studies. However, because of the high cost of these enzymes, it is not suitable for industrial production. To avoid expensive *in vitro* steps, the most straightforward approach is to remove N-sites by substituting Asn with Gln (Dubrovskaja *et al.*, 2017; Wu *et al.*, 2022). However, suppression of N-glycosylation sites by mutagenesis often results in conformational changes. Indeed, besides the amino acid substitution itself, the addition of N-glycan precursors is required for correct protein folding in the ER. Alternatively, recombinant glycoproteins can be deglycosylated *in vivo*. Transient co-expression in *N. benthamiana* leaves of *Flaviobacterium meningosepticum* peptide-N-glycosidase F (PNGase F) with antibodies or antigenic proteins from *Plasmodium falciparum* and *Bacillus anthracis* resulted in the production of deglycosylated recombinant proteins (Mamedov *et al.*, 2012). However, due to the PNGase F substrate specificity (N-glycans without core α 1,3-fucose residues), all proteins (PNGase F and recombinant proteins) had to be fused with the KDEL ER-retention motif to avoid the formation of complex N-glycans that would be PNGase F-resistant. Besides the impact on protein accumulation and storage, the drawback of this strategy is that PNGase F deaminates the Asn residue to Asp, hence modifying the peptide sequences.

Later on, protein antigens from *P. falciparum* and *B. anthracis* (Mamedov *et al.*, 2017), spike protein RBD from SARS-CoV-2 (Mamedov *et al.*, 2021) and the anti-programmed cell death protein-1 antibody nivolumab used in immunotherapy (Kang *et al.*, 2022) were transiently co-expressed with Endo- β -N-acetylglucosaminidase H (EndoH) of *Streptomyces plicatus* in *N. benthamiana* leaves. EndoH hydrolyzes the GlcNAc-GlcNAc bound of high mannose N-glycans, leaving a single GlcNAc on the Asn residue which might be important to maintain the overall protein fold. However, EndoH substrate specificity still required the retention of all recombinant proteins in the ER. PNGase F- and EndoH-*in vivo* deglycosylated antigens were more stable at elevated temperatures, compared to their aberrantly glycosylated counterparts (Mamedov *et al.*, 2017; Mamedov *et al.*, 2019). Moreover, the deglycosylated *P. falciparum* antigen was better recognized by a conformational antibody (Mamedov *et al.*, 2017) and had a stronger activity as a transmission-blocking vaccine against malaria (Mamedov *et al.*, 2019). Deglycosylated SARS-CoV-2 RBD showed specific binding to the SARS-CoV-2 receptor ACE2, induced a potent neutralizing response in mice, resulting in sera with greater neutralizing

activity against SARS-CoV-2 than the glycosylated variant (Mamedov *et al.*, 2021). The immune checkpoint inhibitory function of deglycosylated nivolumab antibody was unaffected while the ADCC, which causes the death of T cells with anticancer immune activity, was reduced (Kang *et al.*, 2022).

To circumvent the limitation of fusing the C-terminus of proteins with an ER-retention signal, it was proposed to express an endo- β -*N*-acetylglucosaminidase in complex glycan-deficient hosts. This strategy was first achieved by expressing the endo- β -*N*-acetylglucosaminidase T (EndoT) from the fungus *Hypocrea jecorina* (which displays similar substrate specificity as EndoH) in the GnTI-inactivated human embryonic kidney (HEK293S *Mgat1*) cell line that produces only high mannose *N*-glycans (Meuris *et al.*, 2014). This system is called GlycoDelete®. *N*-glycans of the GlycoDelete cell line were shortened down to a single GlcNAc, which can be further galactosylated and sialylated, resulting in the production of recombinant glycoproteins with nearly uniform, small *N*-glycans. This elegant and effective GlycoDelete strategy was then successfully implemented in a *A. thaliana cgl1* mutant lacking GnTI activity (Piron *et al.*, 2015). The GlycoDelete *A. thaliana* plant was obtained by expressing the EndoT enzyme specifically in the seeds, thus allowing convenient accumulation and storage of recombinant proteins without impacting plant growth. EndoT was targeted to the medial-Golgi membrane by fusing its catalytic region with the CTS domains of At β 1,2-xylosyltransferase. Expression of activation associated secretory protein 1 from the helminth *Ostertagia ostertagi* in the GlycoDelete seeds showed that the two *N*-glycans had been completely trimmed to a single GlcNAc (Piron *et al.*, 2015). The GlycoDelete technology has also been recently validated in the yeast *Pichia pastoris* (Wang *et al.*, 2020), the filamentous fungus *Aspergillus oryzae* (Li *et al.*, 2020) and CHO cells (Kang *et al.*, 2022). Kang and colleagues also report that their attempts to obtain GlycoDelete *A. thaliana* plants failed, suggesting a lethal effect of this strategy in that organism.

4. CRISPR/Cas9 genome editing tool

To knock down/out plant GT genes described in the previous chapter, four main techniques were used: chemical/radiation/transposon random mutagenesis, RNAi, HR, and CRISPR/Cas9. In this study, we aim at obtaining a complete inactivation of different GTs activity in BY-2 cells. Therefore, RNAi is not a suitable technique because gene silencing is rarely complete and might not be stable over time. Likewise, HR is not considered. Indeed, it is a highly efficient integration technique in *P. patens* (up to 100% HR frequency) but not in flowering plants (10^{-4} - 10^{-5} HR frequency) (Kamisugi *et al.*, 2006). Finally, generating a mutant library by random mutagenesis is highly time-consuming and laborious while the CRISPR/Cas9 technique is relatively quick, precise and straightforward.

CRISPR/Cas9 technology comes from the type II CRISPR/Cas system, a bacterial and archaean adaptive immunity system against viruses. This system comprises two main components: clustered regularly interspaced short palindromic repeats (CRISPR) and the *cas* (CRISPR-associated) genes (Figure 12A). The CRISPR is composed of identical palindromic repeats and unique noncoding short RNAs called spacers (Doudna and Charpentier, 2014). This system uses antisense RNAs as memory of past phage invasions. It is a three-step process: adaptation, expression and interference. The adaptation phase consists of the acquisition of new spacers during a phage infection and their insertion into the CRISPR array, between two repeats. These spacers originate from short sequences in the invading DNA (called protospacers). Next, during the expression phase, the *cas* operon as well as the CRISPR array are transcribed. Transcription of the CRISPR array generates a unique pre-CRISPR RNA (pre-crRNA) (Figure 12B). Then, small trans-activating crRNAs (tracrRNAs) hybridize with the repeats in the pre-crRNA and activate the maturation by Cas9 and ribonuclease III. Individual crRNAs are generated from the pre-crRNA, each composed of a repeat and a spacer. Finally, Cas9 associates with the tracrRNA:crRNA complex to form a ribonucleoprotein (RNP) (Figure 12C). The spacer directs Cas9 to cleave complementary DNA, hence tackling foreign nucleic acid (Demirci *et al.*, 2017).

Each protospacer is associated with a protospacer adjacent motif (PAM). Cas9 needs a PAM sequence upstream of the crRNA-hybridization site to successfully bind to the target DNA for an effective cleavage. As a result, Cas9 RNP is able to recognize and cleave protospacers (viral DNA) but not spacers in the CRISPR cluster in its own genome. As such, the PAM plays an important role during the adaptation and interference steps to discriminate self and non-self DNA (Jinek *et al.*, 2012). Cas9

from different bacterial species need different PAM. The most widely used Cas9 originates from *Streptococcus pyogenes* and its PAM is 5'NGG3'. Once the Cas9 RNP is associated with a complementary target sequence downstream of a PAM, the Cas9 nuclease domains RuvC and HNH each cleave one of the viral DNA strands, creating a double-strand break (DSB) (Figure 12C) (Doudna and Charpentier, 2014).

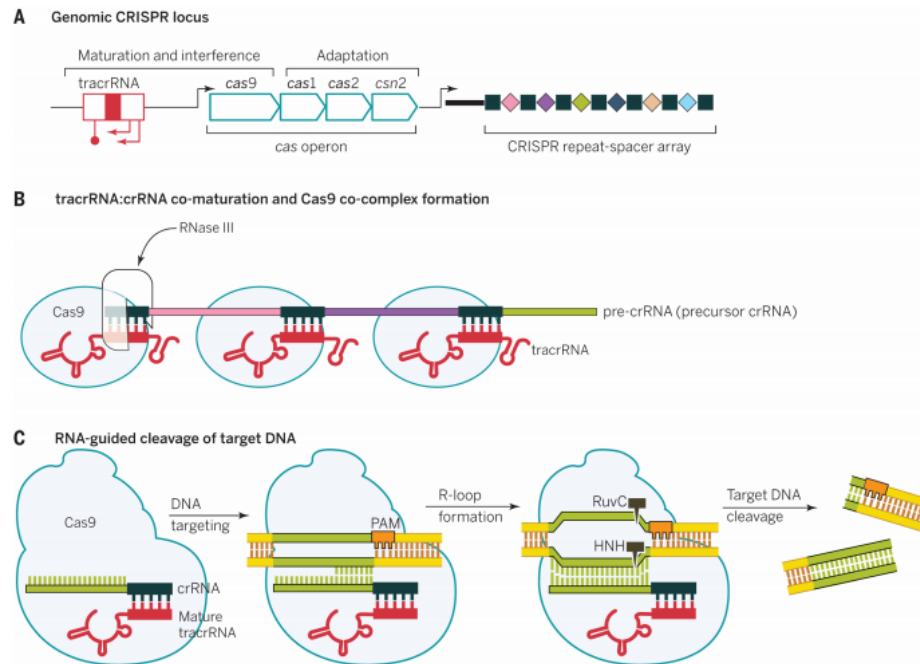


Figure 12. Biology of the type II-A CRISPR/Cas system. The system from *S. pyogenes* is shown as an example. (A) The *cas* gene operon with *tracrRNA* and the CRISPR array. (B) The transcription of the CRISPR array generates a pre-crRNA, which forms duplexes with *tracrRNAs*. Interaction with Cas9 and processing by ribonuclease III generates crRNAs from the pre-crRNA. (C) *tracrRNA:crRNA* associates with Cas9 and guides the cleavage of target DNA (Doudna and Charpentier, 2014).

The CRISPR/Cas9 system has been adapted to be used as a genome-engineering tool. In this setting, the targeted DNA is in the host genome and the spacers are tailored designed and introduced alongside *cas9*. The *tracrRNA:crRNA* guide complex has been simplified into a single guide RNA (sgRNA) by fusing the two RNAs. sgRNAs are composed of the targeted sequence (~20 bp) and the scaffold (117 bp) corresponding to the *tracrRNA* (Demirci *et al.*, 2017).

The usage of promoters for sgRNA and *cas9* expression impacts the efficiency of the tool. In plants, RNA polymerase II-dependent promoters such as the cauliflower mosaic virus (CaMV) 35S promoter are used for efficient *cas9* expression. Indeed, Pol II polymerases synthesize mRNAs and are typically used for heterologous protein expression in plants. On the other hand, RNA polymerase III-dependent promoters such as the U6 promoter are used for sgRNA expression. Pol III polymerases uniquely transcribe small non-coding RNAs, which lack a 5' cap and a 3' polyadenylated tail (Gao *et al.*, 2018; Zhang *et al.*, 2019).

Once a DSB is generated into the plant genome by Cas9 RNP, it can be repaired by the endogenous non-homologous end joining pathway (NHEJ) or homology directed repair (HDR). The most common form of HDR is HR, in which the sequence information interrupted by a DSB is faithfully restored by referencing the intact homologous DNA molecule as a template for repair (Spies *et al.*, 2018). NHEJ is an error-prone process which leads to the formation of small insertions/deletions (indels). When targeting a gene, these indels are very likely to inactivate the corresponding protein either by inducing a frameshift leading to the formation of premature STOP codons or mutating residues that are important for its activity or folding (Figure 13). On the other hand, if the DSB is resolved via the HDR pathway, a homologous DNA sequence on the sister chromatid is used to precisely guide the repair. As a result, no alteration is generated around the DSB and the initial sequence is restored. Alternatively, precise mutations can be obtained by HR, by introducing a mutated homologous sequence, which will serve as a repair template, simultaneously with the CRISPR/Cas9 components. After NHEJ, indels prevent a new hybridization with the sgRNA. On the other hand, after HDR, Cas9 RNP is still able to induce DSB since the sequence is not altered. Hence, the formation of indels is favored whatever the rate of HDR/NHEJ. This system has also been used for site-specific knockin of genes. To this end, a Cas is co-introduced with DNA fragments containing desired sequences to function as HDR templates for the Cas-induced DSBs. This technique is also known as gene targeting (Zhang *et al.*, 2019).

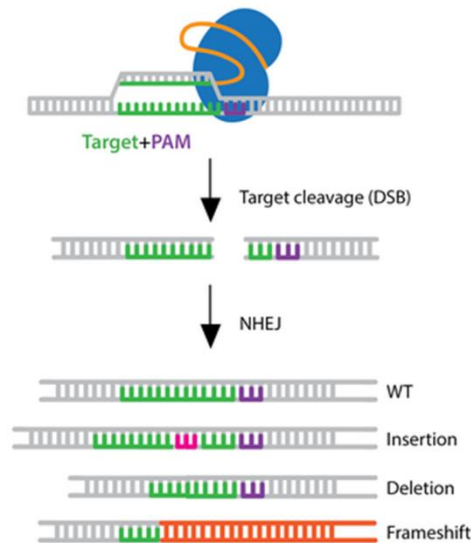


Figure 13. Possible outcomes for a DNA sequence upon repair of a CRISPR-induced DSB by the NHEJ pathway (Adapted from Addgene.com).

With this system, it is theoretically possible to cleave any site in the genome, provided that there is a PAM upstream. Previously used site-specific nucleases such as zinc finger nucleases (ZFN) or Tale effector-like nucleases (TALEN) rely on protein-DNA interactions for sequence recognition. As a result, to target a new site, complicated and time-consuming protein design is required. This is why CRISPR/Cas9 is such a powerful tool. Targeting a new site only requires to change 20 bp in the sgRNA.

However, this is only the tip of the CRISPR/Cas9 iceberg. Indeed, the CRISPR/Cas toolbox is expanding every day with new applications. First, Cas9 orthologues from other bacterial and archaeal species have been used to take advantage of their alternative PAM recognition or smaller size for virus-based delivery (Zhang *et al.*, 2019). Furthermore, genetic engineering of Cas9 led to many variants with new or altered features. Engineering of the PAM-interacting domain produced many Cas9 variants with different PAM requirements. SpCas9 has been engineered to recognize NGA, NGAG, NGCG and even NG PAMs (Zhang *et al.*, 2019). This is of interest to target AT-rich regions or to introduce indels at a very precise location.

Introduction of point mutations to inactive RuvC or HNH domains created the Cas9 nickase (nCas9) that only cleaves one strand. The combined use of a nCas9 with a pair of gRNAs targeting opposite DNA strands in close proximity generates a staggered DSB with overhangs. Since two nCas9 must recognize the target site simultaneously, this strategy strongly reduces the risk for off-target activity (Gopalappa *et al.*, 2018). When both domains are mutated, the nuclease activity is suppressed, resulting in a dead Cas9 (dCas9). dCas9 RNP generates a DNA recognition complex that can specifically interfere with transcriptional elongation, RNA polymerase binding, or transcription factor binding (Qi *et al.*, 2013). Thus, it can be used either for promoting or suppressing gene transcription.

By using tissue-specific promoters to drive *cas9* expression, it is possible to inactivate genes in a tissue-specific manner. This is useful to study the function of essential genes whose ubiquitous knockout is lethal or prevents the generation of homozygous mutants (Lockhart, 2019).

CRISPR/Cas9 can also be used to target multiple genes at once. To this end, one can simply clone and integrate several gRNAs, each being driven by its own promoter. However, more ingenious strategies have been designed to ease the cloning steps and reduce the size of the T-DNA. One of the most efficient and straightforward multiplex editing systems is taking advantage of endogenous transfer RNA (tRNA)-processing enzymes (Xie *et al.*, 2015). In this system, all gRNAs are put under the control of a single Pol III promoter and tRNAs are inserted between gRNAs (Figure 14A). As a result, a polycistronic tRNA-gRNA (PTG) transcript is obtained (Figure 14B). Endogenous P and Z RNases recognize the tRNAs and cleave them at both sides. These cleavages release the gRNAs that can individually associate with Cas9 and target their respective DNA site (Figure 14C). This system was used, *inter alia*, to knock out four or five *at1,3-FucT* genes and two *XylT* genes in *N. benthamiana* plants (Jansing *et al.*, 2019), *N. tabacum* plants (Göritzner *et al.*, 2022) and BY-2 cells (Mercx *et al.*, 2017).

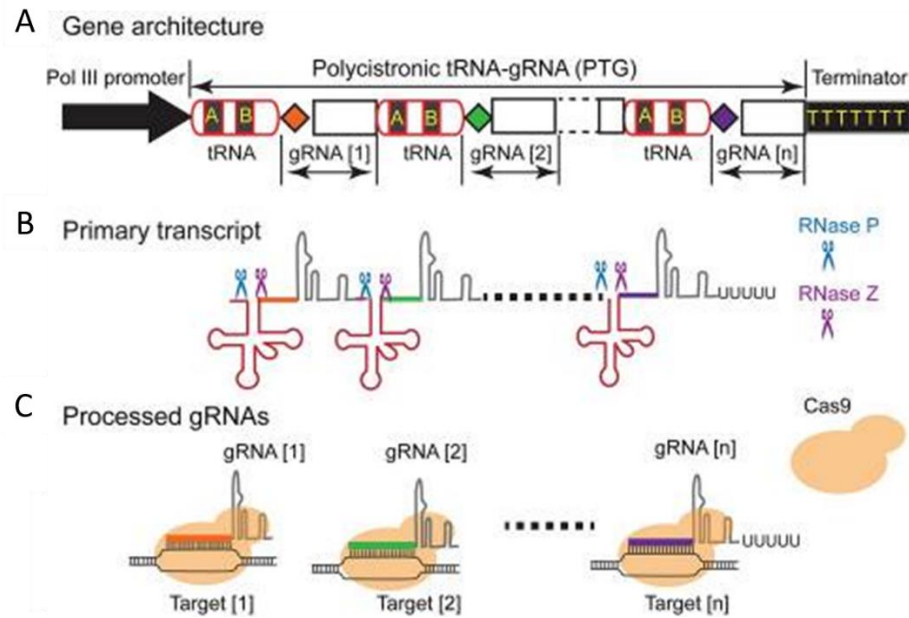


Figure 14. Schematic depiction of the PTG/Cas9 method for simultaneously targeting multiple sites. (A) The synthetic PTG consists of tandemly arrayed tRNA-gRNA units. The tRNA (red) contains A and B elements, responsible for the Pol III complex recruitment. Each gRNA consists of a target-specific spacer (diamond) and the conserved gRNA scaffold (rectangle). (B) Endogenous RNase P and RNase Z (scissors) cleave the primary transcript of PTG, which releases mature gRNAs and tRNAs (red structures). (C) Each mature gRNA directs Cas9 to the spacer-specific target(s) (Xie *et al.*, 2015).

5. Generation of marker-free transgenic plants

5.1. Selectable marker genes used for plant transformation

The generation of transgenic plants or plant cells is a key tool for basic research as well as for applications ranging from molecular farming to crop improvement. It is a necessary process to obtain plant or plant cells stably expressing a recombinant (glyco)protein. However, the transformation step is critical because of its low efficiency for most plant species. As a result, plant transformation usually involves the use of selectable marker genes to isolate transformed cells from non-transformed cells. Markers are usually antibiotic or herbicide resistance genes.

Selectable marker genes can be classified in different groups. They are called positive if their presence in the plant genome allows the growth on selective medium. If their presence in the plant genome results in the inability to grow on selective medium, they are negative markers. They are conditional if their mechanism of action requires the addition of a substrate, otherwise non-conditional. Different conditional positive markers have been successfully used for plant transformation, some of the most prominent are listed in Table 1.

Table 1. Selectable markers commonly used in transgenic plants (Adapted from Barampuram and Zhang, 2011).

Gene	Enzymes encoded	Substrate	Product	Gene source
<i>bar</i>	Phosphinothricin acetyl transferase	Phosphinothricin	N-acetylphosphinothricin	<i>Streptomyces hygroscopicus</i>
<i>BADH</i>	Betaine aldehyde dehydrogenase	Betaine aldehyde	Glycin betaine	<i>Spinacia oleracea</i>
<i>bxn</i>	Bromoxynil nitrilase	Bromoxynil (dibromo-hydroxybenzonitrile) and other oxynils	dibromo-hydroxybenzoic acid	<i>Klebsiella pneumonia</i>
<i>cat</i>	Chloramphenicol acetyl transferase	Chloramphenicol	Chloramphenicol acetate	<i>Escherichia coli</i> Tn5
<i>dhfr</i>	Dihydrofolate reductase	Methotrexate	NA (methotrexate is a plant DHFR inhibitor)	<i>Candida albicans</i>
<i>EPSPS</i>	5-Enolpyruvyl shikimate-3-phosphate synthase	Glyphosate	5-Enolpyruvyl shikimate-3-phosphate	<i>Petunia×hybrida</i>
<i>gox</i>	Glyphosate oxidoreductase	Glyphosate	Aminomethyl phosphonic acid + glyoxylate	<i>Ochrobactrum anthropi</i>
<i>hpt II</i>	Hygromycin phosphotransferase II	Hygromycin B	Phospho-hygromycin B	<i>E. coli</i>
<i>ManA</i>	Phosphomannose isomerase	D-Mannose-6-P	D-Fructose-6-P	<i>E. coli</i>
<i>npt II</i>	Neomycin phosphotransferase II	Kanamycin	Kanamycin 3'-phosphate	<i>E. coli</i> Tn5
<i>xylA</i>	Xylose isomerase	D-Xylose	D-Xylulose	<i>Streptomyces rubiginosus</i>

Conditional negative selectable marker genes are useful tools to identify transformed plants that have lost unwanted transgenes. By placing a negative marker and a positive marker in the same excisable cassette, the excision events can be selected by growing the transformants on medium supplemented with the substrate of the conditional negative marker. Transformants in which no excision occurred will not be able to grow in the presence of the negative selection substrate. The most widely used plant conditional negative markers are listed in Table 2.

Table 2. Chemicals and enzymes for the conditional negative selection of transgenic plant tissues (Adapted from Miki and McHugh, 2004).

Gene	Enzyme encoded	Substrate	Product	Gene source
<i>codA</i>	Cytosine deaminase	5-Fluorocytosine	5-Fluorouracil	<i>Escherichia coli</i>
<i>aux2</i>	Amido hydrolase	Naphthalene acetamide	Naphthalene acetic acid	<i>Agrobacterium rhizogenes</i>
<i>tms 2</i>	Indoleacetic acid hydrolase	Indole-3-acetamide	Indole-3-acetic acid	<i>Agrobacterium tumefaciens</i>
<i>dhlA</i>	Dehalogenase	Dihaloalkanes	Halogenated alcohols	<i>Xanthobacter autotrophicus</i>
<i>CYP105A</i>	Cytochrome P450 mono-oxygenase	Sulfonylurea R7402	Phytotoxic R7402 metabolite	<i>Streptomyces griseolus</i>
<i>cue</i>	Alcohol dehydrogenase	Allyl alcohol	Acrolein	<i>Arabidopsis thaliana</i>

Cytosine deaminase A (*codA*) from *E. coli* can be used as a conditional negative marker in higher eukaryotes. In *E. coli*, there are two pyrimidine biosynthesis pathways: the *de novo* pathway, starting from glutamine, and the salvage pathway, starting from cytosine (Figure 15) (Stougaard, 1993). CodA catalyzes the first step of the salvage pathway: the deamination of cytosine into uracil. It can also catalyze the deamination of the harmless compound 5-fluorocytosine (5-FC) into 5-fluorouracil (5-FU). 5-FU is then converted to 5-fluorodeoxyuridine monophosphate which irreversibly inhibits thymidylate synthase activity and thus DNA synthesis. In addition, 5-FU can also be metabolized into fluorodeoxyuridine triphosphate and fluorouridine triphosphate that are incorporated into DNA and RNA, respectively, preventing the incorporation of pyrimidine bases leading to cell death (Miura *et al.*, 2010).

The salvage pathway is found in bacteria and fungi, but is not present in mammals or plants, which only use the *de novo* pathway. Hence, CodA has no homologue in plants, which cannot naturally convert 5-FC to 5-FU, allowing the use of *codA* as a conditional negative marker (Stougaard, 1993). The addition of *codA* in the excisable fragment of a T-DNA allows selecting excision events in the presence of 5-FC. Indeed, transgenic plants or plant cells in which excision did not occur will metabolize 5-FC into 5-FU through CodA activity (Figure 15). Inhibition of DNA synthesis by 5-FU will result in cell death. *codA* was successfully used as a negative marker gene in different species including tobacco plants (Schlaman and Hooykaas, 1997; Stougaard, 1993) and BY-2 cells (Sheva *et al.*, 2020).

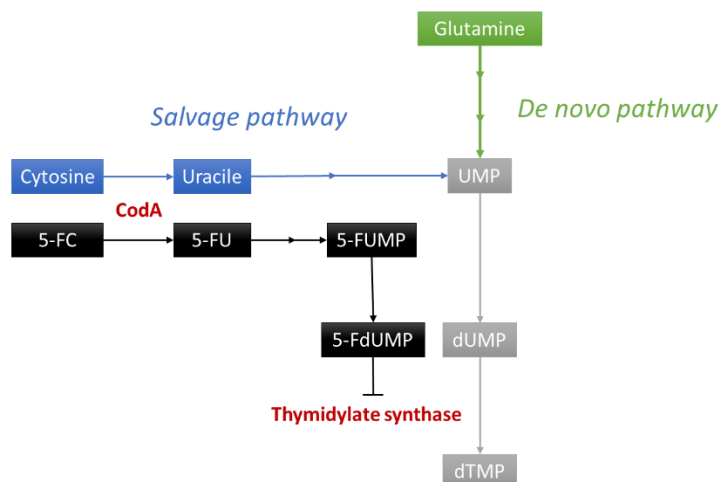


Figure 15. Schematic outline of the salvage (blue) and *de novo* (green) pathways for pyrimidine biosynthesis. CodA-mediated metabolization of 5-FC in transgenic plants and its main effect is represented in black. Enzymes are represented in red. UMP, uridine monophosphate; dUMP, deoxyuridine monophosphate; dTMP, deoxythymidine monophosphate; 5-FUMP, 5-fluorouridine monophosphate; 5-FdUMP, 5-fluorodeoxyuridine monophosphate.

Since most fungi express cytosine deaminase, 5-FC is commonly used as an antifungal agent (Ireton *et al.*, 2003). It also has potential for gene therapy applications against solid tumors (Lv *et al.*, 2008).

5.2. Why remove selectable marker genes?

After selection, marker genes are not required anymore and might raise public and regulatory concerns based on the fear that transgenic crops possessing an antibiotic resistance gene could undergo horizontal transfer with microorganisms and spread antibiotic resistance in pathogens. Likewise, the concern of an escape of herbicide resistance genes to wild relatives through pollen also justifies the interest in transgene removal techniques (Chong-Pérez and Angenon, 2013).

In addition to these concerns, marker gene removal can be motivated by practical reasons. The development of transgenic lines may involve successive transformation events and thus requests the use of different markers. However, few efficient marker genes are available for plant selection. A study carried out on 396 scientific papers revealed that kanamycin, hygromycin and phosphinotricin resistance genes

represented 96% of the markers used in plant transformation (Miki and McHugh, 2004). The availability of markers also depends on the plant species. Their removal allows to re-use the same marker gene in successive transformation events of transgenic plants with new transgenes.

Selectable marker genes are not the only unwanted DNA in transgenic plants. Indeed, constitutively expressed DNA-modifying enzymes such as the Cas nucleases might also need to be removed. Indeed, once the target DNA has been modified by the CRISPR/Cas system, there is no use for gRNAs and Cas anymore. The longer they are produced, the more likely it is that off-target modifications will occur. Again, transgene excision should be recommended to stabilize these cell lines.

In the next chapter, we will go through different types of plant transgene excision systems, with a special emphasis on the *Cre-loxP* excision system that was used in this thesis.

5.3. Transgene removal techniques

5.3.1. Co-transformation

Co-transformation is one of the earliest methods set up to obtain a marker-free transgenic line. Briefly, the plant is co-transformed with the gene of interest (GOI) and the marker gene, delivered separately. For *Agrobacterium*-mediated transformation, the two expression cassettes can either be on two different binary vectors in the same *Agrobacterium* line or in two different T-DNA elements in the same binary vector. To co-deliver the two cassettes by biolistics, the gold particles are coated with a mixture of both genetic constructs. After selection, a large proportion of lines carrying the marker have also integrated the GOI (Darbani *et al.*, 2007). If integrated in unlinked chromosomal sites, the two cassettes can be segregated in the progeny after crossing with a WT plant. Then, F1 plants carrying the GOI but no marker gene can be screened for. However, co-transformed transgenes often integrate in the same locus, which impairs segregation. Moreover, genetic segregation is not suitable for suspension cells since it can only be applied to sexually reproducing plants because it relies on the separation of genes during meiosis (Matsunaga *et al.*, 2002).

5.3.2. Transposon-based systems

To avoid co-integration of the marker gene and the GOI, transposon-mediated repositioning of transgenes can be used. The maize Activator/Dissociation (Ac/Ds)

transposon system was used to generate marker-free transgenic rice (Gao *et al.*, 2015a; Cotsaftis *et al.*, 2002). First, the T-DNA containing the marker gene, Ac transposase and the GOI associated with the Ds elements is integrated in the plant genome (Figure 16). Then, Ac is expressed and recognizes the Ds elements flanking the GOI. The GOI is excised and reinserted somewhere else into the genome. The genetic distance between the reinserted Ds and its donor T-DNA must be large enough to recombine in germinal cells and segregate marker-free progeny. Again, this cannot be applied to plant suspension cells or vegetatively propagated plants. Other disadvantages include the imprecision of transposon excision and the repeated insertion-excision cycles, which may lead to the generation of mutations at numerous unknown loci. This may also lead to genomic instability in transgenic plants (Scutt *et al.*, 2002).

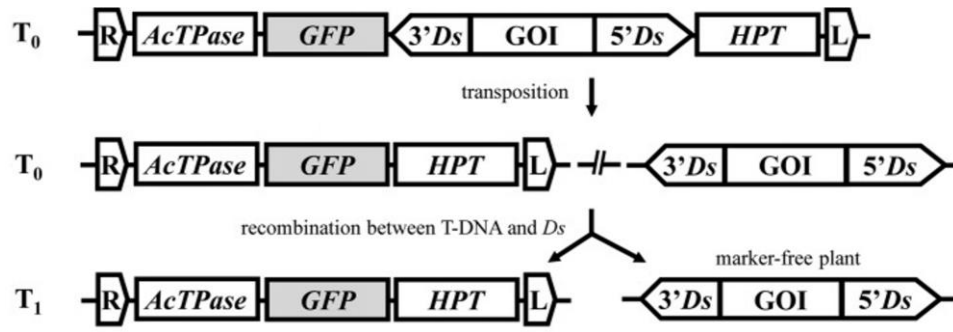


Figure 16. Ac/Ds-mediated transgene reintegration for the generation of marker-free transgenic plants. In the presence of Ac, the Ds-based GOI in transformed cells (T₀) is excised from the T-DNA and reintegrated into an unlinked chromosomal site. Recombination between transposed Ds and its donor T-DNA produces a marker-free T₁ plant (Gao *et al.*, 2015a). *AcTPase*, Ac transposase; *GFP*, green fluorescent protein; *HPT*, hygromycin phosphotransferase.

5.3.3. Excision via site-specific recombinases or nucleases

Expression of microbial site-specific nucleases or recombinases associated with the flanking of the marker gene with their corresponding recognition sequences is the most flexible method developed for transgene excision.

Recombination systems such as *Cre-loxP* from bacteriophage P1 (Cyclization recombination - locus of X-over of P1) or *FLP-FRT* from *S. cerevisiae* and *R-RS* from *Zygosaccharomyces rouxii* have been used for marker gene excision. A first transformation is carried out with the T-DNA containing the GOI and the flanked

marker gene. Next, re-transformation or sexual crossing is required to express the recombinase and precisely remove the marker gene by recombination, leaving only one recognition sequence (Scutt *et al.*, 2002).

Site-specific nucleases such as Cas nucleases, ZFN or TALEN can be used following the same strategy. However, in this case, the nucleases (or the gRNAs for Cas9) are adapted to recognize the flanking sites, not the opposite.

In its simplest form, this type of excision system also relies on transgene segregation to get rid of the second T-DNA, containing the recombinase or nuclease. Prolonged activity of recombinant recombinases and nucleases may also introduce off-target mutations. However, adaptations of these systems have been developed to tackle both concerns. This will be described for Cre-*loxP* and CRISPR/Cas9.

5.3.3.1. Cre-*loxP* recombination system

The Cre-*loxP* system consists of the 38.5 kDa Cre recombinase and the 34 bp DNA target on the bacteriophage P1 genome called *loxP* (Tronche *et al.*, 2002). Cre is a member of the integrase family of site-specific recombinases comprising proteins involved in integration and excision of viral and plasmid DNA, DNA excision, transposition and resolution of interlocked DNA circles in archae, bacteria and yeast (Nunes-Düby *et al.*, 1998).

5.3.3.1.1. Mechanism

During P1 lysogenic cycle, its nucleic acid is not integrated into the bacterium's chromosome. Instead, it persists in the cytosol as a single copy circular episome. During P1 episome duplication, two catenated dsDNA circles are produced. In order to separate them, Cre catalyzes the recombination between the two *loxP* sites in the phage episomes (Marintcheva, 2018).

The 34 bp of the *loxP* site can be further broken down to left and right inverted repeats of 13 bp, and an 8 bp spacer separating them (Figure 17A). The spacer is the recognition site of Cre and provides a direction on the recombination site. First, two Cre recombinases bind to the inverted repeats of one *loxP* site (Tronche *et al.*, 2002, Gilbertson *et al.*, 2003). Once two *loxP* sites are assembled in an antiparallel fashion, the four Cre recombinases catalyze recombination, forming two new *loxP* sites while exchanging strands (Figure 17B). If *loxP* sites are on the same DNA molecule, the portion comprised between them is termed « floxed ». The relative orientation of *loxP* sites will determine the outcome of the reaction. If the two *loxP* sites are in the same orientation, recombination will lead to excision of the floxed DNA. If they are

in the opposite orientation, the floxed sequence will be inverted. If the selectable marker gene is excised via Cre-catalyzed recombination, it will form a circular DNA fragment. This fragment can be integrated back in a site-specific manner due to the remaining presence of a *loxP* site. However, a deletion is more likely to occur as the two *loxP* sites are in close proximity as opposed to the reintegration reaction where the *loxP* sites must find each other again. Furthermore, the small excised fragment will be purged if not quickly reintegrated while reintegration is always reversible (Hare and Chua, 2002; Ow, D.W. 2016).

One of the main applications of Cre-*loxP* system is to create conditional knockout mice. The system first requires the conventional creation of two transgenic mouse lines: one expressing *cre* from an inducible promoter and the other coding for a floxed GOI. Both mouse lines are crossed and *cre* expression is induced at the required developmental stage, resulting in a conditional null mutant.

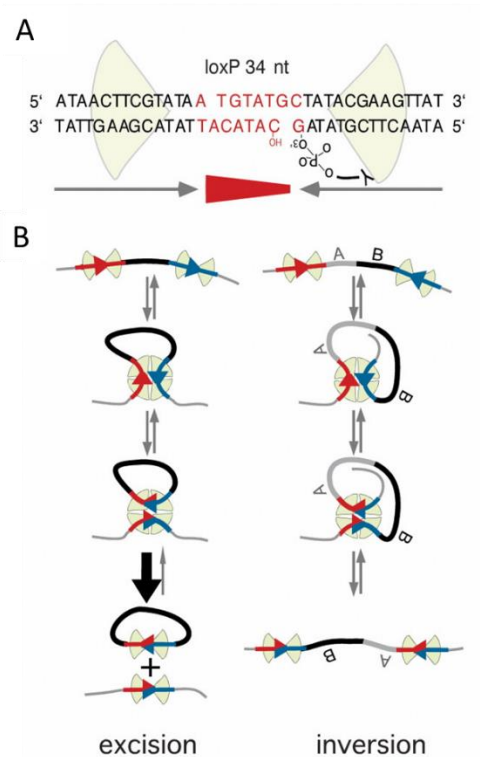


Figure 17. Mechanism of site-specific Cre-loxP recombination (Tronche *et al.*, 2002). (A) The loxP site consists of left and right inverted repeats of 13 bp, and an 8 bp spacer separating them. Two Cre recombinases bind to the inverted repeats of one loxP site. (B) Two loxP sites are assembled by four Cre recombinases. Cre then catalyzes the exchange and separation of the two dsDNA molecules. Depending on the relative orientation of the loxP sites, recombination will result in excision or inversion of the floxed sequence.

5.3.3.1.2. Cre-loxP system applied to plant biotechnology

Adaptation of the Cre-loxP system can produce site-directed excision and inversion of transgenes, integration of exogenous DNA into genomic recombination sites, and the rearrangement of chromosome segments (Ow, 1996). In this chapter, we will focus on transgene excision.

The most common use of Cre-mediated excision in plants is for marker gene removal. First, a transgenic plant expressing the GOI and a floxed marker gene is obtained (Figure 18A). In parallel, a plant expressing *cre* with another marker gene (not represented in Figure 18) is generated. The two plants are then crossed and Cre

triggers the excision of the first marker via recombination (Figure 18B). Meiotic segregation is then necessary to remove *cre* and the second marker (Figure 18C). This technique was used by Monsanto Company to create the LY038 maize line expressing *corda4A*, the key enzyme in lysine biosynthesis. This maize line contains increased lysine content and was designed for poultry and livestock farming (Sekan *et al.*, 2015). However, this is a lengthy process, not applicable to suspension cells.

An alternative strategy is the autoexcision. In this case, the GOI is encoded on the same DNA fragment as Cre and the marker, which are floxed (Figure 18D-C). *cre* expression is under the control of an inducible promoter so that recombination can be triggered after selection. Upon recombination, both the marker and *cre* are removed from the plant genome (Marintcheva, 2018). *cre* is only expressed for a short period of time, which decreases the risk of off-targets. This strategy is applicable to plant suspension cells as it does not require any sexual crossing. It has been successfully applied to tobacco plants (Wang *et al.*, 2005).

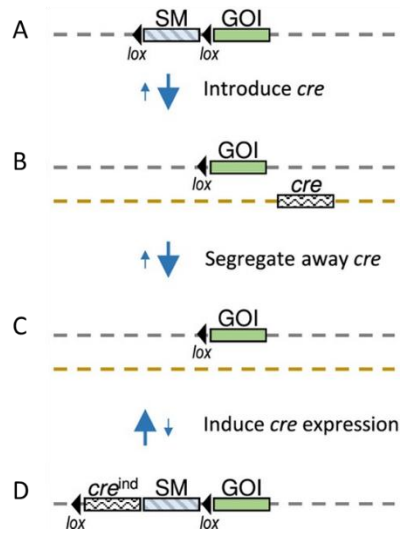


Figure 18. Recombinase-mediated excision of a selectable marker. (A) Plants are transformed with a GOI along with a linked selectable marker flanked by directly oriented *loxP* sites. (B) Cre recombinase is introduced, leading to Cre-mediated excision of the floxed SM. (C) After segregating away *cre*, the host genome is left with the GOI and a *loxP* site, but without a SM. (D) Use of an inducible *cre* gene situated between directly oriented *loxP* sites such that SM and *cre*^{ind} can be simultaneously excised (Ow, 2016). Selectable marker, SM; *cre*^{ind}, inducible *cre* gene.

Another alternative to the crossing/retransformation strategy consists of the transient expression of *cre*. This can be obtained via *Agrobacterium*-mediated transformation. The frequency of transient transformation is at least 1000-fold greater than that of stable transformation, when the T-DNA is integrated into the genome. In most cells, delivered T-DNA is transiently expressed during several days and only rare copies integrate into the plant genome (Janssen *et al.*, 1990). When *Agrobacterium*-mediated transformation is used for stable integration, lengthy selection with antibiotic/herbicide only retains the plant cells with integrated T-DNA. Transient transformation can be associated with the cloning of a negative conditional selectable marker gene such as *codA* in the initial T-DNA to select for recombination events. This strategy was carried out by Gleave *et al.* (1999) in tobacco plants. The first T-DNA contains a GOI with floxed *codA* and *nptII* marker genes. Transgenic plants are generated via classic *Agrobacterium*-mediated transformation. Then, a second transformation is carried out with a plasmid containing *cre* and a marker gene, *hpt*. Transformants that have recombined are selected on 5-FC. After selection, these plants are tested for transgene integration either by PCR on genomic DNA or by screening for hygromycin resistance in the T1 progeny. Plants that are not resistant to hygromycin are the result of recombination mediated by a transiently expressed Cre and thus only contain the GOI.

Transient expression can also be obtained following the aforementioned strategy by the biolistic technique. RNAi and proteolistic techniques, in which gold beads are coated with RNA or proteins instead of DNA present the advantage of avoiding any transgene integration. However, the efficiency of these techniques is lower than that of DNA-based biolistics (Martin-Ortigosa and Wang, 2014).

Plant genetic engineering usually relies on the use of infectious bacteria or harsh conditions owing to the difficulty of delivering biomaterial into plant cells through the cell wall. However, recently, Cre protein delivery by electroporation was carried out in *A. thaliana* suspension cells with intact cell walls (Fuhurata *et al.*, 2019). Protein delivery by electroporation was also attempted in BY-2 cells. However, it could only enter cells in which survivability was compromised due to the electroporation (Cedeño *et al.*, 2017).

One way to work around the limitation of the plant cell wall is to use protoplasts. Protoplasts are plant cells whose cell walls have been enzymatically removed. They can be transfected with DNA, RNA or proteins in the presence of polyethylene glycol (PEG) or by electroporation. However, generating protoplasts is time-consuming, subsequent transfection can induce a loss of viability and plant regeneration is often complicated (Woo *et al.*, 2015; Fuhurata *et al.*, 2019). Other

transfection techniques such as the delivery of RNA via conjugated polymer nanoparticles can lower the toxicity (Silva *et al.*, 2010).

5.3.3.2. Dual targeting by CRISPR/Cas9

CRISPR/Cas9 was first used to inactivate endogenous genes via the introduction of small indels at one or multiple loci as described in chapter 4. To improve the knockout efficiency, multiple gRNAs targeting different regions of the same gene can be used. This multiplex targeting strategy highlights a new application of this technique. Indeed, gRNAs targeting regions close to each other often triggers the deletion of a larger DNA fragment (Xie *et al.*, 2015). The concomitant DSBs induced by two Cas9 proteins associated with two different gRNAs can trigger the excision of the DNA fragment comprised between the two target sites. The two extremities of the DNA molecule are then ligated by NHEJ. This kind of large deletion was observed in our laboratory. Deletions up to 178 bp and 130 bp were identified upon using multiplex editing strategy (Mercx *et al.*, 2016; Mercx *et al.*, 2017, respectively). This outcome suggests that CRISPR/Cas9 could be used to remove unwanted transgenes such as marker genes.

To promote large deletions between two gRNA sites, a high level of expression of both *cas9* and gRNAs is crucial. Indeed, both cleavages must be simultaneous. A high concentration of Cas9 and gRNAs in the cell will increase the probability for this event. Furthermore, to completely excise a marker gene, Cas9-induced deletion of more than 1 kb is required. For example, the *nptII* expression cassette used in this thesis is 1.5 kb long. In general, the fragment deletion efficiency is negatively correlated to the distance between the two gRNA sites (Xie *et al.*, 2015).

Large genomic fragment excision was extensively carried out in several non-plant species such as humans (Cong *et al.*, 2013; Mali *et al.*, 2013; He *et al.*, 2015 ; Zhu *et al.*, 2016), zebrafish (Xiao *et al.*, 2013), yeast (Hao *et al.*, 2016) and mouse (Kraft *et al.*, 2015 ; Zhang *et al.*, 2015 ; Li *et al.*, 2015) by dual-targeting CRISPR/Cas9. Large genomic fragment excision via dual-targeting using CRISPR/Cas9 is also reported in plants.

In rice, *Agrobacterium*-mediated transformation and biolistics were used to excise a 1.6 kb reporter gene (Srivastava *et al.*, 2017). In another study, a 430 bp deletion was obtained at 21% frequency and a 10 kb deletion was achieved at 9% frequency through *Agrobacterium*-mediated transformation (Wang *et al.*, 2017). More strikingly, one of the first attempts to generate large DNA deletions in plants laid impressive results. To improve the probability to generate simultaneous DSBs at both extremities of the DNA fragment, each site was targeted by two gRNAs. This

technique allowed to generate deletions up to 245 kb in rice calli via *Agrobacterium*-mediated transformation and in rice protoplasts via PEG-mediated DNA delivery (Zhou *et al.*, 2014). In light of these experiments, a detailed protocol to generate and screen large chromosomal deletions in rice via CRISPR/Cas9 was recently published (Li *et al.*, 2019).

In soybean, three endogenous loci were targeted. A deletion frequency of 15.5% was achieved when using gRNAs 1.1 kb apart. gRNAs located 4.5 kb apart induced fragment excision at 12% efficiency (Cai *et al.*, 2018).

In *N. tabacum* and *N. benthamiana*, several studies using multiplex gRNAs to inactivate one or several genes identified large genomic deletions (Gao *et al.*, 2015b; Mercx *et al.*, 2017; Jansing *et al.*, 2019), up to 7.2 kb (Herman *et al.*, 2021). CRISPR-mediated marker gene excision in BY-2 cells was recently achieved (Sheva *et al.*, 2020). First, a 14.3 kb construct was integrated into the plant genome via *Agrobacterium*-mediated transformation. The T-DNA contained *codA*, *hptII*, *cas9*, gRNAs targeting *XylT* and *a1,3-FucT* under the control of U6 promoter and a gRNA targeting the T-DNA borders, under the control of a heat-shock promoter. After selection on hygromycin-supplemented medium, *XylT/FucT* knockouts were identified. Then, to induce the expression of the gRNA targeting the borders, knockout lines were exposed to heat treatment. Afterwards, induced cell lines were transferred to a selection medium containing 5-FC and without hygromycin. The addition of 5-FC allowed for the selection of cells for which transgene excision occurred, thanks to the absence of *codA*. The main advantage of this strategy is that the genome editing step and the transgene removal step are achieved in one transformation run. However, NHEJ-mediated repair of Cas9-induced excision leads to unpredictable indels around the break sites. As a result, Sheva *et al.* (2020) were not able to amplify the remaining T-DNA borders and the characterization of the precise sites of the excision failed.

In conclusion, excision of large genomic DNA fragments by multiple gRNAs targeting has been obtained in a large range of organisms, including several plant species (mono- and dicotyledons). Its efficiency varies depending on many parameters such as plant species, gRNAs design, transformation technique and fragment size. Furthermore, for the same experiment, important variations in the length of the excised fragments were observed. This is due to the error-prone NHEJ. The excision is thus generally not precise and this can be of concern, for example for obtaining non-transgenic crops. Another critical point is the delivery of gRNAs and Cas9. If the aim is to remove transgenic material, this must be done without the integration of *cas9* and gRNAs. Moreover, marker gene excision using CRISPR/Cas9 to allow successive transformation cannot be completed if *cas9* and gRNAs are

integrated and stably expressed. As a result, transient expression of *cas9* and gRNAs or direct delivery of RNA or Cas9 RNP is also carried out by the methods described in the previous chapter for Cre recombinase (Woo *et al.*, 2015).

6. Human cytomegalovirus

6.1. Pathogenesis and prophylaxis

Herpesviruses are enveloped, double-stranded DNA (dsDNA) viruses divided in three subfamilies: alpha, beta and gamma-herpesviruses. The HCMV is a beta-herpesvirus. It is endemic in most parts of the world. It establishes a lifelong infection in 40% to nearly 100% of the adult population worldwide (Jean Beltran and Cristea, 2014). Primary infection usually results in subclinical disease. However, like most herpesviruses, HCMV remains silent in the host and undergoes reactivation cycles that contribute to its efficient transmission. These reactivation cycles are easily suppressed by competent immune systems.

Two populations are at risk when facing this virus. First, roughly 10% of newborns infected *in utero* develop long-term neurodevelopmental disabilities such as sensorineural hearing loss, epilepsy, mental retardation and microcephaly (Schleiss *et al.*, 2017). It is the leading cause of congenital infection in the developed world, with an estimated incidence of 0.6–0.7% of all live births (Marsico and Kimberlin, 2017). Immunocompromised people such as transplant patients or those suffering from HIV infection form the second group at risk upon HCMV infection. In this case, HCMV becomes an opportunistic pathogen, causing a large variety of symptoms including gastrointestinal ulceration, hepatitis, pneumonitis, retinitis and blindness (Burke and Heldwein, 2015).

Treatment and prophylaxis are based on the use of the antiviral ganciclovir, often combined with immunoglobulins (Ig) from seropositive individuals. However, ganciclovir cannot be administered to pregnant women because of associated toxicity. Rising ganciclovir resistance is also a concern (Burke and Heldwein, 2015). As a result, the development of a HCMV vaccine is a major public health priority. Despite the progress that has been made toward the goal of development of a HCMV vaccine, none has been licensed yet (Schleiss *et al.*, 2017).

Different types of HCMV vaccines have been -and are still- developed and evaluated. The oldest vaccine development approach is the generation of live attenuated viruses via serial tissue culture passage. This technique allowed the obtaining of the first candidates, the laboratory-adapted HCMV strains AD169 and Towne. These strains have accumulated mutations that affect their replication ability but unfortunately also their immunogenicity. This approach is less and less considered because of the concern that these strains could still carry an unacceptable level of risk of establishing latency and undergoing productive replication in the host (Schleiss *et al.*, 2017).

New vaccine types are considered safer because they rely on the delivery or expression of specific components derived from the pathogen rather than the whole attenuated (or inactivated) virus. These vaccines do not retain the ability to replicate and, although they can still trigger side effects, do not cause pathological effects.

Sub-unit vaccines rely on the injection of a fragment of the virus, typically a surface protein, responsible for inducing protective immunity. Sub-unit vaccines combining recombinant HCMV glycoprotein B (gB) with adjuvants are promising, as several have reached phase I and phase II clinical trials. gB is usually produced as a truncated protein. Removal of its transmembrane region allows its secretion and purification directly from the extracellular medium (Schleiss *et al.*, 2017). The most extensively studied vaccine is a sub-unit vaccine based on Chinese hamster ovary cells-produced gB ectodomain. It is admixed with an oil-in-water emulsion adjuvant, MF59 (Schleiss, M. R. 2018).

More recent techniques are also under investigation, some of them having reached phase I, phase II and even phase III clinical trials. Enveloped virus-like particles (eVLPs) are structures which contain some of the outer surface structural proteins of the virus, such as gB, but not its genome. These structures are expected to give a better immunological response as they better mimic the conformational and morphological characteristics of viral surface glycoproteins. The only vaccine having reached phase III trial is a nucleic acid-based vaccine developed by Moderna, with an estimated study completion date in July 2025 (U.S. National Library of Medicine, NCT05085366). This type of vaccine relies on the injection of DNA or RNA coding for viral proteins into the patient. It is taken up by the patient's cells, which will then synthesize and secrete viral glycoproteins stimulating the host immune system. Finally, vector-based vaccines also rely on viral protein-encoding nucleic acid uptake but a more efficient and tissue-specific delivery is obtained by using heterologous VLPs. Most of sub-unit, eVLP, nucleic acid and vector-based HCMV vaccines rely, at least partially, on gB immunogenicity (Schleiss *et al.*, 2017).

6.2. Glycoprotein B as a vaccine antigen

HCMV dsDNA genome is approximately 235 kb long and encodes many glycoproteins (Dolan *et al.*, 2004). These glycoproteins form the viral capsid, which protects the dsDNA molecule, the tegument, which maintains the cohesion between the capsid and the envelope, and four envelope-anchored glycoprotein complexes responsible for host cell recognition, attachment and entry.

During infection, HCMV attachment is ensured by the binding of envelope glycoprotein complexes to specific cell surface receptors. This recognition triggers membrane fusion and the capsid is released into the host cell. It is believed that the capsid is then brought to the nucleus via interaction between tegument glycoproteins and host microtubules. The viral dsDNA is injected into the nucleus where it is transcribed, replicated and later encapsidated. Once capsids reach the cytoplasm, they hijack the host secretory machinery to acquire their tegument layer and viral envelope before being released (Jean Beltran and Cristea, 2014).

Envelope glycoproteins are the target of neutralizing antibodies and hence represent the only relevant viral antigens for antibody-based vaccines. Of these, gB draws the most attention because of its high immunogenicity. Indeed, up to 70% of the serum virus-neutralizing activity of CMV-immune individuals is directed against gB (Schleiss *et al.*, 2017).

Two independent crystal structures of HCMV gB in the postfusion conformation were obtained (Burke and Heldwein, 2015; Chandramouli *et al.*, 2015). They are very similar to the structures of gB from other herpesviruses (Shama *et al.*, 2013). It is a protein with a large ectodomain, a hydrophobic membrane-proximal region, a transmembrane domain and a cytoplasmic domain (Figure 19A). The ectodomain is divided into five structural domains and forms a homotrimer of more than 300 kDa (Figure 19B).

gB is a type I envelope glycoprotein consisting of 906 amino acid residues. It is a viral fusogen, a glycoprotein essential for membrane fusion. HCMV recognizes and binds to the host cell receptors in a gB-independent fashion. Interaction of host receptors with viral glycoprotein complexes triggers conformational changes of gB: unfurling of the ectodomain and exposure of fusion loops for insertion into the membrane (Figure 19A). In this conformation, gB catalyzes virus-cell membrane fusion. As a result, gB can be found in two conformations: pre- and postfusion (Si *et al.*, 2018).

Posttranslational modifications of gB have a considerable importance for its conformation and function. In the host cell, gB is cleaved into two subunits by the endoprotease furin. The two subunits are connected by disulfide bonds (Figure 19A). Up to eighteen *N*-glycosylation sites and two *O*-glycosylation sites have been identified. This extensive *N*-glycosylation is responsible for nearly half of the glycoprotein mass (Smargiasso *et al.*, 2019). Most of the gB surface is thus shielded by a thick glycan layer. This is a common virus strategy to escape immune recognition. Indeed, it is believed that viral glycosylation protects functionally important regions from immune recognition (Burke and Heldwein, 2015).

Recombinant gB is usually truncated at the transmembrane domain and the furin site is mutated to facilitate purification. However, gB ectodomains form trimers, which have a propensity to aggregate via interaction of hydrophobic residues in the fusion loops. To avoid this, mutations in these loops, which trigger the fusion of the viral membrane with the host cell membrane, can be introduced (Chandramouli *et al.*, 2015).

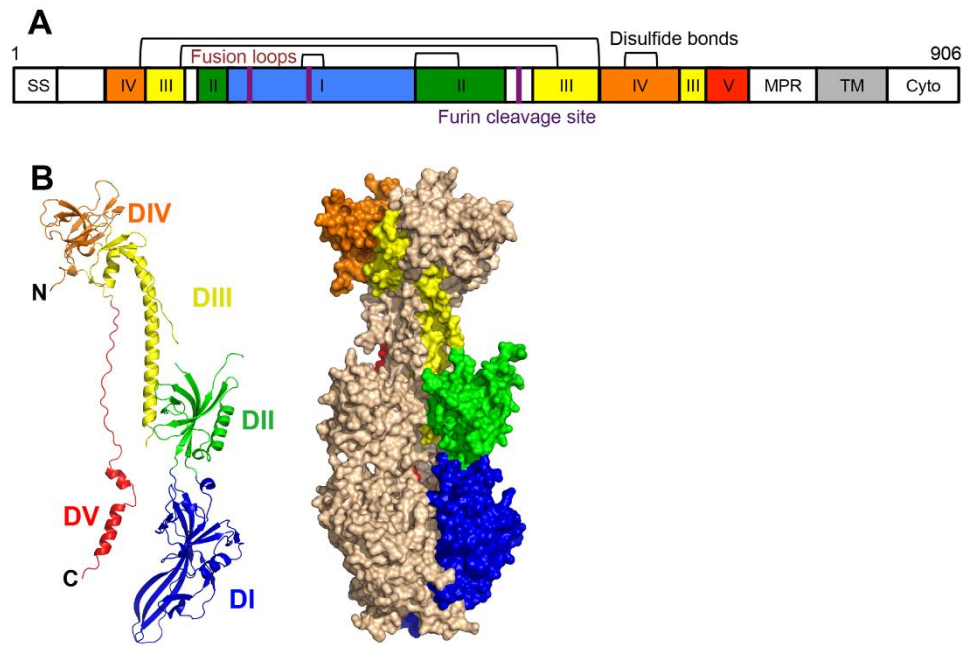


Figure 19. HCMV gB ectodomain structure. (A) Schematic representation of the full-length HCMV gB. Disulfide bonds are represented as black brackets. Structural domains are colored. SS = signal sequence, MPR = membrane proximal region, TM = transmembrane domain and Cyto = cytoplasmic domain. (B) The crystal structure of the HCMV gB ectodomain is shown as a protomer and a trimer in surface representation (Adapted from Burke and Heldwein, 2015).

6.3. Recombinant gB *N*-glycoprofile expressed in mammalian CHO and plant BY-2 cells

Our laboratory has expressed gB from HCMV in BY-2 cells (Smargiasso *et al.*, 2019). The aim was to compare gB *N*-glycosylation profile with that of the same glycoprotein produced in CHO cells by GlaxoSmithKline (GSK). More precisely, the objective was to identify the specific heterogeneity at each *N*-glycosylation site and determine if there is a similarity between the degree of processing in BY-2 and CHO-produced gB.

The gB construct consists of the soluble ectodomain from AD169 HCMV strain in the presumably postfusion conformation (Chandramouli *et al.*, 2015; Liu *et al.*, 2021) with the endogenous signal peptide. The transmembrane domain and the cytoplasmic tail were removed. The furin cleavage site was mutated and a 6xHis-tag was added in C-terminal. Finally, it was codon-optimized for expression in plants or mammals.

Recombinant gB is secreted in the extracellular medium of WT BY-2 cells at a yield of 20 mg/L. gB was directly semi-purified by ammonium sulfate precipitation and size exclusion chromatography. A comprehensive site-specific quantitative *N*-glycosylation analysis of 15 out of the 18 *N*-sites of gB produced in BY-2 cells was carried out by the laboratory of mass spectrometry at the University of Liège (MolSys-ULiège). The *N*-glycosylation profile differs at each site, some sites being occupied exclusively by oligomannosidic type *N*-glycans and others by typical plant-specific complex or hybrid biantennary *N*-glycans (Figure 20). These data were compared to those obtained in parallel for the same HCMV gB variant produced in CHO cells by GSK. The profiles were strikingly similar between BY-2 and CHO-produced gB, indicating that, overall, the gB conformational structure is conserved when expressed in plant cells and that this structure influences the *N*-glycosylation profile at each site. However, as expected, the processed *N*-sites displayed much more heterogeneity in CHO cells with the presence of bi, tri and tetraantennary structures, which can be further elongated with sialic acids (Smargiasso *et al.*, 2019).

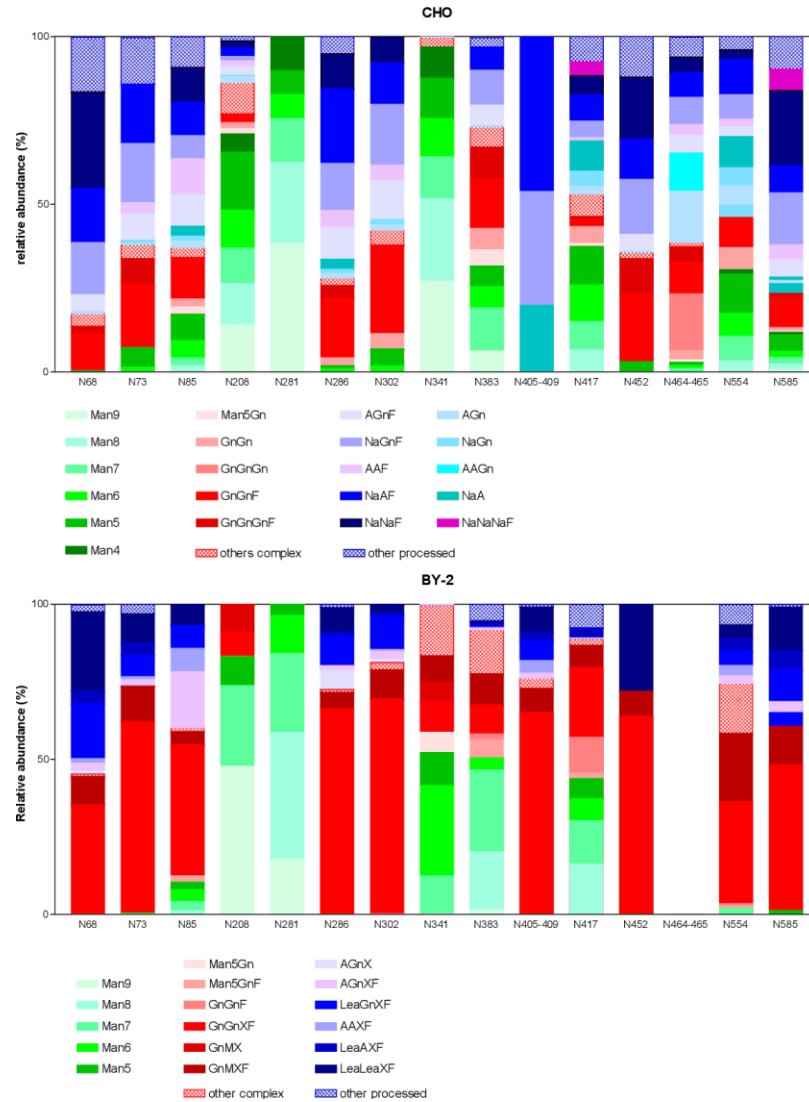


Figure 20. Quantification of the relative abundance of *N*-glycans detected by mass spectrometry at each glycosite on gB ectodomain expressed in CHO and BY-2 cells. *N*-glycans detected (>5%) are represented and abbreviated according to the nomenclature (Proglycan short) from <http://www.proglycan.com>. The quantitative data are derived from a single analysis for both samples. Glycans are assigned into three groups: oligomannosidic structures in green; hybrid/complex structures exhibiting terminal GlcNAc in red; highly processed complex *N*-glycans in blue. Glycans present in less than 5% are grouped in “others” and are hatched (Smargiasso *et al.*, 2019).

Objectives

Plant suspension cells offer an interesting alternative for the expression of recombinant glycoproteins. BY-2 cell line is characterized by a short culture cycle, easy scaling up in bioreactors, low risk of contamination by human pathogens and the possibility to purify secreted recombinant proteins directly from the culture medium (Tekoah *et al.*, 2015; Santos *et al.*, 2016). It has already been used to produce several pharmaceutical glycoproteins such as antibodies (De Muynck *et al.*, 2009; Holland *et al.*, 2010; Magy *et al.*, 2014), HCMV glycoprotein B (Smargiasso *et al.*, 2019) or human enzymes (Hanania *et al.*, 2017; Ilan *et al.*, 2017, Herman *et al.*, unpublished data). Engineering the BY-2 *N*-glycosylation pathway is a powerful strategy to improve the quality of these recombinant glycoproteins, both in terms of biological activity and production processes. However so far, only *XylT* and *FucT* genes have been inactivated in BY-2 cells (Mercx *et al.*, 2017; Hanania *et al.*, 2017) and a few human glycosyltransferases have been expressed (Karg *et al.*, 2010, Kajiura *et al.*, 2011; Navarre *et al.*, 2017).

The first objective of this project is to extend the set of BY-2 cell lines with specific and predefined *N*-glycosylation profiles through new glyco-engineering strategies. To this end, knockouts of GTs as well as genes indirectly involved in the *N*-glycosylation pathway and expression of ectopic genes will be carried out. To generate a cell line depleted of non-human residues β 1,2-xylose, α 1,3-fucose and α 1,4-fucose, *XylT* and *GMD* genes will be inactivated using the CRISPR/Cas9 edition system. Next, a cell line producing only high mannose *N*-glycans will be generated by knocking out the *GnTI* genes. Finally, EndoT will be expressed in the *GnTI* knockout cell line to shorten *N*-glycans to a single GlcNAc stump and hence obtain *in vivo* deglycosylated proteins.

The second main objective is to produce recombinant glycoproteins in the glyco-engineered cell lines and analyze their *N*-glycosylation profile and the afferent properties. An IgG was chosen to validate the new cell lines for molecular farming. To further challenge the glyco-engineered cell lines, the production of a highly glycosylated protein, the ectodomain of HCMV gB, will be carried out. Once produced, the glycosylation status of both recombinant glycoproteins will be investigated by *in vitro* enzymatic assays and mass spectrometry analysis. The influence of the glycan alteration on IgG stability will be determined by a differential scanning fluorimetry.

In parallel, to facilitate successive transformations and ensure genomic stability of the glyco-engineered cell lines, a transgene excision system will be developed to get rid of DNA elements that are not desirable anymore, such as marker genes, nucleases or gRNAs.

Results and Discussion

1. Inactivation of *N*- *acetylglucosaminyltransferase I* and α 1,3- *fucosyltransferase* genes in *N. tabacum* BY-2 cells results in glycoproteins with highly homogeneous, high-mannose *N*-glycans

1.0. Contextualization

The main safety concern associated with plant-made glycoproteins is the presence of non-human residues, susceptible to trigger immune or even allergic reactions. Due to their important abundance, α 1,3-fucose and β 1,2-xylose residues linked to the *N*-glycan core of the bi-antennary complex structures (Figure 5) have drawn most attention in the plant glyco-engineering field. Besides these residues, α 1,4-fucose and β 1,3-galactose forming Le^a structures are also considered as a potential risk. Indeed, although these residues can be synthesized by humans, they are only detected in low amounts. Hence, simultaneous suppression of β 1,2-xylose, α 1,3-fucose and Le^a-associated residues would represent a step further towards humanization of BY-2 cells *N*-glycans. To obtain a BY-2 cell line producing *N*-glycans without β 1,2-xylose and depleted for both α 1,3-fucose and α 1,4-fucose, we targeted the *XyIT* and *GDP-mannose 4,6-dehydratase* (*GMD*) genes via CRISPR/Cas9 technique (Annex 1). However, while the strategy proved to be effective to some extent, the remaining α 1,3-fucose content of the *XyIT/GMD*-KO cell line was too high to consider the use of this cell line instead of the *XyIT/FucT*-KO cell line.

Instead, impairing the synthesis of complex *N*-glycans responsible for most of the *N*-glycan diversity in order to produce only high-mannose *N*-glycans (Man4–Man9) is a powerful strategy to suppress non-human residues, while homogenizing the *N*-glycan repertoire. Moreover, many enzymes used as therapeutics for lysosomal storage diseases require phosphorylated high-mannose *N*-glycans for their targeting to the lysosomes. To obtain a cell line conferring such characteristics to glycoproteins, the enzyme GnTII, essential for the processing of high-mannose to hybrid and complex *N*-glycans, was inactivated.

This chapter corresponds to the following original research article:

Herman Xavier, Far Johann, Courtoy Adeline, Bouhon Laurent, Quinton Loïc, De Pauw Edwin, Chaumont François and Navarre Catherine (2021) Inactivation of *N-Acetylglucosaminyltransferase I* and *α1,3-Fucosyltransferase* Genes in *N. tabacum* BY-2 Cells Results in Glycoproteins With Highly Homogeneous, High-Mannose N-Glycans. *Frontiers in Plant Science* **12** 634023. doi: 10.3389/fpls.2021.634023

XH¹ and CN¹: conceptualization. XH and JF²: methodology. XH, JF, AC¹, LB¹, and CN: investigation. XH, JF, and CN: data curation. XH: writing – original draft preparation. CN, JF, and FC¹: writing – review and editing. CN, LQ², and FC: supervision. FC and EDP²: funding acquisition. All authors have read and agreed to the manuscript.

¹Louvain Institute of Biomolecular Science and Technology, UCLouvain, Louvain-la-Neuve, Belgium

²Mass Spectrometry Laboratory-MolSys, GIGA Proteomics Facility, University of Liège, Liège, Belgium

The original contributions presented in the study are publicly available. This data can be found here: Repository MassIVE: MSV000086640, <ftp://massive.ucsd.edu/MSV000086640/>.

1.1. Identification of *N-acetylglucosaminyltransferase I* genes in *N. tabacum*

N. tabacum is an allotetraploid species resulting from the cross between *N. tomentosiformis* and *N. sylvestris* (Yukawa *et al.*, 2006). The *NtGnTI* orthologs derived from *N. tomentosiformis* and *N. sylvestris* will be referred to as *NtGnTL.A* and *NtGnTL.B*, respectively, nomenclature previously proposed for *NtSBT1* genes (Navarre *et al.*, 2012), and *NtFucT* or *NtXylT* genes (Mercx *et al.*, 2017). The *NtGnTL.A* and *NtGnTL.B* cDNAs have already been identified from *N. tabacum* cDNA leaves libraries (Strasser *et al.*, 1999; Wenderoth and von Schaewen, 2000). Blast analysis on the four *N. tabacum* genome assemblies (TN90, BX, K326, and Nitab 4.5) as well as the *N. tomentosiformis* and *N. sylvestris* genomes¹ identified the *GnTL.A* full-length gene in all four *N. tabacum* cultivars (Ntab-TN90_AYMYSS16267) and in *N. tomentosiformis* (Ntom_KB953441.1). On the other hand, the *GnTL.B* gene was found to be split into three contigs in *N. tabacum* TN90 and BX cultivars (Ntab-TN90_AYMY-SS68155: exons 1-11, Ntab-TN90_AYMYSS133799: exon 12-18, Ntab-TN90_AYMY-SS242176: exon 19), and in two contigs in *N. sylvestris* (Nsyl_KD955459.1: exons 1-11 and Nsyl_KD976298.1: exons 12-19). Both *NtGnTI* genes were also annotated in NCBI database: *NtGnTL.A* (NCBI gene ID: 107789589) consists of 19 exons spanning on a 12.5 kb genomic sequence (Figure 21A), while *NtGnTL.B* (gene ID: 107800221) is incomplete since the annotated contig covers only the exons 1 to 10.

¹ https://solgenomics.net/organism/Nicotiana_tabacum/genome

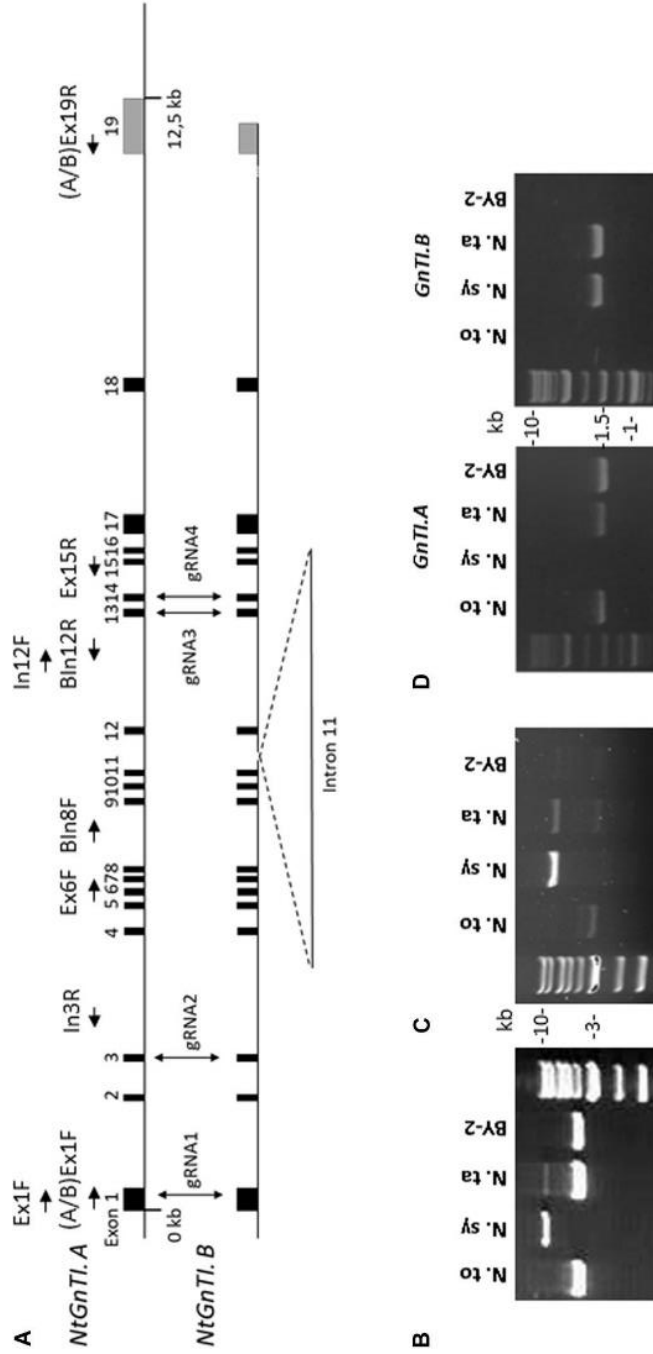


Figure 21. Characterization of Nicotiana *GnTI* genes. (A) Schematic representations of *NtGnTI.A* and *NtGnTI.B* genomic sequences. Black rectangles: exons; gray rectangles: 3'UTR; horizontal arrows: primers; vertical arrows: guide RNA target sites; dotted line: interrupted genomic sequence. Note that intron 11 is extended in *NtGnTI.B*. (B) PCR amplification of genomic DNA with the non-specific *GnTI* primers Ex6F and Ex15R. Bands are expected at 3.8 kb (*GnTI.A*) and 8.6 kb (*GnTI.B*). (C) PCR amplification of genomic DNA with *GnTI.B*-specific primers Bln8F and Bln12R. *GnTI.B*-specific amplification band is detected at 7.2 kb. (D) PCR amplification of cDNAs with *GnTI.A* and *GnTI.B*-specific primers (A/B)Ex1F and (A/B)Ex19R. The amplicon size is 1.56 kb. N.to: *N. tomentosiformis*; N.sy: *N. sylvestris*; N.ta: *N. tabacum*.

To compare the *GnTI.A* and *GnTI.B* genomic regions between exons 11 and 12, genomic DNA was extracted from the three *Nicotiana* plants as well as from BY-2 cells. Next, the *GnTI* region spanning exons 6 to 15 was amplified using non-specific primers Ex6F and Ex15R (Figure 21B, Table S1). The expected 3.8 kb amplicon, corresponding to *GnTI.A*, was detected for *N. tomentosiformis*, *N. tabacum* and BY-2 cells. An 8.6 kb amplicon, corresponding to *GnTI.B*, was detected for *N. sylvestris* and, although less intense, for *N. tabacum* but not for BY-2 cells. Amplification of the region spanning intron 8 to intron 12 was carried out using *GnTI.B* specific primers BIn8F and BIn12R (Figure 21C, Table S1). Sequencing analysis showed that introns 11 of *NtGnTI.B* (Figure S1) and *NsGnTI.B* were 5.4 kb long, while introns 11 of *NtGnTI.A* and *NtoGnTI.A* were only 0.4 kb long. *NtGnTI.A* and *NtGnTI.B* intron 11 sequences were aligned and the 5 kb sequence specific to *NtGnTI.B* was blasted on the *N. tabacum* TN90 genome. Two hundred fifty contigs (with an expect score lower than 10^{-10}) were identified, showing the presence of numerous sequences highly similar to *NtGnTI.B* intron 11 in the *N. tabacum* genome. *In silico* analysis of *NtGnTI.B* intron 11 sequence revealed the presence of three ORFs of 447, 183, and 507 codons (Figure S1), which are associated with the retrovirus-related Pol polyprotein from transposons TNT 1–94, RE1 or RE2 from *Vitis vinifera*. Taken together, these results suggest that the extended intron 11 in *NtGnTI.B* is due to a transposition event.

To assess the expression of both *NtGnTI* orthologs, RNA was extracted from the leaves of the three *Nicotiana* plants and from BY-2 cells, and analyzed by RT-PCR, using primers specific to each ortholog (Figure 21D). *GnTI.A*-specific primers AEx1F and AEx19R (Table S1) amplified the expected 1.56 kb band for *N. tomentosiformis* and *N. tabacum* plants as well as BY-2 cells. On the other hand, *GnTI.B*-specific primers BEx1F and BEx19R (Table S1) only gave the expected 1.56 kb band for *N. sylvestris* and *N. tabacum* plants but not for BY-2 cells. This is corroborated by BY-2 cells RNAseq data (Ahmed *et al.*, 2020) in which no cDNA encoding *GnTI.B* was detected. The five *GnTI* cDNAs were sequenced. The coding nucleotide sequences of *NtGnTI.A* and *NtGnTI.B* from *N. tabacum* leaves were identical to the *N. tabacum* cDNAs A4 and A9 already described in the literature (Wenderoth and von Schaewen, 2000). In addition, *NtGnTI.A* sequences obtained from *N. tabacum* leaves and BY-2 cells were identical to the sequence identified in BY-2 cells RNAseq data. Finally, the *NsGnTI* sequence was identical to the database sequences (XM_009803697 and XM_009776507) whereas *NtoGnTI* displayed a single mutation (Thr175Ala), compared to the database sequence (XM_009610507).

1.2. Generation, screening, and characterization of BY-2 *GnTI*-KO cell lines

Four guide RNAs (gRNAs) were designed to inactivate both *NtGnTI.A* and *NtGnTI.B* by targeting exons 1, 3, 13, and 14 (Table S2). A transfer RNA (tRNA) was placed before each gRNA to generate a polycistronic tRNA-gRNA (*GnTI*.PTG). Once the PTG is transcribed, individual gRNAs are expected to be released by the endogenous tRNA processing machinery (Xie *et al.*, 2015). The *GnTI*.PTG construct was cloned into a pPAM plasmid containing a plant codon-optimized *cas9* (Jansing *et al.*, 2019) to generate the pPAM-*GnTI*-KO plasmid (Figure 22A and Figure S2). BY-2 cells were transformed with the pPAM-*GnTI*-KO construct using biolistics. After selection on kanamycin-supplemented medium, 36 transgenic lines were screened for N-glycosylation profile modification. The extent of complex N-glycan suppression was first assessed by analyzing the N-glycosylation profile of the proteins secreted in the culture medium by Western blotting combining antibodies recognizing β 1,2-xylose and α 1,3-fucose epitopes. Complete inactivation of *GnTI* by genome editing was expected to suppress the synthesis of all complex N-glycans and thereby incorporation of β 1,2-xylose and α 1,3-fucose residues. Thirty-one out of 36 cell lines showed a dramatic and similar reduction of complex N-glycan-associated signal as compared with the WT cell line (Figure S3 and data not shown). However, none of the cell lines displayed a total absence of signals as observed for the *XylT/FucT*-KO cell line used as a control (Mercx *et al.*, 2017). Four *GnTI*-KO lines (*GnTI*-KO #2, 23, 27, 30) were further analyzed using either anti- β 1,2-xylose or anti- α 1,3-fucose antibodies (Figure 23A). While no signal corresponding to β 1,2-xylose could be detected, an α 1,3-fucose specific signal was detected in all four *GnTI*-KO lines, but with a 30 to 100-fold lower intensity compared with the WT cell line. Similar results were obtained when TSCPs were analyzed by Western blotting (Figure 23B).

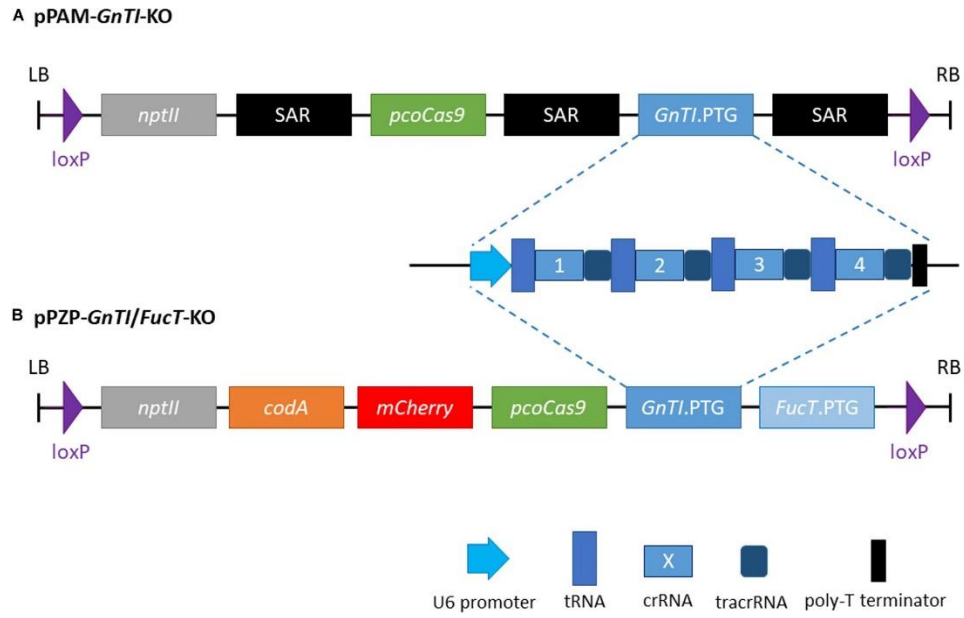


Figure 22. Schematic representation of the two plant binary vectors. (A) pPAM-*GnTI*-KO (B) pPZP-*GnTI*/*FucT*-KO. SAR, *N. tabacum* Rb7 scaffold attachment region (genetic insulator); *nptII*, neomycin phosphotransferase II cassette; *pcoCas9*, plant codon-optimized *cas9* cassette; PTG, polycistronic tRNA-gRNA; *codA*, cytosine deaminase A cassette.

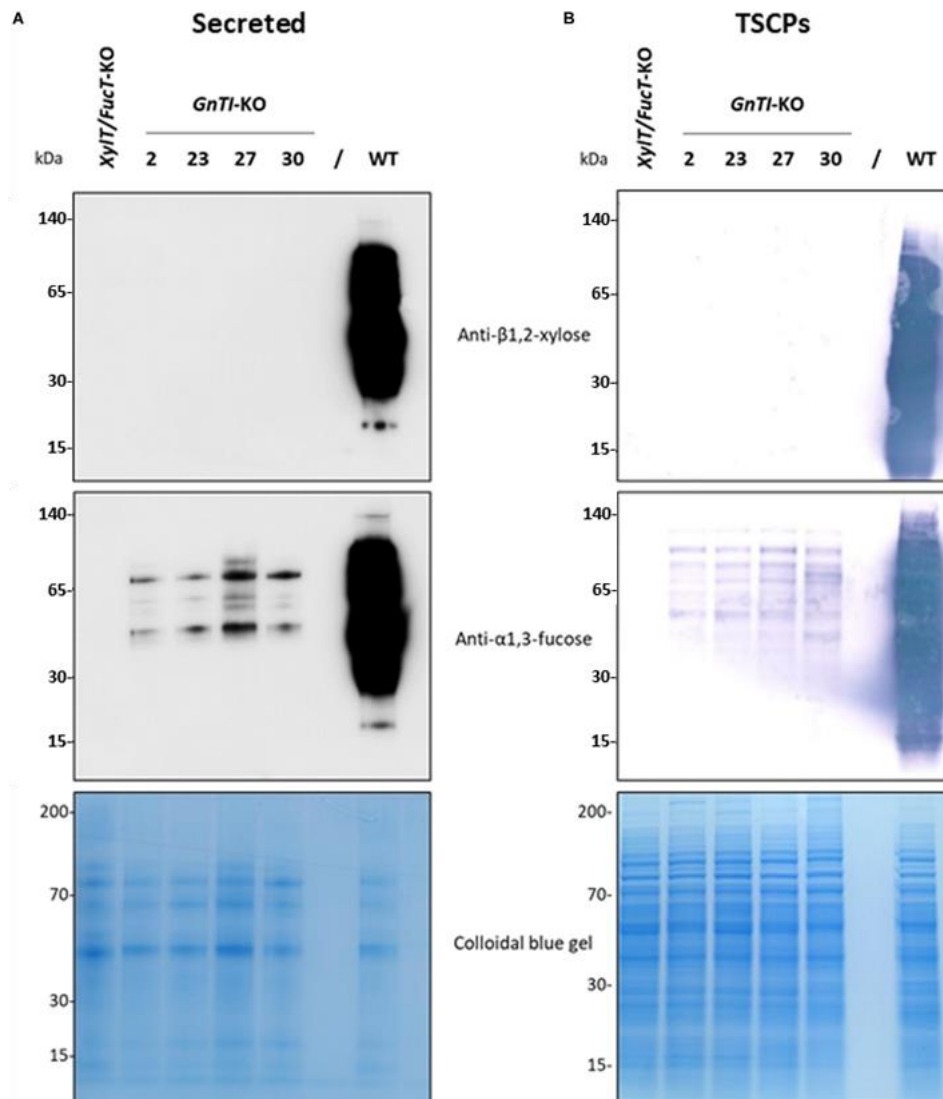


Figure 23. Absence of β 1,2-xylose and reduction of α 1,3-fucose on glycoproteins from the *GnTI-KO* BY-2 cell lines. Secreted proteins (40 μ L culture medium) (A) or total soluble cellular proteins (12.5 μ g) (B) from a WT, *XyIT/FucT-KO*, and the indicated *GnTI-KO* cell lines were separated by gel electrophoresis and analyzed by Western blotting using anti- β 1,2-xylose or anti- α 1,3-fucose antibodies. Horseradish peroxidase-linked or phosphatase alkaline-linked secondary antibodies were used for secreted proteins (A) and total soluble cellular proteins (B), respectively. A colloidal blue gel is displayed as a loading control.

1.3. Generation, screening, and characterization of *GnTI/FucT*-KO cell lines

To suppress both GnTI and FucT activities, the *GnTI*.PTG genetic construct was combined with the *FucT*.PTG construct previously used to obtain the *XylT/FucT*-KO cell line (Mercx *et al.*, 2017) into a pPZP-RCS2 binary plasmid containing the plant codon-optimized *cas9* (Figure 22B and Figure S4). This plasmid is referred to as pPZP-*GnTI/FucT*-KO and was used to transform WT BY-2 via *A. tumefaciens* cocultivation. Preference for *Agrobacterium* transformation was guided by the need for precise genomic integration if further T-DNA removal was needed, using the Cre-*loxP* system. After selection on kanamycin-supplemented medium, 22 *GnTI/FucT*-KO cell lines were screened using the same immunoblotting strategy as for *GnTI*-KO lines. Fifteen cell lines showed an absence or strong reduction of β 1,2-xylose and α 1,3-fucose combined signals on secreted glycoproteins (Figure S5 and data not shown). Four *GnTI/FucT*-KO cell lines (#5, 10, 16, and 18) for which no signal had been observed were chosen for further analysis. Secreted glycoproteins were analyzed by Western blotting using either anti- β 1,2-xylose or anti- α 1,3-fucose antibodies (Figure 24A). No β 1,2-xylose nor α 1,3-fucose residues were detected in any of these lines. Analysis of the TSCPs by Western blotting confirmed the absence of β 1,2-xylose and α 1,3-fucose residues in glycoproteins of cell lines #5, 10, and 16 (Figure 24B), indicating the complete inactivation of both *GnTI* and *FucT* genes. A faint signal for α 1,3-fucose was observed in cell line #18, suggesting that a minor residual FucT activity was still present in that particular cell line.

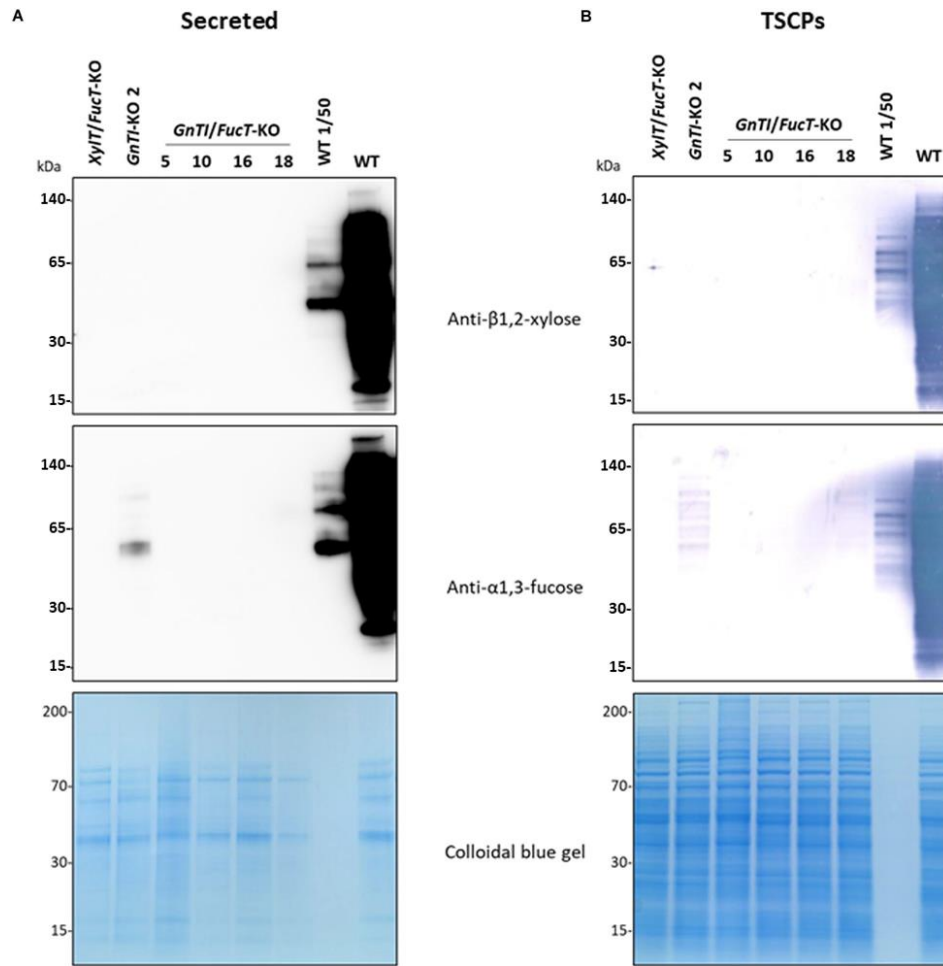


FIGURE 24. Absence of β 1,2-xylose and α 1,3-fucose on glycoproteins from *GnTI/FucT-KO* lines. Secreted proteins (40 μ L culture medium) (A) or total soluble cellular proteins (12.5 μ g) (B) from WT, *XylTI/FucT-KO*, *GnTI-KO #2*, and the indicated *GnTI/FucT-KO* cell lines were separated by gel electrophoresis and analyzed by Western blotting using anti- β 1,2-xylose or anti- α 1,3-fucose antibodies. Horseradish peroxidase-linked or phosphatase alkaline-linked secondary antibodies were used for secreted proteins (A) and total soluble cellular proteins (B), respectively. A 50-fold dilution of WT proteins (equivalent to 0.8 μ L culture medium or 0.25 μ g TSCPs) were loaded in the penultimate well (WT 1/50). A colloidal blue gel is displayed as a loading control.

1.4. Mass spectrometry analysis of total *N*-glycans on glycoproteins secreted by *GnTI*-KO #2 and *GnTI/FucT*-KO #10 cell lines

Secreted glycoproteins of WT, *GnTI*-KO #2 and *GnTI/FucT*-KO #10 cell lines were analyzed by MALDI-FT ICR mass spectrometry after 2-aminobenzamide reductive labeling to determine their respective *N*-glycoprofile. Spectra are displayed in Figure 25 and relative amounts of the *N*-glycan structures are provided in Table 3. The *N*-glycoprofile of bovine RNase B was determined as a positive control (Figure S6 and Table S3). Fourteen *N*-glycan structures were detected on the secreted glycoproteins of WT BY-2 cells. The most abundant structures were GnMXF (27.9%), GnGnXF (20.9%), and GnMX (14.6%). Complex, paucimannosidic and hybrid *N*-glycans accounted for 84.8% of the *N*-glycans on proteins secreted in WT BY-2 cells. High-mannose *N*-glycans made up for only 15.2% of the *N*-glycans. *GnTI* inactivation resulted in a dramatic shift toward high-mannose *N*-glycan structures in *GnTI*-KO #2 and *GnTI/FucT*-KO #10 lines, in which they accounted for 98.4% and 99% of the *N*-glycans, respectively (the remaining being paucimannose structures). Only six *N*-glycan structures, corresponding to Man3-8 were identified in the KO lines. The main *N*-glycan structures detected in *GnTI*-KO #2 and *GnTI/FucT*-KO #10 were Man5 (79.7 and 85.9%, respectively) and Man4 (16.4 and 11%, respectively).

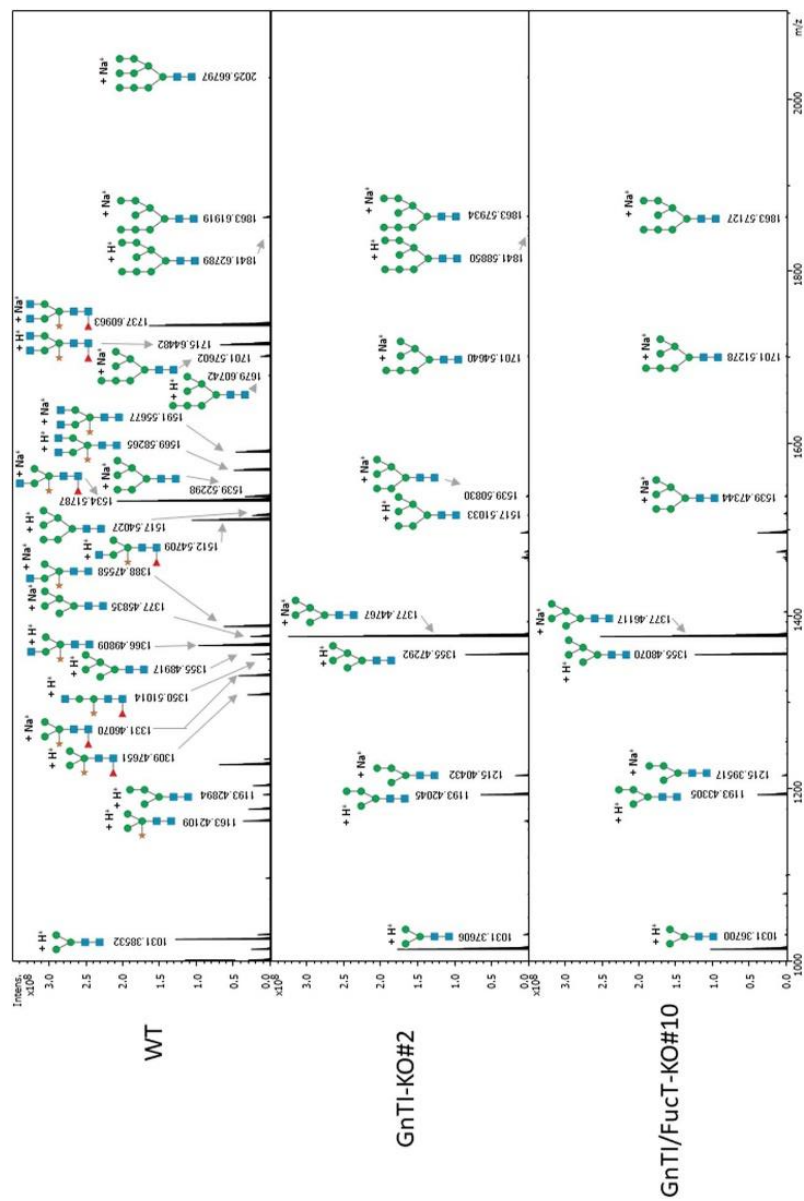


Figure 25. Total 2-aminobenzamide labeled N-glycans released from secreted glycoproteins of indicated cell lines. Peaks identified as N-glycan signals are labeled. Signals were detected as singly positively charged $[M + H]^+$ or $[M + Na]^+$ using MALDI FT-ICR mass spectrometry.

TABLE 3. Relative amounts of total 2-aminobenzamide labeled N-glycans (%) decorating the secreted glycoproteins from WT and KO cell lines as determined by MALDI FT-ICR mass spectrometry.

Glycan type	Composition	Putative structure	Abbreviation ¹	WT	GnT1-KO#2	Relative amounts ²	GnT1/FucT-KO#10
Oligomannose	HexNAc2Hex9		Man9	0.4%	ND ³	ND	ND
	HexNAc2Hex8		Man8	1.4%	0.8%	0.6%	0.6%
	HexNAc2Hex7		Man7	1.7%	0.4%	0.6%	0.6%
	HexNAc2Hex6		Man6	5.6%	1.2%	1.0%	1.0%
	HexNAc2Hex5		Man5	5.0%	79.7%	85.9%	85.9%
	HexNAc2Hex4		Man4	1.1%	16.4%	11.0%	11.0%
Paucimannose	HexNAc2Hex3		MM	1.7%	1.6%	1.0%	1.0%
	HexNAc2Hex3Pent1		MMX	3.4%	ND	ND	ND
	HexNAc2Hex3dHex1Pent1		MMXF	7.0%	ND	ND	ND
	HexNAc3Hex3Pent1		GnMX	14.6%	ND	ND	ND
Complex	HexNAc3Hex3dHex1Pent1		GnMXF	27.9%	ND	ND	ND
	HexNAc3Hex2dHex1Pent1		N/A ⁴	0.4%	ND	ND	ND
	HexNAc4Hex3Pent1		GnGnX	8.9%	ND	ND	ND
	HexNAc4Hex3dHex1Pent1		GnGnXF	20.9%	ND	ND	ND

¹According to the nomenclature (Proglycan short) from <http://www.proglycan.com>. Gn, N-acetylglucosamine; M/Man, mannose; X, xylose; F, fucose. ²Relative amounts were determined by considering only the absolute intensities of the peaks associated to an N-glycan structure. ³Not detected. ⁴Not assigned.

1.5. Genomic analysis of CRISPR/Cas9-induced indels in *GnTI*-KO #2 and *GnTI/FucT*-KO #10 cell Lines

To identify the Cas9-mediated editing of the *GnTI* alleles, genomic DNA was extracted from WT, *GnTI*-KO #2, and *GnTI/FucT*-KO #10 cell lines. The *GnTI* region encompassing the four targeted loci (exon 1 to exon 14) was PCR-amplified using primers hybridizing to conserved regions Ex1F and Ex15R (Figure 21A, Table S1). Gel electrophoresis showed amplicons of the same expected size (7.6 kb) for WT, *GnTI*-KO #2, and *GnTI/FucT*-KO #10 cell lines as well as an extra smaller amplicon (0.44 kb) for *GnTI/FucT*-KO #10 (Figure 26A). Amplicons were cloned and eight different *GnTI.A* alleles in *GnTI*-KO #2 line and two in *GnTI/FucT*-KO #10 were identified by sequencing (Figure 26B). No *GnTI.B* amplicon could be obtained due to the presence of the large 5.4 kb intron 11. To obtain amplicons corresponding to *GnTI.B*, two separate PCR amplifications were carried out to get regions corresponding either to sites 1–2 using primers Ex1F and In3R or to sites 3–4 using primers In12F and Ex15R (Figure 21A, Table S1). Amplification of sites 1–2 resulted in obtaining *GnTI.A* sequences exclusively, whereas amplification of sites 3–4 gave both *GnTI.A* and *GnTI.B* sequences. Three *GnTI.B* alleles were identified for *GnTI*-KO #2 and one for *GnTI/FucT*-KO #10 (Figure 26B). Sequence analysis confirmed that these alleles were effectively inactivated. Indeed, three out of the four target sites were mutated in *GnTI.A* allele 4 while all other *GnTI.A* and *GnTI.B* alleles were modified at each observed target site. Most *GnTI*-KO #2 indels consisted of small insertions or deletions but a deletion of 180 nucleotides occurred between sites 3 and 4 in *GnTI.A* alleles 1, 3, and 7. The presence of more than four alleles indicates that some indels were not induced in the initial transformed BY-2 cell, but later on in daughter cells. The largest deletion was observed for *GnTI/FucT*-KO #10 cell line. Indeed, the small 0.44 kb amplicon observed in gel electrophoresis (Figure 26A) corresponds to *GnTI.A* allele 2, which underwent a 7.16 kb deletion between sites 1 and 4. All sequences obtained for *GnTI.B* sites 3–4 using primers In12F and Ex15R correspond to allele 3, suggesting homozygous mutation. *GnTI/FucT*-KO #10 cell line was likely homogeneous since each *GnTI* ortholog was present in one or two mutated versions.

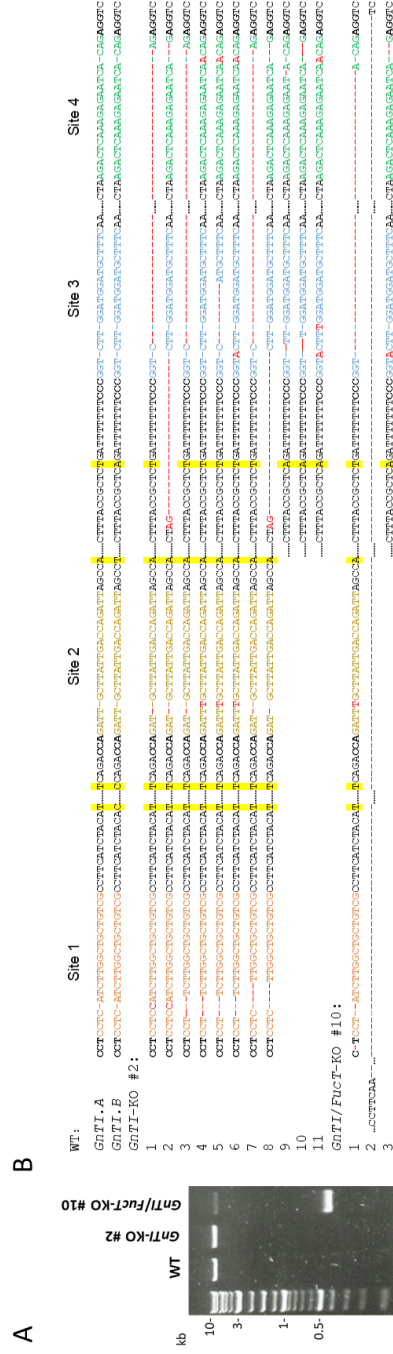


Figure 26. Identification of CRISPR-induced indels in *GnTI* alleles from *GnTI-KO #2* and *GnTI/FucT-KO #10* BY-2 cell lines. (A) PCR amplification of *GnTI.A* from genomic DNA extracted from the indicated cell lines, using primers Ex1F and Ex15R. (B) Sequences of the *GnTI* target sites. Colored nucleotides correspond to the gRNA-hybridizing sites. PAM sequences are in bold. CRISPR/Cas9-induced deletions are represented by red dashes, insertions by red nucleotides. Single nucleotide polymorphisms allowing distinguishing *GnTI.A* from *GnTI.B* alleles are highlighted in yellow.

1.6. Discussion

In this study, we inactivated the two *GnTI* genes in *N. tabacum* BY-2 suspension cells using CRISPR/Cas9 gene edition in order to generate a cell line producing only high-mannose *N*-glycans. GnTI has indeed a key role in the formation of *N*-glycan diversity because it is responsible for the addition of a first GlcNAc, on the α 1,3-arm. Addition of this particular residue on a Man5 is considered as a prerequisite before being further modified by Golgi α -mannosidase II, *N*-acetylglucosaminyltransferase II, β 1,2-xylosyltransferase, and α 1,3-fucosyltransferase to form hybrid, complex and paucimannosidic *N*-glycans (Strasser *et al.*, 2007).

To efficiently knock out *GnTI* in BY-2 cells, we first characterized both *NtGnTI* orthologs. However, *NtGnTI.B* exons 1–11 were detected in *N. tabacum* leaf material but not in BY-2 cells. This suggests that genome modifications impacting the first half of *GnTI.B* occurred in the cell line available in our lab. It is known that plant suspension cells, including BY-2, might be genetically unstable and subjected to chromosome rearrangements (Kovarik *et al.*, 2012). To guarantee the inactivation of both genes, two pairs of gRNAs targeting the first (exons 1 and 3) and second halves (exons 13 and 14) of *GnTI.A* and *GnTI.B* were designed.

Complete inactivation of *GnTI* was expected to result in a cell line producing only high-mannose *N*-glycans. Western blotting analysis of secreted proteins from the *GnTI*-KO cell lines showed absence of β 1,2-xylose residues and an extensive reduction (30-to-100-fold) of α 1,3-fucose residues as compared to WT secreted proteins. Absence of *N*-glycans containing β 1,2-xylose residues demonstrates the suppression of the entire GnTI activity since GnTI-processed *N*-glycans would be recognized as substrates by Golgi-resident glycosyltransferases like xylosyltransferases (Strasser *et al.*, 2004b; Kaulfurst-Soboll *et al.*, 2011). Genomic analysis of the *GnTI*-KO #2 cell line confirmed the *GnTI* inactivation since all *GnTI* alleles sequenced contained Cas9-induced indels. Complete suppression of α 1,3-fucose residues was required because their elimination is desirable to produce glycoproteins lacking non-human epitopes susceptible to trigger allergic response (van Ree *et al.*, 2000; Altmann, 2007; Gomord *et al.*, 2010). To remove the residual α 1,3-fucose, combined inactivation of the two *GnTI* and the four *FucT* genes (twelve alleles) was carried out to generate *GnTI/FucT*-KO cell lines. Secreted and soluble intracellular proteins of three *GnTI/FucT*-KO cell lines (#5, #10, and #16) were totally devoid of any detectable β 1,2-xylose and α 1,3-fucose residues.

Mass spectrometry analysis of *N*-glycan structures of secreted glycoproteins from WT, *GnTI*-KO #2 and *GnTI/FucT*-KO #10 lines identified the major changes that

occurred in the KO lines. While most *N*-glycans in the WT cell line were of complex type, none were detected in either of the KO lines. Man4 and Man5 structures accounted for 96.1 and 96.9% of the *N*-glycans in the *GnTI*-KO #2 and *GnTI/FucT*-KO #10 cell lines, respectively. Unlike the Western blotting data, we could not detect any fucosylated structure in the *GnTI*-KO #2 cell line. However, this can be explained by the fact that the α 1,3-fucose signal observed by Western blotting in *GnTI*-KO #2 cell line was very weak. It probably belongs to fucosylated *N*-glycans distributed between different masses. Their individual MS signal drops below the detection limit of the method.

Detection of α 1,3-fucose on *N*-glycans in *GnTI*-KO cell lines was surprising and can only be explained by a relaxed substrate specificity of FucTs, unlike what is usually reported in the literature. Our hypothesis is that at least one of the four *N. tabacum* FucTs can recognize and process high-mannose *N*-glycans, although with a lower rate than hybrid and complex *N*-glycans. The lack of high-mannose *N*-glycans with α 1,3-fucose reported in most plant and plant cell studies is probably due to their very low levels. It could also be correlated with a higher affinity of FucTs for complex *N*-glycans. However, detection of Man5F *N*-glycans on human glucocerebrosidase and α -iduronidase produced in the *A. thaliana* seeds of the conditional *egl1-1* mutant has been reported several times (He *et al.*, 2012; He *et al.*, 2013; Pierce *et al.*, 2017). The authors linked the presence of Man5F to hexosaminidase activities on the hybrid structure GnMan5F. However, it is highly unlikely that only Man5F would be detected if this structure resulted from hexosaminidase processing. Indeed, neither GnMan5 with or without β 1,2-xylose or α 1,3-fucose nor the corresponding hexosaminidase-processed structures Man5XF and Man5X were detected in these studies. We rather suggest that the Man5F *N*-glycan observed in the *GnTI*-deficient host was not the result of hexosaminidase activity but the product of the α 1,3-FucT-catalyzed addition of a core fucose on some Man5 *N*-glycan structures present in very large amount because *GnTI* activity was lost. Similar loose substrate specificity was described for the mammalian core α 1,6-fucosyltransferase (FUT8). Core α 1,6-fucosylation by FUT8 in mammals was first believed to be strictly *GnTI*-dependent because α 1,6-fucosylated *N*-glycans identified by mass spectrometry were for the overwhelming majority hybrid or complex types. Furthermore, *in vitro* studies showed that FUT8 requires the presence of an α 1,3-arm GlcNAc (but not an α 1,6-arm GlcNAc) to use free *N*-glycans as substrates (Calderon *et al.*, 2016). In contrast to this result, *in vivo* observations showed that, in *GnTI*-deficient CHO and HEK293S cells, high-mannose *N*-glycans could be efficiently core-fucosylated (Lin *et al.*, 1994; Crispin *et al.*, 2006; Meuris *et al.*, 2014). In addition, recent *in vitro* Man5 fucosylation assays, using protein- or peptide-linked *N*-glycans instead of free *N*-glycans showed

that protein- and peptide-linked Man5 could be fucosylated to some extent, according to the glycopeptide or glycoprotein (Yang *et al.*, 2017). We propose a similar loose substrate specificity of some plant α 1,3-FucTs for Man5.

To identify the α 1,3-fucosylated structure(s) detected in *GnTI*-KO spent CM, affinity purification was attempted using agarose bound *Aleuria Aurantia* lectin, which binds to α 1,3-fucose bound to GlcNAc. After purification and concentration of proteins containing α 1,3-fucosylated N-glycans, the next step was deglycosylation and MS analysis. However, Western blotting showed no α 1,3-fucose signal in the elution fractions in contrast to the flow-through, indicating that the binding to this lectin was not efficient (data not shown). WT spent CM did not yield better results. As an alternative way of identifying the origin of the *GnTI*-independent α 1,3-fucose signal, recombinant glycoproteins will be produced both in the *GnTI/FucT*-KO and *GnTI*-KO cell lines. Then, site-specific N-glycosylation will be analyzed by LC-MS/MS in hope that some sites would be sufficiently enriched in α 1,3-fucosylated N-glycans to allow detection of the associated structures.

Man4 N-glycans are usually not mentioned when plant N-glycosylation pathways are described (Lannoo and Van Damme, 2015; Strasser, 2016; Schoberer and Strasser, 2018) and are rarely detected on glycoproteins. However, 16.4% and 11% of Man4 structures were detected in the spent culture media of *GnTI*-KO and *GnTI/FucT*-KO BY-2 cell lines, respectively. In comparison, only 1.1% of Man4 was detected in the WT D11b culture medium in this study and none in WT and *XylT/FucT*-KO MS culture media (Mercx *et al.*, 2017). Man4 content was measured at 4-5% in the TSCPs of WT and *XylT*-KO BY-2 cell lines but none in *FucT*-KO and *XylT/FucT*-KO cell lines (Hanania *et al.*, 2017). Man4 proportion in the TSCPs of rice calli also increased in *GnTI* knockouts, from undetected to 17.4% (Fanata *et al.*, 2013). On the other hand, knocking out *GnTI* in *A. thaliana* plants only increased Man4 proportion in the TSCPs from undetected to 1% (Strasser *et al.*, 2005). Moreover, knocking down *GnTI* in *N. benthamiana* did not result in the detection of Man4 in TSCPs from leaves (Limkul *et al.*, 2016). Surprisingly, Man4 structures were detected on all four N-glycosylation sites of recombinant glucocerebrosidase produced in the *N. benthamiana* *GnTI* knockdown plants, ranging from 3.5% to 23.9% while only complex structures were detected on the WT glycovariant. The presence of Man4 structures is not discussed in any of these articles.

According to the canonical plant N-glycosylation pathway (Figure 5), Man4 could be formed after the action of *GnTI*, from the processing of GlcNAcMan5GlcNAc2 hybrid structure by GMII, into GlcNAcMan4GlcNAc2. Then, the complex antenna can be processed by HEXOs, resulting in Man4GlcNAc2 structure. However, the

presence of Man4 on glycoproteins produced in *GnTI*-KO BY-2 cells, rice and *N. benthamiana* invalidates this hypothesis. Rather, Man4 could arise from an undocumented mannosidase activity on Man5 structure, for example by GMII. The absence or very low amount of Man4 in WT plants would be due to the lesser amount of Man5 substrate as well as a higher affinity of GMII for hybrid structures such as GlcNAcMan5GlcNAc2. Relaxed GMII substrate specificity could also explain the unexpected presence, in our *GnTI*-KO samples, of Man3 structures, which are supposed to be formed by the activity of HEXOs on complex *N*-glycans. This hypothesis could be tested by inactivating *GMII* in the *GnTI*(/*FucT*)-KO BY-2 cell line and monitor the change in *N*-glycan composition. Finally, to rule out any bias due to sample processing, comparative analysis should be carried out on the different *N*-glycan repertoires obtained by analyzing the same glycoprotein samples using different sample preparations and MS techniques.

GnTI-KO and *GnTI*/*FucT*-KO BY-2 cell lines produced highly homogeneous glycoproteins secreted in the culture medium with only two main *N*-glycan structures (Man4 and Man5). This *N*-glycoprofile is strongly recommended for some therapeutic glycoproteins (Grubb *et al.*, 2010; He *et al.*, 2012; He *et al.*, 2013; Tekoah *et al.*, 2015; Limkul *et al.*, 2016; Pierce *et al.*, 2017). Furthermore, a high *N*-glycan homogeneity can ease the protein purification as well as the process reproducibility, leading to more consistent therapeutic efficacy (Meuris *et al.*, 2014). As a result, the development of a BY-2 production platform with simplified and homogenized *N*-glycan repertoire, such as the *XylT*/*FucT*-KO cell lines (Hanania *et al.*, 2017; Mercx *et al.*, 2017) or the two glyco-engineered cell lines we generated, is an important step to produce highly efficient therapeutic glycoproteins and simplify vaccine production in a system free of potential mammalian pathogens.

2. *In vivo* deglycosylation of recombinant glycoproteins in *Nicotiana tabacum* BY-2 cells

2.0. Contextualization

In this chapter, production of *in vivo* deglycosylated proteins in BY-2 cells was attempted. While glycosylation is a critical feature for some recombinant glycoproteins, the presence of *N*-glycans might be dispensable for others. The proteins whose properties are not significantly affected by the presence of *N*-glycans could benefit from a glycosylation shortening strategy. Indeed, removal of *N*-glycans suppresses the presence of non-human residues, increases homogeneity and simplifies downstream processing. In addition, absence of *N*-glycan structures can confer useful properties to the glycoprotein, even if the native form is glycosylated. In particular, vaccine antigens could benefit from the suppression of *N*-glycans, which are abundant on viral envelope proteins and act as a shield, hiding peptide epitopes from the host immune system.

This chapter is in preparation for submission to Plant Biotechnology Journal.

XH, CN and FC: conceptualization. XH and JF: methodology. XH, JF, MP and CN: investigation. XH, JF, and CN: data curation. XH: writing – original draft preparation. CN, JF and FC: writing – review and editing. CN, LQ and FC: supervision and funding acquisition.

Xavier Herman¹, Johann Far², Marie Peeters¹, Loïc Quinton², François Chaumont^{1*}, Catherine Navarre^{1*}

¹Louvain Institute of Biomolecular Science and Technology, UCLouvain, Louvain-la-Neuve, Belgium

²Mass Spectrometry Laboratory-MolSys Research Unit, ULiege, Liège, Belgium.

2.1. AtXylT-EndoT is localized in the Golgi apparatus and propeptide-processed in BY-2 cells

The N-terminal signal sequence of EndoT was swapped for the *medial*-Golgi targeting signal of AtXylT as proposed by Piron *et al.* (2015). The sequence encoding the N- and C-terminal propeptides of EndoT were included in the construct to ensure correct folding of the protein (Figure 27A). This chimeric *EndoT* gene was driven by the enhanced CaMV 35S promoter (*En2p35S*) and cloned into a pPZP-*hpt*

binary vector to obtain transgenic BY-2 cells via *Agrobacterium tumefaciens*-mediated transformation. In addition, to further confirm the localization of EndoT in the Golgi apparatus, the chimeric EndoT was fused to the Venus fluorescent protein at C-terminus.

EndoT-Venus cell lines were imaged by confocal microscopy and the Venus signal was partly detected in typical Golgi-like mobile vesicles localized near the plasma membrane and nucleus as well as in transvacuolar cytoplasmic strands (Figure 27B).

The distribution of EndoT in cell extracts from *EndoT*- and *EndoT-Venus*-expressing cell lines was then investigated by Western blotting using antibodies raised against EndoT (Figure 27C). In the microsomal fractions, bands corresponding to EndoT-Venus and EndoT were detected at around 70 and 43 kDa, respectively, corresponding to the molecular weight of the full-length, unprocessed proteins. A band around 32 kDa was also detected in the total soluble cellular proteins (TSCP) and culture medium (CM) of both lines. Since EndoT and EndoT-Venus contained N- and C-terminal endogenous propeptides, processing of the precursor in BY-2 cells can be expected. This would release soluble EndoT in the Golgi lumen and, eventually, in the CM. EndoT-Venus-containing fractions were also analyzed using antibodies raised against GFP, which also recognize Venus protein (Figure S7). Each fraction yielded the expected signal, *i. e.* a band around 70 kDa in microsomal fractions, corresponding to unprocessed EndoT-Venus and around 27 kDa in the TSCPs and CM, corresponding to soluble Venus, resulting from EndoT propeptide processing. Western blotting with antibodies raised against FLAG tag was not conclusive due to the presence of non-specific signals (data not shown).

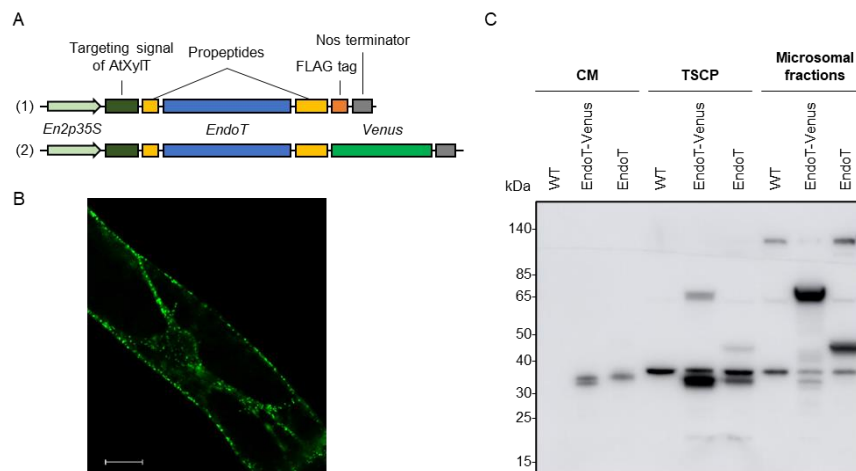


Figure 27. Localization of AtXylT-EndoT and *in vivo* processing in BY-2 cells. (A) Schematic representation of AtXylT-EndoT (1) and AtXylT-EndoT-Venus (2) expression cassettes. (B) Localization of EndoT-Venus fused to the signal-anchor of *A. thaliana* β1,2-xylosyltransferase. λ_{exc} : 514 nm; λ_{em} : 537 nm. Scale bar: 20 μm. (C) Thirty-five μL of 7-day-old culture medium (CM), 35 μg of total soluble cellular proteins (TSCP) and microsomal fractions from wild type (WT), EndoT and EndoT-Venus BY-2 cell cultures were analyzed by Western blotting using an anti-EndoT serum and a horseradish peroxidase-linked secondary antibody. Hollow (○) and full (●) dots indicate bands, the size of which correspond to unprocessed EndoT-Venus and EndoT, respectively, whereas a star (*) indicates processed EndoT and EndoT-Venus. Diamond (◇) mark signals from endogenous proteins of BY-2 cells.

2.2. *N*-glycan composition of *GnTI*-KO and *GnTI/FucT*-KO BY-2 cell lines is suitable for EndoT processing

EndoT has a specific activity towards high mannose *N*-glycans. As a result, to generate mono-GlcNAc proteins, this enzyme had to be expressed in a BY-2 cell line devoid of complex *N*-glycans. Such BY-2 cell lines were previously obtained by knocking out *GnTI* as well as *GnTI* and *FucT* glycosyltransferase genes and mass spectrometry analysis of secreted glycoproteins from one line of each cell type showed a very similar *N*-glycan composition (Herman *et al.*, 2021). To investigate the reproducibility of the *N*-glycan content in several independent transgenic cell lines, four *GnTI*-KO and three *GnTI/FucT*-KO cell lines were grown in D11b production medium, which was optimized for the accumulation of a human antibody in BY-2

cells (Vasilev *et al.*, 2013). The secreted glycoproteins were analyzed by MALDI FT-ICR mass spectrometry (Figure S8). A highly reproducible and similar *N*-glycome was obtained for both *GnTI*-KO and *GnTI/FucT*-KO cell lines, with mainly Man5 (87% and 89%, respectively) and Man4 (9% and 7%, respectively) structures. The remaining structures, accounting for only 4% of *N*-glycans, are all oligomannosidic structures (Man3-9). The *N*-glycan distribution obtained from the *GnTI*-KO cell lines grown in Murashige and Skoog medium (MS) (Figure S8) was very similar to that observed for D11b cultures, with mainly Man5 (80%) and Man4 (14%) glycans. No α 1,3-fucosylated *N*-glycan structure was detected in *GnTI*-KO cell lines. Hence, we concluded that *N*-glycans found on secreted glycoproteins of *GnTI*-KO and *GnTI/FucT*-KO cell lines grown in either MS or D11b media are suitable for EndoT processing.

2.3. Expression of *EndoT* strongly reduces *N*-glycosylation

The *N*-glycans decorating cellular glycoproteins from WT, *GnTI/FucT*-KO and *GnTI/FucT*-KO-*EndoT* cell lines were compared by Western blot analysis using the lectin concanavalin A, which binds high mannose *N*-glycans (Figure 28). The low signal observed in WT samples is expected, considering the presence of only 20% high mannose *N*-glycans in the cellular proteins (Hanania *et al.*, 2017). The intense signal obtained for the *GnTI/FucT*-KO cell lines is in agreement with the high proportion of high mannose *N*-glycans. On the contrary, the strong reduction of the signal observed in *GnTI/FucT*-KO-*EndoT* cell lines suggests that a large part of the high mannose *N*-glycans was processed by EndoT. The reduction was estimated at ~75% by semi-quantitative analysis of Western blot signals.

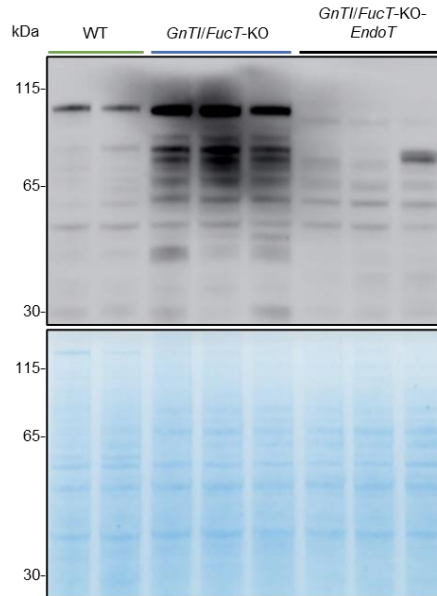


Figure 28. Deglycosylation of endogenous glycoproteins by EndoT. Twenty-five μ g of TSCP from WT cell lines (lane 1: WT-*gB* 23, lane 2: WT), *GnTII/FucT-KO* cell lines (*GnTII/FucT-KO-gB* #1, 5 and 16) and *GnTII/FucT-KO-EndoT* cell lines (*GnTII/FucT-KO-EndoT-gB* #19, 29, 30) were analyzed by Western blotting using concanavalin A coupled to horseradish peroxidase. A colloidal blue gel is displayed as a loading control.

2.4. *N*-glycosylation engineering negatively affect the growth kinetics of BY-2 cell lines

The impact of *GnTII* inactivation as well as *EndoT* expression on cell growth kinetics was explored in MS medium (Figure S9). Depletion of complex *N*-glycans led to a slower growth (detectable already in early exponential phase, from day 4), resulting in a 11% drop in maximal fresh weight (FW) at day 7. Expression of EndoT in high mannose cell lines seems to further reduce the growth rate with a stronger drop (29.5% as compared to WT cell lines) of maximal FW. Furthermore, while cell density is stable between day 7 and day 10 for WT cell lines (1% growth), it drops for high mannose cell lines (-12%) and EndoT-expressing lines (-34%).

2.5. Human IgG is successfully expressed by and purified from glyco-engineered cell lines

The following lines were obtained: WT, *XyIT/FucT*-KO, *GnTI*-KO and *GnTI/FucT*-KO cell lines expressing the human immunoglobulin 2 (IgG2) anti-IgM Lo-BM2; *GnTI*-KO and *GnTI/FucT*-KO cell lines co-expressing IgG2 and EndoT. By expressing IgG2 in a *GnTI*-KO background, we also sought to detect the α 1,3-fucosylated *N*-glycan structures responsible for the very low signal observed by Western blot analysis by Herman *et al.* (2021). To select elite cell lines, IgG2 accumulation in the CM was evaluated by Western blotting (data not shown). The presence of EndoT in the total cellular proteins (TCP) was assessed by Western blotting (Figure S10). EndoT could be detected in most *GnTI*-KO-*EndoT* and *GnTI/FucT*-KO-*EndoT* cell lines, with sizes corresponding to full-length (~43 kDa) and propeptide-processed (~32 kDa) proteins.

The various IgG2 glycovariants were purified from spent D11b CM through protein A affinity chromatography (Figure S11). From now on, purified IgG2 glycovariants will be named according to their expected main *N*-glycan structure: IgG2_{FG} for fully glycosylated (WT-IgG2); IgG2_{ΔX/ΔF} for xylose/ α 1,3-fucose-deleted (*XyIT/FucT*-KO-IgG2); IgG2_{HM} for high mannose (*GnTI*-KO-IgG2 and *GnTI/FucT*-KO-IgG2) and IgG2_{MG} for mono-GlcNAc (*GnTI*-KO-*EndoT*-IgG2 and *GnTI/FucT*-KO-*EndoT*-IgG2). Cell lines, major *N*-glycan structure expected and glycovariants are summarized in Table S4.

2.6. Glyco-engineering strategies confer the expected *N*-glycosylation profile to IgG glycovariants

The glycosylation profile of purified IgG2 was then analyzed by liquid chromatography followed by tandem quadrupole orbitrap mass spectrometry (LC-MS/MS). *N*-glycans detected on the conserved *N*-site EEQFN*STFR of the heavy chain CH2 domain are summarized in Figure 29 and Table S5. No α 1,3-fucose-containing *N*-glycans were detected in *GnTI*-KO cell lines. The *N*-glycosylation profile of IgG2 purified from *GnTI*-KO and *GnTI/FucT*-KO cell lines were highly similar. The EEQFN*STFR site had a *N*-glycosylation occupancy of 98.7% for IgG2_{HM} and 99.7% for IgG2_{MG}. As anticipated, *N*-glycan diversity was low for IgG2_{HM} and IgG2_{MG}, with only five and two different *N*-glycan structures detected at more than 1%, respectively. Man5 was the main structure detected for IgG2_{HM}, accounting for 66% of CH2 *N*-glycans. The majority of *N*-glycans detected for IgG2_{MG} (93%) were trimmed down to mono-GlcNAc structures, indicating a highly

efficient *in vivo* processing of the N-glycans at this site by EndoT. The remaining structures identified were mainly Man5 (6%), along with traces of other oligomannosidic structures.

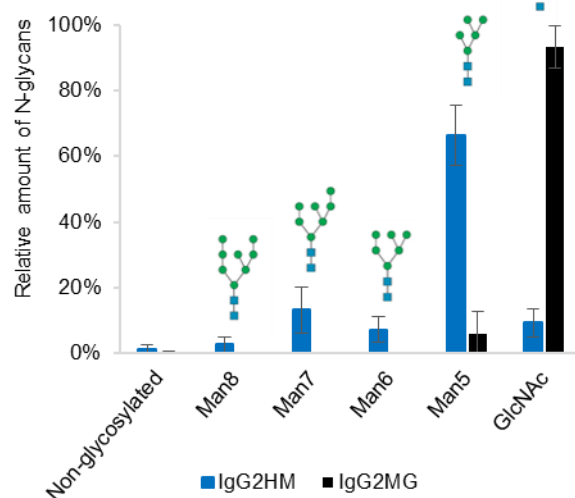


Figure 29. Relative amount of N-glycans on the N-site EEQFN*STFR of secreted IgG2. N-glycosylation was determined by LC-MS/MS of IgG2 samples purified from 10 days-old D11b culture medium. Error bars represent standard deviation of $n = 5$ (IgG2_{HM}: *GnTI*-KO-IgG2 #12, 25, 30, *GnTI/FucT*-KO-IgG2 #17, 20) and $n = 4$ (IgG2_{MG}: *GnTI*-KO-*EndoT*-IgG2 #9, 12; *GnTI/FucT*-KO-*EndoT*-IgG2 #6, 16) independent cell lines. Structures accounting for less than 1% of N-glycans were not represented.

Besides the conserved EEQFN*STFR N-glycosylation site in the CH2 domain, the variable region of the Lo-BM2 heavy chain contained the additional predicted N-glycosylation site, CVRPN*YSGF. It displayed a much lower occupancy rate of only 9% and 11% for IgG2_{HM} and IgG2_{MG}, respectively (Figure S12A and Table S5). N-glycans observed at this site were mostly Man5 for both IgG2_{HM} and IgG2_{MG} (89% and 70% of the glycosylated forms, respectively). A small proportion of mono-GlcNAc structures can be detected on IgG2_{MG} (23% of the glycosylated forms) but not on IgG2_{HM}.

Intriguingly, a third N-glycosylation site was also identified on the non-canonical N-site (TVSWN*SGAL) with a significant occupation rate of 13% (Figure S12B and Table S5). Similar to the two other N-sites, the main structure identified was Man5.

N-glycosylation was similar in *GnTI*-KO and *GnTI/FucT*-KO backgrounds (Figure S13).

2.7. Glycovariants display different thermal stability

Having an antibody with a wide range of *N*-glycosylation profiles allowed us to question the effect of this glycosylation on the stability of the protein. A differential scanning fluorimetry (DSF) assay, also called thermal shift, was carried out to compare the thermal stability of the different IgG2 glycovariants produced in BY-2 cells. Besides the IgG2_{FG}, IgG2_{ΔX/ΔF}, IgG2_{HM} and IgG2_{MG} produced in this study, IgG2_{FG} and IgG2_{ΔX/ΔF} previously purified from independent transgenic cell lines (Magy *et al.*, 2014, Mercx *et al.*, 2017) were also analyzed. The melting temperature (*T*_m) was calculated based on the first derivative of the melting curves (Figure S14A and B). DSF assay with new and old IgG2_{FG} and IgG2_{ΔX/ΔF} samples gave similar results, thereby confirming the robustness of the BY-2 production platform (Figure S15). The lowest *T*_m, 60.5°C, was obtained for IgG2_{FG} while the highest *T*_m, 66.9°C, was observed for IgG2_{ΔX/ΔF}. IgG2_{HM} and IgG2_{MG} resulted in intermediate *T*_m of 63.4°C and 61.9°C, respectively (Figure 30A). A drastic increase in thermal stability (+6.4°C) was thus observed when comparing IgG2_{FG} and IgG2_{ΔX/ΔF}, demonstrating that core-fucosylation and/or xylosylation impact thermal stability of the IgG2. To the best of our knowledge, this has never been reported in the literature. The effect of *N*-glycans on the thermal stability of IgG1 was also assessed, using hIgG1 Lo-BM2 produced in WT and *XyIT/FucT*-KO BY-2 cells. The respective DSF *T*_m, 62.1°C (IgG1_{FG}) and 70.1°C (IgG1_{ΔX/ΔF}), were consistent with those obtained for IgG2 (Figure 30B; Figure S14B and C).

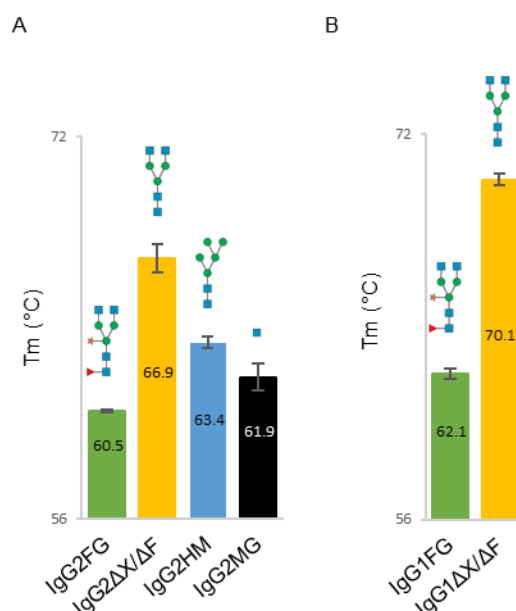


Figure 30. Thermal stability of IgG glycovariants. (A) T_m of IgG2 glycovariants secreted in BY-2 CM measured by differential scanning fluorimetry analysis. (B) T_m of IgG1 glycovariants secreted in BY-2 CM measured by DSF analysis. IgG were purified from two (IgG2_{FG}), two (IgG2_{ΔX/ΔF}), six (IgG2_{HM}) and three (IgG2_{MG}) independent cell lines and n = 6 for technical replicates. T_m was obtained as the mean of the T_m calculated for each independent sample. Error bars represent standard deviation of the mean T_m obtained for the independent cell lines. As an indication, the main N-glycan structure of each glycovariant at the CH2 N-site is represented above the corresponding bar. IgG1 N-glycans are the expected structures based on IgG2, it was not analyzed by MS.

2.8. Glycovariants of glycoprotein B from HCMV display different electrophoretic sizes

The ability to produce gB glycovariants constitutes a powerful tool to study the impact of N-glycosylation on immunogenicity and hence develop better vaccines. The gB ectodomain was already successfully produced as a secreted glycoprotein containing both complex and high mannose structures (gB_{FG}) in WT BY-2 cells (Smargiasso *et al.*, 2019). Due to the eighteen N-glycosylation sites in gB ectodomain,

it exists as a population of glycoforms characterized by an important meta-, macro and micro-heterogeneity (variations of glycan structures across multiple sites of a protein, occupancy and variation of glycans at a given site, respectively (Čaval *et al.*, 2021)). This heterogeneity results in gB glycoforms with different sizes, characterized by a diffuse electrophoretic mobility.

Here, a version of gB devoid of β 1,2-xylose and α 1,3-fucose (gB $_{\Delta X/\Delta F}$) was obtained in the *XyT/FucT*-KO cell line. We also produced gB in *GnTI*-KO and *GnTI/FucT*-KO cell lines (gB $_{HM}$). Expression of gB in TSCP of those cell lines was analyzed by Western blotting (Figure 31A). gB $_{HM}$ was detected as both an intact protein at ~100-105 kDa and degradation fragments, mainly at 25-30 kDa. Intact gB $_{HM}$ was slightly smaller than gB $_{FG}$ and gB $_{\Delta X/\Delta F}$. This variation in size likely resulted from differences in *N*-glycan composition.

Next, the gB ectodomain was co-produced with EndoT in the *GnTI/FucT*-KO cell line (gB $_{MG}$). TCP and TSCP from *GnTI/FucT*-KO-*EndoT*-gB cell lines were analyzed by Western blotting using antibodies against gB or EndoT (Figure 31B and C, respectively). gB was detected in most transgenic lines (e.g., #17, 18, 19, 23, 25, 29, 30, 33, 34) (Figure 31B). EndoT-specific signal was observed, but only for some transgenic cell lines (e.g., #19, 29, 30, 34) (Figure 31C). The bands corresponding to intact gB showed distinct patterns according to the cell lines. In cell lines without EndoT production, the migration behavior was similar to gB $_{HM}$ (e.g., #17, 18, 23, 25, 33), while in EndoT-expressing cell lines (e.g., #19, 29, 30, 34), the bands corresponding to intact gB $_{MG}$ displayed a remarkably shifted mobility in regards to gB $_{HM}$. This mobility shift was attributed to the *in vivo* processing of high mannose *N*-glycans decorating gB by EndoT. It was observed as a smear lower than gB $_{HM}$, compatible with processed glycosylated gB subjected to various levels of *N*-glycan truncation.

To monitor gB $_{HM}$ accumulation in the culture medium, two *GnTI*-KO-gB (#2, 4) and two *GnTI/FucT*-KO-gB (#5, 16) cell lines were grown in D11b medium (Figure 31D). Intact gB $_{HM}$ was already detected after four days of culture and the signal culminates at day seven or ten. Accumulation of secreted gB in *GnTI/FucT*-KO-*EndoT*-gB CM was also investigated by Western blotting, both in MS and D11b media, after 5, 7, 10 and 13 days of culture (Figure S16). However, no intact gB $_{MG}$ could be detected although gB $_{MG}$ degradation fragments were detected in higher amount than in the control *GnTI/FucT*-KO-gB #5 line, suggesting that gB $_{MG}$ was either less stable than gB $_{HM}$ or more subject to protease-mediated proteolysis.



Figure 31. Expression of gB HCMV in glyco-engineered BY-2 cell lines. Twenty μ g of TSCP from *GnTI*-KO-gB and *GnTI/FucT*-KO-gB cell lines (A) and *GnTI/FucT*-KO-*EndoT*-gB (B) cell lines and indicated controls were analyzed by Western blotting using antibodies raised against gB and AP-coupled secondary antibodies. Signal was detected using a precipitating substrate. gB_{FG} from WT-gB cell line is displayed as a positive control. (C) Twenty μ g of TCPs from *GnTI/FucT*-KO-*EndoT*-gB cell lines and indicated controls were analyzed by Western blotting using antibodies raised against EndoT and AP-coupled secondary antibodies. Signal was detected using a chemiluminescent substrate. (D) Accumulation of gB_{HM} in the culture medium of indicated cell lines. Thirty-five μ L of CM from 4, 7 or 10-day-old D11b cultures were analyzed by Western blotting using anti-gB antibody and AP-linked secondary antibody. (E) Deglycosylation gel shift assay for gB glycovariants. Immunodetection of gB in the indicated cell lines without deglycosylation (-) and after denaturation and deglycosylation by EndoH (E) or PNGase F (P). Twenty μ g of TSCP from WT-gB #23 (gB_{FG}), *XyI*/*FucT*-KO-gB #4 (gB _{Δ X/ Δ F}), *GnTI/FucT*-KO-gB #5 (gB_{HM}), *GnTI/FucT*-KO-*EndoT*-gB #29 (gB_{MG}) cell lines were loaded. Primary and secondary antibodies were anti-gB antibodies and AP-linked antibodies, respectively.

2.9. Glycoprotein B glycovariants are differentially processed *in vitro* by glycosidases

To further study and compare the *N*-glycosylation profile of gB_{FG}, gB _{Δ X/ Δ F}, gB_{HM} and gB_{MG}, an *in vitro* deglycosylation assay was carried out using either peptide-*N*-glycosidase F (PNGase F) or endoglycosidase H (EndoH). PNGase F cleaves *N*-glycans that are not α 1,3-fucosylated. EndoH, like EndoT, cleaves high mannose (and some hybrid) *N*-glycans to produce mono-GlcNAc *N*-glycosylated asparagines.

TSCPs from WT-gB, *XyI*/*FucT*-KO-gB, *GnTI/FucT*-KO-gB and *GnTI/FucT*-KO-*EndoT*-gB cell lines were denatured, incubated with PNGase F or EndoH and then resolved by gel electrophoresis and Western blotting (Figure 31E). A slight mobility shift was observed for gB_{FG} after the action of EndoH or PNGase F. This result was expected since most complex structures of gB_{FG} contain α 1,3-fucose and therefore cannot be processed by EndoH or PNGase F (Smargiasso *et al.*, 2019). The small (but still detectable) mobility shift was therefore related to the processing of the few *N*-glycans that were oligomannosidic and were equally cleaved by both glycosidases. EndoH-mediated deglycosylation of gB _{Δ X/ Δ F} resulted in an incomplete shift while PNGase F-mediated deglycosylation was compatible with a completely deglycosylated gB (predicted size: 73 kDa), highlighting the presence of non-fucosylated complex *N*-glycans (such as GnGn), which are substrates for PNGase F but not EndoH. Deglycosylation of gB_{HM} with either EndoH or PNGase F both

resulted in the large mobility shift compatible with complete deglycosylation. This clearly demonstrates the dramatic increase in high mannose *N*-glycan structures decorating gB_{HM} when produced in *GnTI/FucT*-KO cell lines. Finally, complete deglycosylation of gB_{MG} was also obtained with both EndoH and PNGase F, generating highly similar bands as with gB_{HM}.

Deglycosylation assay was also carried out with the culture medium collected by filtration from *GnTI/FucT*-KO-gB cell cultures. Similar results as the assays using TSCPs were obtained (Figure S17).

Finally, to investigate the stability and EndoT-mediated processing of gB_{HM} in the CM, gB_{HM}-containing CM was spiked in CM from *GnTI/FucT*-KO and *GnTI/FucT*-KO-*EndoT* cell lines (Figure S18). After 22 h incubation at 25°C, no change in size nor intensity of gB signal was observed by Western blotting, suggesting that EndoT is poorly active or too diluted in the CM and that most of the degradation occurs either in the secretory pathway or in the apoplasm of BY-2 cells.

2.10. Discussion

In this study, we generated BY-2 cell lines producing deglycosylated proteins via expression of EndoT in a high mannose *N*-glycosylation background and validated them for molecular farming by producing a human IgG2 as well as the ectodomain of the glycoprotein B from HCMV envelope. Lectin blot analysis, *in vitro* deglycosylation assays and site-specific glycoproteomics confirmed the truncation of *N*-glycans decorating endogenous and recombinant glycoproteins by EndoT. Noticeably, more than 93% of *N*-glycans found at the conserved Asn residue in the CH2 domain of IgG2 were effectively shortened to a single GlcNAc. IgG and gB glycovariants were also successfully obtained using a wild type cell line, a cell line devoid of xylose- and α 1,3-fucose non-human residues (*XyII/FucT*-KO) and high mannose cell lines (*GnTI*-KO and *GnTI/FucT*-KO). Mass spectrometry experiments confirmed the expected *N*-glycosylation profile and revealed a high reproducibility. Besides the interest of controlling the type of *N*-glycan present on recombinant proteins to tune their properties, reproducibility and homogenization of *N*-glycan content are major assets to ease quality control for pharmaceutical proteins.

The comprehensive *N*-glycosylation profile of IgG2 glycovariants was determined by LC/MS-MS. Two *N*-glycosylation sites were predicted based on the consensus N-X-S/T sequence (where X is any amino acid but a proline). The extensively studied glycopeptide EEQFN*STFR in the conserved CH2 domain showed a very high

occupancy rate (99%-100%, Figure 29, Table S5). The second putative glycopeptide CVRPN*YSGF located in the variable region of the Lo-BM2 heavy chain, displayed a much lower occupancy rate (9-11%) (Figure S12A, Table S5). This could be related to the presence of a Ser in the +2 position, which was associated with a lower fraction of occupancy than Thr (Kasturi *et al.*, 1995). It could also be linked to an unfavorable amino acid context or steric hindrance. An additional *N*-glycosylation site was detected at the non-canonical sequence TVSWN*SGAL in the conserved CH1 domain of the heavy chain. This atypical *N*-site had already been reported as a backwards consensus *N*-glycosylation motif (S/T)XN (Valliere-Douglass *et al.*, 2010). It had been already reported on antibodies from human serum and had been also detected on human IgG1 and IgG2 produced in CHO and HEK293 cells although the *N*-glycan occupancy level was only 0.5 to 2% (Valliere-Douglass *et al.*, 2009) as compared to the 13% measured in this study for BY-2-produced IgG2. Considering that even low-level glycosylation could have a significant impact on physiological or pharmacological effects, mapping of all *N*-glycosylation sites was an important feature. Detection of the non-canonical *N*-glycosylation site of the IgG was made possible by the use of Multi-Enzymatic Limited Digestion (MELD) instead of the classic trypsin-only treatment (that would result in a 58-residue peptide including TVSWN*SGAL). This MELD technique allows for the generation of a more diverse (glyco)peptides library and exhaustive determination of post-translational modifications.

Analysis of the *N*-glycosylation library found at the CH2 *N*-glycosylation site showed that the glyco-engineered cell lines produced the anticipated glycoforms. The main *N*-glycan found on IgG2_{HM} (Man5) and IgG2_{MG} (mono-GlcNAc) accounted for 66% and 93% of the *N*-glycans, respectively (Figure 29, Table S5). The main *N*-glycans of IgG2_{FG} (GnGnXF) accounted for 74.5% of all *N*-glycans and IgG2_{ΔX/ΔF} (GnGn) for 69%, as determined by Navarre *et al.* (2017) and Mercx *et al.* (2017), respectively. Thus, production of an antibody in glyco-engineered BY-2 cell lines generates glycovariants with one defined, predominant *N*-glycan structure.

Thermal stability of IgG2 and IgG1 glycovariants was measured by DSF. IgG2 showed a single transition in the DSF graphs while IgG1 displayed two transition peaks. A single transition peak indicates an overlap of the different domain denaturation signals. The presence of two peaks, in contrast, suggests that at least two domains denature at different temperatures. The IgG CH3 domain is reported to unfold at a higher temperature than the CH2 domain (Meuris *et al.*, 2014; Akazawa-Ogawa *et al.*, 2018). Hence, we speculated that the first IgG1 peak could correspond to CH2 domain while the second one would be for the CH3 domain. This was further

confirmed with IgG1_{ΔX/ΔF} by the change in T_m observed for the first peak of IgG1_{FG} and IgG1_{ΔX/ΔF}, while the second T_m was stable around 75°C (Figure S14). Indeed, given that the CH2 domain is the one with the highest N-glycan occupancy, it is the most susceptible to be (de)stabilized by glyco-engineering. Since no other transition peak was observed, we hypothesized that the other domains denatured simultaneously with CH2 or CH3.

Relative CH2 T_m measured for IgG2_{ΔX/ΔF}, IgG2_{HM} and IgG2_{MG} in this study were consistent with those previously reported in the literature by DSF, differential scanning calorimetry and intrinsic fluorescence spectroscopy for IgG1 (Zheng *et al.*, 2011; Meuris *et al.*, 2014; Wada *et al.*, 2018). Indeed, these techniques showed that GnGn N-glycosylation results in a higher T_m than Man5, which in turn is more stable than the mono-GlcNAc stump. However, to the best of our knowledge, this is the first thermal stability study of an antibody with typical plant N-glycans (GnGnXF). IgG2_{FG} displayed the lowest T_m among all glycovariants. To consolidate this observation, DSF was also carried out with IgG1_{FG} and IgG1_{ΔX/ΔF} and provided similar results, IgG1_{ΔX/ΔF} being significantly more stable than IgG1_{FG} (ΔT_m = +8°C). This result suggests that the presence of β1,2-xylose and/or α1,3-fucose residues triggers a destabilizing effect on IgG2 and IgG1. By forming non-covalent interactions with hydrophobic residues in Fc region, N-glycans participate in maintaining the structural integrity of the CH2 domain (Wada *et al.*, 2018). Moreover, glycan-glycan interactions have been shown to be crucial to maintain the tertiary structure of CH2 domains as well as Fc quaternary structure (Buck *et al.*, 2013). The Fc apparent molecular hydrodynamic radius (corresponding to the radius of a spherical particule with the same diffusion coefficient) increases after deglycosylation, a probable outcome of the structural deformation caused by the disruption of glycan interactions (Zheng *et al.*, 2011). Hence, N-glycan engineering of IgG not only impacts the function of the antibody, but also its higher order structure and structural dynamics.

The effect of Fc N-glycosylation on IgG1 stability was investigated by Wada and colleagues (2018). In that study, an extensive set of homogeneous IgG1 glycoforms was generated *in vitro* via chemoenzymatic transglycosylation, a technique based on the use of Endoglycosidase S2 to deglycosylate an IgG and Endo S2 D184M mutant to transfer the desired N-glycan from synthetic glycan-oxazoline substrates to the acceptor IgG. The main limitations of this technique are the specificity of Endo S2 for IgG Fc and the costs associated with Endo S enzymes and glycan-oxazolin substrates. Antibody stability (thermal stability, aggregation propensity) and Fc effector function (binding properties to Fc receptor and complement component,

ADCC activity) assays revealed a very specific impact of different Fc N-glycosylations. Consistent with our findings, thermal stability (measure by differential scanning calorimetry and intrinsic fluorescence spectroscopy) of high mannose IgG1 was reduced compared to different complex glycoforms (including GnGn but not GnGnXF) and IgG1_{MG} stability was even lower. The presence of typical mammalian core α 1,6-fucose residues slightly increased the stability of most IgG1 glycoforms, as opposed to our findings on IgG_{FG}. However, the absence of xylose and the difference in fucose linkages should refrain the comparison with the results obtained for IgG_{FG} and IgG _{Δ X/ Δ F}. Generation of IgG _{Δ X} and thermal stability analysis would shed light on the impact of the core-fucose linkage (and core-xylose presence).

The production of a more complex, heavily glycosylated, recombinant glycoprotein in these new cell lines was also evaluated. The ectodomain of gB was chosen because it is a vaccine candidate against HCMV and was already produced in WT BY-2 cells (Smargiasso *et al.*, 2019). Precise N-glycan composition of gB_{FG} had already been determined but this information was not known for the other glycovariants. Interestingly, gB glycovariants showed different electrophoretic mobility on Western blotting because of the different N-glycans decorating them: gB_{FG} and gB _{Δ X/ Δ F} (that lacks β 1,2-xylose and α 1,3-fucose residues) electrophoretic sizes were slightly higher than gB_{HM} (probably decorated mainly with Man5), while an important size shift was detected for gB_{MG} (Figure 31A, B and E).

In vitro deglycosylation assay of denatured gB glycovariants coupled to immunostaining was carried out to obtain a global insight in N-glycan compositions (Figure 31E). The size of electrophoretic mobility shifts was consistent with expected gB glycoforms and IgG2 glycovariant glycoproteomics. Particularly, it confirmed that an important part of gB_{MG} N-glycans, but not all, were deglycosylated *in vivo* by EndoT.

While full-length gB_{HM} was observed in the culture medium (Figure 31D), no gB_{MG} could be detected (Figure S16). This could be the direct result of gB deglycosylation by EndoT. Indeed, removal of N-glycans in the Golgi could unmask protease cleavage sites, decrease stability through conformational changes or favor aggregation by exposing hydrophobic residues. However, since N-glycans of endogenous proteins were also modified, we could not exclude an indirect, more global, effect such as an increase in some protease activity. The change in growth kinetics observed after N-glycan modification (Figure S9) could also impact gB_{MG} accumulation indirectly. Although spiking of gB_{HM} in spent CM did not result in any detectable degradation, spiking of *in vitro* deglycosylated gB_{HM} in BY-2 cell CM and monitoring of the quantity and degradation profile at different time points by

Western blotting would help to determine the impact of deglycosylation on gB_{MG} stability. The lack of impact of the spiking would suggest, as for gB_{HIM}, that most of the degradation occurs intracellularly or in the apoplasm.

Prior GlycoDelete-like strategies reported on a 95-100% efficiency of EndoT *in vivo* processing in HEK293S, *A. thaliana* seeds, *P. pastoris* and *A. oryzae* (Meuris *et al.*, 2014; Piron *et al.*, 2015; Wang *et al.*, 2020; Li *et al.*, 2020). For all four studies, these efficiency rates were calculated based on the *N*-glycosylation content of one or two recombinant proteins containing one to three *N*-glycosylation sites. These EndoT processing rates were comparable to that obtained here for human IgG2. Indeed, CH2 *N*-glycosylation site of IgG2_{MG} was occupied at 93% on average with mono-GlcNAc (up to 98% for cell line *GnTI/FucT-KO-EndoT-IgG2* #16). However, the EndoT processing rate seemed to be highly dependent on the protein context, as illustrated by the variation in mono-GlcNAc content between the two IgG2_{MG} canonical *N*-sites. Indeed, the CH2 site had a processing rate of 93% while the Fab site showed an EndoT processing rate of 23% (2.6% of the 11.2% *N*-glycans were mono-GlcNAc, Figure S12A). Hence, herein, we challenged these results further by assessing the EndoT deglycosylation level of an 18 *N*-glycosylation sites protein, gB (Figure 31E). We suggest that accessibility of the *N*-glycosylation site to EndoT determined the mono-GlcNAc content, in the same way as the *N*-glycan libraries of different *N*-glycosylation sites varied because they were not equally processed by glycosyltransferases in the secretory pathway. Therefore, the efficiency of the GlycoDelete-like strategy is dependent on the recombinant glycoprotein to be produced.

Additional *in vitro* deglycosylation assays of gB glycovariants without prior denaturation should be carried out to compare the extent of deglycosylation with *in vivo* deglycosylated gB. More precisely, it would shed light on the intrinsic deglycosylation limit of gB_{HIM} by EndoT due to *N*-glycosylation site accessibility. Although EndoH and PNGase F-mediated *in vitro* deglycosylation of native gB would be informative, deglycosylation by EndoT is required to accurately determine *N*-site accessibility for this enzyme. However, EndoT is not available commercially. It could be obtained from Prof. N. Callewaert laboratory since Meuris *et al.* (2014) produced it in *P. pastoris* for rabbit immunization and generation of anti-EndoT antibodies. Alternatively, it could be purified in our laboratory from the CM of EndoT-expressing BY-2 cells. Purification of the endogenous enzyme (like commercial PNGase F) or from transgenic *E. coli* (like commercial EndoH) could also be attempted.

If the deglycosylation rate of some recombinant proteins by EndoT needs to be improved, several strategies can be undertaken. First, if it does not affect its folding, expression of EndoT without the two propeptides might be beneficial. Indeed, it is highly likely that EndoT is activated via propeptide cleavage. As a result, not including the propeptides in the genetic construct would result in EndoT being active as soon as it is correctly folded, which is susceptible to increase the deglycosylation rate. Next, use of different promoter and terminator elements, as well as 5' UTRs, could be compared to optimize EndoT expression and production. Optimization of the culture conditions such as temperature, pH or medium composition is susceptible to lead to substantial improvement of the rate of deglycosylation *in vivo*. Alternatively, expression of EndoH instead of EndoT could be considered. Finally, the AtXylT transmembrane domain that is used to target EndoT to the medial-Golgi could be replaced by the targeting domain of other glycosyltransferases such as AtMNSI/II or AtGMII (cis/medial-Golgi), At β 1,3-GalT or rat ST (trans-Golgi) to target EndoT to more up- or downstream Golgi sub-compartments (Dicker *et al.*, 2015; Schoberer and Strasser, 2011). EndoT could also be targeted to the ER using a KDEL retention signal or the 13 amino acids (MTGASRRSARGRI) from the N-terminal end of *A. thaliana* GCSI, which are sufficient for ER-localization (Dicker *et al.*, 2015).

The subcellular localization of XylT-EndoT chimeric protein was evaluated in this study. We confirmed the presence of XylT-EndoT and XylT-EndoT-Venus in the microsomal fractions as well as in the characteristic mobile Golgi vesicles (near the plasma membrane and nucleus as well as in transvacuolar cytoplasmic strands) (Figure 27B, C). The high deglycosylation rate observed for endogenous and recombinant secreted glycoproteins also confirmed that the targeting of XylT-EndoT in the secretory pathway was adequate for the enzymatic activity, regardless whether EndoT was processed or not. To determine the localization more in depth, different strategies can be carried out. First, the Golgi localization can be easily confirmed by treating XylT-EndoT-Venus BY-2 cell line with brefeldin A, which causes disassembly of the Golgi apparatus, resulting in clusters of Golgi stacks (Ito *et al.*, 2014). Next, colocalization of XylT-EndoT-Venus with another well-known cis/medial-Golgi marker such as MNSI/II transmembrane fused to another fluorescent protein would confirm the correct localization (Dicker *et al.*, 2015). Alternatively, immunofluorescence using antibodies directed against Golgi antigens such as Le^a (JIM 84 antibody) or immunogold electron microscopy coupled with anti-EndoT or anti-Venus antibodies (Chevalier *et al.*, 2010).

In conclusion, generation of a set of glyco-engineered BY-2 cell lines allowed for the production of tailor-made recombinant glycoproteins. With the mono-GlcNAc cell line, we produced recombinant proteins with minimal *N*-glycosylation in BY-2 cells. Production of an IgG2 and the vaccine antigen gB in wild type, xylose- and α 1,3-fucose-devoid, high mannose and mono-GlcNAc BY-2 cell lines and subsequent analysis of their *N*-glycosylation profile validated the interest and suitability of glyco-engineered BY-2 cells as a tool for molecular farming.

3. Excision of the transgenes

3.0. Contextualization

Once the targeted genes are effectively inactivated using the CRISPR/Cas9 technique, the transgenes are no longer useful. Even worse, the constitutive expression of Cas9 and gRNAs increases the risk for off-target nuclease activity. The residual presence of the selectable marker gene after selection forces the use of a new marker gene for each subsequent re-transformation. Such markers suitable for BY-2 cell selection are not numerous and require the generation of new genetic constructs for each transformation. Therefore, the development of a transgene excision system for BY-2 cells is of great interest. To remove the transgenes in BY-2 cells, a system combining Cre-*loxP* recombination and the negative selection marker CodA was designed. The Cre-*loxP* recombination system is an efficient and widely used tool. Three challenges were associated with its application to our project. First, to the best of our knowledge, Cre-*loxP* system has never been used in BY-2 suspension cells and therefore we had to evaluate its functionality in this cell system. Second, we sought to develop a strategy to deliver or express the Cre recombinase without chromosomal integration. Finally, we had to find a strategy to select cell lines in which excision of the transgenes occurred.

In this chapter, I performed all the experiments except for the CodA1 cell line generation (Claudio Screpanti, Emmanuel van Cutsem and Catherine Navarre), the regeneration of calli from Multi^R23-CodA1 mixtures (Valentin Tonnel and Catherine Navarre) and the growth tests of CodA-Cre cell lines (Catherine Navarre).

3.1. Sensitivity of BY-2 cells to 5-fluorocytosine

A cell line was previously obtained by *Agrobacterium*-mediated transformation using the pPZP-*loxP*-*nptII*-*codA*-*loxP*-*gB* construct (Figure 32). This cell line will be referred to as CodA1. The plasmid contains the negative selectable marker gene *codA* and the positive marker *nptII*, flanked by *loxP* sites, as well as a *gB* expression cassette outside the *loxP* region. Consequently, after recombination, both genes should be removed but *gB* expression cassette should remain intact. The sensitivity of this cell line to 5-FC was confirmed by comparing growth in MS supplemented with kanamycin, hygromycin or 5-FC with the growth of the Multi^R23 cell line, which contains *hpt*, *nptII* but not *codA* (Figure 33). As expected, both cell lines grew well in kanamycin-supplemented medium. In contrast to Multi^R23, CodA1 cell line did not grow on hygromycin or 5-FC-supplemented media.

Next, selection of recombination events on 5-FC-supplemented medium was assessed. To mimic the low efficiency of this step, Multi^R23 cells, which express *mCherry* but not *codA*, were diluted 1:100, 1:500, 1:1000 into CodA1 cells. Then, they were plated on 5-FC-supplemented medium (100 µg/mL). Calli were identified in all conditions under visible light or after excitation of the mCherry and detection of the fluorescence (Figure 34). Selection on 50, 100 or 200 µg/mL 5-FC-supplemented medium was also evaluated. Multi^R23 calli were obtained on 50 and 100 µg/mL 5-FC but not on 200 µg/mL 5-FC (data not shown).

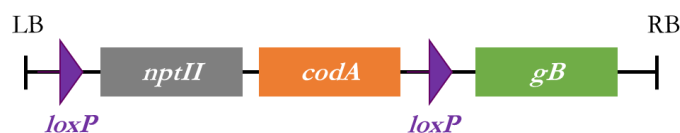


Figure 32. Schematic map of pPZP-loxP-*nptII*-*codA*-loxP-*gB*. LB, left border; RB, right border; loxP, locus of X-over P1; *nptII*, *neomycin phosphotransferase II* cassette; *codA*, *cytosine deaminase A* cassette; *gB*, *glycoprotein B* cassette. *nptII* is under the control of the *nopaline synthase* promoter and *octopine synthase* terminator; *codA* is under the control of the enhanced CaMV 35S promoter and *nopaline synthase* terminator; *gB* is under the control of the enhanced PMA4 promoter and the *nopaline synthase* terminator.

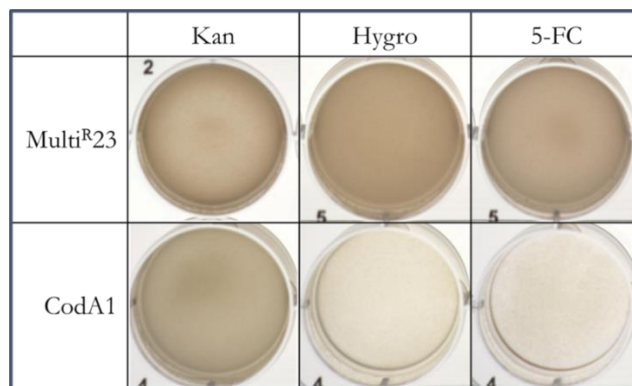


Figure 33. CodA1 cell line growth test. Pictures were taken after 7 days of growth in 4 mL MS medium supplemented with kanamycin (100 µg/mL), hygromycin (50 µg/mL) or 5-FC (100 µg/mL). Multi^R23 cell line is resistant to kanamycin, hygromycin and 5-FC.

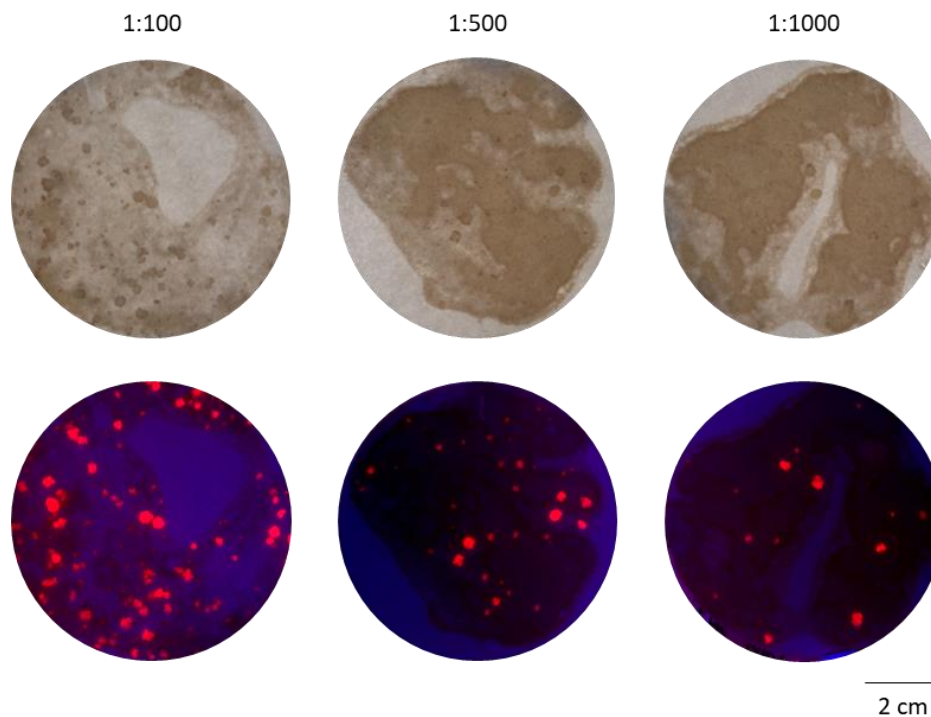


Figure 34. 5-FC selection assay on solid medium. Multi^R23 cells expressing *mCherry* are mixed with the *codA1* line expressing *codA* at the indicated dilution ratios. Mixtures are directly transferred onto 5-FC-supplemented MS-agar medium (100 $\mu\text{g}/\text{mL}$) and observed under visible light (above) or after excitation of the *mCherry* and detection of the fluorescence (below) (Amersham Imager 600). The pictures were taken after 17 days.

3.2. Evaluation of Cre-*loxP* system in BY-2 cells

The next step was to assess the efficiency of Cre-*loxP* recombination in BY-2 suspension cells. A second plasmid, pPZP-*hpt-cre*, was generated. It contains *hpt* and *cre* expression cassettes without any *loxP* site (Figure 35). pPZP-*hpt-cre* vector was used to carry out *Agrobacterium*-mediated transformation of *CodA1* cell line. Twenty-six *CodA*-Cre cell lines were selected on hygromycin-supplemented medium.



Figure 35. Schematic map of pPZP-*hpt-cre*. *hpt*, hygromycin phosphotransferase cassette; *cre*, cyclic recombinase cassette. LB, left border; RB, right border. *hpt* is under the control of the *nopaline synthase* promoter and *octopine synthase* terminator; *cre* is under the control of the enhanced CaMV 35S promoter and *nopaline synthase* terminator.

CodA-Cre calli were then grown in parallel on solid medium supplemented with hygromycin or kanamycin to identify the lines in which *npfII* was excised by recombination. Lines #1, 6, 7, 12, 14, 18, 19, 20, 23 were chosen because they showed the greatest growth delay on kanamycin as compared to hygromycin-supplemented medium (data not shown). Cell lines #13, 15, 21 did not show any growth delay and were chosen as a comparison (data not shown).

To identify the recombination events, a first PCR amplification was carried out using genomic DNA and the primers gBR and LBR (Table S1), located outside the *loxP* sites (Figure 36A). In absence of recombination, the amplicon was 5.5 kb long, while after recombination it was shortened down to 0.9 kb. All tested CodA-Cre cell lines displayed a shortened amplicon, while the CodA1 cell line produced the expected 5.5 kb amplicon (Figure 36B), suggesting that recombination occurred in all the CodA-Cre lines. However, this data did not prove that the transgenes were excised in all cells in each line. Indeed, as the 0.9 kb is amplified faster than the 5.5 kb, a heterogeneous cell line with both recombined and non-recombined transgenes might end up with only a 0.9 kb detectable amplicon.

To test for the presence of non-recombined transgenes, a PCR was carried out with the primers LBR and CodAF (Table S1; Figure 36A). DNA from fully recombined cells cannot be amplified while the presence of non-recombined transgenes will lead to a 3.2 kb amplicon. Cell lines #1, 6, 7, 12, 14, 18, 19, 20 and 23 did not display any amplification signal and were thus considered as fully recombined (Figure 36C). Cell lines #13, 15 and 21 showed a non-recombined profile similar to CodA1. Since these cell lines also showed a shortened amplicon with the primers gBR and LBR, they were either heterogeneous or they contained several transgene copies and only some did recombine.

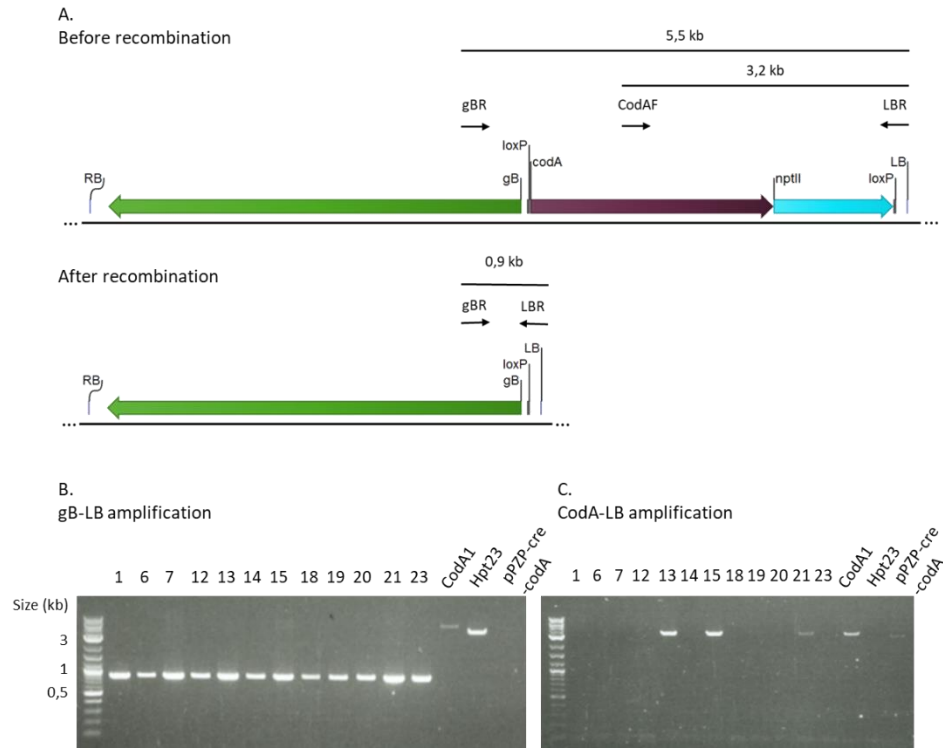


Figure 36. Identification of Cre-loxP recombination events in CodA-Cre cell lines.

(A) Schematic representation of the transgenes before and after recombination, with the different primer hybridization sites. (B) PCR amplification carried out on DNA from CodA-Cre lines with gBR and LBR primers. CodA-Cre lines display a recombined profile at 0.9 kb while CodA1 displays a non-recombined profile at 5.5 kb. Multi^R23 was generated with another plasmid and the expected amplification is 4.1 kb. pPZP-*Cre* plasmid is used as a negative control. (C) PCR amplification carried out on DNA from CodA-Cre lines with CodAF and LBR primers. Cell lines #13, 15 and 21 show the CodA1 non-recombined profile. Cell lines #1, 6, 7, 12, 14, 18, 19, 20 and 23 did not display any non-recombined signal. Multi^R23 is used as a negative control.

To determine the origin of the 3.2 kb amplicon, putative heterogeneous lines #13, 15 and 21 were cultured with 5-FC (100 µg/mL). Cell line #15 was not able to grow in 5-FC supplemented medium. gDNA from cell lines #13 and 21 was extracted after two passages in medium containing or not 5-FC. Cell line #6 was used as a positive control for recombination. Amplifications with the primer pairs gBR-LBR and

CodAF-LBR were carried out. gBR-LBR amplification gave the expected recombined band for all cell lines (Figure 37A). CodAF-LBR amplification showed that, after 5-FC selection, the non-recombined band was not observed anymore (Figure 37B). We concluded that both lines consisted initially of both recombined and non-recombined cells, but further selection on 5-FC allowed the obtaining of homogenous, fully recombined cell lines, demonstrating that 5-FC selection in liquid culture was efficient in mixed cell lines. The cell line #15 was probably also heterogenous and the PCR amplification was sensitive enough to detect the recombination event. However, the proportion of recombined cells was not sufficient to grow in 5-FC-supplemented medium.

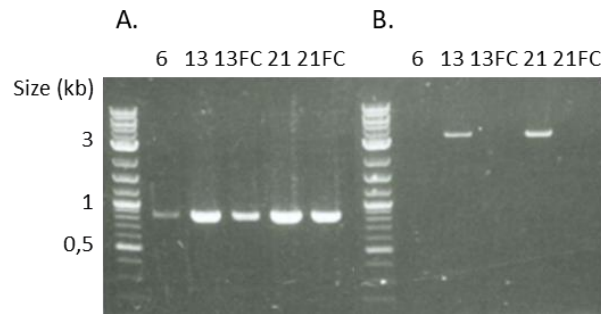


Figure 37. Selection of Cre-loxP recombination events in mixed CodA-Cre BY-2 cell lines. PCR amplification carried out on gDNA from CodA-Cre cell lines growing in media supplemented or not with 5-FC (100 $\mu\text{g}/\text{mL}$). (A) PCR amplification carried out with gBR and LBR primers. CodA-Cre cell lines display a recombined profile at 0.9 kb. (B) PCR amplification carried out with CodAF and LBR primers. Lines #13 and 21 display the non-recombined 5.5 kb band. Lines #6, 13FC, 21FC do not show any band.

Finally, CodA-Cre cell lines were grown in 4 mL hygromycin, kanamycin and 5-FC-supplemented media and their growth was monitored by measuring the FW. Growth of Multi^R23 and CodA1 cell lines were used as controls. The cell line #15 displayed the growth profile of a transformed (hygromycin-resistant) but non-recombined (kanamycin-resistant, 5-FC-sensitive) line (Figure 38). The cell lines #6, 7, 12, 13, 13FC, 21, 21FC displayed the recombined growth profile (hygromycin-resistant, kanamycin-sensitive and 5-FC-resistant). These results were consistent with those obtained via PCR amplification. In conclusion, the Cre-loxP/CodA system can be used to remove floxed transgenes in BY-2 cells.

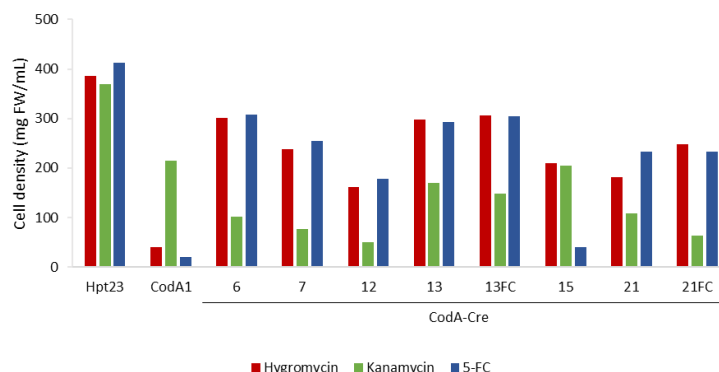


Figure 38. Sensitivity test of the CodA-Cre cell lines. Multi^R23 is resistant to hygromycin, kanamycin and 5-FC; CodA1 is sensitive to hygromycin and 5-FC and resistant to kanamycin. Cell lines XFC were previously cultured in 5-FC-supplemented medium.

3.3. Excision of the T-DNA in *GnTI/FucT*-KO cell lines

To investigate the origin of the puzzling remaining α 1,3-fucose signal observed by Western blot for the *GnTI*-KO samples (See Chapter 2), separate ectopic expression of the tobacco *a1,3-FucT* B and C genes will be carried out in *GnTI/FucT*-KO BY-2 cells (Antoine Mercier ongoing Master's thesis). Thereby, we should be able to identify the isoform(s) responsible for the transfer of α 1,3-fucose in *GnTI*-KO cell lines. Moreover, overexpression of *a1,3-FucT* should give rise to more α 1,3-fucosylated *N*-glycans compared with those obtained by the endogenous expression in *GnTI*-KO cells and, hence, could allow the detection of these structures by mass spectrometry. Altogether, this project aims at better understanding the substrate specificity of tobacco *a1,3-FucTs*.

Direct insertion of *a1,3-FucT* genes cannot be carried out in the *GnTI/FucT*-KO cell lines because the *a1,3-FucT* transgenes would be targeted by the constitutively expressed CRISPR/Cas9 system. Introduction of silent mutations to suppress the PAM sequences in the *a1,3-FucT* transgenes was considered but no such mutations were possible. Instead, it was decided to remove the whole *GnTI/FucT*-KO T-DNA (*nptII*, *mCherry*, *codA*, *cas9*, *GnTI* and *a1,3-FucT*-specific gRNAs) using the *Cre-loxP*/CodA system. For this particular project, integration of a new transgene comprising *hpt* and *cre* genes was not an issue. Indeed, subsequent insertion of each

a1,3-FucT gene will be obtained via kanamycin selection, since the *nptII* gene will be removed together with *cas9* and the *a1,3-FucT*-specific gRNAs.

The pPZP-*hpt-cre* (Figure 35) was used to transform *GnTI/FucT*-KO #10 cell line. Selection on 5-FC (even with high concentrations 250, 500 and 750 µg/mL), failed to generate any callus. Instead, hygromycin selection followed by screening of the survival phenotypes on different media and PCR analysis were carried out. *GnTI/FucT*-KO-*cre* lines were simultaneously grown on media supplemented with hygromycin (20 µg/mL), kanamycin (100 µg/mL) or 5-FC (200 µg/mL). As expected, the parental cell line *GnTI/FucT*-KO #10 did grow on kanamycin-supplemented medium but not on hygromycin- and 5-FC-supplemented media. Only two out of eighteen cell lines (*GnTI/FucT*-KO-*cre* #3, 16) showed a clear reduced growth on kanamycin as compared to hygromycin. Fourteen cell lines (*GnTI/FucT*-KO-*cre* #1, 3, 4, 7, 8, 9, 10, 12, 13, 16, 21, 22, 24, 26) grew well on 5-FC-supplemented medium. Cell lines growing both on kanamycin and 5-FC-supplemented media are likely heterogeneous, as demonstrated for CodA1 transgene excision. To obtain cell lines with a homogeneous excision profile, *GnTI/FucT*-KO-*cre* #1, 3, 4, 8, 12 were cultivated in liquid 5-FC-supplemented medium (200 µg/mL).

A first PCR, targeting *XylI*, was carried out using the primers XylTF and XylTR (Table S1) and confirmed the good quality of all gDNA samples (Figure 39B). Next, the presence of the *cas9* transgene was assessed using the primers Cas9F and TnosR (Table S1; Figure 39A), located inside *cas9* and the *Tnos* sequence, respectively. The expected 2 kb amplicon was detected in all *GnTI/FucT*-KO control samples, while *GnTI/FucT*-KO-*cre* samples did not yield any signal, indicating homogeneous excision of the transgene (Figure 39B). Finally, to precisely characterize the remaining T-DNA sequence consisting of a single *loxP* flanked by LB and RB, it was amplified using the primers RecF and RecR (Table S1) located next to LB and RB (Figure 39A). The use of primers located inside LB and RB is not recommended because the borders can be trimmed during the integration process, preventing PCR amplification. As expected, a 130 bp amplicon was detected in *cre*-containing cell lines (Figure 39B). Sequencing confirmed the suppression of the T-DNA comprised between the two *loxP* sites, leaving only one *loxP* sequence, flanked by RB and LB. Surprisingly, a faint 130 bp amplicon was also observed in DNA samples from *GnTI/FucT*-KO #10 cell line (but not *GnTI/FucT*-KO #16) and sequencing gave the same result as for *cre*-containing cell lines.

In conclusion, efficient Cre-*loxP*-mediated recombination occurred in *GnTII/FucT*-KO-*cre* cell lines, which resulted in the excision of the transgene. These cell lines are now available to knock in *α1,3-FucT* genes and study fucosylation of high mannose N-glycans.

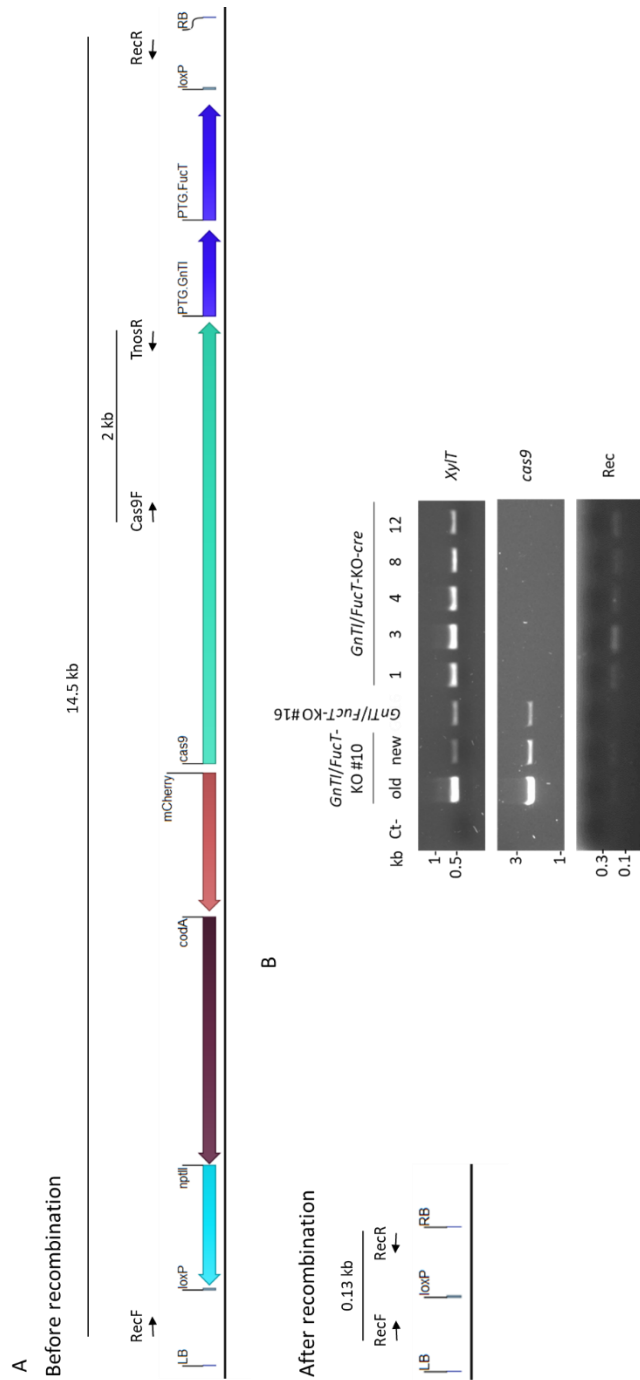


Figure 39. Identification of Cre-loxP recombination events in *GntII/FucT-KO-cre* cell lines. A. PCR amplification carried out on DNA from indicated cell lines with XylTF and XylTR primers. All samples yielded the expected 0.5 kb amplicon. B. Amplification of a portion of the *cas9* expression cassette using primers Cas9F and TnosR. DNA samples from *GntII/FucT-KO* lines yielded the expected 2 kb amplicon. No amplification in *GntII/FucT-KO-cre* cell lines. C. Amplification of the T-DNA sequence remaining after recombination using primers RecF and RecR. *GntII/FucT-KO-cre* samples yielded the expected 0.13 kb amplicon. RecF and RecR primers are also present in unrecombined transgenes, but 14.5 kb apart, effectively preventing amplification. Ct-, water. Old and new samples were harvested in 2020 and 2022, respectively.

3.4. Cre delivery without chromosomal integration

To improve this excision system, we wanted to deliver or express Cre recombinase into BY-2 cells without DNA integration. Two techniques were initially considered.

First, transient *Agrobacterium*-mediated transformation was attempted by using 5-FC selection but did not yield any callus. The second method was proteolistics. It involves the use of a biolistic apparatus to bombard cells with the Cre protein. It is an elegant method because there is no risk of transgene integration since no DNA is involved. The protein is mixed with gold microparticles and the mix placed onto a macrocarrier before being shot into the plant tissue (Martin-Ortigosa and Wang, 2014). There are very few examples of proteolistic method and it was not reported for BY-2 cells. To determine proteolistics efficiency, GFP was used instead of Cre. Onion epidermis was preferred to BY-2 cells to set up the experimental conditions. Fluorescent onion cells were observed 1 to 3h after bombardment (Figure 40). Next, GFP delivery by proteolistics was attempted with BY-2 cells. However, no fluorescent cells could be identified.

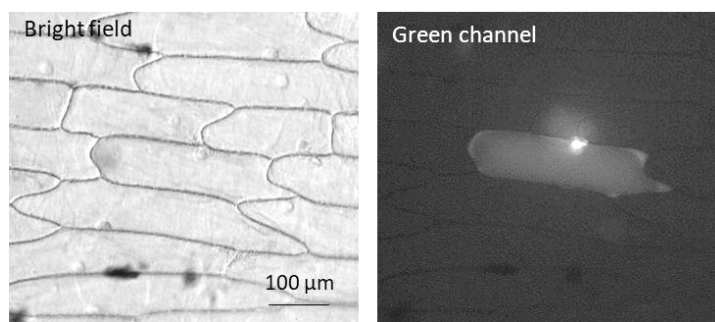


Figure 40. Microscope image of an onion epidermis cell showing green fluorescence after bombardment with GFP protein coated on gold particles. Air-drying 2h; 2.5 µg GFP/120 µg 0.6 µm gold microparticles/shot; 9 cm distance between tissue and macrocarrier; 0.5 cm distance between rupture disk and macrocarrier.

3.5. Discussion

This chapter sums up the efforts to develop a system allowing for the removal of a transgene in BY-2 cells. The initial setup was performed using the CodA1 cell line. Successful removal of the positive and negative marker genes was achieved, by integrating a new transgene containing a novel marker gene (*hpt*) as well as the *cre* recombinase gene. While the ultimate goal of excision of a transgene without integration of new genetic elements was not achieved, the aforementioned method was already useful. Indeed, understanding substrate specificity of *α1,3-FucT* required to remove the CRISPR-associated transgene in *GnTI/FucT*-KO cell line, to be able to express separately the *α1,3-FucT* in a high mannose-only background.

While working on this project, Sheva *et al.* (2020) managed to set up a transgene removal technique sharing substantial similarities with the one described here. They induced transgene auto-excision in BY-2 cells using *codA*, *cas9* and gRNAs targeting sequences close to the transgene borders, under the control of a heat-shock promoter (See Introduction, Chapter 5.3.3.2). This elegant strategy allows for sequential genome editing and transgene excision. Their results will be considered in this section to determine the relevance of our approach and suggest improvements.

The first challenge was to apply Cre-*loxP* recombination system in BY-2 cells. Expression of *cre* via stable *Agrobacterium*-mediated transformation of CodA1 and *GnTI/FucT*-KO cell lines resulted in the effective excision of the floxed sequence. Complete recombination was first confirmed by sensitivity to kanamycin and resistance to 5-FC and then confirmed by PCR amplification and sequencing, which revealed that precise recombination occurred between the two *loxP* sites, leaving only the predicted *loxP* sequence. This is in contrast with the strategy based on CRISPR/Cas9 developed by Sheva *et al.* (2020), associated with unpredictable indels. Indeed, although they confirmed the excision of the transgene by an antibiotic sensitivity test, Southern blotting and PCR, they were not able to amplify and characterize the residual sequence from the genetic construct. This is most likely due to the propensity of NHEJ to generate deletions around the break sites. This result highlights the asset of the Cre-*loxP* system, as compared to CRISPR/Cas9 technique, to precisely remove large transgenic sequences (14.5 kb in *GnTI/FucT*-KO cell line) in BY-2 cells.

Surprisingly, PCR analysis showed the presence of the shortened T-DNA fragment in *GnTI/FucT*-KO #10 cell line, and sequencing revealed that the sequence was identical to that of *cre*-recombined samples, *i. e.* a single *loxP* sequence. Since this cell line was maintained on kanamycin-supplemented medium and gave an intense signal

when amplifying *cas9* (Figure 39B), at least one non-recombined transgene also remained. As *Agrobacterium*-mediated transformation often results in more than one T-DNA insertion, it is possible that some T-DNA, but not all, could have been removed. However, this was not expected since this cell line did not contain the *cre* gene. The presence of a plant recombinase able to recognize *loxP* sequence has never been reported and is unlikely since it is a bacteriophage gene and as Cre-loxP has been successfully implemented in several plant species. However, the presence of such recombinase in *A. tumefaciens* is not completely excluded and could result in the simultaneous integration of recombined and non-recombined T-DNAs in the same BY-2 cell. This hypothesis could be verified by carrying out PCR amplification on plasmid DNA extracted from the *A. tumefaciens* culture used to transform our BY-2 cells. Finally, Cre-independent homologous recombination could have occurred between the two *loxP* sites in some of the *GnTI/FucT*-KO #10 cells. This hypothesis is also compatible with the very faint signal obtained despite the small size of the amplicon (130 bp). Indeed, it could indicate that very few copies of recombined T-DNA were present in the DNA extract and hence in the BY-2 cell population. Subcloning of *GnTI/FucT*-KO #10 cell line (isolation of single cells and generation of separate cell lines from a potentially heterogenous cell line) and observation of the presence of the recombined fragment in only some of the subclones would attest to the post-transformation timing of the recombination and hence suggest that the homologous recombination hypothesis is most likely. Subcloning was indirectly carried out when *GnTI/FucT*-KO #10 cell line was re-transformed with IgG or gB-containing constructs, which could thus be tested for the presence of the 130 bp fragment. To avoid this type of issue in the future, working with single-insertion cell lines should be favoured. Southern blotting, quantitative PCR or digital droplet PCR can be carried out to determine the T-DNA copy number.

Next, two methods were evaluated to deliver Cre without additional chromosomal integration. First, transient expression via *Agrobacterium* transformation, followed by 5-FC selection was attempted but failed to generate any callus. This issue was likely due to the selection process and will be discussed below. The direct delivery of Cre as a protein was also attempted, via proteolistics. The set-up of this technique using GFP-coated gold beads did succeed in generating fluorescent onion epidermal cells but not BY-2 cells. However, only few positive onion cells could be identified after bombardment. One of the critical features was to maintain the integrity of the protein during the coating, drying and shooting. Survival of the cells after transformation could also be an issue. However, since biolistics experiments were successfully carried out with the same microparticles and similar conditions in our laboratory, it is unlikely that cell survival was the main issue. Only few successful attempts at

proteolistics are described in the literature and, despite trying various delivery conditions, we were not able to expand this list to BY-2 cells. Coating the beads with *in vitro*-synthesized RNA instead of protein might lead to better results (Lacroix and Citovsky, 2020).

Presence of the cell wall hinders the implementation of most transfection techniques used for other eukaryotes to plant cells. To circumvent this issue, protoplasts can be generated by enzymatic digestion of the cell wall. Plant protoplasts are usually transfected with DNA using PEG (Mathur and Koncz, 1998) but electroporation technique has also been used (Fromm *et al.*, 1985). DNA can also be replaced by proteins or RNP to remove the likelihood for integration into the host genome. Transformation of protoplasts with CRISPR/Cas RNP successfully generated genome editions in different plant species, including the wild tobacco *Nicotiana attenuata* (Woo *et al.*, 2015; Kim *et al.*, 2017). Although the process of obtaining protoplasts, subsequent transformation, cell wall regeneration and calli formation is fastidious and time-consuming, it is worth considering its use for Cre protein delivery. Finally, PEG-mediated transformation of tobacco protoplasts with DNA is reported to reach efficiency as high as 60% (Locatelli *et al.*, 2003). Hence, this technique might be used to generate cell lines without the integration of a selection marker.

Recently, generation of DNA-free, gene-edited plants was obtained via virus-induced gene editing (Zhang *et al.*, 2022). This technique relies on the infection of the target plant by genetically modified RNA viruses and subsequent transient expression of their genome in plant cells. Due to the relatively small genome of viruses, first generations of virus-induced gene editing relied on the expression of gRNAs in transgenic plants producing Cas9. However, this issue can be overcome by using viruses with large cargo capacities such as potato virus X, barley yellow striate mosaic virus or sonchus yellow net virus, able to deliver the complete CRISPR/Cas9 cassette. Much work remains to optimize these systems, *i. e.* adaptation to commercially important crops, improvement of genome edition heritability and biosafety of the engineered viruses (Gentzel *et al.*, 2022). Sadly, since phytoviruses rely on plasmodesmata to spread infection cell-to-cell, it is unlikely that this technique reaches a high efficiency in suspension cell cultures.

The third challenge was to select the recombination events that occurred without integration of a new transgene. The strategy was based on the excision of *codA*-containing transgene and the resistance to 5-FC this event conferred. Selection on 5-FC after *Agrobacterium*-mediated transformation did not work for *GnTI/FucT*-KO cell line, despite trying different 5-FC concentrations. The first hypothesis, cell death

due to membrane diffusion or uptake of 5-FU produced by non-recombined cells, was ruled out as Multi^R23 calli were obtained after dilution in CodA1 cells and 5-FC selection (Figure 34). A second hypothesis for the unsuccessful 5-FC selection is that CodA, in the form of mRNA or protein, could still be present for a few days after the recombination occurred. As a result, when the 2-day-old co-cultivation was transferred onto 5-FC-supplemented medium, toxic 5-FU could still be produced in recombined cell lines. Noteworthy, a delay of at least two weeks was applied by Sheva *et al.* (2020) between the first induction of transgene excision and 5-FC selection. Increasing the delay between co-cultivation and selection is thus susceptible to resolve this issue.

An approach combining the advantages of our transgene excision technique and that developed by Sheva *et al.* (2020) is currently under development in our laboratory. An intron-containing *cre* gene was placed under the control of a BY-2 heat-shock promoter (pHSP3A) (Navarre *et al.*, 2011). Next, pHSP3A-*cre* was transferred into a pPZP-*loxP-lpt-codA-cas9*-PTG.X-*loxP* (where X is a gene unrelated with this thesis). After BY-2 cell transformation, hygromycin selection and confirmation of the genome edition, repeated heat shock periods will be applied to liquid cultures. Two weeks after the first heat shock, recombination events will be selected by culturing in medium supplemented with increasing 5-FC concentrations.

When the Cre-CodA transgene removal technique is operational, glyco-engineered cell lines will be cured from the unwanted T-DNA (containing *cas9* and gRNAs) to avoid off-targets. In addition, this technique will be tried for successive transformations with the same expression cassette in order to reach high recombinant protein yields.

General Discussion and Perspectives

PTMs such as *N*-glycosylation are major regulators of protein structure, function and immunogenicity. Therefore, the control of *N*-glycosylation profile of recombinant therapeutic and vaccine glycoproteins has been of increasing interest in recent years. In this thesis, the *N*-glycosylation pathway of BY-2 suspension cells was modified in different ways by knocking out glycosyltransferase genes as well as inserting the *EndoT* transgene, to produce pharmaceutical proteins with different types of *N*-glycans. Thereby, the *XyII/GMD*-KO cell line was first generated. Although β 1,2-xyloses were suppressed and a strong reduction of α 1,3-fucoses was detected, this cell line had less potential than the previously generated *XyII/FucT*-KO cell line, which was completely devoid of both immunogenic residues. Next, a cell line producing high mannose *N*-glycans was obtained by knocking out *GnTI* genes, as well as *α 1,3-FucT* genes since traces of α 1,3-fucose remained in *GnTI*-KO cell lines. Finally, to produce *in vivo* deglycosylated proteins, *EndoT* was expressed in the cell lines producing high mannose *N*-glycans. Suitability of these cell lines for molecular farming was confirmed via expression of different types of pharmaceutical glycoproteins: human antibodies (IgG1 and IgG2) and a viral antigen (gB of HCMV). Glyco-engineered cell lines conferred the expected global *N*-glycosylation profile to IgG2 and gB. Hence, the BY-2 cell platform now permits the production of secreted recombinant glycoproteins bearing typical plant complex *N*-glycans, non-xylosylated and non- α 1,3-fucosylated complex *N*-glycans, high mannose *N*-glycans or mono-GlcNAc *N*-glycans. This represents an important step for molecular farming since it allows to easily purify plant-produced pharmaceutical glycoproteins with tailor-made *N*-glycosylation from the culture medium. A transgene excision system was also set up to remove unwanted transgenes such as *cas9* or resistance markers. This system, combining the Cre-*loxP* recombination system with the negative marker gene *codA*, was used to remove the whole *nptII-codA-mCherry-cas9-gGnTI-gFucT* T-DNA. Removal of *cas9* and gRNAs will allow overexpression of individual *α 1,3-FucT* genes in this cell line to identify those responsible for the remaining α 1,3-fucose signal as well as to determine what the α 1,3-fucosylated structures are. Ongoing work focuses on obtaining an improved transgene removal system, based on an inducible *cre* self-excision coupled with more efficient selection with 5-FC.

The glyco-engineered set of BY-2 cell lines can be used to produce glycoproteins known to benefit from a particular *N*-glycosylation profile. For example, *GnTI/FucT*-KO cell line is indicated for enzymes used in replacement therapy for lysosomal storage diseases, instead of the current carrot cell lines in which recombinant enzymes are targeted to the vacuole for terminal glycan trimming by HEXO1 and bear plant-typical *N*-glycan residues. *EndoT*-expressing cell lines are useful for the production of deglycosylated vaccine antigens, antibodies without Fc effector

function or for crystallography as heterogeneous glycosylation is often an issue for this technique. Moreover, the ability to produce different glycovariants is highly valuable as a tool for fundamental research. Indeed, glyco-engineered BY-2 cell lines could also be very useful to uncover new roles for *N*-glycosylation. Indeed, carrying out RNAseq analysis to identify change in gene expression among the different glyco-engineered cell lines might highlight the importance of *N*-glycans in different cellular processes.

In this thesis, the effort was put towards simplification of *N*-glycans. However, complexification to obtain human-like structures might also be a powerful strategy, depending on the requirement for the protein to produce. Addition of $\beta 1,4$ -*GalT* was already carried out in our laboratory and subsequent addition of (poly)sialic acids would generate another valuable BY-2 cell line. The limited plant *N*-glycome (as compared to that of mammals), associated to its plasticity, is an ideal starting point to reconstruct and mimic valuable *N*-glycosylation profiles of other organisms, such as parasitic helminths, which produce glycoproteins with therapeutic applications (van der Kaaij *et al.*, 2022). Although the glyco-engineering strategies affect all endogenous *N*-glycosylated proteins, no morphological changes were observed and the growth rate of BY-2 cells was only slightly reduced. This demonstrates an important flexibility of the BY-2 *N*-glycosylation pathway. Working with cell cultures is an asset for molecular farming since they can support genome engineering affecting developmental or reproductive processes, whereas whole plants may require sequestration strategies to be viable, as illustrated by the expression of seed-specific transgene required for the implementation of GlycoDelete strategy in *A. thaliana* (Piron *et al.*, 2015).

Recently, a reverse cross-species engineering strategy was carried out by introducing *XylT* and *α1,3-FucT* genes in CHO and knocking out *α1,6-FucT* and $\beta 1,4$ -*GalT*, arguing a potential immune-engaging effect of plant *N*-glycans on vaccine antigens (Larsen *et al.*, 2020). This clearly shows that the study of the impact of *N*-glycans on vaccine antigens is still in its infancy. Similarly, although studies discussed in Introduction Chapter 3.3.5. show promising results for deglycosylated vaccine antigens, absence of *N*-glycans could negatively impact immunogenicity. Indeed, epitopes recognition can be glycan-dependent. For example, a glycopeptide epitope on glycoprotein 120 of HIV elicits potent cellular and humoral immune responses (Sun *et al.*, 2020). Moreover, the peptidic epitopes exposed by removing the glycan shield might not trigger useful immunogenic response since they are hidden during virus infection. Finally, tertiary and quaternary structure modification due to the removal of *N*-glycans might generate new conformational epitopes that are irrelevant

in case of viral infection. Production of deglycosylated antigens and subsequent immunogenicity analyses should clarify these matters, although the impact could vary greatly depending on the antigen.

The outcome of many *N*-glyco-engineering strategies is often reported only via analysis of the profile of a recombinant glycoprotein with one-to-three *N*-glycosylation sites. In this work, our analysis was much more thorough. Indeed, we analyzed the global *N*-glycosylation profile of all secreted glycoproteins (endogenous and recombinant), combining Western blotting and MS analysis. Thereby, we generated an overview of the *N*-glycosylation status in these cell lines, which is more predictive for the future expression of new recombinant glycoproteins. We also analyzed the *N*-glycosylation profile of two recombinant glycoproteins. Again, we used different techniques to confirm the outcome. Western blotting, *in vitro* deglycosylation assay and LC-MS/MS were carried out. We first produced a very well-studied glycoprotein with only few *N*-glycosylation sites, an IgG. Then, production of the highly glycosylated, complex, HCMV antigen gB was carried out. Working with gB showed that glyco-engineered BY-2 cells were capable of producing complex recombinant proteins while conferring the global expected *N*-glycosylation profile. It also highlighted some of the current limits of the deglycosylation strategy (incomplete processing by EndoT, reduced stability) and the expression system (protein degradation).

Throughout this project, emphasis was placed on designing experiments in a way that allows detection of unexpected outcomes, which occurred in several occasions. Using MELD prior to LC-MS/MS, instead of the usual trypsin digestion, identified a non-canonical *N*-glycosylation site on the CH1 domain of IgG2. Reports for such *N*-glycosylation sites are very scarce. However, since trypsin digestion is the gold standard for MS analysis of antibody, it is likely that this and other unexpected PTMs were missed in numerous reports. Effort was also made to identify both *NtGnTI* homeologs, instead of blindly targeting *GnTI* sequences available in *N. tabacum* genome databases, which turned out to be incomplete. In this way, disruption of *NtGnTI.B*, probably by a transposition event, was identified. Another unexpected result that could have been easily overlooked, was the residual presence of α 1,3-fucoses in *GnTI*-KO cell lines. Indeed, it was detected only after extended exposure time of the Western blotting membranes. In fact, as described in the discussion of Chapter 1, a few studies identified α 1,3-fucosylated high mannose *N*-glycans but did not discuss or highlight their presence. We believe this observation is worth investigating and have designed a new project aiming at uncovering the substrate specificity of α 1,3-FucTs. Similarly, significant proportions of Man4 *N*-glycans were

detected in high mannose cell lines, while this type of *N*-glycan was not expected on the base of the established plant *N*-glycosylation pathway. These data obtained from *N*-glycosylation mutants can be used as a starting point to uncover unexpected glycosyltransferase substrate specificities.

Purification of gB and site-specific MS analysis would provide new insights on the accessibility of those sites for EndoT. Moreover, it would be informative to compare the type of *N*-glycan structures observed at each *N*-site of gB_{FG} by Smargiasso *et al.* (2019) with the extent of EndoT truncation at the same site of gB_{MG}. Purified gB glycovariants could also be used in immunological assays to assess a role for specific *N*-glycans. Particularly, evaluation of the efficiency of deglycosylated gB in vaccination assays would shed light on the relevance of glycan shielding suppression strategies. Stability of vaccine proteins is also an important issue. DSF should be carried out on the gB glycovariants once purified, as well as spiking experiments under different culture conditions.

Although this work focuses mainly on qualitative aspects of recombinant glycoproteins, quantitative analyses are required in the scope of commercialization. Quantification of IgG and gB glycovariants in the CM of the different glyco-engineered cell lines is currently being carried out by ELISA. Furthermore, correct binding of IgG LoBM2 to its antigen (IgM) is under investigation by the same technique. Indeed, it is crucial to prove that binding affinity is not disrupted by structural modifications induced by extensive *N*-glycosylation engineering such as deglycosylation. Preliminary data confirm that all the glycovariants bind to IgM. Modulation of the binding properties to the Fc receptor and complement components, as well as the resulting ADCC of the IgG glycovariants should be confirmed by affinity chromatography, surface plasmon resonance and ADCC reporter bioassay.

Although antibody production by BY-2 cells should be cost-effective as compared to hybridomas (Magy, 2014), it is not possible at this stage to determine the commercial feasibility of gB production by BY-2 cells as vaccine antigen. Our results show that gB can be produced in BY-2 cell culture and previous constructs accumulated up to 60 mg/L. The low cost and demand of quality control of plant cell culture medium are an advantage over conventional mammalian cell cultures. However, downstream processing represents an important portion of the costs of recombinant protein production and has not yet been implemented. Moreover, the amount of gB required for immunization will directly impact the costs. The protective effect of BY-2-produced gB, which could be enhanced for some glycovariants, will determine this dose. Regulatory issues associated with uncommon

production hosts such as plant cell cultures may impede progress toward commercialization. However, as more and more plant-produced pharmaceuticals are being accepted, the importance of this parameter should decrease. In conclusion, antigen production in BY-2 cell is feasible but it is premature to rule on cost-effectiveness, as well as vaccine efficiency. The suitability of this and other expression systems needs to be considered on a case-by-case basis, once data on accumulation, downstream processing and immunization efficiency are available.

To further improve the BY-2 cell expression system, the *N*-glycosylation engineering approach should be combined with other genome editions affecting quality or quantity of recombinant glycoproteins. Among these, the modification of the *O*-glycosylation pathway by inactivation of endogenous enzymes and/or expression of human *O*-glycosyltransferases is susceptible to bring the same advantages as *N*-glyco-engineering, *i. e.* greater homogeneity and enhanced protein function. Next, co-expression of recombinant proteins with human chaperones such as calreticulin might improve the folding efficiency as well as their production and decrease the ER stress response (Margolin *et al.*, 2020). The final protein yield is also negatively affected by the high degree of proteolysis typically observed in plants. As a result, reducing specific protease activity towards recombinant proteins, by inactivation of said proteases or via optimization of the culture medium and conditions, is a promising strategy to improve the BY-2 cell system (Navarre *et al.*, 2012). Finally, suppression of gene silencing might allow obtaining cell lines producing higher yields of recombinant proteins. Indeed, this mechanism has a viral defense role but as it is activated when a transcript is produced in unusually important amounts, it also affects transgene expression (Baulcombe, 2004). In this context, the transgene excision system will be an important tool, allowing for easy and quick sequential genome editions to combine different approaches.

Supplementary Figures and Tables

ATGATATGGAAATTTGGCGCTGGATTTTCTTGAGCTTTTGGAGCTGGAGCTATGCTTCTTGAGCGAGCAAGTAAGGCCGGATGTTGCGTGTGTTTACTTTCAAAGTGTGAGAAATGCAC
 GGGAGAAAATATATTGGATATATTCTCTTATATTAACACAATACAATAGGCCCTATATATATATATATATATATATATACATAACATGTCCTACTCTTAATACATATAGAGATTGAG
 GTTATTTACTTCCTATTTACATAAATATTTCTTAATATCTTCCCTCAAGATGTGGCATATAAATCATATGTACCGAGCTTGTTCATATATGTAGTTAATACAGGAGCACTGAGGAGCTTGG
 TGAANAATATCTCGAAGCTGTATCATTTGACTTTCACAACTTTGTAGCAATATCTCCCGAATAATCTTTTCTCTTCACAAAGTGACAGTCAATCTCAATGTTTTAGTTCTCTGATGAAG
 ACTGGATTGGAGCAATATGAAGAGCAACTTGAATATCACACACAAGTCCCATCTGACTAATCACACAAATTTCAACTCTTGGAGAACTGTTTGAATCCAAATAGCTCACAATGTCTG
 CCATAGCCATCTGCTTCGATATTTCTGCTTCTGCACTAGACCCGAGCAACCAATTTTGTCTTCTGCTCTCCAGAACAAACAATTTCTCTCTACTAAGATACAATTTCCAGACGTAGAACG
 CCTATCAGAAGGTGATCTTGCCCAATCAGCATCTGAGTATCCCAACGATCTGCTCATGGCCTCGATCCTCAAACAATAAACCTTTGCTTGGAGCTGATTTATATACCCAGAAATGCAAA
 ACAACTGCATCCCAATGACTATCACACGGGAGAAATCCATAAACTGACTCACAACATTCACAGAAAAGGAAATGTCAAGTCTTGTCACTGTAAAGTAAATTTAATTTACCAACCAACCGTCT
 TATATCTTGCAGGATCGCTAAGAGACTCCCCCTGTTCTGGCAGAAGTTTAGAATTCGGATCCATCGGAGTGTCAACAGGTCTACAACCTGTCAATCTCTCTCAAGAATATCTAA
 GGCATATCTTTCGTTGTGAGATCACAATACCTGAGCTAGACTGAGCGACCTCAATACCCAGAAAAATACTTTAGTTTGGCCAGATCCTTAGTCTGAAGGTGCTGAAGAGATGTTGCTTC
 AATTTAGTAATACTATCTCGTTCATTTGCCGTTAATAACAATGTGCTCAACATAGACCACAGATAAATACAGAGATCCGAGCAGAATGCCGATAGAATACTGAGTGATCAGTTTCAC
 TCCGAATCATGCCAAATTTCTGGATACTGTGCTGAACCTTACAAACAGACTCGAAGAGACTGCTTTAGACCATAAAGTGACCGGCGCAAGCGACATACAAGGCCACAGACTCCCC
 CTGAGCAACAAACCCAGGTGGTGTCTCATATAAACTTCAATTCACAAAGGTACCGGTGTAATAATGCATTTCTAATGTCCTGCTGATAGAGGGGCCAATGGCGAACCAACCACTGGAT
 AGAAAAAGGCGAACTGATGCTATCTTAGCCACGGGGGAGAGGTATCATCTGAATCGAGCCCAAAATATTTAGTATATCCCTTGGCAACAGACGAGCTTAAAGCATCAACCTTAC
 CATCCAGACCAACTTGAATGCATACACCAACGACCAACCAAGAGATTACTTGAAGAAAGAGGAGCGAGCTCCCAAGTACCCTCGTATGTAAAGCAGACATCTCGTCAATCAT
 AGCCTGTGCCATCTTAGTAGAGCAATGCTTACCTATAGACTTAGGGATGGAACAGAGGACAAAAAGATAGAAAAGCATATAAGGGTGTAGCAAAACGATGATACTCAAAACCG
 ACATAATGGGATTAGGATTAAGGTGGTCCGTATACCTTTCCGAAGTGCATCGGTGTACTAGGAAGAGCAAGTCCGAGGAGGAGCATGCTCAGGTGCAGGACGTGAATCAGCT
 GGGCTGTAGTCTGGGTGGCGAGCAGCATGATATGTCAAGAGTGGTGTCTCTGGCGAGGGAATCTAGGAGGAGCCACACTAGACTCCCAACCGGTTGGAATGGGTAAAGACTTTTGTAG
 CGAAGGTGAAGGAGGAGCATATAGTAGACTCTGGTAAAGACTTCTGTGGCGGAAAGTGAAGGAGGAGCTATAGTAGACTCTTAAAGGTCTGGTATAGGTAAAGACCTCAGATATATCAAG
 GTGCTCAAAATGCTTAGAAAAGGTAAAGAAAGCTTTAGACTCMAAAATGTGACCTCAGAGGACATTAAGTACCTACGAAATCAGGTGATTAACAAAGATATCCCTTCTGAACACGA
 GAATAACCAAGGAAGACACACTTGAGAGCAGGATGAGCTAACTTATCTTTCCAGGGCTAAGTTATGAACGAAACAAGTCTCCCAAAAACAGGAGGAGAACAGAGATATAAAGGTG
 ATTGGGGAACAATACTCGATACGGTATCTGATTTGGAATGGGAGAGGAGGAAATCCGATTAACTAAATAACAGCTGTGGAATGCAATCGCCCCAAACCGCAACGGAATATGAGA
 TTCAATGAGAACTGTGGAGCAGTCTCAATGATGTGCTTATCTTCTCTGCAACCCCATTTTGTCTAAAGGGTATAAGGACAGAGGCTCAATGAATAATCTCTTAAGAAATCAATA
 AACTGCTGAATTTGAGAGGATAAATATTCTAAGGCATATCATCTGCGAAAGTGTGCAATAGAAAACACCAAAATGATTTTTAATTTCCAGCAACAAAATCTTGGAAATAGAAAACAAT
 CAGAACGATCTTTCAATTAAGAAAATCCAAGTACATCTTGCATGATCATCAATGAACATAACAAAATAACGAAATCCCAAGTTGAACGTACTCTACTAGGACCCCATATATCAGAATG
 AACTAAAGAAAACAGACTATGCAATGACTTCAATATCTACGAGGAAGGAGGCTCGGGTATGTTTCCCGAGCTGCACTGACTCACACTCTAGTGTAGATAAATAGACAAATAGGCC
 ACCATCTTCTGAAGCTTGGATAAGCTTGGATGTCTTAAAGCTCTATGAATTAGGTCCGGAGTATCTGTAACTAGACATGCCTTGGAGGAATTAGTGAGTTAAGGTAGTAAAGGCGTT
 CTGATTTCAAGTCTGTTCATATCGCTGCCCCGTACTCGCGTCTCGATATAAAGGAATCATCAATAAAATATATACCAAAATGGAGGGCAGAGTCAACGACTAACAATGCAAG
 ATTTAAAGGACATCCAGGACATAGAGAACCGGAATCTAGAGTGACAGAGGTTAGGGGATTCGCTTGTCCAACTCCTTTTGTCTTGGTTTGGAGCCATTTGGCTAAAGTAAACATGTTGGA
 AGAGACTGTGAATACGCAATATTGACAAATGTGATTATTAACAGAGATATGATCAGAAGCGGCGAGTCCCAACCCATTTGTCCAAGAGTACTAGACTGGAAAAACAAAAACAAAAG
 AATTACCAGCAATAGAAATATCAGTCTGAGCAACAGAGGCTATTGTGGAGATGTCTGCTTACTTGTCTGATACCTGAAGGAATCATATATATCCCTTCAGATAAAGAAAAACCCCTG
 GTTACCTGTAGTCTCAGTCTGAGCATGTAAAAAATAGCACACGTCAACAAGTGTCTCAAGTTTATGACAATAAGAGCACTTGGGTCTAGATCTTCCAAAACGACATCTCTCTGCTCTA
 TTCTCCATAGTATGAGATGCCGATTTGTCACCTGACTGGGATACAGGACAGATGAGTCAAGTGTCTGTGATGAGCTCACTGGGTGACTTGGTGTCTGCAGCAGAGGCGAATTCGAG
 AGAATAATTTCATCACTGTGGGACAGTCCGACTAGCCAAAATCTGGTCACTGACTGAATCAAGTGTCTGTGATGAGCTCACTGGGTGACTTGGTGTCTGCAGCAGAGGCGAATTCGAG
 TTGTGTCTTTCAATACTAGCAGAAAATGGTATCAATGTCTCAAAATCTTCCATGACTGCATGTAATTTGTCCCAAGTAAAGTAGACATATCCAAATCTCTGTTTCTTCAAGATTTGTCATT
 CGCGATATTACATCATAGAAAACGAGATATGTCAATTAGTGTATAAAATACGAGCCTTTTCCCAAACTAAATAACATCTCTGGAATGGAGGGAACAAAGGGCATCAACTTGGAAATCAATAG
 ATCGCCACAGGATATCATATAACTGAGCATCGACCTTCTCCCAAGCTGTTTTGGCCTTTTCATCACTCTCGCTAGCTTTTATTAATTAATGATCTTGAATCTCTTGAATCTTACACCA
 CAACTCCAGAGCGAAGCCCAAGTAAATAGTGTGAACCTTCCCAATTAAGGTTCTGAGGTAATCATACACCGGAACTGCCAGAACCCGGTTTTTGAATCCGAAAGCATCGACTCCG
 AATATAGATCTCGAAACTCTCGTGAACAGATGAAGGAACACTGTTTTGCCCGGAAATTTACTGTAGTCCCGGAAACCTCAAAATGCTCGGAATCAACAGAAAACAGGATGGGTA
 GGGTCAGAAATTTAGTACCGACCCAACCTGTTCTGAAGATCCGCTGCAAAAATGGCCGAAAGGGCTCCACTCCCGCGCGCTGAAGCAGATCTCGCCCGGAAAAAATCTCTCCGGC
 GGCGCGTGGCGCGCTGAGAGGTTGTCTGCCGAGGTGTTTGTATGGGTTTTGTGCTCCCGGGGTTGGGGGATGTTGTGGTGGTGTGGTTTTTGCACAAACACGATGGAATTTGATT
 TTTTCTCGAAAAACAGCCCTAAAAGGTCTCGGAATTTGATGCACGATGACTTTTTGGGTGGGTTTTCTTCTCCGGATATTTCTCACTGCTGCTCTGATACCATGTGAGAATGCACG
 GAGAGAAATATTATGATATATTCTCTATATATTAACACAATAACAATGAGCCCTATATATATATATATATATATACATAACATGCTCTACTCCGAATACATATGGGATTTAGGGTTATTT
 ACTCCTTATTTACATAACTATTCTAACACAAAGAAATATTCAATTTAATCTAGTCTGAGGAAAAATTAATATTTTTTACTCGTGTCCAACTCCCGCTCAATTTTCTTTAAAGAACCAAC
 ATAAATCAGATTTCGACAGCATCCAAGATCTCTGCTCTTCCAGGCTTGTGATAGGAGAAAAATCTGATTGACGAGGGGGATAGATTGAATTCCAATTTGGTTATATAACAAATCTTAAAG
 AAAAGGATTAAGAGCTTTTCCCTCATGGACTGATGTCCAAATATGCTAAGTGGTATAGTGAACGAATAGAATGGGATTAGCCTAAAAACATGGGGATAAAAAAGCTGTCTCAATATGTC
 CCAAGTATGTTACAGAATTTCTCAATTTGTTATGGTGAACATCCCAAGTTCGATTTATGGCTATTCTCTTGGAAATGACAATGGACAAATGCAGTTTGTCCAAGATCTCT

Figure S1. Sequence of *NtGnTL.B* intron 11. Exon 11, Exon 12, transposon-
 related ORFs.

136

CGTTTAATATTTCAGGATATTATTAGTTTAACGTTGTCAMTATCTCGGAATTTTGCAMATGAATCAAGCCTATATGGCTGTAATGAATTAAAAGCAGCTCGATGTGGTGTAATGATGATTACTTGATTTCAAAAAAATCCCAAGTATTAATTAATTTCTCGTAGGAAGAAGGTAGCTACGATTTACAGCAAGACGAGATACAAAGAACCAATAAATGATTGAACCTCGGAATATTCAGGAAGCAAAATATTTAAAAAATATACGCAAGTCTGGAAACAAGAAGTGATATATTTTGCTTAAACAGACATCCCCCTCAAGAAAGTGGCAATTTTCGTCATGATCAATTAATGCTCCGTCACAAAAATTTGGACATCATATGGCGGGTGGGGGGGATCAGATTTGCTGTTCCCGCCCTCAGTTTAAACATACTCTGATATATGTTATGCTATACGAAGTATTTCCGCGGGGGTGGGAGGGAAGTTTAAATCATCAGTTTTCAGGAGATATATGGCGGGTAAAC

Figure S2. Sequence of pAM-*GnTI*-KO comprised between left and right borders. LB and RB, *nptII*, SAR, *pcocas9*, *GnTI*.PTG.

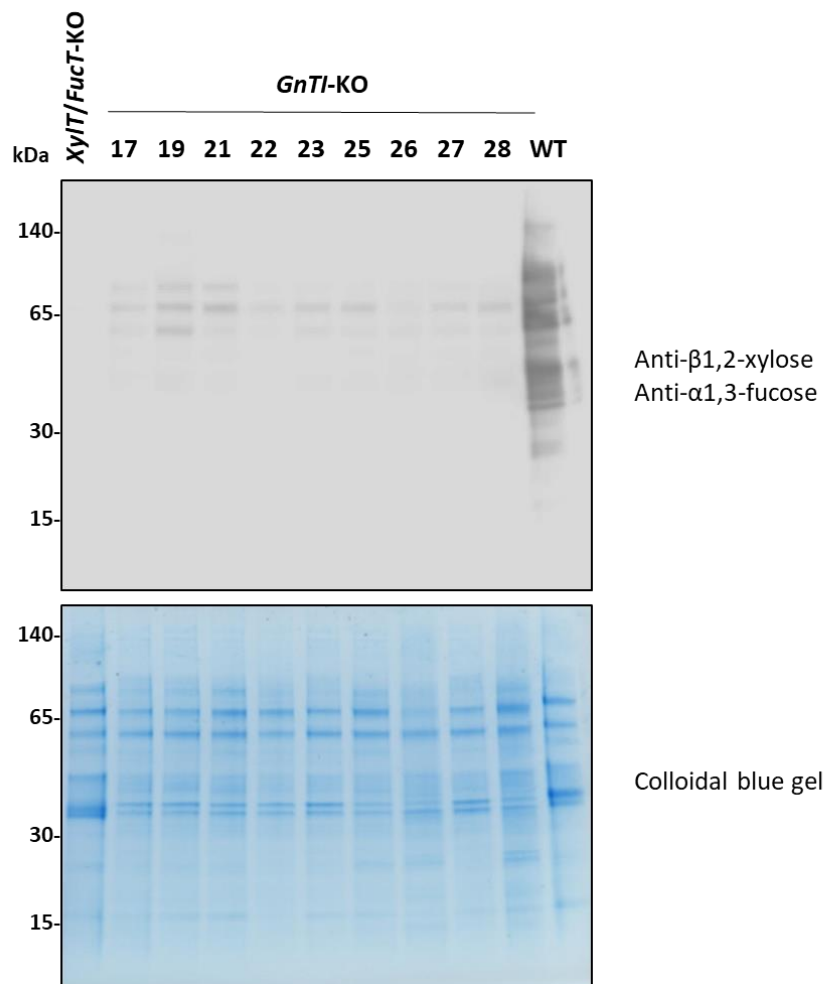


Figure S3. Reduction of protein-associated β 1,2-xylose and α 1,3-fucose residues in *GnTI-KO* BY-2 cell lines. Top. One representative Western blotting membrane is shown. Secreted proteins (40 μ L culture medium) from WT, *XylIT/FucT-KO*, and from the indicated *GnTI-KO* cell lines were separated by gel electrophoresis and analyzed by Western blotting using anti- β 1,2-xylose and anti- α 1,3-fucose antibodies. Bottom. A colloidal blue gel is displayed as a loading control.

139

Figure S4. Sequence of pPZP-GnTI/FucT-KO comprised between left and right borders. LB and RB, *nptII*, *codA*, mCherry, *pcocas9*, *GnTI*.PTG, *FucT*.PTG, homing endonuclease site.

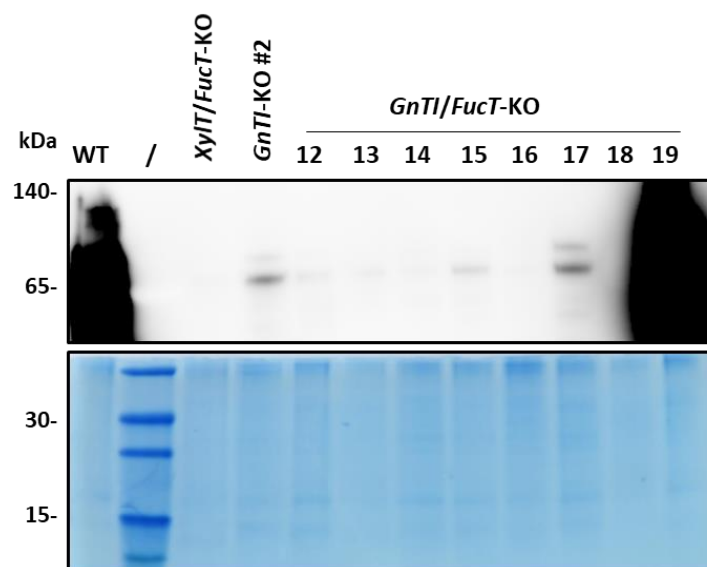


Figure S5. Extensive reduction of complex *N*-glycan-associated signal on secreted glycoproteins from the indicated *GnTI/FucT*-KO BY-2 cell lines. One representative Western blotting membrane is shown. Secreted proteins (40 μ L culture medium) from WT, *XylT/FucT*-KO, *GnTI*-KO #2, and from the indicated *GnTI/FucT*-KO cell lines were separated by gel electrophoresis. The gel was horizontally cut and the upper half was analyzed by Western blotting using anti- β 1,2-xylose and anti- α 1,3-fucose antibodies. The lower half was stained with colloidal blue and is displayed as a loading control. Lane 2 contains a protein ladder.

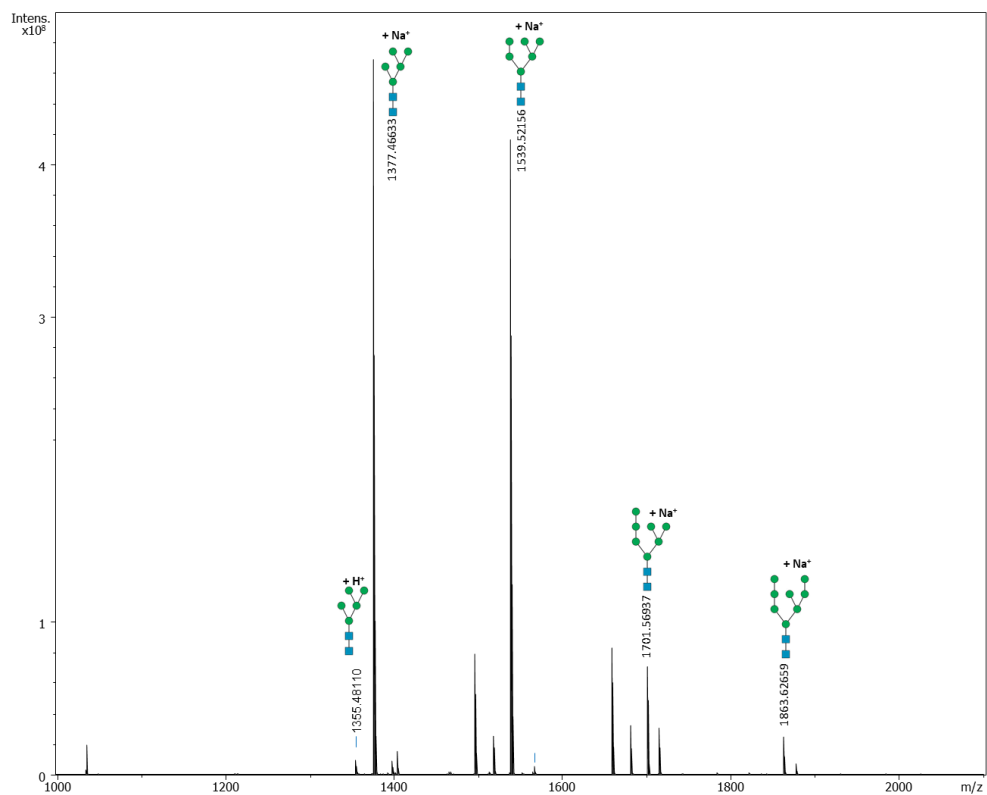


Figure S6. Total *N*-glycans released from RNase B. Peaks identified as *N*-glycan signals were labeled. Signals were detected as singly positively charged $[M+H]^+$ or $[M+Na]^+$.

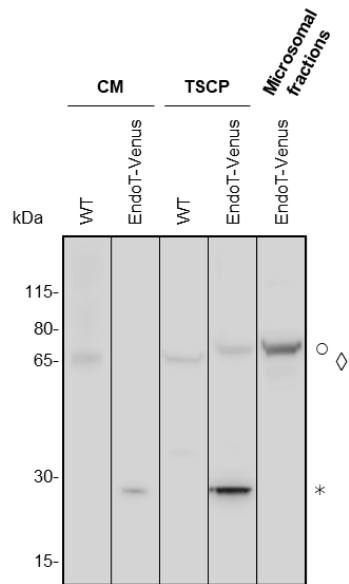


Figure S7. *In vivo* processing of EndoT-Venus in BY-2 cells. Fourty μ L of 7-day-old culture medium (CM), 40 μ g of total soluble cellular proteins (TSCPs) and microsomal fractions from wild type (WT) and EndoT-Venus BY-2 cell cultures were analyzed by Western blotting using anti-GFP antibodies and a horseradish peroxidase-linked secondary antibody. Hollow dots (○) indicate bands, the size of which correspond to unprocessed EndoT-Venus, whereas a star (*) indicates processed, soluble, Venus. Diamond (◇) mark signals from endogenous proteins of BY-2 cells.

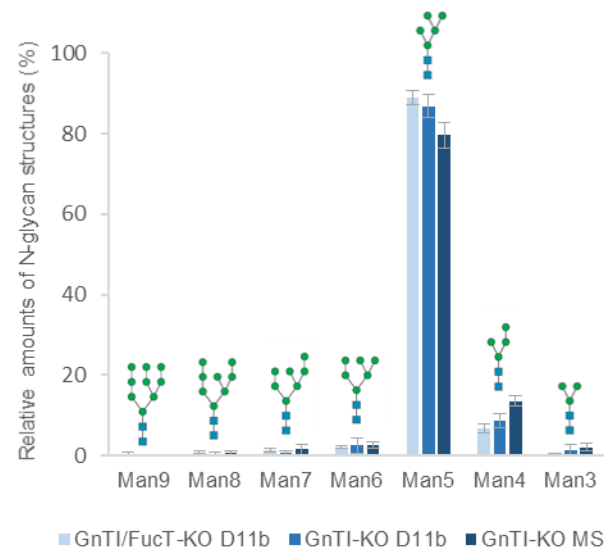


Figure S8. Relative amounts of total *N*-glycans from secreted glycoproteins of *GnTI*-KO and *GnTI/FucT*-KO cell lines. Relative amounts of *N*-glycans are determined by MALDI FT-ICR mass spectrometry. Error bars represent standard deviation of $n = 3$ for *GnTI/FucT*-KO (#5, 10, 16) and $n = 4$ for *GnTI*-KO (#2, 23, 27, 30) cell lines. Cultures were either grown for 10 days in D11b medium (*GnTI/FucT*-KO and *GnTI*-KO) or for 7 days in MS medium (*GnTI*-KO).

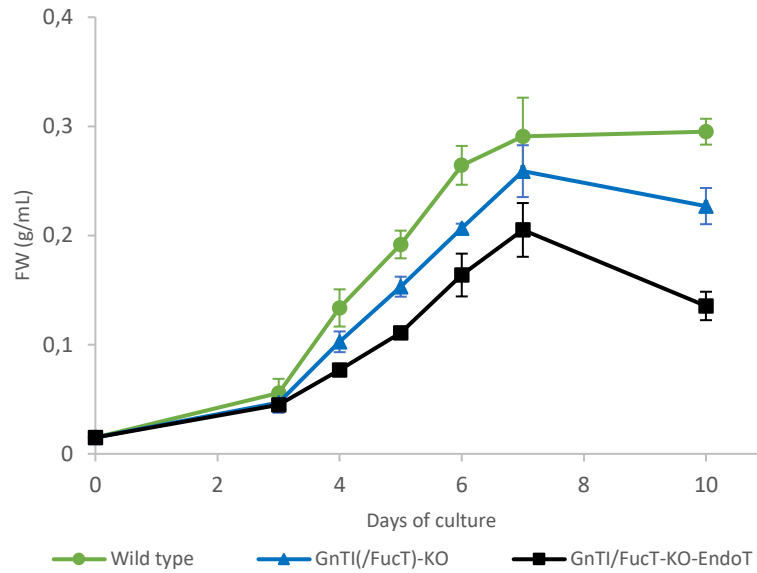


Figure S9. Growth kinetics of WT, *GnTI/(FucT)*-KO and *GnTI/FucT*-KO-*EndoT* cell lines. Independent transgenic cell lines from each type were inoculated at 0.015 g/mL in 50 mL MS medium. Error bars represent standard deviation of $n = 3$. Wild type: WT-gB #1, 6 and 17; *GnTI/(FucT)*-KO: *GnTI/FucT*-KO-gB #5 and 16 and *GnTI*-KO-gB #4; *GnTI/FucT*-KO-*EndoT*: *GnTI/FucT*-KO -*EndoT*-gB #19, 29 and 30.

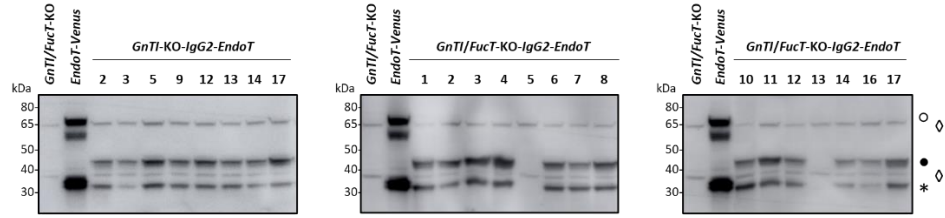


Figure S10. EndoT content of *GnTI-KO-EndoT-IgG2* and *GnTI/FucT-KO-EndoT-IgG2* cell lines. Thirty-five μ g of TCP were analyzed by Western blotting using an anti-EndoT serum and an alkaline phosphatase (AP)-linked secondary antibody. *GnTI/FucT-KO* and *EndoT-Venus* samples were loaded as controls. Hollow (○) and full (●) dots indicate bands, the size of which correspond to unprocessed EndoT-Venus and EndoT, respectively, whereas a star (*) indicates bands, the size of which correspond to processed EndoT and EndoT-Venus. Diamonds (◇) mark BY-2 endogenous signals. Screening for EndoT content was carried out with similar results for other EndoT-expressing cell lines (data not shown).

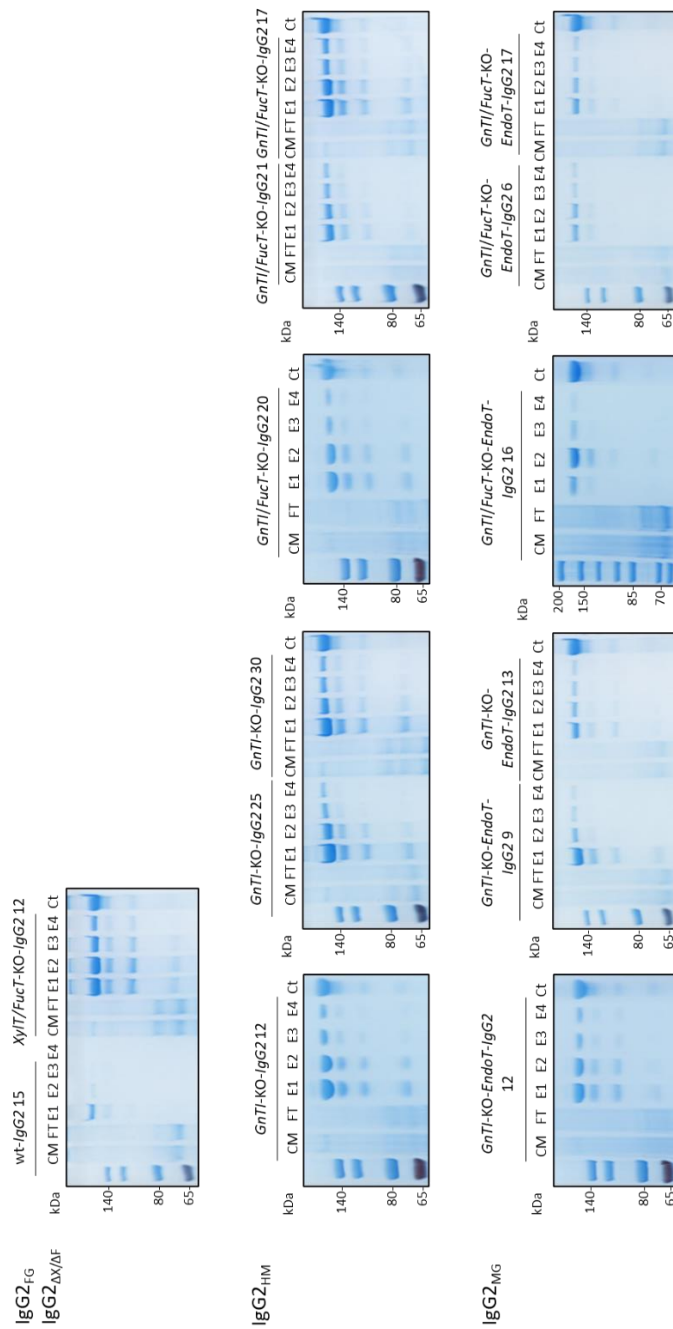


Figure S11. Purification of IgG2 from BY-2 cell cultures. Non-reducing SDS-PAGE of fractions eluted from protein A, colloidal blue stained. Twenty μ L of each fraction were loaded. CM: culture medium; FT: flow-through; E1-4: Elution fractions 1-4; Ct: 5 μ g of polyclonal mouse IgG. Molecular mass of IgG2 with no glycosylation is expected to be around 144 kDa.

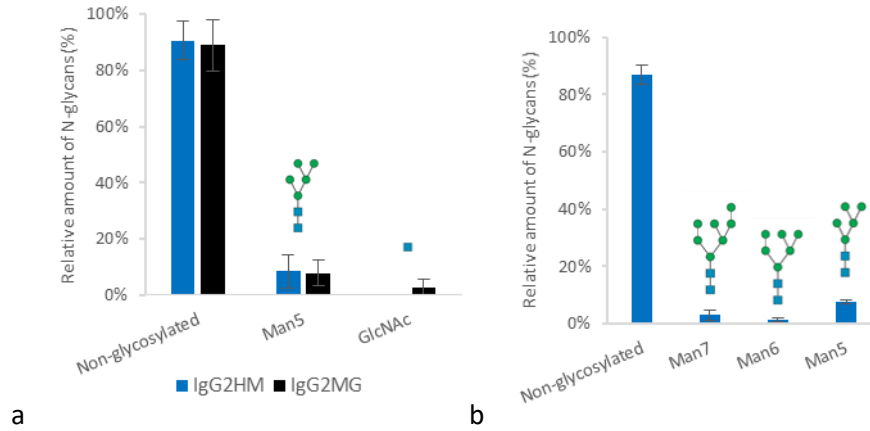


Figure S12. Relative amount of *N*-glycans on the *N*-site CVRPN*YSGF (A) and TVSWN*SGAL (B) of secreted IgG2. *N*-glycosylation was determined by LC-MS/MS of IgG2 samples purified from 10-day-old D11b culture medium. Structures accounting for less than 1% of *N*-glycans were not represented. (A) Error bars represent standard deviation of $n = 4$ (IgG2_{HM}: *GnTI*-KO-IgG2 #12, 30 and *GnTI/FucT*-KO-IgG2 #17, 20) and $n=3$ (IgG2_{MG}: *GnTI*-KO-*EndoT*-IgG2 #12 and *GnTI/FucT*-KO-*EndoT*-IgG2 #6 and 16) independent cell lines. (B) Error bars represent standard deviation of $n = 3$ (IgG2_{HM}: *GnTI*-KO-IgG2 #12, 30 and *GnTI/FucT*-KO-IgG2 #17) independent cell lines. Occupation of site TVSWN*SGAL was not determined for IgG2_{MG}.

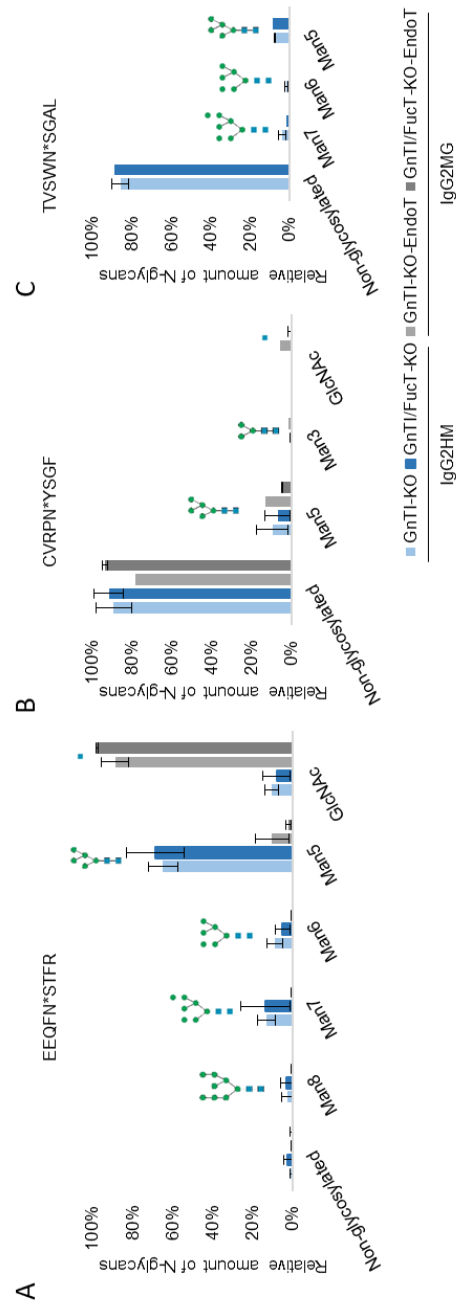


Figure S13. Relative amount of N-glycans on the N-sites of secreted IgG2, per type of cell line. N-glycosylation was determined by LC-MS/MS of IgG2 samples purified from 10-day-old D11b culture medium. Structures accounting for less than 1% of N-glycans were not represented. (A) EEQFN*STFR data were obtained from the cell lines described in Figure 29. (B) CVRPN*YSGF data were obtained from the cell lines described in Figure S12A. (C) TVSWN*SGAL data were obtained from cell lines described in Figure S12B.

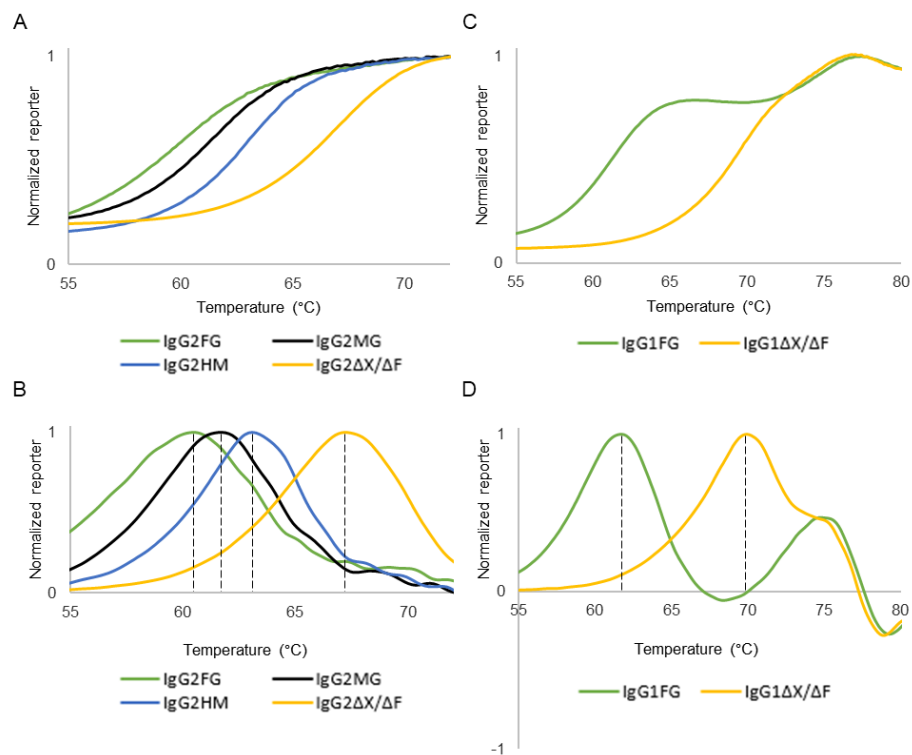


Figure S14. Analysis of IgG thermal stability by differential scanning fluorimetry. Melt curves and first derivatives of IgG2 (A and B, respectively) and hIgG1 (C and D, respectively) glycovariants secreted in the culture medium, obtained by differential scanning fluorimetry assay. Representative curves for each glycovariant are shown. Dotted lines represent T_m.

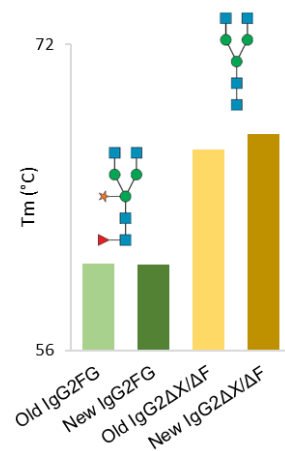


Figure S15. Thermal stability of old and new IgG2_{FG} and IgG2_{ΔX/ΔF}. T_m of IgG2 glycovariants secreted in BY-2 CM measured by differential scanning fluorimetry analysis. As an indication, the main *N*-glycan structure of each glycovariant at the CH2 *N*-site is represented above the corresponding bar.

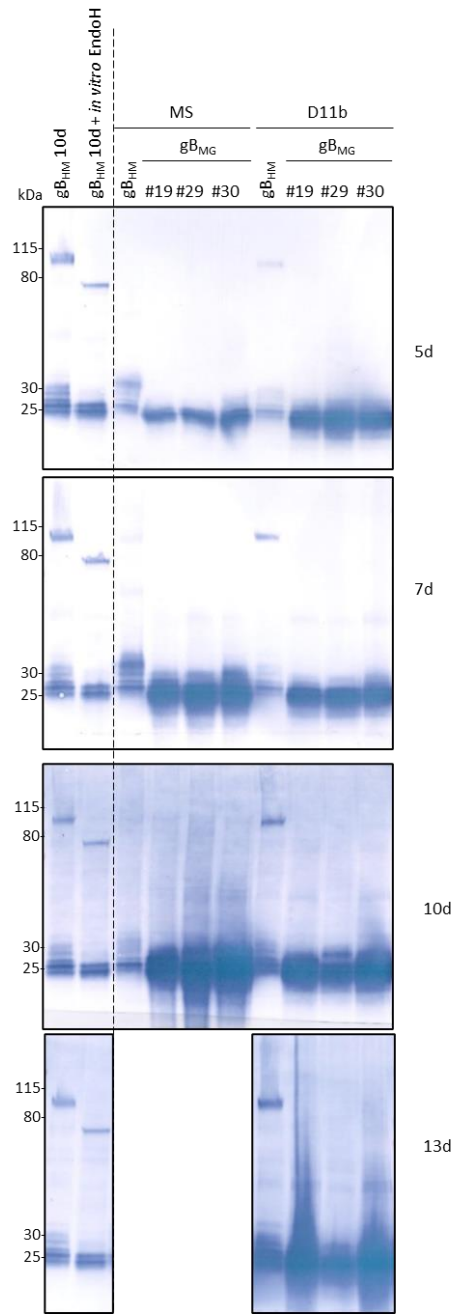


Figure S16. Kinetics of accumulation of gB in the culture medium of glyco-engineered BY-2 cell lines. Thirty-five μ L of CM from MS and D11b cultures at indicated time points were analyzed by Western blotting using anti-gB antibody and AP-linked secondary antibody. For the samples in D11b, which grow slower, an additional time point (13 days) was added. The first two lanes of each membrane contain the same standard for gB_{HM} and gB_{MG} deglycosylated: 10-day-old D11b *GnTII/FucT-KO-gB* #5 and that sample *in vitro* deglycosylated with EndoH, respectively. gB_{HM} and gB_{MG} samples correspond to CM from *GnTII/FucT-KO-gB* #5 and *GnTII/FucT-KO-EndoT-gB* #19, 29 and 30 cell lines, respectively.

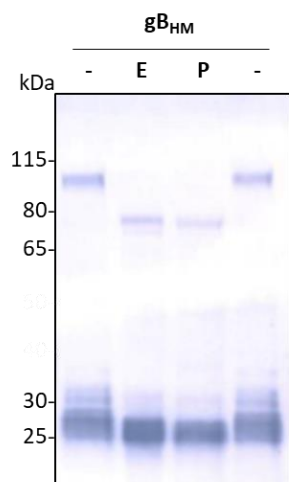


Figure S17. Deglycosylation gel shift assay of secreted gB_{HM}. Immunodetection of gB_{HM} in the *GnTII/FucT*-KO-*gB* #5 cell line without deglycosylation (-) and after deglycosylation by EndoH (E) or PNGase F (P). The wells were loaded with 18 μ L of culture medium. Primary and secondary antibodies were anti-gB antibodies and AP-linked antibodies, respectively. Membranes were revealed using a precipitating substrate.

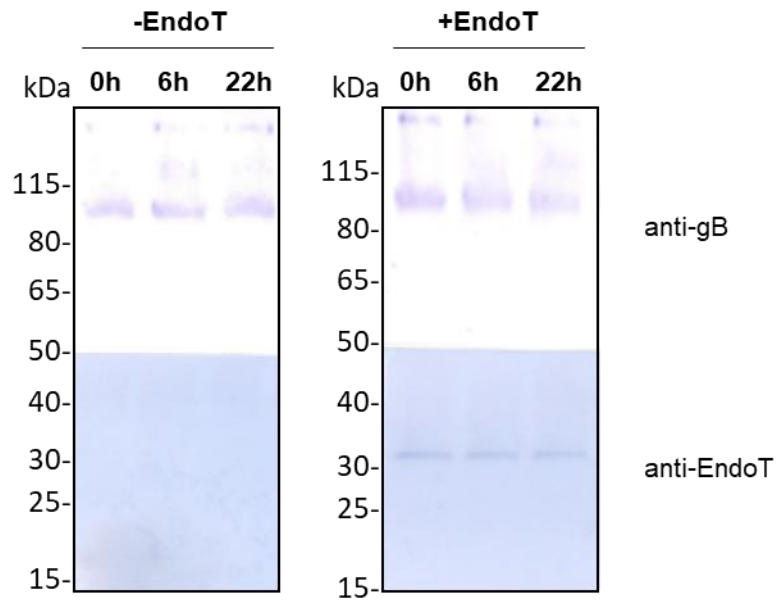


Figure S18. Spiking of gB_{HM} in BY-2 cell culture medium. Twenty μL of *GnTI/FucT*-KO-gB CM were mixed with 20 μL of *GnTI/FucT*-KO (-EndoT) or *GnTI/FucT*-KO-*EndoT*-IgG2 (+EndoT) and directly frozen or incubated 6 h or 22 h at 25°C. Then, samples were separated on SDS-PAGE and immunodetection was carried out using anti-gB (upper part) and anti-EndoT (lower part) primary antibodies followed by AP-linked secondary antibodies. Membranes were revealed using a precipitating substrate. EndoT revelation required a longer incubation, resulting in a higher background signal.

Table S1. Primers used in this thesis. The name of a primer indicates if it is specific to *GnTL.A* (A) or *GnTL.B* (B) or if it is not ortholog-specific (no A/B prefix), which intron (In#) or exon (Ex#) it hybridizes to, and its sense: forward (F) or reverse (R).

Name	Sequence (5'-3')
Chapter 2	
AEx1F	CCTTCAATCGAAATCGCACG
BEx1F	CCTTCAATCGAATTCGCACG
Ex6F	ACCTGGAAAAGACTATTAAATCCATC
BIn8F	CCTCTCTCTCCATAAGTCC
BIn12R	CAAATATTCAATCGAAGCCCC
Ex15R	TTAATTGGCTCAAGATACTGCTTG
AEx19R	CACTTGGCTTAAAAAAAATAGGTTGTGA
BEx19R	TGTTCTGAACACTTAAAATAATAGGTTGTGT
Ex1F	AGACTATTCGCTTTCTCCTAAA
In3R	GGTATTGGTTAGCTCATGC
In12F	GTATATCCTCCATGGAGGT
Chapter 4	
gBR	CCGTCCAATTTGGATTGCTAG
LBR	TGGCAGGATATATTGTGGTGTAAC
CodAF	ATGTCGAATAACGCTTTACAA
XylTF	ATGAACAAGAAAAAGCTGAAA
XylTR	ACTCCTACCAATAACCGACT
Cas9F	GCTTGGATCTCAAATCTTGAAG
TnosR	GATCTAGTAACATAGATGACACC
RecF	GACAACTTAATAACACATTGCGGAC
RecR	GAAACGACAATCTGATCCAAGC

Table S2. Characteristics of crRNAs used in this thesis.

Name	Sequence (5'-3')	Sense	Exon targeted
Target: <i>GMD</i>			
gRNA1	CGCAAAATTGCACTGATCAC	Forward	1
gRNA2	CGGAAAGTAGCTCTGATCAC	Forward	1
gRNA3	AACAAAAACTACGAAGTTCA	Forward	1
gRNA4	GACAAAGGCTACGAAGTCCA	Forward	1
gRNA5	GTAAATGTGATTAATGCGTT	Reverse	1
gRNA6	GCCTTTGTCAAGGAGGAATT	Reverse	1
Target: <i>GnTI</i>			
gRNA1	CGACAGCAGCCAAGATGAGG	Reverse	1
gRNA2	AATCTGGTCAATAAGCAATC	Reverse	3
gRNA3	GAAAGCATCCATCCAAGACC	Reverse	13
gRNA4	AGACTCAAAGAGAATCACAG	Forward	14

Table S3. Relative amounts of total *N*-glycans (%) of RNase B. Relative amounts were determined by considering only the absolute intensities of the peaks associated to a *N*-glycan structure. *According to the nomenclature (Proglycan short) from <http://www.proglycan.com>.

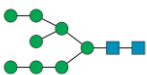
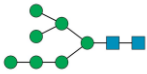
Composition	Putative structure	Abbrev.*	Rel. Amount
HexNAc2Hex8		Man8	2.5%
HexNAc2Hex7		Man7	7.2%
HexNAc2Hex6		Man6	42.2%
HexNAc2Hex5		Man5	48.1%

Table S4. Summary of glyco-engineered BY-2 cell lines, N-glycans and glycovariants nomenclature.
Abbreviations according to the nomenclature (Proglycan short) from <http://www.proglycan.com>.

BY-2 cell lines	Wild type	Xyl/ITIFucT-KO	GnT4-KO	GnTIIIFucT-KO	GnT4-KO-EndoT	GnTIIIFucT-KO-EndoT
Major N-glycan structure expected	 GnGnXF	 GnGn	 Man5	 Mono-GlcNAc		
Glycovariant (abbrev.)	Fully glycosylated (FG)	Xylose- and α 1,3-fucose-devoid (Δ X/ Δ F)	High mannose (HM)		Mono-GlcNAc (MG)	

N-acetylglucosamine (GlcNAc, Gn)
 Mannose (Man, M)
 Xylose (Xyl, X)
 Fucose (Fuc, F)

Table S5. Relative amount of *N*-glycans on the *N*-sites of secreted IgG2. *N*-glycosylation was determined by LC-MS/MS of IgG2 samples purified from 10-day-old D11b culture medium. Abbreviations according to the nomenclature (Proglycan short) from <http://www.proglycan.com>. Relative amounts were determined by considering only the absolute intensities of the peaks associated to an *N*-glycan structure. ND, not detected; NA, not assigned.

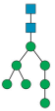
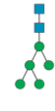
Glycan type	Composition	Putative structure	Abbrev.	Relative amounts					
				EEQFN*STFR		CVRPN*YSGF		TVSWN*SGAL	
				IgG _{HM}	IgG _{MG}	IgG _{HM}	IgG _{MG}	IgG _{HM}	IgG _{MG}
Non-glycosylated	/		NA	1.3 ± 1.4%	0.3 ± 0.4%	90.6 ± 6.9%	88.8 ± 9.0%	86.9 ± 3.6%	
Oligomannose	Hex8HexNAc2		Man8	2.7 ± 2.5%	0.1 ± 0.3%	<0.1%	ND	0.2 ± 0.3%	
	Hex7HexNAc2		Man7	13.1 ± 7%	0.1 ± 0.3%	0.2 ± 0.4%	0.1 ± 0.2%	3 ± 1.7%	
	Hex6HexNAc2		Man6	7.2 ± 4%	0.1 ± 0.1%	0.2 ± 0.4%	<0.1%	1.2 ± 0.7%	
	Hex5HexNAc2		Man5	66.3 ± 9.1%	5.9 ± 6.8%	8.4 ± 6%	7.8 ± 4.6%	7.8 ± 0.7%	
	Hex4HexNAc2		Man4	<0.1%	ND	0.2 ± 0.2%	0.1 ± 0.2%	<0.1%	
	Hex3HexNAc2		Man3	<0.1%	<0.1%	0.3 ± 0.4%	0.5 ± 0.8%	0.6 ± 1%	
Deglycosylated	HexNAc		Mono-GlcNAc	9.3 ± 4.4%	93.3 ± 6.4%	<0.1%	2.5 ± 3.1%	0.3 ± 0.1%	

Table S6. Custom plant N-glycan exact mass table used for the data processing with PEAKS Xpro. ^a Exact masses were computed from the empiric formula and the exact masses obtained from [https://www.ciaaw.org/] the Commission on the Isotopic Abundances and Atomic Weight (CIAAW) of the International Union of Pure and Applied Chemistry (IUPAC), minus the exact mass of H₂O due to the linkage between the N-glycan and the Asn residue of the glycopeptide, as required by Peaks Xpro to build the custom post-translational modification table. * Glycans sharing the same empiric formulae and being perfectly isobars are labeled using the same capital letter index

N-glycan type	Formula	Exact Mass (m/z) ^a	*
Oligomannose	GlcNAc(1)	203,079374	
	GlcNAc(1)Fuc(1)	349,137284	
	Gal(1)GlcNAc(1)	365,132199	
	Man(3)GlcNAc(2)	892,317223	
	Man(4)GlcNAc(2)	1054,370048	
	Man(5)GlcNAc(2)	1216,422873	
	Man(6)GlcNAc(2)	1378,475698	
	Man(7)GlcNAc(2)	1540,528523	
	Man(8)GlcNAc(2)	1702,581348	
	Man(9)GlcNAc(2)	1864,634173	
Fucosylated oligomannoses	Man(3)GlcNAc(2)Fuc(1)	1038,375133	
	Man(4)GlcNAc(2)Fuc(1)	1200,427958	
	Man(5)GlcNAc(2)Fuc(1)	1362,480783	
	Man(6)GlcNAc(2)Fuc(1)	1524,533608	
	Man(7)GlcNAc(2)Fuc(1)	1686,586433	
	Man(8)GlcNAc(2)Fuc(1)	1848,639258	
	Man(9)GlcNAc(2)Fuc(1)	2010,692083	
Hybrid/complex	GlcNAc(1)Man(5)GlcNAc(2)	1419,502247	A
	GlcNAc(1)Man(5)GlcNAc(2)Fuc(1)	1565,560157	B
	GlcNAc(1)Man(5)GlcNAc(2)Xyl(1)	1551,544507	C
	GlcNAc(1)Man(5)GlcNAc(2)Xyl(1)Fuc(1)	1697,602417	D
	GlcNAc(1)Man(4)GlcNAc(2)	1257,449422	E
	GlcNAc(1)Man(4)GlcNAc(2)Fuc(1)	1403,507332	F
	GlcNAc(1)Man(4)GlcNAc(2)Xyl(1)	1389,491682	G

GlcNAc(1)Man(4)GlcNAc(2)Xyl(1)Fuc(1)	1535,549592	H
GlcNAc(1)Man(3)GlcNAc(2)	1095,396597	
GlcNAc(1)Man(3)GlcNAc(2)Fuc(1)	1241,454507	
GlcNAc(1)Man(3)GlcNAc(2)Xyl(1)	1227,438857	
GlcNAc(1)Man(3)GlcNAc(2)Xyl(1)Fuc(1)	1373,496767	
GlcNAc(2)Man(3)GlcNAc(2)	1298,475971	
GlcNAc(2)Man(3)GlcNAc(2)Fuc(1)	1444,533881	
GlcNAc(2)Man(3)GlcNAc(2)Xyl(1)	1430,518231	
GlcNAc(2)Man(3)GlcNAc(2)Xyl(1)Fuc(1)	1576,576141	
Gal(1)GlcNAc(1)Man(5)GlcNAc(2)	1581,555072	
Gal(1)GlcNAc(1)Man(5)GlcNAc(2)Fuc(1)	1727,612982	I
Gal(1)GlcNAc(1)Man(5)GlcNAc(2)Xyl(1)	1713,597332	
Gal(1)GlcNAc(1)Man(5)GlcNAc(2)Xyl(1)Fuc(1)	1859,655242	J
Gal(1)GlcNAc(1)Man(4)GlcNAc(2)	1419,502247	A
Gal(1)GlcNAc(1)Man(4)GlcNAc(2)Fuc(1)	1565,560157	B
Gal(1)GlcNAc(1)Man(4)GlcNAc(2)Xyl(1)	1551,544507	C
Gal(1)GlcNAc(1)Man(4)GlcNAc(2)Xyl(1)Fuc(1)	1697,602417	D
Gal(1)GlcNAc(1)Man(3)GlcNAc(2)	1257,449422	E
Gal(1)GlcNAc(1)Man(3)GlcNAc(2)Fuc(1)	1403,507332	F
Gal(1)GlcNAc(1)Man(3)GlcNAc(2)Xyl(1)	1389,491682	G
Gal(1)GlcNAc(1)Man(3)GlcNAc(2)Xyl(1)Fuc(1)	1535,549592	H
Gal(1)GlcNAc(2)Man(3)GlcNAc(2)	1460,528796	
Gal(1)GlcNAc(2)Man(3)GlcNAc(2)Fuc(1)	1606,586706	K
Gal(1)GlcNAc(2)Man(3)GlcNAc(2)Xyl(1)	1592,571056	
Gal(1)GlcNAc(2)Man(3)GlcNAc(2)Xyl(1)Fuc(1)	1738,628966	L
Gal(2)GlcNAc(2)Man(3)GlcNAc(2)	1622,581621	
Gal(2)GlcNAc(2)Man(3)GlcNAc(2)Fuc(1)	1768,639531	M
Gal(2)GlcNAc(2)Man(3)GlcNAc(2)Xyl(1)	1754,623881	
Gal(2)GlcNAc(2)Man(3)GlcNAc(2)Xyl(1)Fuc(1)	1900,681791	N
Fuc(1)Gal(1)GlcNAc(1)Man(5)GlcNAc(2)	1727,612982	I
Fuc(1)Gal(1)GlcNAc(1)Man(5)GlcNAc(2)Fuc(1)	1873,670892	
Fuc(1)Gal(1)GlcNAc(1)Man(5)GlcNAc(2)Xyl(1)	1859,655242	J
Fuc(1)Gal(1)GlcNAc(1)Man(5)GlcNAc(2)Xyl(1)Fuc(1)	2005,713152	
Fuc(1)Gal(1)GlcNAc(1)Man(4)GlcNAc(2)	1565,560157	B

Fuc(1)Gal(1)GlcNAc(1)Man(4)GlcNAc(2)Fuc(1)	1711,618067	
Fuc(1)Gal(1)GlcNAc(1)Man(4)GlcNAc(2)Xyl(1)	1697,602417	D
Fuc(1)Gal(1)GlcNAc(1)Man(4)GlcNAc(2)Xyl(1)Fuc(1)	1843,660327	
Fuc(1)Gal(1)GlcNAc(1)Man(3)GlcNAc(2)	1403,507332	F
Fuc(1)Gal(1)GlcNAc(1)Man(3)GlcNAc(2)Fuc(1)	1549,565242	
Fuc(1)Gal(1)GlcNAc(1)Man(3)GlcNAc(2)Xyl(1)	1535,549592	H
Fuc(1)Gal(1)GlcNAc(1)Man(3)GlcNAc(2)Xyl(1)Fuc(1)	1681,607502	
Fuc(1)Gal(1)GlcNAc(2)Man(3)GlcNAc(2)	1606,586706	K
Fuc(1)Gal(1)GlcNAc(2)Man(3)GlcNAc(2)Fuc(1)	1752,644616	
Fuc(1)Gal(1)GlcNAc(2)Man(3)GlcNAc(2)Xyl(1)	1738,628966	L
Fuc(1)Gal(1)GlcNAc(2)Man(3)GlcNAc(2)Xyl(1)Fuc(1)	1884,686876	
Fuc(1)Gal(2)GlcNAc(2)Man(3)GlcNAc(2)	1768,639531	M
Fuc(1)Gal(2)GlcNAc(2)Man(3)GlcNAc(2)Fuc(1)	1914,697441	O
Fuc(1)Gal(2)GlcNAc(2)Man(3)GlcNAc(2)Xyl(1)	1900,681791	N
Fuc(1)Gal(2)GlcNAc(2)Man(3)GlcNAc(2)Xyl(1)Fuc(1)	2046,739701	P
Fuc(2)Gal(2)GlcNAc(2)Man(3)GlcNAc(2)	1914,697441	O
Fuc(2)Gal(2)GlcNAc(2)Man(3)GlcNAc(2)Fuc(1)	2060,755351	
Fuc(2)Gal(2)GlcNAc(2)Man(3)GlcNAc(2)Xyl(1)	2046,739701	P
Fuc(2)Gal(2)GlcNAc(2)Man(3)GlcNAc(2)Xyl(1)Fuc(1)	2192,797611	
Man(3)GlcNAc(2)Xyl(1)	1024,359483	
Man(3)GlcNAc(2)Xyl(1)Fuc(1)	1170,417393	

Materials and Methods

1. *N. tabacum* BY-2 cell and *Nicotiana* plants culture

N. tabacum cultivar BY-2 suspension cells (Nagata et al., 1992) were grown in the dark at 25°C with agitation on a rotary shaker (90 rpm) in liquid MS medium [4.4 g/L Murashige and Skoog salts (MP BIOMEDICALS, Solon, OH, United States), 30 g/L sucrose, 0.2 g/L KH₂PO₄, 2.5 mg/L thiamine, 50 mg/L myo-inositol, and 0.2 mg/L 2,4-dichlorophenoxyacetic acid, pH 5.8 (KOH)]. Cultures were grown in 50 mL of medium in a 250 mL Erlenmeyer flask and an 8% inoculum was transferred each week into fresh medium. Transformed cells were grown on MS agar plates supplemented with 100 µg/mL kanamycin or 50 µg/mL hygromycin. Samples used for genomic analyses and Western blot screening were harvested from liquid MS cultures obtained after at least two passages on solid selective medium (1–2 months) and at least two passages in liquid selective medium. Lines used for MS analysis were checked by Western blotting and the same profile was invariably observed. D11b cultures (Vasilev *et al.*, 2013) were grown for ten days in 50 mL of medium without cyclodextrin, in 250 mL Erlenmeyer flasks.

N. tomentosiformis, *N. sylvestris*, and *N. tabacum* seeds were purchased from Bergerac Seed and Breeding (France). *Nicotiana* plants were germinated and grown in soil in a phytotron (8h dark/16h light regime at 25°C (day)/18°C (night) with a light intensity of approximately 200 µmol photon m⁻² s⁻¹). *N. tomentosiformis*, *N. sylvestris*, and *N. tabacum* seeds were first vernalized at 4°C for 3 days.

2. BY-2 cell culture growth kinetics

Three independent transgenic cell lines from each cell type were inoculated at 0.015 g/mL in 50 mL MS medium in a 250 mL Erlenmeyer flask. At each time point, 2 mL of culture were filtered on four layers of Miracloth (Calbiochem, 475855, Billerica, MA, USA) and the cell pack fresh weight was measured after centrifugation for 3 min at 2,700 *g*.

3. BY-2 cell culture sensitivity test

CodA-Cre BY-2 cell lines as well as lines Multi^R23 and CodA-Cre were grown in 4 mL MS medium supplemented with hygromycin (50 µg/mL), kanamycin (100 µg/mL) and 5-FC (100 µg/mL). FW was measured at seven days of culture, after two passages.

4. Binary plasmid construction

4.1. Cas9 and gRNAs plasmids

GnTI-specific guide RNAs (gRNAs) were obtained by fusing the CRISPR RNA (crRNA) sequence (Table S2) to the *trans*-activating crRNA (tracrRNA) optimized by Dang *et al.* (2015). These gRNAs target exons at sites conserved in all identified transcript variants of the orthologs. crRNAs were chosen to minimize the risk of off-targets using the Cas-OFFinder tool (Bae *et al.*, 2014). Moreover, crRNAs designed in this study were selected to have a high minimum free energy (−1 to 0 kcal/mol) since crRNA secondary structures can greatly hinder edition efficiency. A transfer RNA (tRNA) was placed before each gRNA to generate a polycistronic tRNA-gRNA (PTG) under the control of the *A. thaliana* U6 promoter (Xie *et al.*, 2015).

The polycistronic tRNA-gRNA targeting *GnTI* (*GnTI*.PTG) flanked with *StuI* and *XhoI* restriction sites, was synthesized (Genscript) and cloned into a pUC57 vector. The *FucT*/*XylT*.PTG of pPAM-FX-KO (Jansing *et al.*, 2019) was replaced by the *GnTI*.PTG using *StuI*-*XhoI* digestion to generate the pPAM-*GnTI*-KO plasmid, which also contains a plant codon optimized *cas9* with a potato IV2 intron controlled by a hybrid *35SPPDK* promoter (Li *et al.*, 2013) and the selectable marker gene *nptII* (Figure 25A). Cassettes are separated from each other by a *N. tabacum* Rb7 scaffold attachment region (SAR) genetic insulator. Complete sequence between left and right borders is given in Figure S2.

To generate *XylT*/*GMD*-KO and *GnTI*/*FucT*-KO cell lines, a pPZP-RCS2 binary plasmid (Goderis *et al.*, 2002) containing the knockout constructs was prepared (Figure 22, 25B). The pPZP-*GnTI*/*FucT*-KO contains *pcas9*, *nptII*, *mCherry* reporter gene, the negative selectable marker *cytosine deaminase A* (Stougaard, 1993), the two PTGs and *loxP* sites at both ends of the T-DNA. *GnTI*.PTG obtained above and *FucT*.PTG (Mercx *et al.*, 2017) were cloned into I-CeuI and PI-PspI restriction sites, respectively. Complete sequence between left and right borders is given in Figure S3. The same construct was obtained by insertion of the composite, single, *XylT*/*GMD*.PTG designed by S. Mercx into the I-CeuI restriction site. Minimum free energy was not taken into account at that time.

4.2. gB plasmids

gB-encoding constructions were based on the pPZP-RCS2-*nptII*-*mCherry*-gBHis plasmid used in previous studies (Smargiasso *et al.*, 2019). A few modifications were introduced to improve gB production. First, as it is not detected in extracellular protein samples, the 6xHis-tag encoding sequence was removed from the *gB*

expression cassette. Second, to reduce the formation of rosette-like gB aggregates and hence favor monodispersed gB trimers, hydrophobic residues in both fusion loops were mutated for more hydrophilic residues. Third, to improve the yield, two gB expression cassettes were introduced into the I-SceI and I-CeuI cloning sites of pPZP-RCS2 (Goderis *et al.*, 2002) and a *N. tabacum* Rb7 scaffold attachment region (*SAR*) genetic insulator was cloned in the PI-PspI site. Finally, the *neomycin phosphotransferase II* (*nptII*) resistance gene was replaced by *hygromycin phosphotransferase* (*hpt*) because *nptII* was already used to select *GnTI*(/*FucT*)-KO cell lines.

4.3. IgG plasmids

To express the human IgG1 and IgG2 Lo-BM2 (Dehoux *et al.*, 2000), the plasmids pPZP-*nptII*-*hIgG1*-LoBM2 and pPZP-*nptII*-*hIgG2*-LoBM2 described in De Muynck *et al.* (2009) and Magy *et al.* (2014) were used to transform WT and *XylT*/*FucT*-KO cell lines via *A. tumefaciens*. To transform *GnTI*(/*FucT*)-KO cell lines, *nptII* was replaced by *hpt* resistance marker (pPZP-*hpt*-*hIgG2*-LoBM2 described in Magy *et al.* (2014)).

4.4. EndoT plasmids

EndoT from *H. jecorina* (UniProtKB-C4RA89), without the sequence encoding its endogenous signal sequence (the 17 first amino acids), was optimized based on *A. thaliana* codon usage, fused at 5' to the sequence encoding the first 32 amino acids of AtXylT (which consists of the cytoplasmic tail as well as the transmembrane region) and at 3' to the sequence encoding a FLAG-tag (DYKDDDDK) and synthesized (Genscript, Piscataway, NJ, USA). *EndoT-Venus* chimeric gene was obtained by fusing *AtXylT-EndoT* to *Venus* cDNA at 3', instead of the FLAG-tag encoding sequence. *AtXylT-EndoT* and *AtXylT-EndoT-Venus* were placed under the control of the enhanced *CaMV 35S* promoter (*En2p35S*) and the nopaline synthase terminator (*Tnos*). Then, *AtXylT-EndoT* was cloned in the PI-PspI site of pPZP-*hpt*-*hIgG2*-LoBM2 and pPZP-*hpt*-*mCherry*-gB. *AtXylT-EndoT-Venus* was cloned in the PI-PspI site of pPZP-*nptII*-*mCherry*-gB.

5. *A. tumefaciens* transformation

Electrocompetent *A. tumefaciens* strain LBA4404 VirG were transformed using 1 μ L *E. coli* DNA miniprep. The conditions for the electroporation were 25 μ F, 400 Ω and 1.25 kV. Cells were regenerated in 2xYT medium (10 g/L bacto-yeast extract, 16 g/L tryptone, 5 g/L NaCl, 2 g/L glucose, 0.4 g/L MgSO₄) for 4 hours at 28°C and plated on 2xYT agar plates supplemented with 40 μ g/mL gentamycin, 20 μ g/mL rifampicin, and 100 μ g/mL spectinomycin. Colonies were visible after three days of incubation at 28°C.

6. Plant cell transformation

6.1. *Agrobacterium*-mediated transformation

A. tumefaciens LBA4404 VirG-mediated stable transformation of BY-2 cells was carried out as described by Navarre and Chaumont (2022). *nptII* resistance marker was used to transform WT and *XylT/FucT*-KO BY-2 cell lines, *hpt* resistance gene was used to transform *GnTI*-KO and *GnTI/FucT*-KO (*GnTI*/*FucT*)-KO cell lines. Selection of transformed cell lines was carried out on MS-agar plates containing 500 µg/mL cefotaxime, 400 µg/mL carbenicillin and 100 µg/mL kanamycin or 50 µg/mL hygromycin, respectively.

6.2. Biolistics

Due to an incompatibility of pPAM vector (which necessitates the presence of plasmid replication initiation proteins *in trans*) with *Agrobacterium* LBA4404 VirG and the very low BY-2 transformation efficiency observed in our laboratory with the pPAM-compatible strain GV3101 PMP90RK (Koncz and Schell, 1996), we opted for the biolistics transformation technique.

Biolistic particle delivery was carried out with the plasmid pPAM-*GnTI*-KO. Four milliliters of three-day-old BY-2 cells were vacuum filtered on a n°4 filter (Whatman, 1004 090, Maidston, UK) and the cells were transferred on solid MS medium. The next day, plasmid DNA (2 µg) was precipitated on 600 µg of 0.6 µm diameter gold microcarriers (Bio-Rad, 1652262, Hercules, CA, USA) and the cells on the filter were bombarded under a 28-inch Hg vacuum (95 kPa), using a 1,100 PSI (7,600 kPa) rupture disk, at a shooting distance of 4 cm with a Biolistic PDS1000/He device (Bio-Rad). The filter was kept on solid MS medium for three days at 25°C under dark conditions. Filter was then transferred onto kanamycin-supplemented MS medium. After four to five weeks, the growing calli were transferred on new solid kanamycin-supplemented MS medium.

7. Proteolistics

Proteolistic delivery of GFP was carried out based on the method developed by Martin-Ortigosa and Wang (2014). Briefly, epidermis tissue from white onion bulbs was peeled off, cut into ~2.5 cm² pieces and placed on a MS-agar Petri dishes just before bombardment, peeled side upwards. Alternatively, the day before bombardment, 4 mL of BY-2 cells were filtered on n°4 filters (Whatman, 1004 090). They were then placed on a MS-agar Petri dish. Per shot, 5 µL of 0.5 µg/µL eGFP protein in phosphate-buffered saline, produced in *E. coli* (in-house) were mixed with 2 µL of gold microcarriers (Bio-Rad, 1652262) at 30 µg/µL in water. The mixture

was placed on a macrocarrier and air-dried for 2h. Shooting was carried out as described for biolistics. Best results with onion epidermis were obtained using 9 cm distance between tissue and macrocarrier; 0.5 cm distance between rupture disk and macrocarrier. Samples were imaged with a LEICA DMR microscope coupled to a micro color liquid crystal tunable RGB filter (CRI instruments, Woburn, USA) and a HAMAMATSU camera (Hamamatsu photonics, Shizuoka, Japan) using the Hokawo™ software (version 2.1, release pf01, Hamamatsu photonics).

8. Light microscopy

BY-2 cell samples were imaged with a 34-channel Zeiss LSM 710 AxioObserver confocal microscope equipped with Airyscan superresolution system and with a Plan Aplanachromat 25x numerical aperture 0.8 water immersion objective.

9. RNA extraction and RT-PCR

RNA extraction was carried out using the Spectrum Plant Total RNA Kit (Sigma-Aldrich, STRN250, St Louis, MO, USA). cDNAs were obtained using M-MLV Reverse Transcriptase (Promega, M1705, Madison, WI, USA). PCR amplification was performed with Q5 High-Fidelity DNA Polymerase (New England Biolabs, M0491, Ipswich, MA, USA) and the primers listed in Table S1. PCR products were purified and sequenced (Microsynth, Balgach, Switzerland). GenBank mRNA accessions are: *NtGnTLA* (AJ249882), *NtGnTLB* (AJ249883), *NsGnTI* (XM_009803697 and XM_009776507), *NtoGnTI* (XM_009610507), *FucTA* (XM_016657530), *FucTB* (XM_016620229), *FucTC* (NM_001324945), and *FucTD* (XM_016585847).

10. Genomic DNA extraction

Genomic DNA was extracted from 7-day-old BY-2 cells filtered on four layers of Miracloth (Calbiochem, 475855) or from plant leaf material using the Wizard Genomic DNA Purification Kit (Promega, A1120). All samples were first frozen in liquid nitrogen.

11. Characterization of *GnTI* orthologs and analysis of genome modifications

Genomic DNA was amplified by PCR using Q5 High-Fidelity DNA Polymerase (New England Biolabs, M0491) and the primers listed in 2. The PCR products were separated by gel electrophoresis on 1% agarose gel, purified, cloned into the pGEM-T Easy vector (Promega, A1360), and sequenced.

12. Protein extraction

Two mL of the indicated BY-2 culture were filtered on four layers of Miracloth (Calbiochem, 475855). The flow-through was then centrifuged 10 min at 7,200 *g* at 4°C. The supernatant was transferred and corresponds to the culture medium (CM). Filtered BY-2 cells were used to harvest total cellular proteins (TCP), total soluble cellular proteins (TSCP) and microsomal fractions. To do so, cell packs were transferred into a 2 mL micro tube containing 0.5 g of glass beads (0.85–1.23 mm) and frozen in liquid nitrogen. Then, 700 µL of homogenization buffer (250 mM sorbitol, 60 mM Tris-HCl, 2 mM Na₂EDTA, pH 8.0) supplemented with 1 mM phenylmethylsulfonylfluoride (PMSF), and protease inhibitor cocktail (leupeptin, aprotinin, antipain, pepstatin, and chymostatin, each at 2 µg/mL) was added to each sample. Cell grinding was performed for 3 × 40 s at 5 000 rpm (Bertin Technologies, Precellys 24 tissue homogenizer, Montigny-le-Bretonneux, France) with 2 min pauses on ice. The samples were centrifuged first for 5 min at 2,800 *g* at 4°C then for 7 min at 10,000 *g* at 4°C. TCP aliquots were retrieved after this step. To separate TSCP from the microsomal fractions, remaining sample was centrifuged for 15 min at 130,000 *g* at 4°C. The supernatant, corresponding to TSCP, was retrieved and the pellet, corresponding to the microsomal fractions, was resuspended in 50 µL resuspension buffer [3 mM KH₂PO₄, 330 mM sucrose, 3 mM KCl, pH 7.8 (KOH)]. The protein concentrations in the TCP, TSCP and microsomal fractions were quantified by Bradford assay.

13. SDS-PAGE and Western blotting analysis of proteins

Indicated amount of culture medium (µL) or TCP/TSCP/microsomal fractions (µg) were analyzed by SDS-PAGE (reducing, 4–20% polyacrylamide if not stated otherwise) after denaturation (5 min at 100°C for CM and TSCP; 15 min at 56°C for TCP and microsomal fractions). Gels were either stained with Coomassie Brilliant Blue G-250 (SERVA, 17524, Heidelberg, Germany) or transferred onto a PVDF membrane (Millipore, 1704156, Billerica, MA, USA) for Western blotting. The PVDF membrane was incubated with:

- Rabbit polyclonal antibodies against β1,2-xylose (Agrisera, AS07 267, Vännäs, Sweden, 1:5,000)
- Rabbit polyclonal antibodies against α1,3-fucose (Agrisera, AS07 268, 1:5,000)

- Rabbit polyclonal antibodies against the EndoT enzyme (CER groupe, Département Santé, Marloie, Belgium, 1:5,000, custom generated for Meuris *et al.* (2014), kindly provided by UGent)
- Rabbit monoclonal antibodies against HCMV gB (Sino Biological, 10202-R038, Beijing, China, 1:2,000)
- Rabbit polyclonal antibodies against GFP (generated by Duby *et al.* (2001), 1:1,000),

followed by anti-rabbit horseradish peroxidase (HRP) polyclonal antibodies (SynAbs, LO-RG-1-HRP, Gosselies, Belgium, 1:10,000,) or anti-rabbit alkaline phosphatase (AP) polyclonal antibodies (Sigma-Aldrich, A3687, 1:10,000). Alternatively, the PVDF membrane was incubated with:

- Goat polyclonal antibodies against human IgG (Fc-specific) coupled to HRP (Sigma-Aldrich, A0170, dilution 1:5,000)
- Concanavalin A coupled to HRP (Sigma-Aldrich, L6397, 1/200)
- HisProbe coupled to HRP (Thermo Fisher Scientific, 15165, 1:1,000)

Western blots using HRP-conjugated antibodies (or lectin) were revealed with BM Chemiluminescence Blotting Substrate (Roche, 11500694001, Basel, Switzerland). Western blots using AP-conjugated antibodies were revealed with Immun-Star AP chemiluminescent substrate (Bio-Rad, 1705018) or BM Purple AP substrate, precipitating (Roche, 11442074001). Chemiluminescence was detected using the Amersham™ Imager 600 (GE Healthcare, Chicago, IL, USA). The relative intensity of immunoblot signals was quantified with Amersham Imager 600 Analysis Software.

14. Protein purification

14.1. IgG

IgG purification was carried out as described by Mercx *et al.* (2017), from 10-day-old D11b cultures.

14.2. α 1,3-fucosylated glycoprotein purification

Fifty mL of 10-day-old D11b CM was incubated 16h at 4°C with 400 μ L of Agarose bound *Aleuria Aurantia* lectin (Vector laboratories, AL-1393, Newark, CA, USA). Elution was carried out on chromatography columns (Bio-Rad, 732-6008) using 1.5 mL of 100 mM L-fucose (Vector laboratories, S-9007).

15. Mass spectrometry analysis of the total *N*-glycans from secreted proteins

15.1. *GnTI*-KO #2 and *GnTI/FucT*-KO #10 (chapter 2)

15.1.1. Sample preparation for MALDI mass spectrometry *N*-glycan identification

For the mass spectrometry analysis, cultures were grown in liquid D11b medium (Vasilev *et al.*, 2013) without cyclodextrin for 10 days in 50 mL of medium in 250 mL Erlenmeyer flasks. Extracellular proteins were collected by filtration on four layers of Miracloth (Calbiochem, 475855), then centrifuged at 8,000 *g* for 30 min, and precipitated by salting-out as previously described in Smargiasso *et al.* (2019) with the following minor adaptations: solid $(\text{NH}_4)_2\text{SO}_4$ was slowly added to the culture medium with stirring at room temperature to 55% saturation and the solution was then kept for 2 h at 4°C and centrifuged at 3,000 *g* for 40 min at 4°C. The pellet was solubilized in 140 μL of PBS buffer and desalted on a PD-10 filtration column (GE Healthcare, 28918004) equilibrated with PBS buffer.

Ribonuclease B from bovine pancreas was used as a positive control sample and was prepared using the same procedure as for the glycoprotein samples from BY-2 cells. The samples were first centrifuged at 8,000 *g* for 5 min at 4°C, the supernatant was filtered on Amicon 3 kDa and the protein content assessed using the RC DC protein assay kit (Bio-Rad, 500-0019). Reduction of 350 μg proteins was performed in 50 mM $(\text{NH}_4)_2\text{CO}_3$ and 10 mM of dithiothreitol at 56°C for 40 min at 650 rpm in a thermoshaker. Then alkylation of cysteinyl residues was performed using 20 mM $\text{C}_2\text{H}_4\text{INO}$ at RT for 40 min at 650 rpm using a thermoshaker. 2D clean up purification kit (GE Healthcare, 80648451) was used while the extra washing step (using wash additive) was skipped. Pellets containing proteins were resolubilized in 50 mM $(\text{NH}_4)_2\text{CO}_3$ (pH 7.5) before trypsin digestion (protein:trypsin ratio of 50:1) at 37°C for 16 h at 600 rpm. The samples were freeze-dried using SpeedVac centrifugal evaporator and a second trypsin digestion was performed at 37°C in 80% acetonitrile using 100:1 protein:trypsin ratio for 3 h at 600 rpm. Trypsin was inactivated at 90°C for 2 min.

In order to guarantee the removing of all *N*-glycans of interest, a two-step enzymatic release of the *N*-glycans was performed on the trypsin digested samples using first 3 U per 100 μg proteins of PNGase F (Roche, discontinued product) dissolved in 50 mM $(\text{NH}_4)_2\text{CO}_3$ (pH 8) at 37°C for 16 h. Samples were freeze-dried and resolubilized in 48.3 mM $\text{C}_6\text{H}_8\text{O}_7$ and 103.3 mM $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ buffer (pH 5). The second

deglycosylation step was performed with 0.2 mU of PNGase A (Roche, discontinued product) per 100 µg proteins at 37°C for 16 h at 650 rpm. Samples were stored at −20°C after freeze-drying using a SpeedVac centrifugal evaporator.

The *N*-glycans fraction was separated from the peptide fraction using Waters Sep-Pack SPE C₁₈ cartridges after being reconditioned using 5 mL of pure methanol, then 5 mL of 5% acetic acid, 5 mL of pure isopropanol and finally 15 mL of 5% acetic acid. Samples were loaded on the cartridge after solubilizing the dried residue in 200 µL of 5% acetic acid. *N*-glycans were eluted in the 3 mL of 5% acetic acid mobile phase. All the samples were freeze-dried and kept at −20°C. The *N*-glycans containing fraction was desalted using H-cartridges (PROzyme, Agilent, Santa Clara, CA, USA) after being reconditioned following the manufacturer recommendations. Desalted *N*-glycans elution was performed as follows: 3 mL of ultrapure water (discarded), 3 mL of 5% acetonitrile and 0.1% trifluoroacetic acid (discarded), and 4 × 0.5 mL of 50% acetonitrile and 0.1% trifluoroacetic acid. These collected fractions were freeze-dried and kept at −20°C.

The labeling of *N*-glycans was then performed using a mixture of 31.5 mg of 2-aminobenzamide (2-AB) and 31.5 mg of NaBH₃CN in 650 µL of 10 volumes DMSO: 3 volumes glacial acetic acid. Ten microliters of this mixture were added to the dried *N*-glycan fraction and incubated for 2 h at 65°C under 1,200 rpm agitation using a thermoshaker. The 2-AB labeled *N*-glycans were purified using S-cartridges (PROzyme, Agilent) following the manufacturer recommendations and eluted using 3 × 0.5 mL of ultrapure water in Pyrex tubes. The samples were freeze-dried using centrifugal evaporator and kept at −20°C.

15.1.2. MALDI-mass spectrometry identification of 2-AB labeled *N*-glycans

The purified 2-AB labeled *N*-glycans were resolubilized in 10 µL of 50% acetonitrile spiked with 0.1% formic acid and vigorously vortexed in 0.6 mL Eppendorf tube. One microliter of 2-AB labeled *N*-glycan and 1 µL of 2,5-dihydroxybenzoic acid solution (20 mg/mL of 2,5-dihydroxybenzoic acid in 50% acetonitrile and 0.1% formic acid) were mixed and spotted (dried-droplet method) on a MALDI plate Anchorchip384BC fitted on its plate adapter (Bruker, Billerica, MA, USA) and dried at RT. MALDI mass spectra were obtained using a Solarix XR 9.4T FT-ICR (Bruker) fitted with the dual ESI/MALDI source and Smartbeam LASER. The mass range was set from 150 to 3,500 *m/z* using 4 megawords acquisition, resulting in an average mass resolving power (FWMH) of about 150,000 at *m/z* ≈ 1,100. Two hundred laser shots per scan in small focus mode were selected at the rate of 2,000 Hz. Red phosphorus (Sladkova *et al.*, 2009) was used to optimize the ion transmission of *m/z*

going from 800 to 3,000. The MALDI FT-ICR was mass calibrated using Red P and the peptide calibration standard II (Bruker, 8222570). Mass calibration was checked using the m/z determination of high mannose *N*-glycans released from ribonuclease B. Several replicates of mass spectra were acquired using 12 scan accumulation mode to attempt the detection of minor *N*-glycans. The data were manually processed using DataAnalysis v5.0 software (Bruker). Only peaks with an intensity higher than 1.4×10^6 were considered. The ion mass lists were manually generated based on the GlycoMod tool (Cooper *et al.*, 2001; Cooper *et al.*, 2003) using proton and sodium adducts as potentially detected *N*-glycans². Indeed, proton adducts are very frequent when 2-AB-labelled *N*-glycans are in acidic conditions such as those used in this study for spotting on the MALDI plate (Far, J., personal communication). Mass accuracy (0.1 Da) was used to refine the list of identified *N*-glycans and to manually annotate the mass spectra. Since the FT-ICR mass analyzer reached high mass accuracy on the sub-ppm range (with internal calibration), no cesium spiking assay was carried out to confirm the data.

15.2. *GnTI*-KO and *GnTI/FucT*-KO cell lines (chapter 3)

The determination of the *N*-glycan library was performed using the procedure described above, with minor modifications: The PNGase F-released *N*-glycans and the 2-AB-labeled *N*-glycans were loaded on GlycoClean H and S cartridge (PROzyme Agilent, GKI-4025 and GKI-4726), respectively. The dual ESI/MALDI source solariX xR 9.4T FT-ICR (Bruker) operating in MALDI positive ionization mode was mass calibrated using a dried spot of red phosphorus in acetone and the 2-AB labeled *N*-glycans obtained from bovine ribonuclease B, offering sub ppm mass accuracy. Mass range was set from 150 to 3500 and the resolving mass power was around 300'000 in the m/z of the 2-AB labeled *N*-glycans.

16. Site-specific IgG *N*-glycosylation

16.1. Desalting, reduction/alkylation and clean-up procedures of protein samples

Desalting of the IgG samples was done using Amicon 3 kDa (Millipore, Burlington, MA, USA). The proteins were recovered in ammonium carbonate buffer 50 mM pH 7.5. Protein content was assessed using the RC DC protein assay kit (Bio-Rad, 500-0019) against bovine serum albumin fraction V (Sigma-Aldrich, BSAV-RO) as standard using the recommended manufacturer procedure. Multi-enzymatic limited

² <https://web.expasy.org/glycomod/>

digestion (MELD) on RNase B and *N. tabacum* samples was then performed using the method described by Morsa *et al.* (2019).

Simultaneously, to ensure correct identification of glycopeptides, a fraction of some samples (*GnTI*-KO-IgG2 #25, *GnTI*-KO-*EndoT*-IgG2 #9 and RNase B) were deglycosylated according to the procedure described by Smargiasso *et al.* (2019) and Herman *et al.* (2021). These fractions were then submitted to the MELD followed by LC-MS. The absence of glycopeptides was checked as well as the presence of Asp instead of Asn in the identified *N*-glycosylation sites, especially to confirm the non-consensus WN*SG site. Persistence of the PNGase-resistant mono-GlcNAc after deglycosylation was also confirmed.

16.2. Identification and relative quantification of the glycopeptide by LC-MS/MS

The MELD proteolysis products were reconstituted at 15 pmol in 9 μ L water + 0.1% trifluoroacetic acid. M-Class nanoUPLC pump (Waters, Milford, MA, USA) was used for the separation of (glyco)peptides by reverse phase liquid chromatography (LC). The LC and the MS methods used are those described by Morsa *et al.* (2019), with minor modifications. The LC flow-rate was set at 0.5 μ L/min. The MS system was a Q-Exactive quadrupole-Orbitrap instrument (Thermo Fisher Scientific, Waltham, MA, USA) using NCE = 27 (normalized collision energy).

16.3. Glycopeptides data processing

The raw LC-MS/MS data were submitted to PEAK X pro (v10.6 build 20201221 from Bioinformatics Solutions Inc., Canada; Ma *et al.* (2003)) in dedicated workstation (20 cores 40 threads 2.4 GHz Intel Xeon Gold 6148 fitted with 128 Go RAM running Windows 10 pro 64b 21H1) for the identification of the peptides and proteins including database identification on a custom FASTA file as well as *de novo* identification and PTM (post-translational modification) identification using SPIDER algorithm. FASTA files including all genetically encoded proteins in WT *N. tabacum* BY-2 genome plus the IgGs (light and heavy chains), either a reduced FASTA file including only the sequence of IgGs, RNaseB, and all the four proteins from the MassPREP Digestion Standard Mix 1 (Waters, 186002865) used as quality control. Carbamidomethylation of cysteine was set as fixed modification while methionine oxidation, glutamine and asparagine deamidation were set as variable modifications. Two custom PTM mass tables were created to include either all potential *N*-glycans produced by plant cells, in addition to the mono-GlcNAc stump, either mono-GlcNAc and oligomannoses only (Table S6). To reduce the computational cost, the

submitted jobs in PEAKS X pro was done first on the full FASTA of *N. tabacum* using the reduced custom PTM mass table of oligomannoses and mono-GlcNAc. Then the jobs were resubmitted using the reduced FASTA with the complete plant N-glycan custom library. The data from the resubmitted jobs (sequence coverage, generated ion lists, sequence redundancy, scores, identified PTM) were explored in PEAKS X pro then exported in CSV files. The peptide ion lists were manually curated and the peak area of the (glyco)peptide sequences from the same glycosylation site were summed (assuming the ionization efficiency was similar for the detected peptides). The relative abundance of the glycoforms were computed in a spreadsheet software (LibreOffice v6.4.7).

17. Differential scanning fluorimetry

DSF was carried out based on the protocol described by Moggridge *et al.* (2017). Briefly, purified IgG samples were diluted twice in phosphate buffered saline (100 mM phosphate buffer supplemented with 150 mM NaCl, pH 7.0) containing SYPRO Orange Protein Gel Stain (Invitrogen, S6650, Waltham, MA, USA). Samples were transferred into a MicroAmp™ Optical 96-Well Reaction Plate (Applied Biosystems, N8010560, Waltham, MA, USA). Each sample was run as a technical triplicate and successive dilutions were performed. DSF measurements were performed using a temperature increase from 15°C to 90°C with a ramp of 1 °C/min (Applied Biosystems, StepOnePlus™ Real-Time PCR System). Mean melting temperatures were calculated as the temperature corresponding to the maximum value of the mean derivative curves. Mean derivative curves were obtained as the first derivative of the mean melt curves. For graphing, the raw data was normalized (maximal value = 100%).

18. *In vitro* deglycosylation assay

Twenty µg of TSCP or 20 µL CM were denatured by heating for 10 min at 100°C in the manufacturer's glycoprotein denaturing buffer, to ensure optimal N-site accessibility. Samples were then incubated for 1 h at 37°C with EndoH 50 U (New England Biolabs, P0703) or PNGase F 25 U (New England Biolabs, P0704) using the manufacturer's recommended conditions. Control samples were run in parallel but without the EndoH and PNGase F enzymes. Samples were then visualized by reducing SDS-PAGE followed by Western blotting. PNGase F cleaves between the innermost GlcNAc and Asn residues, which results in Asn deamidation into Asp. Unlike PNGase F and similar to EndoT, EndoH cleaves the N-glycan at the bond joining the two GlcNAc of the chitobiose core and hence leaves a single GlcNAc attached to the asparagine. *In vitro* deglycosylation with PNGase A (New England Biolabs, P0707 or Roche, 11642995001) was also attempted. This enzyme cleaves all

high mannose and complex *N*-glycans, regardless of the fucosylation status. Only sialylated *N*-glycans, that are absent in plants, are reported to be PNGase A-resistant. However, no mobility shift was observed despite trying different denaturing and incubation conditions (data not shown). On the other hand, PNGase A was successfully used in this study to deglycosylate MELD-generated glycopeptides for mass spectrometry purposes. Accordingly, some manufacturers (Agilent, Sigma-Aldrich) report that PNGase A requires glycopeptides, not glycoproteins, as substrates. Moreover, it has been shown that the removal of oligosaccharides from native glycoproteins by PNGase A is typically much slower than their removal from derived glycopeptides (Taga *et al.*, 1984). Hence, we conclude that this enzyme is not adequate for *in vitro* deglycosylation of gB.

19. Representation of *N*-glycan structures

Schematic *N*-glycan structures were generated with DrawGlycan-SNFG (Cheng *et al.*, 2017).

Annex

Inactivation of β 1,2-xylosyltransferase and *GDP-mannose 4,6-dehydratase* genes to obtain a BY-2 cell line producing *N*-glycans without β 1,2-xylose and depleted for α 1,3-fucose and α 1,4-fucose

I designed and performed all the experiments, except for the gRNA design, which was done by Sébastien Mercx.

1. Generation and screening of BY-2 *XylT*/*GMD*-KO cell lines

Suppression of the major non-human residues in BY-2 cells was first obtained by knocking out *XylT* and *α*1,3-*FucT* genes (Mercx *et al.*, 2017; Hanania *et al.*, 2017). In parallel, a strategy was initiated to simultaneously suppress β 1,2-xylose, α 1,3-fucose and the Le^a-associated α 1,4-fucose, through inactivation of *XylT* and *GMD* genes. There are four *N*/*GMD* genes identified, consisting of one or two exons (Figure 1).

XylT and *GMD*-specific gRNAs were combined in a single PTG construct, under the control of U6 promoter, and transferred into pPZP-*loxP*-*nptII*-*codA*-*mCherry*-*cas9*-*loxP* (Figure 2). It contained six *GMD*-specific gRNAs (Table S2) so that each *GMD* gene was targeted by three gRNAs (Figure 1). The three *XylT*-specific gRNAs corresponded to those designed to generate the *XylT*/*FucT*-KO cell line (Mercx *et al.*, 2017). *A. tumefaciens* was transformed with the resulting plasmid. *Agrobacterium*-mediated transformation of BY-2 cells was carried out and eighteen calli were isolated. They were cultured in liquid MS medium and the culture medium (CM) was screened for the depletion of β 1,2-xylose and α 1,3-fucose by Western blotting (Figure 3A). Four lines (#7, 14, 17, 22) exhibited a clear signal reduction. Line *XylT*/*GMD*-KO #22 was totally devoid of β 1,2-xylose and contained around five times less α 1,3-fucose than the WT line (Figure 3B).

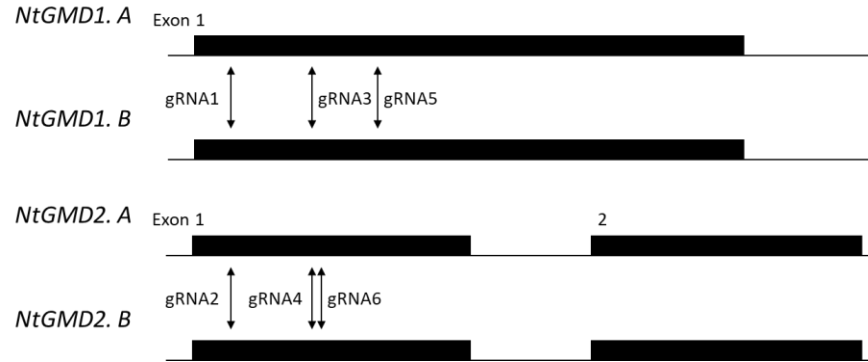


Figure 1. Schematic representations of *NtGMD* genomic sequences. Black rectangles: exons; vertical arrows: guide RNA target sites. *N. tabacum* is an allotetraploid species resulting from the cross between *N. tomentosiformis* and *N. sylvestris* (Yukawa *et al.*, 2006). The *NtGMD* orthologs derived from *N. tomentosiformis* and *N. sylvestris* will be referred to as *NtGMD.A* and *NtGMD.B*, respectively, nomenclature previously proposed for *NtSBT1* genes (Navarre *et al.*, 2012), and *NtFucT* or *NtXylT* genes (Mercx *et al.*, 2017). *NtGMD1.A*: NCBI Gene ID: 107832048, SolGenomics sequence ID: Ntab-TN90_AYMY-SS1096; *NtGMD1.B*: Gene ID: 107810228, Sequence ID: Ntab-TN90_AYMY-SS12399 *NtGMD2.A*: Gene ID: 107827017, Sequence ID: Ntab-TN90_AYMY-SS415 *NtGMD2.B*: Gene ID: 107776909, Sequence ID: Ntab-TN90_AYMY-SS1155



Figure 2. Schematic representation of the plant binary vector pPZP-*XylT/GMD-KO*. *nptII*, neomycin phosphotransferase II cassette; *codA*, cytosine deaminase A cassette; PTG, polycistronic tRNA-gRNA. *nptII* is under the control of the *nopaline synthase* promoter and *octopine synthase* terminator; *codA* and *mCherry* are under the control of the enhanced CaMV 35S promoter and *nopaline synthase* terminator; *cas9* is plant codon optimized, contains a potato IV2 intron and its expression is controlled by a hybrid 35SPPDK promoter; PTG-*XylT/GMD* is under the control of the U6 promoter and a poly-A terminator.

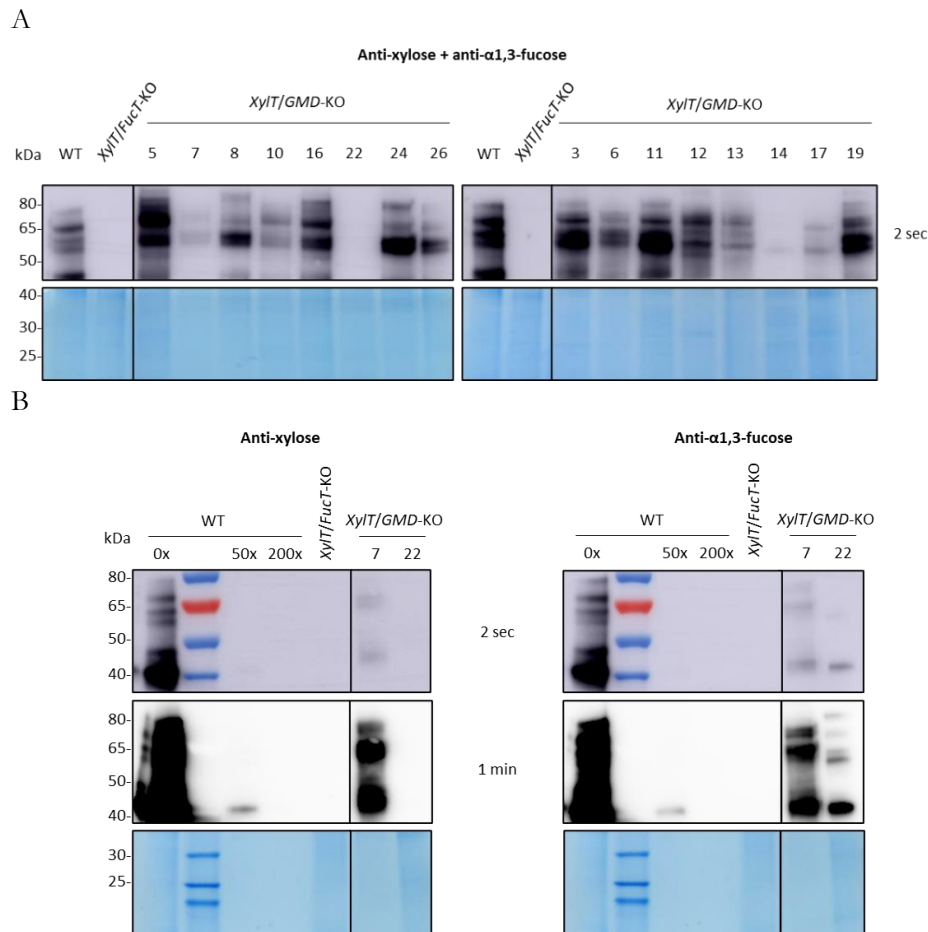


Figure 3. Reduction of protein-associated β 1,2-xylose and α 1,3-fucose residues in *XylIT/GMD-KO* BY-2 cell lines. (A) Secreted proteins (40 μ L CM) from WT, *XylIT/FucT-KO*, and from the indicated *XylIT/GMD-KO* cell lines were separated by gel electrophoresis and analyzed by Western blotting using a mix of anti- β 1,2-xylose and anti- α 1,3-fucose polyclonal antibodies (Agrisera), both obtained by affinity purification of serum from immunized rabbits. A colloidal blue gel is displayed as a loading control. (B) Anti- β 1,2-xylose and anti- α 1,3-fucose antibodies were used separately to characterize elite cell lines *XylIT/GMD-KO* #7 and #22. *XylIT/FucT-KO* and WT CM diluted 0, 50 and 200-fold were loaded as comparisons. Exposure times are indicated next to Western blotting signals, for comparison.

2. Discussion

To unambiguously conclude on the impact of knocking out *GMD* genes on fucose incorporation in BY-2 cell *N*-glycans, genomic analysis should be carried out to assess complete inactivation of all alleles. However, while our results show the strategy is valid to reduce non-human residues, this project was discontinued due to the substantial α 1,3-fucose content detected in the elite lines. The lower quality of *XyIT/GMD*-KO lines *N*-glycan content as compared to that of the already obtained *XyIT/FucT*-KO lines, combined to the time-consuming determining of the inactivation of all eight *GMD* alleles justified the abandonment of this strategy.

α 1,4-fucose content in *XyIT/GMD*-KO cell lines was not assessed. However, since α 1,4-fucose originates from the same pool of GDP-fucose as α 1,3-fucose, it is likely that they were not completely depleted either. Regarding potential allergic reactions to plant-produced pharmaceuticals, the main concerns are β 1,2-xylose and α 1,3-fucose. α 1,4-fucose is far less abundant than these residues in tobacco and hence its reduction only makes sense if it is associated with a strong or complete suppression of β 1,2-xylose and α 1,3-fucose. Obtaining of *XyIT/GMD*-KO lines from plant species containing high amounts of α 1,4-fucose, such as rice, might still be of interest.

However, obtaining plant lines completely devoid of non-human residues should rather be achieved by knocking out *XyIT*, *α 1,3-FucT* and *β 1,3-GalT* (+*HEXOs*). Indeed, this strategy should result in the complete suppression of β 1,2-xylose, α 1,3-fucose as well as Le^a-associated residues. Inactivation of *α 1,4-FucT* should not be necessary due to the already mentioned strict acceptor substrate specificity of α 1,4-FucT for Gal β 1-3GlcNAc (Parsons *et al.*, 2012; Strasser *et al.*, 2007). Moreover, this alternative strategy does not interfere with the GDP-fucose pool, the depletion of which could be detrimental for optimal growth. Indeed, GDP-fucose is incorporated in the cell wall components xyloglucans, rhamnogalacturonan II and arabinogalactans. The depletion of this pool triggers a dwarf phenotype in *A. thaliana* *GMD*-KO and *N. benthamiana* *GMD* knockdown plants (O'Neill *et al.*, 2001; Matsuo *et al.*, 2010).

References

- Altmann, F. (2007). The role of protein glycosylation in allergy. *Int. Arch. Allergy Immunol.* 142 99–115. doi: 10.1159/000096114
- Bae, S., Park, J., and Kim, J. S. (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30 1473–1475. doi: 10.1093/bioinformatics/btu048
- Bakker H., Bardor M., Molthoff J. W., Gomord V., Elbers I., Stevens L. H., Jordi W., Lommen A., Faye L., Lerouge P., Bosch D. (2001). Galactose-extended glycans of antibodies produced by transgenic plants. *Proc. Natl. Acad. Sci. USA* 98 2899–2904. doi: 10.1073/pnas.031419998
- Bakker, H., Rouwendal, G. J., Karnoup, A. S., Florack, D. E., Stoop, G. M., Helsper, J. P., Van Ree, R., Van Die, I. and Bosch, D. (2006). An antibody produced in tobacco expressing a hybrid beta-1,4-galactosyltransferase is essentially devoid of plant carbohydrate epitopes. *Proc. Natl. Acad. Sci. USA* 103 7577–7582. doi: 10.1073/pnas.0600879103
- Barampuram, S., Zhang, Z. J. (2011). Recent Advances in Plant Transformation. In: Birchler, J. (eds) Plant Chromosome Engineering. *Methods Mol. Biol.* 701 1-35. https://doi.org/10.1007/978-1-61737-957-4_1
- Baulcombe, D. (2004). RNA silencing in plants. *Nature* 431 356–363. <https://doi.org/10.1038/nature02874>
- Bohlender, L. L., Parsons, J., Hoernstein, S. N., Rempfer, C., Ruiz-Molina, N., Lorenz, T., Rodriguez-Jahnke, F., Figl, R., Fode, B., Altmann, F., Reski, R. and Decker, E. L. (2020). Stable protein sialylation in *Physcomitrella*. *Front. Plant Sci.* 11 610032. <https://doi.org/10.3389/fpls.2020.610032>
- Bohlender, L. L., Parsons, J., Hoernstein, S. N., Bangert, N., Rodríguez-Jahnke, F., Reski, R. and Decker, E. L. (2022). Unexpected arabinosylation after humanization of plant protein N-glycosylation. *Front. Bioeng. Biotechnol.* 10 838365. doi: 10.3389/fbioe.2022.838365
- Bosch, D., Castilho, A., Loos, A., Schots, A., Steinkellner, H. (2013). N-glycosylation of plant-produced recombinant proteins. *Curr Pharm Design* 19 5503-5512. doi: 10.2174/1381612811319310006
- Bosch, D., Castilho, A., Loos, A., Schots, A. and Steinkellner, H. (2013). N-glycosylation of plant-produced recombinant proteins. *Curr. Pharm. Des.* 19(31) 5503-5512. doi: 10.2174/1381612811319310006

- Brazil, J. C., Sumagin, R., Stowell, S. R., Lee, G., Louis, N. A., Cummings, R. D. and Parkos, C. A. (2017). Expression of Lewis-a glycans on polymorphonuclear leukocytes augments function by increasing transmigration. *J. Leukoc. Biol.* 102(3) 753-762. doi: 10.1189/jlb.1MA0117-013R
- Brooks, S. A. (2009). Strategies for analysis of the glycosylation of proteins: current status and future perspectives. *Mol. Biotechnol.* 43(1) 76. doi: 10.1007/s12033-009-9184-6
- Buck, P. M., Kumar, S. and Singh, S. K. (2013). Consequences of glycan truncation on Fc structural integrity. *MAbs.* 5(6) 904-16. doi: 10.4161/mabs.26453
- Burke, H. G. and Heldwein, E. E. (2015). Crystal structure of the human cytomegalovirus glycoprotein B. *PLOS Pathog.* 11(10) e1005227. doi: 10.1371/journal.ppat.1005227
- Buschmann, H., Green, P., Sambade, A., Doonan, J. H. and Lloyd, C. W. (2010). Cytoskeletal dynamics in interphase, mitosis and cytokinesis analysed through Agrobacterium-mediated transient transformation of tobacco BY-2 cells. *New Phytol.* 190(1) 258-267. doi: 10.1111/j.1469-8137.2010.03587.x
- Büssow, K. (2015). Stable mammalian producer cell lines for structural biology. *Curr. Opin. Struct. Biol.* 32 81-90. doi: 10.1016/j.sbi.2015.03.002.
- Byrne, G., O'Rourke, S. M., Alexander, D. L., Yu, B., Doran, R. C., Wright, M., Chen, Q., Azadi, P. and Berman, P. W. (2018). CRISPR/Cas9 gene editing for the creation of an MGAT1-deficient CHO cell line to control HIV-1 vaccine glycosylation. *PLOS Biol.* 16(8) e2005817. doi: 10.1371/journal.pbio.2005817
- Cai, Y., Chen, L., Sun, S., Wu, C., Yao, W., Jiang, B., Han, T. and Hou, W. (2018). CRISPR/Cas9-Mediated Deletion of Large Genomic Fragments in Soybean. *Int. J. Mol. Sci.* 19(12) 3835. doi: 10.3390/ijms19123835
- Calderon, A. D., Liu, Y., Li, X., Wang, X., Chen, X., Li, L., Wang, P. G. (2016). Substrate specificity of FUT8 and chemoenzymatic synthesis of core-fucosylated asymmetric N-glycans. *Org. Biomol. Chem.* 14 4027–4031. doi: 10.1039/c6ob00586a
- Castilho, A., Pabst, M., Leonard, R., Veit, C., Altmann, F., Mach, L., Glössl, J., Strasser, R. and Steinkellner, H. (2008). Construction of a functional CMP-sialic acid biosynthesis pathway in Arabidopsis. *Plant Physiol.* 147(1) 331-339. doi: 10.1104/pp.108.117572

- Castilho, A., Strasser, R., Stadlmann, J., Grass, J., Jez, J., Gattinger, P., Kunert, R., Quendler, H., Pabst, M., Leonard, R., Altmann, F. and Steinkellner H. (2010). In planta protein sialylation through overexpression of the respective mammalian pathway. *J. Biol. Chem.* 285(21) 15923–15930. doi: 10.1074/jbc.M109.088401
- Castilho, A., Gattinger, P., Grass, J., Jez, J., Pabst, M., Altmann, F., Gorfer, M., Strasser, R., Steinkellner, H. (2011). *N*-glycosylation engineering of plants for the biosynthesis of glycoproteins with bisected and branched complex *N*-glycans. *Glycobiology* 21 813–823. doi: 10.1093/glycob/cwr009
- Castilho, A., Neumann, L., Gattinger, P., Strasser, R., Vorauer-Uhl, K., Sterovsky, T., Altmann, F., Steinkellner, H. (2013). Generation of biologically active multi-sialylated recombinant human EPOFc in plants. *PLOS ONE* 8(1) e54836. doi: 10.1371/journal.pone.0054836
- Čaval, T., Heck, A. J. R. and Reiding, K. R. Meta-heterogeneity: Evaluating and Describing the Diversity in Glycosylation Between Sites on the Same Glycoprotein. *Mol. Cell Proteomics*. 20 100010. doi: 10.1074/mcp.R120.002093
- Cedeño, C., Pauwels, K. and Tompa, P. (2017) Protein delivery into plant cells: toward in vivo structural biology. *Front. Plant Sci.* 8 519. doi: 10.3389/fpls.2017.00519
- Chandramouli, S., Ciferri, C., Nikitin, P. A., Caló, S., Gerrein, R., Balabanis, K., Monroe, J., Hebner, C., Lilja, A. E., Settembre, E. C. and Carfi, A. (2015). Structure of HCMV glycoprotein B in the postfusion conformation bound to a neutralizing human antibody. *Nat. Commun.* 6 8176. doi: 10.1038/ncomms9176
- Chen, J. R., Yu, Y. H., Tseng, Y. C., Chiang, W. L., Chiang, M. F., Ko, Y. A., Chiu, Y. K., Ma, H. H., Wu, C. Y., Jan, J. T., Lin, K. I., Ma, C. and Wong, C. H. (2014a). Vaccination of monoglycosylated hemagglutinin induces cross-strain protection against influenza virus infections. *Proc. Natl. Acad. Sci. USA* 111(7) 2476–2481. doi: 10.1073/pnas.1323954111
- Chen, W. H., Du, L., Chag, S. M., Ma, C., Tricoche, N., Tao, X., Seid, C. A., Hudspeth, E. M., Lustigman, S., Tseng, C. T., Bottazzi, M. E., Hotez, P. J., Zhan, B., Jiang, S. (2014b). Yeast-expressed recombinant protein of the receptor-binding domain in SARS-CoV spike protein with deglycosylated forms as a SARS vaccine candidate. *Hum Vaccin Immunother.* 10(3) 648–58. doi: 10.4161/hv.27464
- Cheng, K., Zhou, Y. and Neelamegham, S. (2017). DrawGlycan-SNFG: a robust tool to render glycans and glycopeptides with fragmentation information. *Glycobiology* 27(3) 200–205. doi: 10.1093/glycob/cww115

- Chevalier, L., Bernard, S., Ramdani, Y., Lamour, R., Bardor, M., Lerouge, P., Follet-Gueye, M.L., Driouich, A. (2010). Subcompartment localization of the side chain xyloglucan-synthesizing enzymes within Golgi stacks of tobacco suspension-cultured cells. *Plant J.* 64(6) 977-89. doi: 10.1111/j.1365-313X.2010.04388.x
- Chong-Pérez, B. and Angenon, G. (2013). Strategies for generating marker-free transgenic plants. *Genet. Eng.* 1-32. doi: 10.5772/55573
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A. and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339 819–823. doi: 10.1126/science.1231143
- Cooper, C. A., Gasteiger, E. and Packer, N. H. (2001). GlycoMod—a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* 1 340–349. doi: 10.1002/1615-9861(200102)1:2<340::AID-PROT340>3.0.CO;2-B
- Cooper, C. A., Gasteiger, E., and Packer, N. H. (2003). Predicting glycan composition from experimental mass using glycomod. In *Handbook of Proteomic Methods*, ed. P. M. Conn (Totowa, NJ: Humana Press), 225–231.
- Cotsaftis, O., Sallaud, C. and Breitler, J. C. (2002). Transposon-mediated generation of marker free rice plants containing a Bt gene conferring insect resistance. *Mol. Breeding* 10(3) 165-180. doi: 10.1023/A:1020380305904
- Cox, K. M., Sterling, J. D., Regan, J. T., Gasdaska, J. R., Frantz, K. K., Peele, C. G., Black, A., Passmore, D., Moldovan-Loomis, C., Srinivasan, M., Cuisin, S., Cardarelli, P. M. and Dickey, L. F. (2006). Glycan optimization of a human monoclonal antibody in the aquatic plant *Lemna minor*. *Nat. Biotechnol.* 24(12) 1591-1597. doi: 10.1038/nbt1260
- Crispin, M., Harvey, D. J., Chang, V. T., Yu, C., Aricescu, A. R., Jones, E. Y., Davis, S. J., Dwek, R. A., Rudd, P. M. (2006). Inhibition of hybrid- and complex-type glycosylation reveals the presence of the GlcNAc transferase I-independent fucosylation pathway. *Glycobiology* 16 748–756. doi: 10.1093/glycob/cwj119
- Dahmen, A. C., Fergen, M. T., Laurini, C., Schmitz, B., Loke, I., Thaysen-Andersen, M. and Diestel, S. (2015). Paucimannosidic glycoepitopes are functionally involved in proliferation of neural progenitor cells in the subventricular zone. *Glycobiology* 25(8) 869-880. doi: 10.1093/glycob/cwv027

- Dang, Y., Jia, G., Choi, J., Ma, H., Anaya, E., Ye, C., Shankar, P., Wu, H. (2015). Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol.* 16 280. doi: 10.1186/s13059-015-0846-3
- Darbani, B., Eimanifar, A., Stewart Jr, C. N. and Camargo, W. N. (2007). Methods to produce marker-free transgenic plants. *Biotechnol. J.* 2(1) 83-90. doi: 10.1002/biot.200600182.
- Dehoux J. P., Hori S., Talpe S., Bazin H., Latinne D., Soares M. P., Gianello P. (2000) Specific depletion of preformed IgM natural antibodies by administration of anti-mu monoclonal antibody suppresses hyperacute rejection of pig to baboon renal xenografts. *Transplantation* 70 935–946. doi:10.1097/00007890-20000927
- Demirci, Y., Zhang, B. and Unver, T. (2018). CRISPR/Cas9: an RNA-guided highly precise synthetic tool for plant genome editing. *J. Cell. Physiol.* 233(3) 1844-1859. doi: 10.1002/jcp.25970
- De Muynck, B., Navarre, C., Nizet, Y., Stadlmann, J. and Boutry, M. (2009). Different subcellular localization and glycosylation for a functional antibody expressed in *Nicotiana tabacum* plants and suspension cells. *Transgenic Res.* 18(3) 467-482. doi: 10.1007/s11248-008-9240-1
- Dicker, M., Schoberer, J., Vavra, U. and Strasser, R. (2015). Subcellular targeting of proteins involved in modification of plant N- and O-glycosylation. In *Glyco-Engineering* (pp. 249-267). Humana Press, New York, NY.
- Dolan, A., Cunningham, C., Hector, R. D., Hassan-Walker, A. F., Lee, L., Addison, C., Dargan, D. J., McGeoch, D. J., Gatherer, D., Emery, V. C., Griffiths, P. D., Sinzger, C., McSharry, B. P., Wilkinson, G. W. and Davison, A. J. (2004). Genetic Content of Wild-Type Human Cytomegalovirus. *J. Gen. Virol.* 85 (5) 1301-1312. doi: 10.1099/vir.0.79888-0
- Doudna, J. A. and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science* 346(6213) 1258096. doi: 10.1126/science.1258096.
- Dubrovskaya, V., Guenaga, J., de Val, N., Wilson, R., Feng, Y., Movsesyan, A., Karlsson Hedestam, G. B., Ward, A. B. and Wyatt, R. T. (2017). Targeted N-glycan deletion at the receptor-binding site retains HIV Env NFL trimer integrity and accelerates the elicited antibody response. *PLOS Pathog.* 13(9) e1006614. doi: 10.1371/journal.ppat.1006614

- Duby, G., Oufattole, M. and Boutry, M. (2001). Hydrophobic residues within the predicted N-terminal amphiphilic α -helix of a plant mitochondrial targeting presequence play a major role in in vivo import. *Plant J.* 27(6) 539-549. doi: 10.1046/j.1365-313X.2001.01098.x
- Ebert, B., Rautengarten, C. and Heazlewood, J. L. (2017). GDP-L-Fucose transport in plants: the missing piece. *Channels* 11(1) 8-10. doi: 10.1080/19336950.2016.1222760
- Fanata, W. I. D., Lee, K. H., Son, B. H., Yoo, J. Y., Harmoko, R., Ko, K. S., Ramasamy, N. K., Kim, K. H., Oh, D. B., Jung, H. S., Kim, JY., Lee, S. Y. and Lee, K. O. (2013). N-glycan maturation is crucial for cytokinin-mediated development and cellulose synthesis in *Oryza sativa*. *Plant J.* 73(6) 966-979. doi: 10.1111/tpj.12087
- Fischer, R., Holland, T., Sack, M., Schillberg, S., Stoger, E., Twyman, R. M. and Buyel, J. F. (2018). Glyco-engineering of plant-based expression systems. *Adv. Biochem. Eng. Biotechnol.* 175 137-166. doi: 10.1007/10_2018_76
- Fitchette-Lainé, A. C., Gomord, V., Cabanes, M., Michalski, J. C., Saint Macary, M., Foucher, B., Cavellier, B., Hawes, C., Lerouge, P. and Faye, L. (1997). N-glycans harboring the Lewis x epitope are expressed at the surface of plant cells. *Plant J.* 12(6) 1411-1417. doi: 10.1046/j.1365-313x.1997.12061411.x
- Frank, J., Kaulfürst-Soboll, H., Rips, S., Koiwa, H., and Von Schaewen, A. (2008). Comparative analyses of Arabidopsis complex glycan1 mutants and genetic interaction with staurosporin and temperature sensitive3a. *Plant Physiol.* 148 1354–1367. doi: 10.1104/pp.108.127027
- Kaulfürst-Soboll, H., Rips, S., Koiwa, H., Kajiura, H., Fujiyama, K. and von Schaewen, A. (2011). Reduced immunogenicity of Arabidopsis hybrid glycosylation1 (hgl1) N-glycans due to altered accessibility of xylose and core fucose epitopes. *J. Biol. Chem.* 286 22955-22964. doi: 10.1074/jbc.M110.196097
- Frank, M., Kaulfürst-Soboll, H., Fischer, K. and Von Schaewen, A. (2021). Complex-type N-glycans influence the root hair landscape of arabidopsis seedlings by altering the auxin output. *Front. Plant Sci.* 12 635714. doi: 10.3389/fpls.2021.635714
- Fromm, M., Taylor, L. P., Walbot, V. (1985). Expression of genes transferred into monocot and dicot plant cells by electroporation. *Proc. Natl. Acad. Sci. U S A.* 82(17) 5824-8. doi: 10.1073/pnas.82.17.5824

- Furuhata, Y., Sakai, A., Murakami, T., Morikawa, M., Nakamura, C., Yoshizumi, T., Fujikura, U., Nishida, K., Kato, Y. (2019). A method using electroporation for the protein delivery of Cre recombinase into cultured Arabidopsis cells with an intact cell wall. *Sci. Rep.* 9 (1) 1-10. doi: 10.1038/s41598-018-38119-9
- Fujiyama, K., Misaki, R., Katsura, A., Tanaka, T., Furukawa, A., Omasa, T. and Seki, T. (2006). N-linked glycan structures of a mouse monoclonal antibody produced from tobacco BY2 suspension-cultured cells. *J. Biosci. Bioeng.* 101(3) 212-218. doi: 10.1263/jbb.101.212
- Fujiyama, K., Furukawa, A., Katsura, A., Misaki, R., Omasa, T. and Seki, T. (2007). Production of mouse monoclonal antibody with galactose-extended sugar chain by suspension cultured tobacco BY2 cells expressing human β (1, 4)-galactosyltransferase. *Biochem. Biophys. Res. Commun.* 358(1) 85-91. doi: 10.1016/j.bbrc.2007.04.054
- Fujiyama, K., Misaki, R., Sakai, Y., Omasa, T., and Seki, T. (2009). Change in glycosylation pattern with extension of endoplasmic reticulum retention signal sequence of mouse antibody produced by suspension-cultured tobacco BY2 cells. *J. Biosci. Bioeng.* 107 165–172. doi: 10.1016/j.jbiosc.2008.09.016
- Gao, X., Zhou, J., Li, J., Zou, X., Zhao, J., Li, Q., Xia, R., Yang, R., Wang, D., Zuo, Z., Tu, J., Tao, Y., Chen, X., Xie, Q., Zhu, Z., Qu, S. (2015a). Efficient generation of marker-free transgenic rice plants using an improved transposon-mediated transgene reintegration strategy. *Plant Physiol.* 167(1) 11-24. doi: 10.1104/pp.114.246173
- Gao, J., Wang, G., Ma, S., Xie, X., Wu, X., Zhang, X., Wu, Y., Zhao, P. and Xia, Q. (2015b). CRISPR/Cas9-mediated targeted mutagenesis in *Nicotiana tabacum*. *Plant Mol. Biol.* 87(1-2) 99-110. doi: 10.1007/s11103-014-0263-0
- Gao, Z., Herrera-Carrillo, E. and Berkhout, B. (2018). RNA polymerase II activity of type 3 Pol III promoters. *Mol. Ther. Nucleic Acids* 12 135-145. doi: 10.1016/j.omtn.2018.05.001
- Gentzel, I. N., Ohlson, E. W., Redinbaugh, M. G. and Wang, G.-L. (2022). VIGE: virus-induced genome editing for improving abiotic and biotic stress traits in plants. *Stress Biology* 2, 2. <https://doi.org/10.1007/s44154-021-00026-x>
- Gilbertson, L. (2003). Cre-lox recombination: Cre-active tools for plant biotechnology. *Trends Biotechnol.* 21(12) 550-555. doi: 10.1016/j.tibtech.2003.09.011

- Gleave, A. P., Mitra, D.S., Mudge, S. R., Morris, B.A. (1999). Selectable marker-free transgenic plants without sexual crossing: transient expression of cre recombinase and use of a conditional lethal dominant gene. *Plant Mol. Biol.* 40(2) 223-35. doi: 10.1023/a:1006184221051
- Goderis, I. J. W. M., De Bolle, M. F. C., Francois, I. E. J. A., Wouters, P. F. J., Broekaert, W. F. and Cammue, B. P. A. (2002). A set of modular plant transformation vectors allowing flexible insertion of up to six expression units. *Plant Mol. Biol.* 50 17–27. doi: 10.1023/a:1016052416053
- Gomord, V., Fitchette, A. C., Menu-Bouaouiche, L., Saint-Jore-Dupas, C., Plasson, C., Michaud, D., Faye, L. (2010). Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnol. J.* 8, 564–587. doi: 10.1111/j.1467-7652.2009.00497.x
- Gopalappa, R., Suresh, B., Ramakrishna, S. and Kim, H. (2018). Paired D10A Cas9 nickases are sometimes more efficient than individual nucleases for gene disruption. *Nucleic Acids Res.* 46(12) e71. doi: 10.1093/nar/gky222
- Göritzer, K., Grandits, M., Gruber-Grünwald, C., Figl, R., Mercx, S., Navarre, C., Ma, J. K., The, A.Y. (2022). Engineering the N-glycosylation pathway of *Nicotiana tabacum* for molecular pharming using CRISPR/Cas9. *Front. Plant Sci.* 13 1003065. doi: 10.3389/fpls.2022.1003065.
- Gottlieb, C., Skinner, A.M. and Kornfeld, S. (1974). Isolation of a clone of Chinese hamster ovary cells deficient in plant lectin-binding sites. *Proc. Natl. Acad. Sci. USA* 71 1078–1083. doi: 10.1073/pnas.71.4.1078
- Grubb, J. H., Vogler, C. and Sly, W. S. (2010). New strategies for enzyme replacement therapy for lysosomal storage diseases. *Rejuvenation Res.* 13 229–236. doi: 10.1089/rej.2009.0920
- Hanania, U., Ariel, T., Tekoah, Y., Fux, L., Sheva, M., Gubbay, Y., Weiss, M., Oz, D., Azulay, Y., Turbovski, A., Forster, Y. and Shaaltiel, Y. (2017). Establishment of a tobacco BY2 cell line devoid of plant-specific xylose and fucose as a platform for the production of biotherapeutic proteins. *Plant Biotechnol. J.* 15(9) 1120-1129. doi: 10.1111/pbi.12702
- Hao, H., Wang, X., Jia, H., Yu, M., Zhang, X., Tang, H. and Zhang, L. (2016). Large fragment deletion using a CRISPR/Cas9 system in *Saccharomyces cerevisiae*. *Anal. Biochem.* 509 118–123. doi: 10.1016/j.ab.2016.07.008

- Hare, P. D. and Chua, N. H. (2002). Excision of selectable marker genes from transgenic plants. *Nat. Biotechnol.* 20(6) 575-580. doi: 10.1038/nbt0602-575
- Harmoko, R., Yoo, J. Y., Ko, K. S., Ramasamy, N. K., Hwang, B. Y., Lee, E. J., Kim, H. S., Lee, K. J., Oh, D. B., Kim, D. Y., Lee, S., Li, Y., Lee, S. Y., Lee, K. O. (2016). N-glycan containing a core α 1, 3-fucose residue is required for basipetal auxin transport and gravitropic response in rice (*Oryza sativa*). *New Phytol.* 212(1) 108-122. doi: 10.1111/nph.14031
- He, X., Galpin, J. D., Tropak, M. B., Mahuran, D., Haselhorst, T., von Itzstein, M., Kolarich, D., Packer, N. H., Miao, Y., Jiang, L., Grabowski, G. A., Clarke, L. A., Kermode, A. R. (2012). Production of active human glucocerebrosidase in seeds of *Arabidopsis thaliana* complex-glycan-deficient (cgl) plants. *Glycobiology* 22 492–503. doi: 10.1093/glycob/cwr157
- He, X., Pierce, O., Haselhorst, T., von Itzstein, M., Kolarich, D., Packer, N. H., Gloster, T. M., Voadlo, D. J., Qian, Y., Brooks, D., Kermode, A. R. (2013). Characterization and downstream mannose phosphorylation of human recombinant α -L-iduronidase produced in *Arabidopsis* complex glycan-deficient (cgl) seeds. *Plant Biotechnol. J.* 11 1034-1043. doi: 10.1111/pbi.12096
- He, Z., Proudfoot, C., Mileham, A. J., McLaren, D. G., Whitelaw, C. B. A. and Lillico, S. G. (2015). Highly efficient targeted chromosome deletions using CRISPR/Cas9. *Biotechnol. Bioeng.* 112 1060–1064. doi: 10.1002/bit.25490
- Herman, X., Far, J., Courtoy, A., Bouhon, L., Quinton, L., De Pauw, E., Chaumont, F. and Navarre, C. (2021). Inactivation of *N-acetylglucosaminyltransferase I* and *α 1,3-fucosyltransferase* genes in *N. tabacum* BY-2 cells results in glycoproteins with highly homogeneous, high-mannose N-glycans. *Front. Plant Sci.* 12 634023. doi: 10.3389/fpls.2021.634023
- Hesselink, T., Rouwendal, G. J., Henquet, M. G., Florack, D. E., Helsper, J. P. and Bosch, D. (2014). Expression of natural human β 1, 4-GalT1 variants and of non-mammalian homologues in plants leads to differences in galactosylation of N-glycans. *Transgenic Res.* 23(5) 717-728. doi: 10.1007/s11248-014-9806-z
- Hintze, S., Limmer, S., Dabrowska-Schlepp, P., Berg, B., Krieghoff, N., Busch, A., Schaaf, A., Meinke, P. and Schoser, B. (2020). Moss-derived human recombinant GAA provides an optimized enzyme uptake in differentiated human muscle cells of pompe disease. *Int. J. Mol. Sci.* 21(7) 2642. doi: 10.3390/ijms21072642

- Huang, H. Y., Liao, H. Y., Chen, X., Wang, S. W., Cheng, C. W., Shahed-Al-Mahmud, M., Liu, Y. M., Mohapatra, A., Chen, T. H., Lo, J. M., Wu, Y. M., Ma, H. H., Chang, Y. H., Tsai, H. Y., Chou, Y. C., Hsueh, Y. P., Tsai, C. Y., Huang, P. Y., Chang, S. Y., Chao, T. L., Kao, H. C., Tsai, Y. M., Chen, Y. H., Wu, C. Y., Jan, J. T., Cheng, T. R., Lin, K. I., Ma, C. and Wong, C. H. (2022). Vaccination with SARS-CoV-2 spike protein lacking glycan shields elicits enhanced protective responses in animal models. *Sci. Transl. Med.* 14(639) eabm0899. doi: 10.1126/scitranslmed.abm0899
- Hundleby, P. A., D'Aoust, M. A., Finkle, C., Atkins, J. and Twyman, R. M. (2022). Regulation of molecular farming products. *Methods Mol. Biol.* 2480 313-333. doi: 10.1007/978-1-0716-2241-4_17
- Hwang, H. H., Yu, M. and Lai, E. M. (2017). Agrobacterium-mediated plant transformation: biology and applications. *Arabidopsis Book* 15 e0186. doi: 10.1199/tab.0186
- Ioffe, E. and Stanley P. (1994). Mice lacking N-acetylglucosaminyltransferase I activity die at mid-gestation, revealing an essential role for complex or hybrid N-linked carbohydrates. *Proc. Natl. Acad. Sci. USA* 91 728–732. doi: 10.1073/pnas.91.2.728
- Ireton, G. C., Black, M. E. and Stoddard, B. L. (2003). The 1.14 Å crystal structure of yeast cytosine deaminase: evolution of nucleotide salvage enzymes and implications for genetic chemotherapy. *Structure* 11(8) 961-972. doi: 10.1016/s0969-2126(03)00153-9
- Ito, Y., Uemura, T. and Nakano, A. (2014). Formation and maintenance of the Golgi apparatus in plant cells. *Int. Rev. Cell. Mol. Biol.* 310 221-287. doi: 10.1016/B978-0-12-800180-6.00006-2
- Jansing, J., Sack, M., Augustine, S. M., Fischer, R. and Bortesi, L. (2019). CRISPR/Cas9-mediated knockout of six glycosyltransferase genes in *Nicotiana benthamiana* for the production of recombinant proteins lacking β -1, 2-xylose and core α -1, 3-fucose. *Plant Biotechnol. J.* 17(2) 350-361. doi: 10.1111/pbi.12981
- Janssen, B. J., Gardner, R. C. (1990) Localized transient expression of GUS in leaf discs following cocultivation with *Agrobacterium*. *Plant Mol. Biol.* 14 61-72. doi: 10.1007/BF00015655

- Jean Beltran, P. M. and Cristea, I. M. (2014). The life cycle and pathogenesis of human cytomegalovirus infection: lessons from proteomics. *Expert Rev. Proteomics* 11(6) 697-711. doi: 10.1586/14789450.2014.971116
- Jiao, Q. S., Niu, G. T., Wang, F. F., Dong, J. Y., Chen, T. S., Zhou, C. F. and Hong, Z. (2020). N-glycosylation regulates photosynthetic efficiency of *Arabidopsis thaliana*. *Photosynthetica* 58 72–79. doi: 10.32615/ps.2019.153
- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096) 816-821. doi: 10.1126/science.1225829
- Ju, M. S. and Jung, S. T. (2014). Aglycosylated full-length IgG antibodies: steps toward next-generation immunotherapeutics. *Curr. Opin. Biotechnol.* 30 128-139. doi: 10.1016/j.copbio.2014.06.013
- Jung, J.-W., Huy, N.-X., Kim, H.-B., Kim, N.-S., Van Giap, D. and Yang, M.-S. (2017). Production of recombinant human acid α -glucosidase with high-mannose glycans in gnt1 rice for the treatment of Pompe disease. *J. Biotechnol.* 249 42–50. doi: 10.1016/j.jbiotec.2017.03.033
- Jung, J.-W., Choi, H.-Y., Huy, N.-X., Park, H., Kim, H. H., Yang, M.-S., Kang, S. H., Kim, D. I., Kim, N.S. (2019). Production of recombinant human acid β -glucosidase with high mannose-type N-glycans in rice gnt1 mutant for potential treatment of Gaucher disease. *Protein Expr. Purif.* 158 81–88. doi: 10.1016/j.pep.2019.02.014
- Jung, J. W., Shin, J. H., Lee, W. K., Begum, H., Min, C. H., Jang, M. H., Oh, H. B., Yang, M. S. and Kim, S. R. (2021). Inactivation of the β (1, 2)-xylosyltransferase and the α (1, 3)-fucosyltransferase gene in rice (*Oryza sativa*) by multiplex CRISPR/Cas9 strategy. *Plant Cell Rep.* 40(6) 1025-1035. doi: 10.1007/s00299-021-02667-8
- Kajiura, H., Misaki, R., Fujiyama, K. and Seki, T. (2011). Stable coexpression of two human sialylation enzymes in plant suspension-cultured tobacco cells. *J. Biosci. Bioeng.* 111(4), 471-477. doi: 10.1016/j.jbiosc.2010.11.018
- Kallolimath, S., Castilho, A., Strasser, R., Grünwald-Gruber, C., Altmann, F., Strubl, S., Galuska, C. E., Zlatina, K., Galuska, S. P., Werner, S., Thiesler, H., Werneburg, S., Hildebrandt, H., Gerardy-Schahn, R., Steinkellner H. (2016). Engineering of complex protein sialylation in plants. *Proc. Natl. Acad. Sci. USA* 113(34) 9498-9503. doi: 10.1073/pnas.1604371113

- Kamisugi, Y., Schlink, K., Rensing, S. A., Schween, G., von Stackelberg, M., Cumming, A. C., Reski, R. and Cove, D. J. (2006). The mechanism of gene targeting in *Physcomitrella patens*: homologous recombination, concatenation and multiple integration. *Nucleic Acids Res.*, 34(21), 6205-6214. doi: 10.1093/nar/gkl832
- Kang, J. S., Frank, J., Kang, C. H. Kajiura, H., Vikram, M., Ueda, A., Kim, S., Bahk, J. D., Triplett, B., Fujiyama, K., Lee, S. Y., von Schaewen, A., Koiwa, H. (2008) Salt tolerance of *Arabidopsis thaliana* requires maturation of N-glycosylated proteins in the Golgi apparatus. *Proc. Natl. Acad. Sci. USA* 105, 5933–5938. doi: 10.1073/pnas.0800237105
- Kang, C. E., Lee, S., Ahn, T., Seo, D. H., Ko, B. J., Jung, M., Lee, J., Kim, J. Y., Kim, W. T. (2022). Antibody-dependent cellular cytotoxicity-null effector developed using mammalian and plant GlycoDelete platform. *Sci. Rep.* 12(1) 19030. doi: 10.1038/s41598-022-23311-9
- Karg, S. R., Frey, A. D. and Kallio, P. T. (2010). Reduction of N-linked xylose and fucose by expression of rat β 1, 4-N-acetylglucosaminyltransferase III in tobacco BY-2 cells depends on Golgi enzyme localization domain and genetic elements used for expression. *J. Biotechnol.*, 146(1-2), 54-65. doi: 10.1016/j.jbiotec.2010.01.005
- Kasturi, L., Eshleman, J. R., Wunner, W. H. and Shakin-Eshleman, S. H. (1995). The Hydroxy Amino Acid in an Asn-X-Ser/Thr Sequon Can Influence N-Linked Core Glycosylation Efficiency and the Level of Expression of a Cell Surface Glycoprotein*. *J. Biol. Chem.* 270(24) 14756-14761. doi: 10.1074/jbc.270.24.14756
- Kaulfürst-Soboll, H., Mertens, M., Brehler, R., and von Schaewen, A. (2011). Reduction of Cross-reactive Carbohydrate Determinants in plant foodstuff: elucidation of clinical relevance and implications for allergy diagnosis. *PLOS ONE* 6 e17800. doi: 10.1371/journal.pone.0017800
- Kaulfürst-Soboll, H., Mertens-Beer, M., Brehler, R., Albert, M. and von Schaewen, A. (2021). Complex N-glycans are important for normal fruit ripening and seed development in tomato. *Front. Plant Sci.*, 12, 635962. doi: 10.3389/fpls.2021.635962
- Kim, H., Kim, S. T., Ryu, J., Kang, B. C., Kim, J. S., Kim, S. G. (2017). CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nat. Commun.* 8 14406. doi: 10.1038/ncomms14406
- Kittur, F. S., Hung, C. Y., Zhu, C., Shajahan, A., Azadi, P., Thomas, M. D., Pearce, J. L., Gruber, C., Kallolimath, S. and Xie, J. (2020). Glycoengineering tobacco plants

to stably express recombinant human erythropoietin with different *N*-glycan profiles. *Int. J. Biol. Macromol.* 157 158-169. doi: 10.1016/j.ijbiomac.2020.04.199

Kiyoshi, M., Tsumoto, K., Ishii-Watabe, A. and Caaveiro, J. M. (2017). Glycosylation of IgG-Fc: a molecular perspective. *Int. Immunol.* 29(7) 311-317. doi: 10.1093/intimm/dxx038

Kizuka, Y. and Taniguchi, N. (2016). Enzymes for *N*-Glycan Branching and Their Genetic and Nongenetic Regulation in Cancer. *Biomolecules.* 6(2) 25. doi: org/10.3390/biom6020025

Koncz, C. and Schell, J. (1986). The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.* 204(3) 383-396. doi: 10.1007/BF00331014

König-Beihammer, J., Vavra, U., Shin, Y. J., Veit, C., Grünwald-Gruber, C., Gillitschka, Y., Huber, J., Hofner, M., Vierlinger, K., Mitteregger, D., Weinhäusel, A. and Strasser, R. (2021). In planta production of the receptor-binding domain from SARS-CoV-2 with human blood group A glycan structures. *Front. Chem.* 9 816544. doi: 10.3389/fchem.2021.816544

Koprivova, A., Altmann, F., Gorr, G., Kopriva, S., Reski, R. and Decker, E. L. (2003). *N*-glycosylation in the moss *Physcomitrella patens* is organized similarly to that in higher plants. *Plant Biol.* 5(06) 582-591. doi:10.1055/s-2003-44721

Koprivova, A., Stemmer, C., Altmann, F., Hoffmann, A., Kopriva, S., Gorr, G., Reski, R., Decker, E. L. (2004). Targeted knockouts of *Physcomitrella* lacking plant-specific immunogenic *N*-glycans. *Plant Biotechnol. J.* 2(6) 517-523. doi: 10.1111/j.1467-7652.2004.00100.x.

Kovarik, A., Lim, K. Y., Soucková-Skalická, K., Matyasek, R. and Leitch, A. R. (2012). A plant culture (BY-2) widely used in molecular and cell studies is genetically unstable and highly heterogeneous. *Bot. J. Linn.* 170(3) 459-471. doi: 10.1111/j.1095-8339.2012.01280.x

Kozak, S., Bloch, Y., De Munck, S., Mikula, A., Bento, I., Savvides, S. N. and Meijers, R. (2020). Homogeneously *N*-glycosylated proteins derived from the GlycoDelete HEK293 cell line enable diffraction-quality crystallogenes. *Acta Crystallogr. D Struct. Biol.* 76(12) 1244-1255. doi: 10.1107/S2059798320013753

Kraft, K., Geuer, S., Will, A.J., Chan, W.L., Paliou, C., Borschiwer, M., Harabula, I., Wittler, L., Franke, M., Ibrahim, D. M., Kragesteen, B. K., Spielmann, M., Mundlos,

- S., Lupiáñez, D. G., Andrey, G. (2015). Deletions, inversions, duplications: Engineering of structural variants using CRISPR/Cas in mice. *Cell Rep.* 10 833–839. doi: 10.1016/j.celrep.2015.01.016
- Kriechbaum, R., Ziaee, E., Grunwald-Gruber, C., Buscaill, P., van der Hoorn, R.A.L. and Castilho, A. (2020). BGAL1 depletion boosts the level of beta-galactosylation of N- and O-glycans in *N. benthamiana*. *Plant Biotechnol. J.* 18 1537–1549. doi: 10.1111/pbi.13316
- Lacroix, B., Citovsky, V. (2020). Biolistic approach for transient gene expression studies in plants. *Methods Mol. Biol.* 2124 125-139. doi: 10.1007/978-1-0716-0356-7_6
- Lannoo, N. and Van Damme, E. J. (2015). Review/*N*-glycans: The making of a varied toolbox. *Plant Sci.* 239 67-83. doi: 10.1016/j.plantsci.2015.06.023
- Larsen, J. S., Karlsson, R. T. G., Tian, W., Schulz, M. A., Matthes, A., Clausen, H., Petersen, B.L., Yang, Z. (2020). Engineering mammalian cells to produce plant-specific N-glycosylation on proteins. *Glycobiology.* 30(8) 528-538. doi: 10.1093/glycob/cwaa009
- Lee, N. E., Kim, S. H., Yu, D. Y., Woo, E. J., Kim, M. I., Seong, G. S., Lee, S. M., Ko, J. H. and Kim, Y. S. (2020). Aglycosylated antibody-producing mice for aglycosylated antibody-lectin coupled immunoassay for the quantification of tumor markers (ALIQUT). *Commun. Biol.* 3(1) 636. doi: 10.1038/s42003-020-01363-9
- Lhernould, S., Labrousse, P., Lejolly, D., Léonard, R., Carlué, M. and Costa, G. (2008). In vivo activity of recombinant human lewis fucosyltransferase III in leaves of *Nicotiana tabacum* L. *Biol. Plant.* 52(2) 267-274. doi: 10.1007/s10535-008-0057-9
- Li, Y., Park, A. I., Mou, H., Colpan, C., Bizhanova, A., Akama-Garren, E., Joshi, N., Hendrickson, E. A., Feldser, D., Yin, H., Anderson, D. G., Jacks, T., Weng, Z. and Xue, W. (2015) A versatile reporter system for CRISPR-mediated chromosomal rearrangements. *Genome Biol.* 16(1) 111. doi: 10.1186/s13059-015-0680-7
- Li, R., Char, S. N. and Yang, B. (2019). Creating Large Chromosomal Deletions in Rice Using CRISPR/Cas9. *Methods Mol. Biol.* 1917 47-61. doi: 10.1007/978-1-4939-8991-1_4
- Li, Q., Higuchi, Y., Tanabe, K., Katakura, Y. and Takegawa, K. (2020). Secretory production of N-glycan-deleted glycoprotein in *Aspergillus oryzae*. *J. Biosci. Bioeng.* 129(5) 573-580. doi: 10.1016/j.jbiosc.2019.12.006

- Liebminger, E., Veit, C., Pabst, M., Batoux, M., Zipfel, C., Altmann, F., Mach, L. and Strasser, R. (2011). β -N-Acetylhexosaminidases HEXO1 and HEXO3 are responsible for the formation of paucimannosidic N-glycans in *Arabidopsis thaliana*. *J. Biol. Chem.* 286(12) 10793-10802. doi: 10.1074/jbc.M110.178020
- Limkul, J., Misaki, R., Kato, K., and Fujiyama, K. (2015). The combination of plant translational enhancers and terminator increase the expression of human glucocerebrosidase in *Nicotiana benthamiana* plants. *Plant Sci.* 240, 41–49. doi: 10.1016/j.plantsci.2015.08.018
- Limkul, J., Iizuka, S., Sato, Y., Misaki, R., Ohashi, T., Ohashi, T. and Fujiyama, K. (2016). The production of human glucocerebrosidase in glyco-engineered *Nicotiana benthamiana* plants. *Plant Biotechnol. J.* 14(8) 1682-1694. doi: 10.1111/pbi.12529
- Lin, A. I., Philipsberg, G. A., and Haltiwanger, R. S. (1994). Core fucosylation of high-mannose-type oligosaccharides in GlcNAc transferase I-deficient (Lec1) CHO cells. *Glycobiology* 4, 895–901. doi: 10.1093/glycob/4.6.895
- Liu, C. P., Tsai, T. I., Cheng, T., Shivatare, V. S., Wu, C. Y., Wong, C. H. (2018). Glycoengineering of antibody (Herceptin) through yeast expression and in vitro enzymatic glycosylation. *Proc. Natl. Acad. Sci. U S A.* 115(4):720-725. doi: 10.1073/pnas.1718172115
- Liu, Y., Heim, K. P., Che, Y., Chi, X., Qiu, X., Han, S., Dormitzer, P. R., Yang, X. (2021). Prefusion structure of human cytomegalovirus glycoprotein B and structural basis for membrane fusion. *Sci. Adv.* 7(10) eabf3178. doi: 10.1126/sciadv.abf3178
- Locatelli, F., Vannini, C., Magnani, E., Coraggio, I., Bracale, M. (2003). Efficiency of transient transformation in tobacco protoplasts is independent of plasmid amount. *Plant Cell Rep.* 21(9) 865-71. doi: 10.1007/s00299-003-0593-x.
- Lockhart, J. (2019). Introducing CRISPR-TSKO: A Breakthrough in Precision Gene Editing. *Plant Cell.* 31(12) 2831-2832. doi: 10.1105/tpc.19.00753
- Loos, A. and Steinkellner, H. (2014). Plant glyco-biotechnology on the way to synthetic biology. *Front. Plant Sci.*, 5, 523. doi: 10.3389/fpls.2014.00523
- Lv, S.-Q., Zhang, K. B., Zhang, E. E., Gao, F. Y., Yin, C. L., Huang, C. J., He, J. Q. and Yang, H. (2009). Antitumor efficiency of the cytosine deaminase/5-fluorocytosine suicide gene therapy system on malignant gliomas: an in vivo study. *Med Sci Monit.* 15(1) 13-20. PMID: 19114960

Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., Lajoie, G. (2003). PEAKS: Powerful Software for Peptide De Novo Sequencing by MS/MS. *Rapid Commun. Mass Spectrom.* 17(20) 2337-2342. doi: 10.1002/rcm.1196

Magy, B. (2014). Accumulation of secreted antibodies in plant cell cultures varies according to the isotype, host species, and culture conditions, PhD thesis, UCLouvain, Louvain-la-Neuve.

Magy, B., Tollet, J., Laterre, R., Boutry, M. and Navarre, C. (2014). Accumulation of secreted antibodies in plant cell cultures varies according to the isotype, host species and culture conditions. *Plant Biotechnol. J.* 12(4) 457-467. doi: 10.1111/pbi.12152

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339 823–826. doi: 10.1126/science.1232033

Mamedov, T., Ghosh, A., Jones, R. M., Mett, V., Farrance, C. E., Musiyuchuk, K., Horsey, A. and Yusibov, V. (2012). Production of non-glycosylated recombinant proteins in *Nicotiana benthamiana* plants by co-expressing bacterial PNGase F. *Plant Biotechnol. J.*, 10(7), 773-782. doi: 10.1111/j.1467-7652.2012.00694.x

Mamedov, T., Cicek, K., Gulec, B., Ungor, R. and Hasanova, G. (2017). In vivo production of non-glycosylated recombinant proteins in *Nicotiana benthamiana* plants by co-expression with Endo- β -N-acetylglucosaminidase H (Endo H) of *Streptomyces plicatus*. *PLOS ONE*, 12(8), e0183589. doi: 10.1371/journal.pone.0183589

Mamedov, T., Cicek, K., Miura, K., Gulec, B., Akinci, E., Mammadova, G. and Hasanova, G. (2019). A Plant-Produced *in vivo* deglycosylated full-length Pfs48/45 as a Transmission-Blocking Vaccine Candidate against malaria. *Sci. Rep.* 9 9868. doi: 10.1038/s41598-019-46375-6

Mamedov, T., Yuksel, D., Ilgin, M., Gurbuzaslan, I., Gulec, B., Yetiskin, H., Uygut, M. A., Islam Pavel, S. T., Ozdarendeli, A., Mammadova, G., Say, D., Hasanova, G. (2021). Plant-produced glycosylated and In vivo Deglycosylated receptor binding domain proteins of SARS-CoV-2 induce potent neutralizing responses in mice. *Viruses* 13(8) 1595. doi: 10.3390/v13081595

Margolin, E., Oh, Y. J., Verbeek, M., Naude, J., Ponndorf, D., Meshcheriakova, Y. A., Peyret, H., van Diepen, M. T., Chapman, R., Meyers, A. E., Lomonossoff, G. P., Matoba, N., Williamson, A. L., Rybicki, E. P. (2020). Co-expression of human

calreticulin significantly improves the production of HIV gp140 and other viral glycoproteins in plants. *Plant Biotechnol. J.* 18(10) 2109–17. doi: 10.1111/pbi.13369

Marintcheva, B. (2018) Viral Tools for Genome Manipulations In Vivo. In *Harnessing the Power of Viruses*, Academic Press: Cambridge, MA, USA. ISBN 978-0-12-810514-6.

Marsico, C. and Kimberlin, D. W. (2017). Congenital Cytomegalovirus infection: advances and challenges in diagnosis, prevention and treatment. *Ital. J. Pediatr.* 43(1) 38. doi: 10.1186/s13052-017-0358-8

Martin-Ortigosa, S., and Wang, K. (2014). Proteolistics: a biolistic method for intracellular delivery of proteins. *Transgenic Res.* 23(5) 743–756. doi: 10.1007/s11248-014-9807-y

Mathieu-Rivet, E., Mati-Baouche, N., Walet-Balieu, M. L., Lerouge, P. and Bardor, M. (2020). N- and O-Glycosylation pathways in the microalgae polyphyletic group. *Front. Plant Sci.*, 11, 609993. doi: 10.3389/fpls.2020.609993

Mathur, J. and Koncz, C. (1998). PEG-mediated protoplast transformation with naked DNA. *Methods Mol. Biol.* 82 267–76. doi: 10.1385/0-89603-391-0:267

Matsunaga, E., Sugita, K. and Ebinuma, H. (2002). Asexual production of selectable marker-free transgenic woody plants, vegetatively propagated species. *Mol. Breed.* 10 95–106. doi: 10.1023/A:1020308216747

Matsuo, K. and Matsumura, T. (2011). Deletion of fucose residues in plant N-glycans by repression of the GDP-mannose 4, 6-dehydratase gene using virus-induced gene silencing and RNA interference. *Plant Biotechnol. J.* 9(2) 264–281. doi: 10.1111/j.1467-7652.2010.00553.x

Matsuo, K., Kagaya, U., Itchoda, N., Tabayashi, N., Matsumura, T. Deletion of plant-specific sugar residues in plant N-glycans by repression of GDP-D-mannose 4,6-dehydratase and β -1,2-xylosyltransferase genes. *J. Biosci. Bioeng.* 118(4) 448–454. doi: 10.1016/j.jbiosc.2014.04.005

Mercx, S., Tollet, J., Magy, B., Navarre, C. and Boutry, M. (2016). Gene inactivation by CRISPR-Cas9 in *Nicotiana tabacum* BY-2 suspension cells. *Front. Plant Sci.* 7 40. doi: 10.3389/fpls.2016.00040

Mercx, S., Smargiasso, N., Chaumont, F., De Pauw, E., Boutry, M., Navarre, C. (2017). Inactivation of the β (1,2)-xylosyltransferase and the α (1,3)-fucosyltransferase

genes in *Nicotiana tabacum* BY-2 cells by a multiplex CRISPR/Cas9 strategy results in glycoproteins without plant-specific glycans. *Front. Plant Sci.* 8 403. doi: 10.3389/fpls.2017.00403

Meuris, L., Santens, F., Elson, G., Festjens, N., Boone, M., Dos Santos, A., Devos, S., Rousseau, F., Plets, E., Houthuys, E., Malinge, P., Magistrelli, G., Cons, L., Chatel, L., Devreese, B. and Callewaert, N. (2014). GlycoDelete engineering of mammalian cells simplifies *N*-glycosylation of recombinant proteins. *Nat. Biotechnol.* 32(5) 485-489. doi: 10.1038/nbt.2885

Metzler M., Gertz A., Sarkar M., Schachter H., Schrader J. W. and Marth J. D. (1994). Complex asparagine-linked oligosaccharides are required for morphogenic events during post-implantation development. *EMBO J.* 13 2056–2065. doi: 10.1002/j.1460-2075.1994.tb06480.x

Miki, B., and McHugh, S. (2004). Selectable marker genes in transgenic plants: applications, alternatives and biosafety. *J. Biotechnol.* 107(3) 193-232. doi: 10.1016/j.jbiotec.2003.10.011

Misaki, R., Kimura, Y., Palacpac, N. Q., Yoshida, S., Fujiyama, K. and Seki, T. (2003). Plant cultured cells expressing human β 1, 4-galactosyltransferase secrete glycoproteins with galactose-extended N-linked glycans. *Glycobiology* 13(3) 199-205. doi: 10.1093/glycob/cwg021

Misaki, R., Fujiyama, K. and Seki, T. (2006). Expression of human CMP-N-acetylneuraminic acid synthetase and CMP-sialic acid transporter in tobacco suspension-cultured cell. *Biochem. Biophys. Res. Commun.* 339(4) 1184-1189. doi: 10.1016/j.bbrc.2005.11.130

Miura, K., Kinouchi, M., Ishida, K., Fujibuchi, W., Naitoh, T., Ogawa, H., Ando, T., Yazaki, N., Watanabe, K., Haneda, S., Shibata, C., Sasaki, I. (2010). 5-FU metabolism in cancer and orally-administrable 5-FU drugs. *Cancers* 2(3) 1717-1730. doi: 10.3390/cancers2031717

Moggridge, J., Biggar, K., Dawson, N. and Storey, K. B. (2017). Sensitive detection of immunoglobulin G stability using in real-time isothermal differential scanning Fluorimetry: Determinants of protein stability for antibody-based therapeutics. *Technol. Cancer Res Treat.* 16(6) 997-1005. doi: 10.1177/1533034617714149

Montero-Morales, L. and Steinkellner, H. (2018). Advanced plant-based glycan engineering. *Front. Bioeng. Biotechnol.* 6 81. doi: 10.3389/fbioe.2018.00081

- Morsa, D., Baiwir, D., La Rocca, R., Zimmerman, T. A., Hanozin, E., Grifnée, E., Longuespée, R., Meuwis, M. A., Smargiasso, N., De Pauw, E. and Mazzucchelli, G. (2019). Multi-enzymatic limited digestion: the next-generation sequencing for proteomics?. *J. Proteome Res.* 18(6) 2501-2513. doi: 10.1021/acs.jproteome.9b00044
- Moustafa, K., Makhzoum, A. and Trémouillaux-Guiller, J. (2016). Molecular farming on rescue of pharma industry for next generations. *Crit. Rev. Biotechnol.* 36(5) 840-850. doi: 10.3109/07388551.2015.1049934
- Nagata, T., Nemoto, Y. and Hasezawa, S. (1992). Tobacco BY-2 cell line as the “HeLa” cell in the cell biology of higher plants. *Int. Rev. Cytol.* 132 1-30. doi: 10.1016/S0074-7696(08)62452-3
- Nagels, B., Van Damme, E. J., Pabst, M., Callewaert, N. and Weterings, K. (2011). Production of complex multiantennary N-glycans in *Nicotiana benthamiana* plants. *Plant Physiol.* 155(3) 1103-1112. doi: 10.1104/pp.110.168773
- Nagels, B., Van Damme, E. J., Callewaert, N., Weterings, K. (2012). Introduction of tri-antennary N-glycans in *Arabidopsis thaliana* plants. *Plant Sci.* 185 161-8. doi: 10.1016/j.plantsci.2011.10.002
- Nandakumar, K. S., Collin, M., Happonen, K. E., Croxford, A. M., Lundström, S. L., Zubarev, R. A., Rowley, M. J., Blom, A. M., & Holmdahl, R. (2013). Dominant suppression of inflammation by glycan-hydrolyzed IgG. *Proc. Natl. Acad. Sci. USA* 110(25) 10252-10257. doi: 10.1073/pnas.1301480110
- Navarre, C., Delannoy, M., Lefebvre, B., Nader, J., Vanham, D. and Boutry, M. (2006). Expression and secretion of recombinant outer-surface protein A from the Lyme disease agent, *Borrelia burgdorferi*, in *Nicotiana tabacum* suspension cells. *Transgenic Res.* 15(3) 325. doi: 10.1007/s11248-006-0002-7
- Navarre, C., Sallets, A., Gauthy, E., Maîtrejean, M., Magy, B., Nader, J., Pety de Thozée, C., Crouzet, J., Batoko, H. and Boutry, M. (2011). Isolation of heat shock-induced *Nicotiana tabacum* transcription promoters and their potential as a tool for plant research and biotechnology. *Transgenic Res.* 20(4) 799-810. doi: 10.1007/s11248-010-9459-5
- Navarre, C., De Muynck, B., Alves, G., Vertommen, D., Magy, B. and Boutry, M. (2012). Identification, gene cloning and expression of serine proteases in the extracellular medium of *Nicotiana tabacum* cells. *Plant Cell Rep.* 31(10) 1959-1968. doi: 10.1007/s00299-012-1308-y

- Navarre, C., Smargiasso, N., Duvivier, L., Nader, J., Far, J., De Pauw, E. and Boutry, M. (2017). N-glycosylation of an IgG antibody secreted by *Nicotiana tabacum* BY-2 cells can be modulated through co-expression of human β -1, 4-galactosyltransferase. *Transgenic Res.* 26(3) 375-384. doi: 10.1007/s11248-017-0013-6
- Navarre C. and Chaumont, F. (2022). Production of Recombinant Glycoproteins in *Nicotiana tabacum* BY-2 Suspension Cells. *Methods Mol. Biol.* 2480 81-88. doi: 10.1007/978-1-0716-2241-4_5
- Nunes-Düby, S. E., Kwon H. J., Tirumalai, R. S., Ellenberger T. and Landy A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.* 26(2) 391-406. doi: 10.1093/nar/26.2.391
- O'Neill, M. A., Eberhard, S., Albersheim, P. and Darvill, A. G. (2001). Requirement of borate cross-linking of cell wall rhamnogalacturonan II for Arabidopsis growth. *Science* 294(5543) 846-849. doi: 10.1126/science.1062319
- Ow, D. W. (1996). Recombinase-directed chromosome engineering in plants. *Curr. Opin. Biotechnol.* 7 181-186. doi: 10.1016/S0958-1669(96)80010-3
- Ow, D. W. (2016). The long road to recombinase-mediated plant transformation. *Plant Biotechnol. J.* 14(2) 441-447. doi: 10.1111/pbi.12472
- Paccalet, T., Bardor, M., Rihouey, C., Delmas, F., Chevalier, C., D'Aoust, M. A., Faye, L., Vézina, L., Gomord, V. and Lerouge, P. (2007). Engineering of a sialic acid synthesis pathway in transgenic plants by expression of bacterial Neu5Ac-synthesizing enzymes. *Plant Biotechnol. J.* 5(1) 16-25. doi: 10.1111/j.1467-7652.2006.00211.x
- Palacpac, N. Q., Yoshida, S., Sakai, H., Kimura, Y., Fujiyama, K., Yoshida, T. and Seki, T. (1999). Stable expression of human β 1, 4-galactosyltransferase in plant cells modifies N-linked glycosylation patterns. *Proc. Natl. Acad. Sci. USA*, 96(8), 4692-4697. doi: 10.1073/pnas.96.8.4692
- Pan, Y., Yang, J., Wu, J., Yang, L. and Fang, H. (2022). Current advances of *Pichia pastoris* as cell factories for production of recombinant proteins. *Front. Microbiol.* 13 1059777. doi: 10.3389/fmicb.2022.1059777
- Parshenkov, A. and Hennet, T. (2022). Glycosylation-Dependent Induction of Programmed Cell Death in Murine Adenocarcinoma Cells. *Front. Immunol.* 13 797759. doi: 10.3389/fimmu.2022.797759

- Parsons, J., Altmann, F., Arrenberg, C. K., Koprivova, A., Beike, A. K., Stemmer, C., Gorr, G., Reski, R. and Decker, E. L. (2012). Moss-based production of asialo-erythropoietin devoid of Lewis A and other plant-typical carbohydrate determinants. *Plant Biotechnol. J.* 10(7) 851-861. doi: 10.1111/j.1467-7652.2012.00704.x
- Pedersen, C. T., Loke, I., Lorentzen, A., Wolf, S., Kamble, M., Kristensen, S. K., Munch, D., Radutoiu, S., Spillner, E., Roepstorff, P., Thaysen-Andersen, M., Stougaard, J., Dam, S. (2017). *N*-glycan maturation mutants in *Lotus japonicus* for basic and applied glycoprotein research. *Plant J.* 91(3) 394-407. doi: 10.1111/tpj.13570
- Pierce, O. M., McNair, G. R., He, X., Kajiura, H., Fujiyama, K., and Kermode, A. R. (2017). *N*-glycan structures and downstream mannose-phosphorylation of plant recombinant human alpha-l-iduronidase: toward development of enzyme replacement therapy for mucopolysaccharidosis I. *Plant Mol. Biol.* 95 593-606. doi: 10.1007/s11103-017-0673-x
- Piron, R., Santens, F., De Paepe, A., Depicker, A. and Callewaert, N. (2015). Using GlycoDelete to produce proteins lacking plant-specific *N*-glycan modification in seeds. *Nat. Biotechnol.* 33(11) 1135-1137. doi: 10.1038/nbt.3359
- Pottier, M., Laterre, R., Van Wessel, A., Ramirez, A. M., Herman, X., Boutry, M. and Hachez, C. (2020). Identification of two new trichome-specific promoters of *Nicotiana tabacum*. *Planta* 251(3) 58. doi: 10.1007/s00425-020-03347-9
- Puthalakath, H., Burke, J. and Gleeson, P. A. (1996). Glycosylation defect in Lec1 Chinese hamster ovary mutant is due to a point mutation in *N*-acetylglucosaminyltransferase I gene. *J. Biol. Chem.* 271(44) 27818-27822. doi: 10.1074/jbc.271.44.27818
- Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P. and Lim, W. A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152(5) 1173-1183. doi: 10.1016/j.cell.2013.02.022
- Ramazi, S. and Zahiri, J. (2021). Posttranslational modifications in proteins: resources, tools and prediction methods. *Database-Oxford* baab012. doi: 10.1093/database/baab012
- Reeves, P. J., Callewaert, N., Contreras, R. and Khorana, H. G. (2002). Structure and function in rhodopsin: high-level expression of rhodopsin with restricted and homogeneous *N*-glycosylation by a tetracycline-inducible *N*-

acetylglucosaminyltransferase I-negative HEK293S stable mammalian cell line. *Proc. Natl. Acad. Sci. USA* 99(21) 13419-13424. doi: 10.1073/pnas.212519299

Rosenlöcher, J., Sandig, G., Kannicht, C., Blanchard, V., Reinke, S. O. and Hinderlich, S. (2016). Recombinant glycoproteins: The impact of cell lines and culture conditions on the generation of protein species. *J Proteomics* 134 85-92. doi: 10.1016/j.jprot.2015.08.011

Rouwendaal, G. J., Wührer, M., Florack, D. E., Koeleman, C. A., Deelder, A. M., Bakker, H., Stoopen, G. M., van Die, I., Helsper, J. P., Hokke, C. H. and Bosch, D. (2007). Efficient introduction of a bisecting GlcNAc residue in tobacco N-glycans by expression of the gene encoding human N-acetylglucosaminyltransferase III. *Glycobiology* 17 334–344. doi: 10.1093/glycob/cwl078

Santos, R. B., Abranches, R., Fischer, R., Sack, M., Holland, T. (2016). Putting the Spotlight Back on Plant Suspension Cultures. *Front. Plant Sci.* 7 297. doi: 10.3389/fpls.2016.00297

Sariyatun, R., Kajiura, H., Ohashi, T., Misaki, R. and Fujiyama, K. (2021). Production of human acid-alpha glucosidase with a paucimannose structure by glycoengineered Arabidopsis cell culture. *Front. Plant Sci.* 12 703020. doi: 10.3389/fpls.2021.703020

Schachter, H. (2010). Mgat1-dependent N-glycans are essential for the normal development of both vertebrate and invertebrate metazoans. *Semin. Cell Dev. Biol.* 21(6) 609-615. doi: 10.1016/j.semcdb.2010.02.010

Schlaman, H. R. and Hooykaas, P. J. (1997). Effectiveness of the bacterial gene *codA* encoding cytosine deaminase as a negative selectable marker in Agrobacterium-mediated plant transformation. *Plant J.* 11(6) 1377-1385. doi: 10.1266/jjg.70.409

Schleiss, M. R., Permar, S. R. and Plotkin, S. A. (2017). Progress toward development of a vaccine against congenital cytomegalovirus infection. *Clin. Vaccine Immunol.* 24(12) e00268-17. doi: 10.1128/CVI.00268-17

Schleiss, M. R. (2018). Recombinant cytomegalovirus glycoprotein B vaccine: Rethinking the immunological basis of protection. *Proc. Natl. Acad. Sci. USA* 115(24) 6110-6112. doi: 10.1073/pnas.1806420115

Schmidt, F. R. (2004). Recombinant expression systems in the pharmaceutical industry. *Appl. Microbiol. Biotechnol.* 65(4) 363-372. doi: 10.1007/s00253-004-1656-9

- Schoberer, J. and Strasser, R. (2011). Sub-compartmental organization of Golgi-resident N-glycan processing enzymes in plants. *Mol. Plant* 4(2) 220-228. doi: 10.1093/mp/ssq082
- Schoberer, J. and Strasser, R. (2018). Plant glyco-biotechnology. *Semin Cell Dev Biol*. 80 133-141. doi: 10.1016/j.semcdb.2017.07.005
- Schön, K., Lepenies, B. and Goyette-Desjardins, G. (2021). Impact of protein glycosylation on the design of viral vaccines. *Adv Biochem Eng Biotechnol*. 175 319-354. doi: 10.1007/10_2020_132
- Scutt, C. P., Zubko, E. and Meyer, P. (2002). Techniques for the removal of marker genes from transgenic plants. *Biochimie* 84(11) 1119-1126. doi: 10.1016/s0300-9084(02)00021-4
- Sekan, A. S., Isayenkov, S. V. and Blume, Y. B. (2015). Development of marker-free transformants by site-specific recombinases. *Cytol. Genet.* 49 397–407. doi: 10.3103/S0095452715060080
- Sharma, S., Wisner, T. W., Johnson, D. C., Heldwein, E. E. (2013). HCMV gB shares structural and functional properties with gB proteins from other herpesviruses. *Virology* 435 239-24. doi: 10.1016/j.virol.2012.09.024
- Shaul, O. (2017). How introns enhance gene expression. *Int. J. Biochem. Cell Biol.* 91 145-155. doi: 10.1016/j.biocel.2017.06.016
- Shen, J. S., Busch, A., Day, T. S., Meng, X. L., Yu, C. I., Dabrowska-Schlepp, P., Fode, B., Niederkrüger, H., Forni, S., Chen, S., Schiffmann, R., Frischmuth, T. and Schaaf, A. (2016). Mannose receptor-mediated delivery of moss-made α -galactosidase A efficiently corrects enzyme deficiency in Fabry mice. *J. Inherit. Metab. Dis.* 39(2) 293-303. doi: 10.1007/s10545-015-9886-9
- Sheva, M., Hanania, U., Ariel, T., Turbovski, A., Rathod, V. K. R., Oz, D., Tekoah, Y. and Shaaltiel, Y. (2020). Sequential genome editing and induced excision of the transgene in *N. tabacum* BY2 cells. *Front. Plant Sci.* 11 607174. doi: 10.3389/fpls.2020.607174
- Shin, Y. J., Chong, Y. J., Yang, M. S. and Kwon, T. H. (2011). Production of recombinant human granulocyte macrophage-colony stimulating factor in rice cell suspension culture with a human-like N-glycan structure. *Plant Biotechnol. J.* 9(9) 1109-1119. doi: 10.1111/j.1467-7652.2011.00636.x

- Shin, Y. J., Castilho, A., Dicker, M., Sádio, F., Vavra, U., Grünwald-Gruber, C., Kwon, T. H., Altmann, F., Steinkellner, H. and Strasser, R. (2017). Reduced paucimannosidic N-glycan formation by suppression of a specific β -hexosaminidase from *Nicotiana benthamiana*. *Plant Biotechnol. J.*, 15(2), 197-206. doi: 10.1111/pbi.12602
- Si, Z., Zhang, J., Shivakoti, S., Atanasov, I., Tao, C. L., Hui, W. H., Zhou, K., Yu, X., Li, W., Luo, M., Bi, G. Q. and Zhou, ZH (2018). Different functional states of fusion protein gB revealed on human cytomegalovirus by cryo electron tomography with Volta phase plate. *PLOS Pathog.* 14(12) e1007452. doi: 10.1371/journal.ppat.1007452
- Sierro, N., Battey, J. N., Ouadi, S., Bovet, L., Goepfert, S., Bakaher, N., Peitsch, M. C. and Ivanov, N. V. (2013). Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. *Genome Biol.* 14(6) R60. doi: 10.1186/gb-2013-14-6-r60
- Silva, A. T., Nguyen, A., Ye, C., Verchot, J. and Moon, J. H. Conjugated polymer nanoparticles for effective siRNA delivery to tobacco BY-2 protoplasts. *BMC Plant Biol.* 10 291 (2010). doi: 10.1186/1471-2229-10-291
- Sim, J. S., Kesawat, M. S., Kumar, M., Kim, S. Y., Mani, V., Subramanian, P., Park, S., Lee, C. M., Kim, S. R. and Hahn, B. S. (2018). Lack of the α 1, 3-fucosyltransferase gene (OsFucT) affects anther development and pollen viability in rice. *Int. J. Mol. Sci.* 19(4) 1225. doi: 10.3390/ijms19041225
- Simmons, C. W., VanderGheynst, J. S. and Upadhyaya, S. K. (2009). A model of *Agrobacterium tumefaciens* vacuum infiltration into harvested leaf tissue and subsequent in planta transgene transient expression. *Biotechnol. Bioeng.* 102(3) 965-970. doi: 10.1002/bit.22118
- Sladkova, K., Houska, J. and Havel, J. (2009). Laser desorption ionization of red phosphorus clusters and their use for mass calibration in time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 23 3114–3118. doi: 10.1002/rcm.4230
- Smargiasso, N., Nader, J., Rioux, S., Mazzucchelli, G., Boutry, M., De Pauw, E., Chaumont, F. and Navarre, C. (2019). Exploring the N-Glycosylation Profile of Glycoprotein B from Human Cytomegalovirus Expressed in CHO and *Nicotiana tabacum* BY-2 Cells. *Int. J. Mol. Sci.*, 20(15), 3741. doi: 10.3390/ijms20153741
- Sourrouille, C., Marquet-Blouin, E., D'Aoust, M. A., Kiefer-Meyer, M. C., Seveno, M., Pagny-Salehabadi, S., Bardor, M., Durambur, G., Lerouge, P., Vezina, L. and

- Gomord, V. (2008). Down-regulated expression of plant-specific glycoepitopes in alfalfa. *Plant Biotechnol. J.* 6(7) 702-721. doi: 10.1111/j.1467-7652.2008.00353.x
- Spies, M. and Malkova, A. (2018). Mechanisms of DNA Recombination and Genome Rearrangements: Methods to Study Homologous Recombination. *Academic Press*. ISBN: 9780128144305
- Srba, M., Černíková, A., Opatrný, Z. and Fischer, L. (2016). Practical guidelines for the characterization of tobacco BY-2 cell lines. *Biol. Plant.* 60 13–24. doi: 10.1007/s10535-015-0573-3
- Srivastava, V., Underwood, J. L. and Zhao, S. (2017). Dual-targeting by CRISPR/Cas9 for precise excision of transgenes from rice genome. *Plant Cell Tissue Organ Cult.* 129(1) 153–160. doi:10.1007/s11240-016-1166-3
- Stanley, P., S. Narasimhan, L. Siminovitch, and H. Schachter. (1975). Chinese hamster ovary cells selected for resistance to the cytotoxicity of phytohemagglutinin are deficient in a UDPN-acetylglucosamine-glycoprotein N-acetylglucosaminyltransferase activity. *Proc. Natl. Acad. Sci. USA* 72 3323-3327. doi: 10.1073/pnas.72.9.3323
- Stanley, P. (1983). Lectin-resistant CHO cells: selection of new mutant phenotypes. *Somat. Cell Genet.* 9(5) 593-608. doi: 10.1007/BF01574260
- Stanley, P. (2016). What have we learned from glycosyltransferase knockouts in mice?. *J. Mol. Biol.* 428(16) 3166-3182. doi: 10.1016/j.jmb.2016.03.025
- Stoger, E., Ma, J. K., Fischer, R. and Christou, P. (2005). Sowing the seeds of success: pharmaceutical proteins from plants. *Curr. Opin. Biotechnol.*, 16(2), 167-173. doi: 10.1016/j.copbio.2005.01.005
- Stougaard, J. (1993). Substrate-dependent negative selection in plants using a bacterial cytosine deaminase gene. *Plant J.* 3(5) 755-761. doi: 10.1111/j.1365-313X.1993.00755.x
- Strasser, R., Mucha, J., Schwihla, H., Altmann, F., Glössl, J. and Steinkellner, H. (1999). Molecular cloning and characterization of cDNA coding for β 1, 2N-acetylglucosaminyltransferase I (GlcNAc-TI) from *Nicotiana tabacum*. *Glycobiology* 9(8) 779-785. doi: 10.1093/glycob/9.8.779
- Strasser, R., Altmann, F., Glössl, J. and Steinkellner, H. (2004a). Unaltered complex N-glycan profiles in *Nicotiana benthamiana* despite drastic reduction of β 1,2-N-

acetylglucosaminyltransferase I activity. *Glycoconj. J.* 21(5) 275. doi: 10.1023/B:GLYC.0000045099.29038.04

Strasser, R., Altmann, F., Mach, L., Glossl, J., and Steinkellner, H. (2004b). Generation of *Arabidopsis thaliana* plants with complex N-glycans lacking beta1,2-linked xylose and core alpha1,3-linked fucose. *FEBS Lett.* 561 132–136. doi: 10.1016/S0014-5793(04)00150-4

Strasser, R., Stadlmann, J., Svoboda, B., Altmann, F., Glössl, J. and Mach, L. (2005). Molecular basis of N-acetylglucosaminyltransferase I deficiency in *Arabidopsis thaliana* plants lacking complex N-glycans. *Biochem. J.* 387(2) 385–391. doi: 10.1042/BJ20041686

Strasser, R., Schoberer, J., Jin, C., Glössl, J., Mach, L. and Steinkellner, H. (2006). Molecular cloning and characterization of *Arabidopsis thaliana* Golgi α -mannosidase II, a key enzyme in the formation of complex N-glycans in plants. *Plant J.* 45(5) 789–803. doi: 10.1111/j.1365-3113X.2005.02648.x

Strasser, R., Bondili, J. S., Vavra, U., Schoberer, J., Svoboda, B., Glössl, J., Léonard, R., Stadlmann, J., Altmann, F., Steinkellner, H. and Mach, L. (2007). A unique β 1, 3-galactosyltransferase is indispensable for the biosynthesis of N-glycans containing Lewis a structures in *Arabidopsis thaliana*. *Plant Cell* 19(7) 2278–2292. doi: 10.1105/tpc.107.052985

Strasser, R., Stadlmann, J., Schähs, M., Stiegler, G., Quendler, H., Mach, L., Glössl, J., Weterings, K., Pabst, M. and Steinkellner, H. (2008). Generation of glyco-engineered *Nicotiana benthamiana* for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure. *Plant Biotechnol. J.*, 6(4), 392–402. doi: 10.1111/j.1467-7652.2008.00330.x

Strasser, R., Castilho, A., Stadlmann, J., Kunert, R., Quendler, H., Gatteringer, P., Jez, J., Rademacher, T., Altmann, F., Mach, L. and Steinkellner, H. (2009). Improved virus neutralization by plant-produced anti-HIV antibodies with a homogeneous beta1,4-galactosylated N-glycan profile. *J. Biol. Chem.* 284, 20479–20485. doi: 10.1074/jbc.M109.014126

Strasser, R. (2016). Plant protein glycosylation. *Glycobiology* 26(9) 926–939. doi: 10.1093/glycob/cww023

Strasser, R. (2022). Recent Developments in Deciphering the Biological Role of Plant Complex N-Glycans. *Front. Plant Sci.*, 13, 897549. doi: 10.3389/fpls.2022.897549

- Sun, L., Paschall, A. V., Middleton, D. R., Ishihara, M., Ozdilek, A., Wantuch, P. L., Aceil, J., Duke, J. A., LaBranche, C. C., Tiemeyer, M., Avci, F. Y. (2020) Glycopeptide epitope facilitates HIV-1 envelope specific humoral immune responses by eliciting T cell help. *Nat Commun.* 11(1):2550. doi: 10.1038/s41467-020-16319-0
- Taga, E. M., Waheed, A., Van Etten, R. L. (1984). Structural and chemical characterization of a homogeneous peptide *N*-glycosidase from almond. *Biochemistry.* 23(5) 815-22. doi: 10.1021/bi00300a006
- Takahashi, S., Kojo, K. H., Kutsuna, N., Endo, M., Toki, S., Isoda, H. and Hasezawa, S. (2015). Differential responses to high-and low-dose ultraviolet-B stress in tobacco Bright Yellow-2 cells. *Front. Plant Sci.* 6 254. doi: 10.3389/fpls.2015.00254
- Takano, S., Matsuda, S., Funabiki, A., Furukawa, J. I., Yamauchi, T., Tokuiji, Y., Nakazono, M., Shinohara, Y., Takamure, I., and Kato, K. (2015). The rice RCN11 gene encodes β 1, 2-xylosyltransferase and is required for plant responses to abiotic stresses and phytohormones. *Plant Sci.*, 236, 75-88. doi: 10.1016/j.plantsci.2015.03.022
- Tekoah, Y., Shulman, A., Kizhner, T., Ruderfer, I., Fux, L., Nataf, Y., Bartfeld, D., Ariel, T., Gingis-Velitski, S., Hanania, U., Shaaltiel, Y. (2015). Large-scale production of pharmaceutical proteins in plant cell culture-the protalix experience. *Plant Biotechnol. J.* 13 1199-1208. doi: 10.1111/pbi.12428
- Tjondro, H. C., Loke, I., Chatterjee, S. and Thaysen-Andersen, M. (2019). Human protein paucimannosylation: cues from the eukaryotic kingdoms. *Biol. Rev.* 94(6) 2068-2100. doi: 10.1111/brev.12548
- Tronche, F., Casanova, E., Turiault, M., Sahly, I. and Kellendonk, C. (2002). When reverse genetics meets physiology: the use of site-specific recombinases in mice. *FEBS Lett.* 529(1) 116-21. doi: 10.1016/s0014-5793(02)03266-0
- U.S. National Library of Medicine (2021, -). A Phase 3, Randomized, Observer-Blind, Placebo-Controlled Study to Evaluate the Efficacy, Safety, and Immunogenicity of mRNA-1647 Cytomegalovirus (CMV) Vaccine in Healthy Participants 16 to 40 Years of Age. Identifier NCT05085366. <https://clinicaltrials.gov/ct2/show/NCT05085366>
- Uthailak, N., Kajiura, H., Misaki, R. and Fujiyama, K. (2022). Production of recombinant β -glucocerebrosidase in wild-type and glycoengineered transgenic *Nicotiana benthamiana* root cultures with different *N*-glycan profiles. *J. Biosci. Bioeng.* 133(5) 481-488. doi: 10.1016/j.jbiosc.2022.01.002

- Valliere-Douglass, J. F., Kodama, P., Mujacic, M., Brady, L. J., Wang, W., Wallace, A., Yan, B., Reddy, P., Treuheit, M. J. and Balland, A. (2009). Asparagine-linked oligosaccharides present on a non-consensus amino acid sequence in the CH1 domain of human antibodies. *J. Biol. Chem.* 284(47) 32493-32506. doi: 10.1074/jbc.M109.014803
- Valliere-Douglass, J. F., Eakin, C. M., Wallace, A., Ketchum, R. R., Wang, W., Treuheit, M. J. and Balland, A. (2010). Glutamine-linked and non-consensus asparagine-linked oligosaccharides present in human recombinant antibodies define novel protein glycosylation motifs. *J. Biol. Chem.* 285(21) 16012-16022. doi: 10.1074/jbc.M109.096412
- van der Kaaij, A., van Noort, K., Nibbering, P., Wilbers, R. H. P. and Schots, A. (2022). Glyco-Engineering Plants to Produce Helminth Glycoproteins as Prospective Biopharmaceuticals: Recent Advances, Challenges and Future Prospects. *Front. Plant Sci.* 13:882835. doi: 10.3389/fpls.2022.882835
- Van Landuyt, L., Lonigro, C., Meuris, L. and Callewaert, N. (2019). Customized protein glycosylation to improve biopharmaceutical function and targeting. *Curr. Opin. Biotechnol.* 60 17-28. doi: 10.1016/j.copbio.2018.11.017
- van Ree, R., Cabanes-Macheteau, M., Akkerdaas, J., Milazzo, J. P., Loutelier-Bourhis, C., Rayon, C., Villalba, M., Koppelman, S., Aalberse, R., Rodriguez, R., Faye, L., Lerouge, P. (2000). Beta(1,2)-xylose and alpha(1,3)-fucose residues have a strong contribution in IgE binding to plant glycoallergens. *J. Biol. Chem.* 275 11451–11458. doi: 10.1074/jbc.275.15.11451
- Vasilev, N., Grömping, U., Lipperts, A., Raven, N., Fischer, R. and Schillberg, S. (2013). Optimization of BY-2 cell suspension culture medium for the production of a human antibody using a combination of fractional factorial designs and the response surface method. *Plant Biotechnol. J.* 11, 867–874. doi: 10.1111/pbi.12079
- Vermeer, J. E. M., von Wangenheim, D., Barberon, M., Lee, Y., Stelzer, E. H. K., Maizel, A. and Geldner, N. (2014). A spatial accommodation by neighboring cells is required for organ initiation in *Arabidopsis*. *Science* 343 178-183. doi: 10.1126/science.1245871
- Vézina, L. P., Faye, L., Lerouge, P., D'Aoust, M. A., Marquet-Blouin, E., Burel, C., Lavoie, P. O., Bardor, M., Gomord, V. (2009). Transient co-expression for fast and high-yield production of antibodies with human-like N-glycans in plants. *Plant Biotechnol. J.* 7(5) 442–455. doi:10.1111/j.1467-7652.2009.00414.x

- von Schaewen, A., Sturm, A., O'Neill, J. and Chrispeels, M. (1993). Isolation of a mutant Arabidopsis plant that lacks N-acetyl glucosaminyl transferase I and is unable to synthesize Golgi-modified complex N-linked glycans. *Plant Physiol.* 102 1109–1118. doi: 10.1104/pp.102.4.1109
- Wada, R., Matsui, M. and Kawasaki, N. (2019). Influence of N-glycosylation on effector functions and thermal stability of glycoengineered IgG1 monoclonal antibody with homogeneous glycoforms. *MAbs* 11(2) 350-372. doi: 10.1080/19420862.2018.1551044
- Wang, Y., Chen, B., Hu, Y., Li, J. and Lin, Z. (2005). Inducible excision of selectable marker gene from transgenic plants by the Cre/lox site-specific recombination system. *Transgenic Res.* 14(5) 605-614. doi: 10.1007/s11248-005-0884-9
- Wang, C. C., Chen, J. R., Tseng, Y. C., Hsu, C. H., Hung, Y. F., Chen, S. W., Chen, C. M., Khoo, K. H., Cheng, T. J., Cheng, Y. S., Jan, J. T., Wu, C. Y., Ma, C. and Wong, C. H. (2009). Glycans on influenza hemagglutinin affect receptor binding and immune response. *Proc. Natl. Acad. Sci. USA* 106(43) 18137-18142. doi: 10.1073/pnas.0909696106
- Wang, X., Jiang, D., Shi, J. and Yang, D. (2017). Expression of α -1, 6-fucosyltransferase (FUT8) in rice grain and immunogenicity evaluation of plant-specific glycans. *J. Biotechnol.* 242 111-121. doi: 10.1016/j.jbiotec.2016.12.017
- Wang, S., Rong, Y., Wang, Y., Kong, D., Wang, P. G., Chen, M. and Kong, Y. (2020). Homogeneous production and characterization of recombinant N-GlcNAc-protein in *Pichia pastoris*. *Microb. Cell Fact.* 19(1) 1-11. doi: 10.1186/s12934-020-1280-0
- Wenderoth, I. and von Schaewen, A. (2000). Isolation and characterization of plant N-acetyl glucosaminyltransferase I (GntI) cDNA sequences. Functional analyses in the Arabidopsis cgl mutant and in antisense plants. *Plant Physiol.* 123(3) 1097-1108. doi: 10.1104/pp.123.3.1097
- Wilbers, R., Westerhof, L., van Noort, K., Obieglo, K., Driessen, N. N., Everts, B., Gringhuis, S. I., Schramm, G., Goverse, A., Smant, G., Bakker, J., Smits, H. H., Yazdanbakhsh, M., Schots, A., Hokke, C. H. (2017). Production and glyco-engineering of immunomodulatory helminth glycoproteins in plants. *Sci. Rep.* 7 45910. doi: 10.1038/srep45910
- Woo, J. W., Kim, J., Kwon, S. I., Corvalán, C., Cho, S. W., Kim, H., Kim, S. G., Kim, S. T., Choe, S. and Kim, J. S. (2015). DNA-free genome editing in plants with

preassembled CRISPR-Cas9 ribonucleoproteins. *Nat. Biotechnol.*, 33(11), 1162-1164. doi: 10.1038/nbt.3389

Wu, Y., Feng, Z., Li, P. and Yu, Q. (2020). Relationship between ABO Blood Group Distribution and Clinical Characteristics in Patients with COVID-19. *Clinica Chim. Acta* 509 220–223. doi:10.1016/j.cca.2020.06.026

Wu, C. Y., Cheng, C. W., Kung, C. C., Liao, K. S., Jan, J. T., Ma, C. and Wong, C. H. (2022). Glycosite-deleted mRNA of SARS-CoV-2 spike protein as a broad-spectrum vaccine. *Proc. Natl. Acad. Sci. USA* 119(9) e2119995119. doi: 10.1073/pnas.2119995119

Xiao, A., Wang, Z., Hu, Y., Wu, Y., Luo, Z., Yang, Z., Zu, Y., Li, W., Huang, P., Tong, X., Zhu, Z., Lin, S. and Zhang, B. (2013). Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Res.* 41 e141. doi: 10.1093/nar/gkt464

Xie, K., Minkenberg, B. and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. USA* 112(11) 3570–3575. doi:10.1073/pnas.1420294112

Yang, Q., Zhang, R., Cai, H., and Wang, L. X. (2017). Revisiting the substrate specificity of mammalian α 1, 6-fucosyltransferase reveals that it catalyzes core fucosylation of *N*-glycans lacking α 1, 3-arm GlcNAc. *J. Biol. Chem.* 292 14796-14803. doi: 10.1074/jbc.M117.804070

Yao, J., Weng, Y., Dickey, A. and Wang, K. Y. (2015). Plants as factories for human pharmaceuticals: applications and challenges. *Int. J. Mol. Sci.* 16(12) 28549-28565. doi: 10.3390/ijms161226122

Yee, C. M., Zak, A. J., Hill, B. D. and Wen, F. (2018). The coming age of insect cells for manufacturing and development of protein therapeutics. *Ind. Eng. Chem. Res.* 57(31) 10061-10070. doi: 10.1021/acs.iecr.8b00985

Yin, B. J., Gao, T., Zheng, N. Y., Li, Y., Tang, S. Y., Liang, L. M. and Xie, Q. (2011). Generation of glyco-engineered BY2 cell lines with decreased expression of plant-specific glycoepitopes. *Protein Cell.* 2(1) 41-47. doi: 10.1007/s13238-011-1007-4

Zhang, K., Baeckstrom, D. and Hansson, G.C. (1994). A secreted mucin carrying sialyl-Lewis a from colon carcinoma cells binds to E-selectin and inhibits HL-60 cell adhesion. *Int. J. Cancer.* 59 823–829. doi: 10.1002/ijc.2910590619

- Zhang, L., Jia, R., Palange, N.J., Satheka, A.C., Togo, J., An, Y., Humphrey, M., Ban, L., Ji, Y., Jin, H., Feng, X. and Zheng, Y. (2015). Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. *PLOS ONE* 10(3) e0120396. doi: 10.1371/journal.pone.0120396
- Zhang, Y., Malzahn, A. A., Sretenovic, S. and Qi, Y. (2019). The emerging and uncultivated potential of CRISPR technology in plant science. *Nat. Plants*. 5(8) 778-794. doi: 10.1038/s41477-019-0461-5
- Zheng, K., Bantog, C. and Bayer, R. (2011). The impact of glycosylation on monoclonal antibody conformation and stability. *MAbs* 3(6) 568-576. doi: 10.4161/mabs.3.6.17922
- Zhou, H., Liu, B., Weeks, D. P., Spalding, M. H. and Yang, B. (2014). Large chromosomal deletions and heritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acids Res.* 42(17) 10903-14. doi: 10.1093/nar/gku806
- Zhu, S., Li, W., Liu, J., Chen, C.-H., Liao, Q., Xu, P., Xu, H., Xiao, T., Cao, Z., Peng, J., Yuan, P., Brown, M., Liu, X. S. and Wei, W. (2016). Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR library. *Nat. Biotechnol.* 34 1279–1286. doi: 10.1038/nbt.3715
- Zhu, Q., Tan, J. and Liu, Y. G. (2022). Molecular farming using transgenic rice endosperm. *Trends Biotechnol.* 40(10) 1248-1260. doi: 10.1016/j.tibtech.2022.04.002
- Zubko, E., Scutt, C. and Meyer, P. (2000). Intrachromosomal recombination between attP regions as a tool to remove selectable marker genes from tobacco transgenes. *Nat. Biotechnol.* 18(4) 442-5. doi: 10.1038/74

