Faculté des Bioingénieurs

Louvain Institute of Biomolecular Science and Technology

Engineering *O*-glycosylation in *Nicotiana tabacum* BY-2 cells

Thèse présentée par Nicolas Bailly en vue de l'obtention du grade de docteur en Sciences Agronomiques et Ingénierie Biologique

Promoteur.rice.s :	Prof. François CHAUMONT
	Dr. Catherine NAVARRE
Président du jury :	Prof. Bernard HALLET
Membres du jury :	Prof. Henri BATOKO
	Prof. Cédric DELPORTE
	Prof. Patrice SOUMILLION
	Prof. Richard STRASSER

Février 2025

Remerciements

Ce travail est l'accomplissement de quatre années de recherche en laboratoire, mais également d'une aventure personnelle riche en apprentissages. Ces années ont été marquées par des succès et des moments de fierté, que j'ai eu la chance de partager, mais aussi par des difficultés et des échecs, que j'ai pu surmonter grâce au soutien de mon entourage. Je tiens à exprimer ma gratitude à tous ceux qui m'ont accompagné dans le cadre de mon travail ou en dehors.

Tout d'abord, je souhaite remercier les personnes qui m'ont accueilli et encadré durant ces cinq années. François, merci pour ton accueil chaleureux, ton soutien et tes conseils avisés. Catherine, merci pour ta gentillesse, ton expertise scientifique et ta disponibilité, malgré les nombreux projets que tu mènes en parallèle, ainsi que pour les moments que nous avons passés au labo et pendant les conférences.

Je remercie également les membres du jury et du comité d'accompagnement pour leur suivi et leurs commentaires constructifs. Les discussions que nous avons eues ont enrichi ma réflexion et, je l'espère, amélioré la qualité de ce travail.

L'atmosphère chaleureuse et l'entraide au sein du laboratoire ont toujours contribué à créer un environnement positif et productif. Merci à tous les membres, passés et présents, de l'équipe FARM : Adeline, Jules, Mathilde, Romane, Antoine, Camille, Christophe et Rick. Une mention spéciale à Marie, Xavier et Laurent pour les super moments passés ensemble à Cape Town et à Rome. Je souhaite également remercier ma mémorante, Marie, pour son enthousiasme et sa détermination malgré les difficultés rencontrées lors de son projet.

Un grand merci à Joseph pour son humour, sa bienveillance et ses précieux conseils. Merci également à tous les membres de FYMO, en particulier Maxime, Estelle, Quentin, Louis, Manon, Alice, Jérôme et Joséphine, pour nos échanges, qu'ils soient scientifiques ou plus légers, et pour tous les moments et rires partagés.

Je suis profondément reconnaissant envers ma famille et mes amis, qui m'ont entouré et soutenu tout au long de ce parcours. Merci à ceux qui se sont intéressés à mon travail, qui ont écouté, posé des questions ou partagé des moments comme nos fameux business lunches. Vous êtes de ceux qui rendent la vie plus joyeuse.

Enfin, merci à toi, Emi, pour ton soutien infaillible et ta présence constante à mes côtés. Tu me pousses chaque jour à devenir une meilleure version de moi-même, et je ne te remercierai jamais assez pour cela.

Table of contents

Table of Contents

SUMMARY ..... **IX**

ABBREVIATIONS ..... **XIII**

INTRODUCTION **1**

1 CONTEXTUALIZATION **3**
2 METAZOAN GALNAC-*O*-GLYCOSYLATION **9**
2.1 BIOSYNTHESIS 9
2.1.1 Initiation 9
2.1.2 The multi-layered specificity of GalNAcTs 17
2.1.3 Elongation and capping 19
2.2 MOLECULAR FUNCTIONS OF *O*-GLYCANS 20
2.3 *O*-GLYCOENGINEERING OF PHARMACEUTICAL SYSTEMS 23
3 PLANT HYDROXYPROLINE *O*-GLYCOSYLATION **27**
3.1 PROLINE HYDROXYLATION 28
3.1.1 P4Hs in humans 32
3.1.2 P4Hs in plants 32
3.1.3 Plant proline hydroxylation code 34
3.1.4 Specificities of plant P4Hs 35
3.2 *O*-GLYCOSYLATION 37
3.2.1 AGP type 42
3.2.2 Extensin type 46
3.2.3 Other glycosylation types 48
3.3 IMPLICATION OF PLANT *O*-GLYCOSYLATION IN PMF 50
3.3.1 Downside 50
3.3.2 Upside 52

OBJECTIVES **55**

RESULTS AND DISCUSSION **61**

1 INTRODUCING GALNAC-*O*-GLYCOSYLATION IN BY-2 CELLS **63**
1.1 GENETIC CONSTRUCTS AND BY-2 CELL TRANSFORMATION 63

1.2 GLYCOSYLATION ANALYSIS ON ENDOGENOUS PROTEINS 64
2 GALNAC-*O*-GLYCOSYLATION OF THERAPEUTIC PROTEINS 69
2.1 IMMUNOGLOBULIN A1 69
2.1.1 Genetic constructs 70
2.1.2 Production and purification of IgA1 in ΔXF BY-2 cells 71
2.1.3 Analysis of glycosylation profile by lectin blotting 75
2.1.4 Analysis of the glycosylation profile by *in vitro* deglycosylation 76
2.1.5 Analysis of the glycosylation profile by mass spectrometry 76
2.2 EPO 85
2.2.1 Genetic constructs and BY-2 cell transformation 86
2.2.2 Accumulation and purification of EPO-Fc in the EM of in BY-2 cells 87
2.2.3 Analysis of EPO-Fc *N*-glycosylation profile by *in vitro* deglycosylation 88
2.2.4 Generation of *O*-glycosylated EPO-Fc and analysis by lectin blotting 89
2.2.5 Characterization of protein A purified EPO-Fc 89
2.3 SARS-COV-2 SPIKE 94
2.3.1 Genetic constructs and expression in BY-2 cells 94
2.3.2 Nickel affinity purification of S1 from intracellular protein extracts 95
2.3.3 Analysis of the *N*-glycosylation profile by *in vitro* deglycosylation 95
2.4 DISCUSSION 98
3 PROLINE HYDROXYLATION IN BY-2 CELLS 101
3.1 IDENTIFICATION AND ANALYSIS OF P4HS IN *NICOTIANA TABACUM* 101
3.2 CO-EXPRESSION ANALYSIS OF *NTP4HS* 109
4 SILENCING OF P4H ACTIVITY IN BY-2 CELLS 113
4.1 GENERATION OF REPORTER CELL LINES 113
4.2 DESIGN OF CRISPR/CAS9 GUIDE RNAS AND VECTOR CONSTRUCTION 116
4.3 KNOCK OUT OF *P4H4* 119
4.4 KNOCK OUT OF *P4H1* AND *P4H3/10* 122
4.5 DISCUSSION 123

GENERAL DISCUSSION AND PERSPECTIVES 127

SUPPLEMENTARY FIGURES AND TABLES 133

MATERIALS AND METHODS 149

1 *N. TABACUM* BY-2 CELL AND NICOTIANA PLANTS CULTURE 151
2 BY-2 CELL CULTURE GROWTH KINETICS 151

3 BINARY PLASMID CONSTRUCTION 151
3.1 GalNAc-*O*-glycosylation plasmids 152
3.2 IgA1 plasmids 153
3.3 EPO plasmids 153
3.4 Spike glycoprotein plasmids 153
3.5 Reporter GFP plasmids 154
4 *A. TUMEFACIENS* TRANSFORMATION 155
5 *AGROBACTERIUM*-MEDIATED PLANT CELL TRANSFORMATION 156
6 GENOMIC DNA EXTRACTION 156
7 CHARACTERIZATION OF *P4H* HOMEOLOGS AND ANALYSIS OF GENOME MODIFICATIONS 156
8 PROTEIN EXTRACTION 156
9 SDS-PAGE, WESTERN AND LECTIN BLOTTING ANALYSIS OF PROTEINS 157
10 IGA AND EPO-FC PROTEIN PURIFICATION 158
11 MASS SPECTROMETRY ANALYSIS IGA1 GLYCOPEPTIDES 158
12 MASS SPECTROMETRY ANALYSIS OF EPO-FC 159
13 *IN VITRO* DEGLYCOSYLATION ASSAY 161
14 REPRESENTATION OF *N*-GLYCAN STRUCTURES 161

REFERENCES 163

Summary

Glycosylation has significant impacts on the properties of therapeutic proteins and is described as a critical quality attribute by the industry. Glycoengineering has therefore proved to be a powerful strategy for producing safer and improved therapeutics. Plant cells suspensions such as *Nicotiana tabacum* BY-2 cells are a particularly attractive recombinant expression system in this context, as they display a simple *N*-glycosylation profile that can be easily engineered to produce proteins with a homogeneous and humanised glycosylation profile.

However, the *O*-glycosylation capacity of BY-2 cells is completely different from that of human cells. Indeed, GalNAc-*O*-glycosylation, which is the most abundant type of *O*-glycosylation in mammalian cells, is absent in plants. In contrast, plant *O*-glycosylation consists of the hydroxylation of proline residues and their glycosylation with arabinose chains or arabinogalactan polysaccharides. These residues, which are absent in humans and potentially immunogenic, have been observed on several plant-produced therapeutics.

To improve the BY-2 cells system developed in our laboratory, we engineered a BY-2 cell line to produce GalNAc-*O*-glycans. Starting from a cell line with a humanised *N*-glycosylation profile, we expressed the enzymes responsible for initiating glycosylation and generating the core 1 structure, GalNAcT2 and C1GalT1, respectively, and showed that this cell line can produce the immunoglobulin A1 with humanised *O*-glycosylation of the hinge region. However, the hinge was also modified with hydroxyprolines and pentoses. Human EPO fused to an IgG2 Fc and the S1 glycoprotein from SARS-CoV-2 were also used as reporters for GalNAc-*O*-glycosylation but were subject to proteolysis in BY-2 cells, hampering downstream analysis.

We identified twenty-one putative *P4H* genes in the genome of *N. tabacum*, responsible for the first step of the plant-specific *O*-glycosylation, and we began a CRISPR/Cas9 knockout screen of *P4Hs*. We generated reporter cell lines based on the fusion of hyperglycosylated Ser-Pro repeats modules with GFP to enable a simple phenotypic characterisation by western blotting. This screening has not yet identified P4Hs that are required for *O*-glycosylation of these modules.

This work paved the way for the humanisation of *O*-glycosylation profile in BY-2 cells and a better understanding of the different eukaryotic *O*-glycosylation pathways.

Abbreviations

2ODO	2-oxoglutarate dependent oxygenase
2OG	2-oxoglutarate
AAT	Alpha1 antitrypsin
AG	Arabinogalactan
AGM	AG glucuronic acid methyltransferases
AGP	Arabinogalactan protein
Ara	Arabinose
C1GalT1	Core 1 Galactosyltransferase 1
C2GnT	Core 2 N-acetylglucosaminyltransferase
C3GnT	Core 3 N-acetylglucosaminyltransferase
CLE-RS2	CLAVATA3 associated peptides – root signal 2
CMP	Cytosine-monophosphate
COSMC	Core 1 beta3-Gal-T-specific molecular chaperone
CP4H	collagen P4H
CSF	Colony stimulating factor
CW	cell wall
CWP	cell wall protein
DDA	Data dependent acquisition
EM	Extracellular medium
EPO	Erythropoietin
EPOR	EPO receptor
ER	Endoplasmic reticulum
EXaD	Extensin Arabinose Deficient transferase
EXT	Extensin
Fc	Fraction crystallisable
FGF23	Fibroblast Growth Factor 23
FLA4	Fascilin like AGP 4
Fuc	Fucose
FUT	Fucosyltransferase
Gal	Galactose
GalNAc	N-acetylgalactosamine
GalNAcT	N-acetylgalactosaminyltransferase
GalT	Galactosyltransferase
G-CSF	Granulocyte colony stimulating growth factor
GH	Glycosylhydrolase
GlcA	Glucuronic acid

GlcAT	Glucuronic acid transferase
GlcNAc	N-acetylglucosamine
GMP	Good manufacturing practices
gRNA	Guide RNA
GT	Glycosyltransferase
HC	Heavy chain
HDMSE	Ion Mobility Separation High-Definition Enhanced MS
HER2	Human epidermal growth factor receptor 2
HIF	Hypoxia-inducible transcription factor
HPAT	Hdroxyproline O-arabinosyltransferase
HPGT	Hydroxyproline galactoyltransferase
HPL	*Helix pomatia* lectin
HRGP	Hydroxyproline-rich glycoproteins
HRP	Horseradish peroxidase
Hyp	Hydroxyproline
IDP	Intrinsically disordered proteins
IgA1	Immunoglobulin A1
LB	Left border
LC	Light chain
MS	Mass Spectrometry
NeuAc	N-acetylneuramidic acid
P4H	Prolyl-4-hydroxylase
PDI	Protein disulfide isomerase
PHD	Prolylhydroxylase domain protein
PMF	Plant Molecular Farming
PMP	plant made pharaceutical
PNL	Peanut agglutinin
ProtA	Protein A
PRP	Proline rich protein
PTM	PostTranslational modification
RB	Right border
RGF	Root growth factor
Rha	Rhamnose
RRA	Reduced residual arabinose
S1	Spike glycoprotein subunit 1
SAR	Sacffold attachment region

SEC	Size exclusion chromatography
SGT	Serine Galactosyltranferase
SP	Signal peptide
ST3GalT	ST3 beta-galactoside alpha-2,3-sialyltransferase
ST6GalNAcT	ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase
TNF	Tumor Necrosis Factor α
TSP	Total soluble proteins
UDP	Uracil-diphosphate
VLP	Virus like particle
WT	Wild-type
Xyl	Xylose

Introduction

1 Contextualization

Recombinant proteins are potent and specific therapeutic products that have become one of the most important classes of therapeutics. Initially, they were produced using unicellular organisms such as bacteria or yeasts, due to low costs and easy manipulation, while maintaining high yields. However, the increasing complexity of the proteins produced, including the folding of large proteins, the assembly of multimers and their post-translational modifications, has led to the increasing use of mammalian cell lines, such as like the CHO cells. While mammalian cell lines offer a cellular physiology closer to that of human proteins, they present several drawbacks, mainly in terms of cost and safety. In the recent years, other platforms based on insect cells or plants have been investigated.

Plant molecular farming (PMF) is developed as an alternative to produce recombinant proteins, including therapeutic products. Since the successful expression of an antibody in transgenic tobacco in 1989 (Hiatt et al., 1989), many research groups and industrials have used and developed different plant-based systems to produce proteins from all classes, including antibodies, enzymes, growth hormones and vaccines antigens (reviewed in (Eidenberger et al., 2023; Ghag et al., 2021; Karki et al., 2021; J. Lee et al., 2023; Santos et al., 2016; Zahmanova et al., 2023)). Several products performed well in clinical trials (e.g., Medicago influenza VLP vaccine (Ward et al., 2020)) and even reached marketing approval (Dow AgroSciences' Newcastle disease vaccine (Nurzijah et al., 2022) and Protalix's glucocerebrosidase (Tekoah et al., 2015) and α-galactosidase (Wallace et al., 2024)). Of note, due to recent outbreaks and the ability of *Nicotiana benthamiana* to rapidly produce high amounts of proteins, some plant-made therapeutics also received emergency use (Zmapp antibody cocktail for treatment of Ebola (Davey et al., 2016)) or accelerated approval (Medicago's SARS-Cov-2 VLP vaccine Covifenz (Hager et al., 2022)). One of the advantages of the plant systems is their ability to fold and carry out the posttranslational modifications (PTM) of complex proteins. For instance, the core *N*-glycan is conserved between eukaryotes. However, there are differences that need to be considered to produce safer and better pharmaceuticals. One of the major divergences in the glycosylation capacities of plants and humans is *O*-glycosylation. Indeed, *O*-glycosylation denotes very different pathways in plants and animals. On the one hand, GalNAc-*O*-glycosylation, which is the most abundant form of *O*-glycosylation in metazoans, is absent in plants. On the other hand, Hydroxyproline-*O*-glycosylation is only found in plants. This is a major drawback for PMF as glycosylation can have a profound impact on the properties of proteins and it is thought that 80% of proteins passing through the human secretory pathway are GalNAc-*O*-glycosylated (de las Rivas et al., 2019). Table I.1.1 summarises the GalNAc-*O*-glycoproteins that have been produced in plant systems and lack this important modification. In addition, plants have been extensively used to produce

viral antigens (reviewed in (Rybicki, 2014; Su et al., 2023)) where *O*-glycosylation is being increasingly studied for its implication in viral biology (Ming et al., 2024).

Thus, engineering plant cells to produce proteins with a human *O*-glycosylation profile is a necessary step towards improving plant-based therapeutics and facilitating their regulatory approval. In addition, the ability to introduce GalNAc-*O*-glycosylation in plants from scratch could (1) provide fundamental insights on this complex PTM, which particularly difficult to study and not yet understood at the functional and regulatory levels and (2) enable plants to produce improved therapeutics compared with traditional systems. Indeed, due to the lack of understanding of this modification, it is often neglected in industrial settings, which could result in less stable and less efficient products. For example, many approved products are GalNAc-*O*-glycoproteins (Table I.1.2) that include some of the top pharmaceuticals such as insulin, EPO or Etanercept (Tumor Necrosis Factor α (TNF) receptor fused to IgG Fc) (Biel et al., 2022; Walsh, 2018; Walsh & Walsh, 2022).

In addition, the presence of non-human glycans on therapeutic proteins can pose immunological problems. This concern was first raised by the presence of non-human structures on plant *N*-glycans, and solutions have been thoroughly investigated (for instance: (Jansing et al., 2019; Mercx et al., 2017; Strasser et al., 2008)). However, the presence of plant typical *O*-glycans on pharmaceuticals has not been fully explored, and only few reports have detected it (Cardon et al., 2019; Karnoup et al., 2005; Pinkhasov et al., 2011; Weise et al., 2007). Despite many unknowns about the regulation of the initiation of this pathway, 80% of human proteins are thought to carry sequences recognised by the plant *O*-glycosylation machinery (Gomord et al., 2010). In addition, the Hyp-*O*-glycosylation sites are found close to Ser and Thr while GalNac-*O*-glycosylation sequons are often rich in Pro. This may lead to competition between the two pathways. Thus, suppression of this modification seems necessary to improve the quality of plant-made therapeutics.

These issues can be addressed using *Nicotiana tabacum* BY-2 cell suspension cultures. Indeed, these cells can be used to produce therapeutic glycoproteins in the safe, controlled environment of the biofermenter. This combines the cost and safety advantages of plant systems with the containment and processing and regulatory know-how of the current industry's Good Manufacturing Practices (GMP). In addition, BY-2 cells are amenable to glycoengineering without loss of productivity or adverse phenotypes.

Table I.1.1: Overview of plant-produced GalNAc-*O*-glycoproteins for therapeutic use. The number of *O*-sites is indicated when experimental determination was made. Prediction of glycosylation site was made using NetOGlyc4.0 (Steentoft et al., 2013). *O-glycosylation of human serum albumin occurs only in a variant of the protein.

Protein	Clinical relevance	Experimental *O*-glycoprotein	Predicted *O*-glycoprotein	Plant System
Insulin	Diabetes	2 sites (Madsen et al., 2020)	1 site	Tobacco and lettuce (Ruhlman et al., 2007), Soybean (Cunha et al., 2010), Maize (Farinas et al., 2007), Safflower (Moloney et al., 2009)
Interferon α 2b	Antiviral and immunomodulator	1 site (Adolf et al., 1991)	1 site	BY-2 cells (Xu et al., 2007)
Interferon γ	Immunoregulation	No	2 sites	*N. benthamiana* (Jiang et al., 2019)
Interleukin 2	Immunoregulation	1 site (Thompson & Wakarchuk, 2022)	6 sites	BY-2 cells (Magnuson et al., 1998)
Interleukin 4	Immunoregulation	No	2 sites	BY-2 cells (Magnuson et al., 1998)
Interleukin 12	Immunoregulation	No	7 sites	BY-2 cells (T. H. Kwon et al., 2003)
Growth hormone (GH)	Growth hormone	1 site (Bustamante et al., 2009)	6 sites	BY-2 cells (Xu et al., 2010)
Erythropoietin (EPO)	Erythropoietic factor	1 site (Darling et al., 2002)	3 sites	Arabidopsis, *Medicago truncatula* and BY-2 cells (Pires et al., 2012), *P. patens* (Weise et al., 2007), *N. benthamiana* (Castilho et al., 2012)
Granulocyte colony stimulating growth factor (G-CSF)	Production of white cells	4 sites (Thompson & Wakarchuk, 2022)	2 sites	BY-2 (James et al., 2000; J.-H. Lee et al., 2002), rice cells (T.-G. Kim et al., 2008; Y. Shin et al., 2011)
Acetylcholinesterase	Biodefense	No	4 sites	Carrot cells (Atsmon et al., 2015)
Blood coagulation factor FVIII	Haemophilia	8 sites (King et al., 2017; J. Qu et al., 2020)	>100 sites	*N. benthamiana* (Sherman et al., 2014), lettuce (K. Kwon et al., 2018)

Table I.1.1: Continued

Protein	Clinical relevance	Experimental *O*-glycoprotein	Predicted *O*-glycoprotein	Plant System
α-l-iduronidase	Replacement enzyme	No	7 sites	BY-2 cells (Fu et al., 2009), *Brassica rapa* (Cardon et al., 2019)
TNFα receptor (TNFR)-Fc fusion	Autoimmune diseases	13 sites (Biel et al., 2022)	41 sites	BY-2 cells (Almon et al., 2017; Ilan et al., 2017)
Stem cell factor	Erythropoietic factor	3 sites (Lu et al., 1992)	14 sites	BY-2 cells (X. Wang et al., 2021)
GA733-2	Carcinoma antigen	No	4 sites	Tomato (Park et al., 2022), tobacco and *N. benthamiana* (Park et al., 2021)
IgA1	Passive immunisation	3-5 sites (de Haan et al., 2020)	9 sites	Maize (Karnoup et al., 2005), *N. benthamiana* (Dicker et al., 2016), tobacco (Ma et al., 1994)
Serum albumin	Hypoalbuminemia	1 site* (Minchiotti et al., 2001)	9 sites	Rice cells (Huang et al., 2005), BY-2 cells (Q.-Y. Sun et al., 2011)
Prostaglandin D2 synthase	Clinical marker	1 site (Halim et al., 2013)	No	*M. truncatula* cells (Pires et al., 2014)
DNAse I	Mucoviscidosis	No	1 site	BY-2 cells (Hanania et al., 2017)
Lactoferrin	Immunosuppressive, immunomodulator	41 sites (S. Yang et al., 2020)	6 sites	Siberian Ginseng cells (Jo et al., 2006), Korean Ginseng cells (S.-Y. Kwon et al., 2003), Sweet potato cells (Min et al., 2006)
α1-antitrypsin	Emphysema	6 sites (King et al., 2017)	5 sites	BY-2 (N. Zhang, Wright, Caraway, et al., 2019), rice cells (McDonald et al., 2008)
β-glucocerebrosidase	Replacement enzyme	No	3 sites	Carrot cells (Tekoah et al., 2015)
Thrombin	Fibrin formation	5 sites (King et al., 2017)	31 sites	*N. benthamian* (Laguia-Becher et al., 2019)

Table I.1.2: Non-exhaustive overview of *O*-glycoproteins in marketed therapeutics (Walsh & Walsh, 2022)**.**

Therapeutic class	**Approved products**	***O*-glycoprotein examples**
Clotting factors	32	Factor VIII (19 products), Von Wilderbrand factor (2 products), Factor IX (6 products)
Other blood related products	13	Plasma protease C1 inhibitor (1 product), Plasmin (1 product)
Hormones	89	Insulin (35 products), Growth hormone (17 products)
Growth factors	46	EPO (14 products), Granulocyte colony stimulating growth factor (26 products)
Interferons and interleukins	21	Interferon α 2 (2 products), Interleukin 2 (1 product)
Vaccines	43	HBV vaccines based on recombinant S antigen (19 products), SARS-CoV-2 based on recombinant Spike antigen (2 products)
Monoclonal antibodies	190	/
Enzymes	25	DNAse I (predicted, 1 product), Glucocerebrosidase (predicted, 3 products
Fusion proteins	15	Etanercept (5 products)

2 Metazoan GalNAc-*O*-glycosylation

GalNAc-*O*-glycosylation is an important PTM of metazoans, both in terms of function and abundance. Due to the intrinsic difficulties in studying this modification, the exact extent of *O*-glycosylation is not clear, but it is found on more than 50% of proteins passing through the secretory pathway (Gill et al., 2011), while some authors predicting as much as 80% (de las Rivas et al., 2019). In addition, it is found abundantly on mucins where it can account for more than half of the protein weight and is thus often referred to as mucin-type *O*-glycosylation.

2.1 Biosynthesis

O-glycosylation is takes place throughout the secretory pathway. There are two competing determinants to the final *O*-glycoprofile of proteins: density and elaboration. It results from the interplay between the enzymes that act cooperatively or in competition for the same substrates. *N*-acetylgalactosamine transferases (GalNAcTs) act together to initiate the *O*-glycosylation by transferring a single GalNAc residue to Ser or Thr. The initiation controls the sites and density of *O*-glycosylation. Then, other glycosyltransferases (GTs) generate different core structures from the GalNAc residue. The core structures are then elongated and capped by terminal saccharides (Figure I.2.1). The result is a complex mixture of glycoforms. However, cell-type specific regulation of GalNAcTs isoenzymes and other glycosyltransferases enables controlled structures to be generated.

2.1.1 Initiation

O-glycosylation is initiated by the addition of α-GalNAc on Ser and Thr. This process is mediated by a large family of 20 GalNAcTs in human. Fifteen members have been demonstrated to be active and one is thought to be a pseudogene (*GALNT20*) (Bennett et al., 2012). Figure I.2.2 depicts a phylogenic tree of human GalNAcTs. On the basis of this analysis and the elucidation of the specificity of each isoform, GalNAcTs are classified according to different criteria. A first grouping method takes into account the preference of each enzyme for peptide or glycopeptide substrates (groups I and II). Next, close homologs with similar specificity are grouped together (subgroups a-g). These related isoforms are often expressed in different cell-types giving them a non-redundant function in the organism (Bennett et al., 1999; Festari et al., 2017). Finally, all enzymes from these groups belong to the W family, as opposed to the Y family. Members of the Y family have a Trp to Tyr mutation in the UDP-GalNAc binding site. This mutation, among others, makes them unable of binding this substrate and inactivates the enzymes (Li et al., 2012). This family has

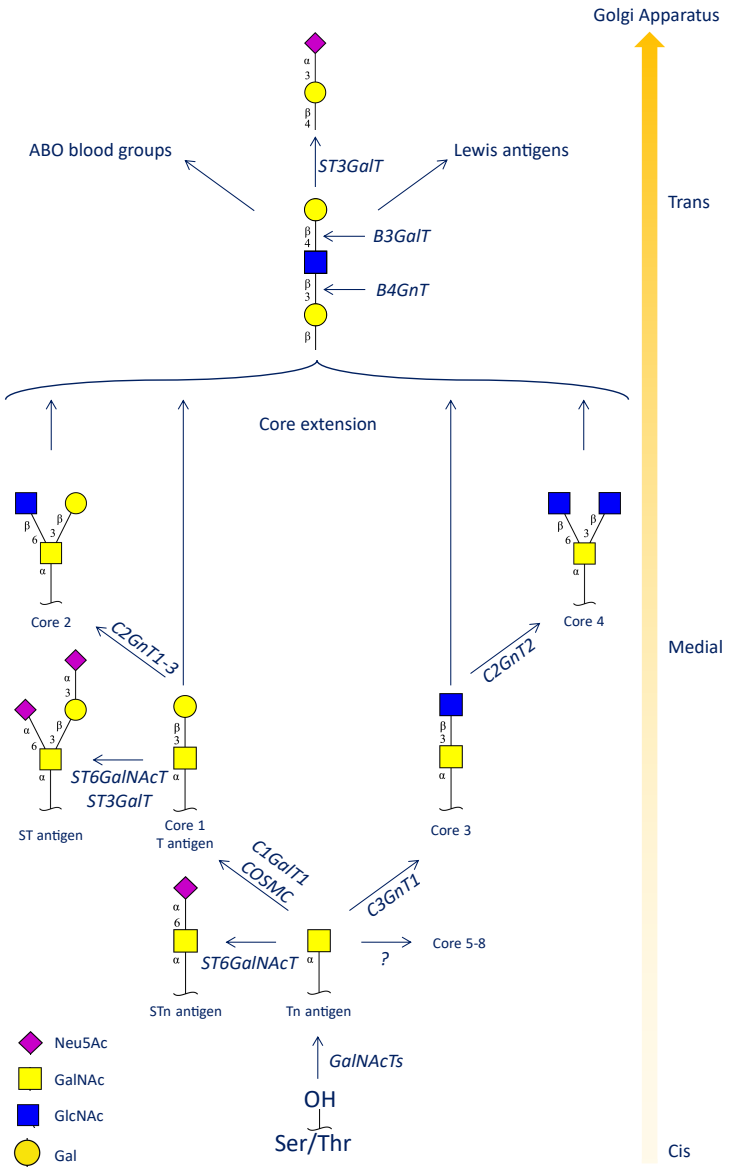


Figure I.2.1: GalNAc-*O*-glycosylation pathway. The hydroxyl group of Ser or Thr is modified by the addition of a single GalNAc residue. The GalNAc-*O*-glycan is then modified into different core structures that are elongated with Gal and GlcNAc. Finally, the glycans are capped with sialic acid or with complex structures such as ABO blood groups and Lewis antigens.

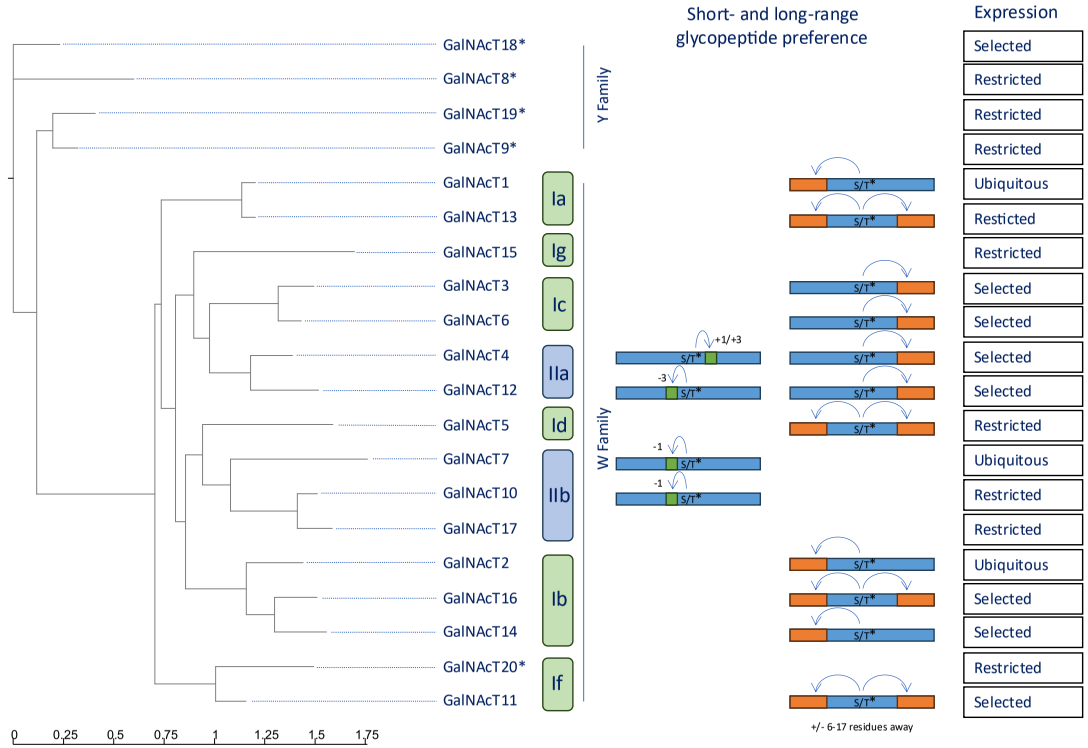


Figure I.2.2: The human GalNAcTs. From left to right: phylogenic tree analysis of GalNAcT proteins in humans allows to class them in different subgroups. The known glycosylation preferences of each enzyme are shown (S/T* indicates a previously modified residue). The enzymes which are not active are denoted with an asterisk. Finally, the expression pattern of each gene is indicated if the gene is expressed in more than 70% of tissues (ubiquitous), between 30 and 70% (selected) or below 30% (restricted) (Bennett et al., 2012).

not been studied in detail and its members have an unknown function, although they are suspected of modulating the activity of other GalNAcTs, possibly acting as chaperones (Li et al., 2012).

The GalNAcT family is the largest family of glycosyltransferases responsible for a single glycosidic linkage, and it is highly conserved throughout animal evolution (Bennett et al., 2012). Initially, GalNAcTs were thought to be highly redundant. However, many studies have shown that this redundancy is misleading: (1) different isoenzymes have site preferences and act together to define *O*-glycosylation patterns on highly glycosylated substrates and (2) some isoenzymes act very specifically on several substrates. For instance, GalNAcT3 is solely responsible of the site-specific *O*-glycosylation of Fibroblast Growth Factor 23 (FGF23)(Kato et al., 2006). Defect in these specific isoenzymes can lead to several diseases such as familial tumoral calcinosis in the case of GalNAcT3. Furthermore, knockout (KO) of specific GalNAcTs has major impacts on development in *Drosophila melanogaster* and *Mus musculus* and is sometimes lethal (Tran & Ten Hagen, 2013).

The different GalNAcT isoforms have various expression profiles throughout the organism (Figure I.2.2). Some are ubiquitously expressed, while others have a more restricted or even cell-type specific expression profile. For instance, *GALNT13* expression is restricted to the brain (Festari et al., 2017). At the subcellular level, GalNAcTs are located in the Golgi apparatus. Some studies also suggest that different isoenzymes may have subtle differences in localisation in the Golgi (Röttger et al., 1998). Anterograde transport to the endoplasmic reticulum (ER) can be activated by Src kinases, which is often observed in cancers (Gill et al., 2010). This leads to an altered *O*-glycosylation profile characterised by denser, less complex glycans. Denser glycosylation could be due to (1) better access to substrates resulting from less competition with Golgi resident enzymes, (2) more access to glycosylation sites on unfolded proteins in the ER, (3) altered activity due to change in pH (Gill et al., 2010). This dense glycosylation is responsible for shorter glycans as it inhibits the activity of elongating enzymes *in vivo* (Granovsk et al., 1994; Schwientek et al., 2000). This mechanism may also promote *O*-glycosylation in normal conditions. This is corroborated by the observation of *O*-glycosylation of ER resident proteins such as calnexin in a Mass Spectrometry (MS) *O*-glycoproteome analysis (Steentoft et al., 2011). Relocation of GalNAcTs into the ER resulted in efficient *O*-glycosylation, further supporting that all components required for initiation of *O*-glycosylation are present in the ER, with the exception of GalNAcTs (Röttger et al., 1998).

Like other Golgi glycosyltransferases, GalNAcTs display a type II membrane structure: a short N-terminal cytoplasmic tail, a hydrophobic membrane-spanning domain, a stem region, and a catalytic domain (Figure I.2.3). They also have a ricin-like lectin domain linked to the catalytic domain by a linker of variable length. Several human isoforms have been crystalised (GalNAcT1 (Fritz et al., 2004), GalNAcT2 (Fritz et al., 2006; Lira-Navarrete et al., 2015), GalNAcT4 (de las Rivas, Daniel, et al., 2018), GalNAcT7 (Yu et al., 2019) and GalNAcT10 (Kubota et al., 2006)), providing a lot of information on the mechanisms controlling catalysis and specificity. The catalytic domain contains a conserved Mn^{2+} binding site composed of the AspXaaHis motif and another His (Asp224, His226, His359, numbers corresponding to GalNAcT2). The UDP-GalNAc donor is hydrogen bound by Arg208, Asp224, Gly308, Trp331, Gly332 and Glu334. A flexible loop from Arg362 to Ser373 controls the enzyme activity and its mutation inactivates the enzyme (de las Rivas, Coelho, et al., 2018; Fritz et al., 2006; Lira-Navarrete et al., 2014). In the open conformation, the enzyme is inactive but allows the reversible binding of UDP-GalNAc. Interaction of GalNAc with Trp331 stabilises the active closed-loop state of the enzyme. Following the transfer, the GalNAc-modified peptide leaves the active site, and the unrestrained loop can open to release UDP (Figure I.2.4, steps 1 and 6).

The catalysis mechanism is proposed as a front-face S_Ni (Figure I.2.4, steps 3 and 4) (Lira-Navarrete et al., 2014). The S_Ni reaction (substitution nucleophilic internal) is a nucleophilic reaction in which the bond configuration is preserved. First, the bond between the phosphate group and GalNAc is broken. This is assisted by the monosaccharide N-acetyl group and the formation of a hydrogen bond between the phosphate and the hydroxyl group acceptor. The oxygen on the hydroxyl of the acceptor then proceeds to a nucleophilic attack of the oxocarbenium ion formed on the C1 of GalNAc. Simultaneously, the proton from the acceptor group is transferred to the phosphate (Lira-Navarrete et al., 2014).

The ricin-like lectin domain is built from three repeated units (α, β and γ) that can each potentially bind GalNAc-modified Ser/Thr. Each repeat is characterised by the presence of GlnXaaTrp and CysLeuAsp motifs and six cysteine residues that form disulphide bridges in otherwise poorly conserved sequences (Wandall et al., 2007). The lectin domains studied so far have a rather high specificity for GalNAc and GalNAc-glycopeptides, but not for glycopeptides carrying elongated *O*-glycans (Bennett et al., 2012). In general, not all repeats are active GalNAc binding sites: GalNAcT1 has two (α and β) whereas GalNAcT2, T3 and T4 (α) and GalNAcT10 (β) have only one (de las Rivas et al., 2019; Tenno et al., 2002). The binding is made by contacts with Asp458, Tyr471, His474 and Asn479 in GalNAcT2 α repeat that

are conserved in almost all enzymes. Mutations in these residues results in repeats with little or no binding activity (Tenno et al., 2002). A flexible linker connects the lectin and catalytic domains. The length of this linker varies between different GalNAcTs and its functions is to reposition the lectin domain relative to the catalytic domain.

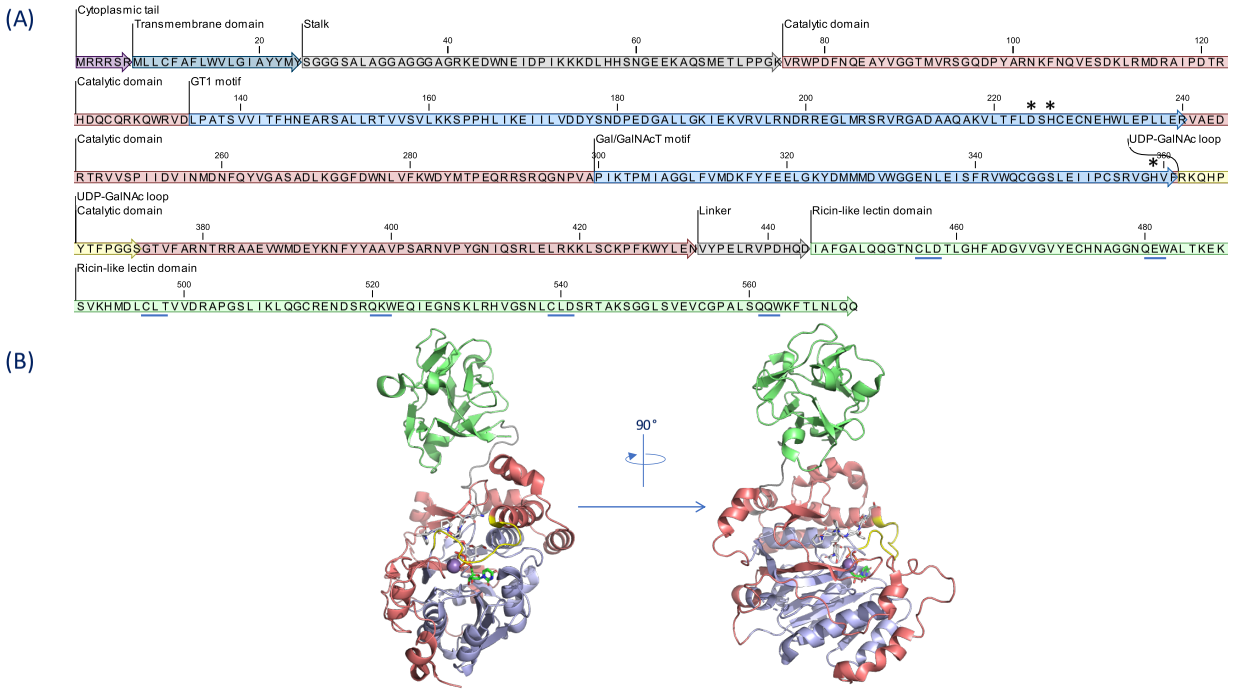


Figure I.2.3: Amino acid sequence and structure of GalNAcT2. (A) The different features of GalNAcT2 are indicated on its amino acid sequence. The residues, Asp224, His226, His359, that bind the Mn2+ ion are denoted with an asterisk and the lectin domain motifs are underlined. (B) Crystal structure of soluble GalNAcT2 in the closed state in complex with Mn^{2+}, UDP and AspSerThrThrProAlaPro (PDB: 4D0T), the different features follow the colour code of (A). The catalytic domain (red) possesses the GT1 motif implicated in the catalytic transfer of the glycan and the Gal/GalNAc motif implicated in the binding of the donor nucleotide-sugar (both in light blue). The flexible UDP-GalNAc loop is shown in yellow. The lectin domain (green) is connected to the catalytic domain by a linker (grey).

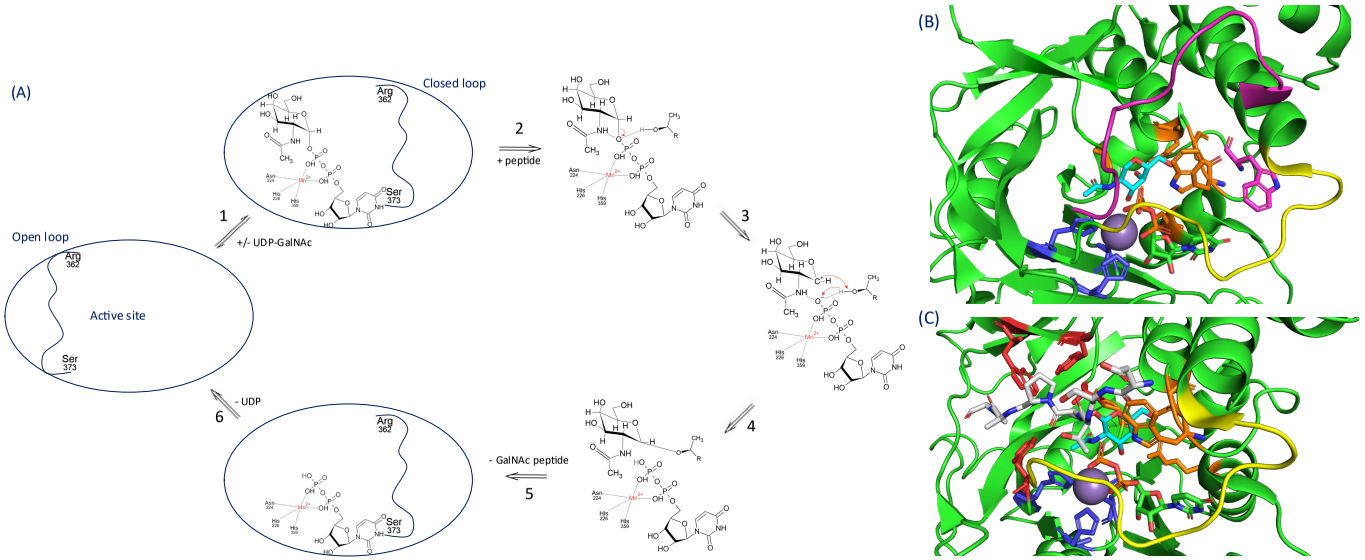


Figure I.2.4: Catalysis of the transfer of GalNAc to Ser/Thr by GalNAcTs. (A) Catalytic mechanism of GalNAc transfer by GalNAcTs. 1) The open loop conformation allows the reversible binding of UDP-GalNAc which stabilises the closed loop conformation by interaction with Trp331. 2), 3) and 4) The peptide binds to the active site and interactions with the phosphate oxygen lead to the rupture of the bond between UDP and GalNAc. The oxocarbenium ion formed on the C1 of GalNAc then proceed to a nucleophilic attack of the Ser/Thr oxygen which lead to the hydrogen rebound to the phosphate group. 5) and 6) The GalNAc-peptide exits the active site, and the unrestrained loop opens allowing release of UDP. (B) Aligned crystal structures of GalNAcT2 in open- (5AJO) and closed-loop (4D0T) conformations. (C) Crystal structure of the active site of GalNAcT2 complexed with UDP and DSTT(GalNAc)PAP glycopeptide. The full crystal structure of 4D0T is respresented in green cartoon with a yellow loop. Only the loop and Trp331 from 5AJO are shown in purple. UDP is shown as green sticks with orange phosphate groups, GalNAc as cyan sticks and the glycopeptide as white sticks. Residues interacting with UDP-GalNAc (Arg208, Asp224, Gly308, Trp331, Gly332 and Glu334) are represented by orange sticks. Mn^{2+} atom is represented as a grey sphere bound to Asp224, His226 and His359 in blue sticks. The hydrophobic pocket residues (Phe280, Trp282, Ala307 and Phe361) are shown in red sticks.

2.1.2 The multi-layered specificity of GalNAcTs

GalNAcTs have a broad substrate specificity and yet many of the isoenzymes act on specific proteins (such as GalNAcT3 and FGF23 or GalNAcT2 and angiopoietin-like protein 3 (Kato et al., 2006; Schjoldager et al., 2010)). This is also illustrated by the many diseases associated with defect in specific GalNAcTs (Magalhães et al., 2021). In fact, GalNAcT isoforms have distinct specificities towards peptides but also GalNAc-glycopeptides (Daniel et al., 2020). This is important because *O*-glycosylation occurs both on isolated and clustered sites. The subset of GalNAcTs present in each cell and their dynamic regulation can have an impact on the patterns of *O*-glycosylation and provide high heterogeneity in response to cellular cues during development and diseases. For instance, GalNAcT4 is the only enzyme capable of completing glycosylation of the mucin MUC1 after it has been initiated by GalNAcT1 or T2 (Hassan et al., 2000). In addition, the initial site of GalNAc attachment can direct the protein towards different *O*-glycosylation "pathways", resulting in different order of attachment to subsequent sites and possibly in a different final profile (Kato et al., 2001).

Many studies have used short synthetic peptides and GalNAc-glycopeptides to study the preferences of the different isoforms *in vitro*. The peptide designs were random (Daniel et al., 2020; Gerken et al., 2011, 2013; Perrine et al., 2009; Revoredo et al., 2016) or derived from known *O*-glycoproteins (Kong et al., 2015; Pratt et al., 2004; Wandall et al., 2007). This made it possible to identify some preferred amino acid sequences and to classify different subfamilies according to peptide or glycopeptide substrates. In the case of glycopeptides, the acceptor site can be located within 1-3 residues from the previous GalNAc (short-range) or 6-17 residues away (long-range) (Figure I.2.2) In general, GalNAcT1, T2 and T3 are the major contributors to peptides and the family II has the most restricted activity (Kong et al., 2015). Indeed, family IIa has mixed preference for close- and long-range glycopeptide, while family IIb has a strict preference for close-range glycopeptide (Revoredo et al., 2016). The specificity is regulated at several levels: by the catalytic domain, by the lectin domain or by the interaction between the two, which is mediated by the flexible linker. It should be noted that one study reports the association of GalNAcT2 in dimers that disappear after addition of substrate peptide (Lira-Navarrete et al., 2015). It remains to be determined whether these dimers are biologically relevant and if the formation of GalNAcTs homo- and heterodimers can regulate the specificity and the *O*-glycosylation process as a whole.

Firstly, the specificity is mediated by the catalytic domain. GalNAcTs show a preference towards Thr over Ser residues. This preference depends on each

isoenzyme with Thr/Ser preference ratios varying from 2 to 18 among the 12 enzymes studied (Daniel et al., 2020). The ratio is constant between close isoforms and between studies suggesting that it is an intrinsic characteristic of each enzyme. Structurally, a hydrophobic pocket (composed of Phe280, Trp282, Ala307 and Phe361 (Lira-Navarrete et al., 2014)) is able to accommodate the Thr methyl group acceptor. Analysing crystal structure in complex with the acceptor Thr and molecular dynamics simulations showed that the hydrophobic interaction and the steric hinderance are responsible for a more stable and better orientation of the Thr hydroxyl group for catalysis compared with Ser (Bennett et al., 2012; Daniel et al., 2020).

Analysis of GalNAcTs crystal structures in complex with a peptide acceptor has identified different substrate-selective residues. However, these residues vary considerably between isoforms leading to distinct peptide specificities, as shown by the *in vitro* peptide studies. Several GalNAcTs have a preference for the (Thr/Ser)ProXaaPro motif (where Xaa is usually a small hydrophobic residue such as Gly or Ala) and other sequences with Pro in -1 or +3 from the acceptor position (Bennett et al., 2012; Brockhausen et al., 2009; Gerken et al., 2011; W. Yang et al., 2018). This is explained by the hydrophobic pocket (Phe280 and Trp282 and Phe361) that can accommodate these residues. Consistently, all the isoforms that possess an equivalent pocket can modify these motifs (such as GalNAcT1, T2 or T3)(Bennett et al., 2012).

However, it is not the case for others such as GalNAcT7 and T10. These enzymes have a strong preference for *O*-GalNac containing residues at the +1 position (Perrine et al., 2009). This preference for glycopeptides is also shared by GalNAcT4 and T12, which modify substrates with *O*-glycans at the -1 and +3 position, respectively. This is corroborated by the GalNAcT4 crystal structure complexed with peptides, which shows several interactions (Thr283, Gln285, Lys366) with the modified GalNAc-Thr/Ser not found in other enzymes (de las Rivas, Daniel, et al., 2018). This GalNAc pocket is responsible for the short-range glycosylation. On the other hand, the presence of 1-3 residues neighbouring glycosylation is generally inhibitory for GalNAcT1 and T2 (Gerken et al., 2002; Perrine et al., 2009).

Secondly, the lectin domain also alters the specificities of GalNAcTs by mediating the GalNAc-dependent long-range glycosylation. Indeed, binding of glycopeptides to both the lectin and catalytic domains promotes catalysis. This promotes modification of otherwise inaccessible or unfavoured acceptor sites. This is corroborated by mutagenesis experiments on the lectin domain, which can reduce and even abrogate the activity of some isoforms on glycopeptides (Fritz et al., 2006;

Kubota et al., 2006; Raman et al., 2008; Wandall et al., 2007). Depending on the orientation of the lectin domain, each GalNAcTs has a C- or N-terminal long-range preference, or both (Gerken et al., 2013). The flexible linker controls this orientation of the lectin domain in GalNAcT2 and T4 (de las Rivas et al., 2017). This finding directly correlates the opposing orientation of the lectin domains with the N- or C-terminal glycopeptide activity of these enzymes. In addition, mutations or swapping the linker of GalNAcT2 with that of GalNAcT3 or T4 can modulate or even reverse the long-range preference (de las Rivas et al., 2017). The association of the lectin domain with the substrate could also favour the complete *O*-glycosylation initiation before elongation can begin, as evidenced by the intermediate GalNAc glycoforms of certain proteins (Sheehan et al., 2004).

Direct interactions between the lectin and catalytic domains also influence the activity on peptide substrates. For instance, GalNAcT10 has no activity on peptides but swapping its lectin domain with that of GalNAcT2 enables this activity (Kubota et al., 2006; Raman et al., 2008). Binding of the lectin domain of GalNAcT10 to GalNAc-glycopeptides is required for activity, as mutations in the lectin domain inactivate the enzyme (Kubota et al., 2006). These results suggest partial masking of the catalytic site by the unbound lectin domain. In DmGalNAcT9, a long-charged loop of the lectin domain can modulate the catalytic activity for charged substrates (Ji et al., 2018). The linker can influence these interactions between the lectin and catalytic domains. Indeed, in GalNAcT2 structure, the enzyme has been found in a compact or extended state (Lira-Navarrete et al., 2015). In addition, mutations in the linker alter the kinetic properties of GalNAcT4 (Lira-Navarrete et al., 2015).

Several reports also provide evidence of Tyr residues *O*-glycosylation (Kong et al., 2015; Steentoft et al., 2011). The enzymes responsible for this modification are still unknown. However, if this modification is carried out by GalNAcTs, it would raise several questions, namely how they can accommodate this acceptor that is very different from Thr/Ser and whether particular isoforms are responsible for this modification.

2.1.3 Elongation and capping

GalNAc-modified proteins are then the substrate for further enzymatic activities that elongate the glycan. This begins with the formation of four possible core structures by the addition of Gal or GlcNAc (Figure I.2.1). The most commonly observed structure is core 1 which is composed β(1-3)-Gal. The enzyme that catalyses the formation of core 1 is C1GalT1 (Granovsk et al., 1994; Ju et al., 2002). It requires a specific molecular chaperone termed COSMC to fold correctly and localise in the

Golgi (Ju & Cummings, 2002). Core 1 serves as the basis for formation of core 2, which is obtained by addition of β(1,6)-GlcNAc to the GalNAc residue. This step is operated by three C2GnT1-3 isoenzymes with distinct but overlapping activities (Schwientek et al., 2000; Yeh et al., 1999). Core 3 is formed by the transfer of β(1,3)-GlcNAc onto an unmodified GalNAc residue, and core 4 by the addition of a β(1,6)-GlcNAc on the core 3. These activities are mediated by C3GnT and C2GnT2, respectively (Büll et al., 2020; Yeh et al., 1999). Core structures 5-8 also exist and are composed of Gal, GlcNAc or GalNAc but are rarely encountered and their formation has not been studied.

The different core structures are then elongated by addition of β(1,4)-Gal or β(1,3)-GlcNAc by β4GalTs and β3GnTs, respectively (Brockhausen et al., 2006, 2009; Narimatsu, Joshi, Nason, et al., 2019; Schjoldager et al., 2020). The core structures or their elongated counterparts can be capped with sialic acids (ST6GalNAc, ST3Gal and ST6Gal)(Gill et al., 2011; Narimatsu, Joshi, Schjoldager, et al., 2019; Schjoldager et al., 2020) or more complex structures such as the Lewis antigens and ABO-blood groups (Hang & Bertozzi, 2005; Kudelka et al., 2015). Sometimes, sulphation or acylation of *O*-glycans is also encountered (J. Liu et al., 2015; Schjoldager et al., 2020). These enzymes are shared with the *N*- and lipid-glycosylation pathways.

The formation of complex *O*-glycans is highly regulated. The core forming enzymes each have their own specificities. For example, C1GalT1 has a broad specificity but show a slight preference for adjacent charged residues, while Pro negatively affects catalysis (González-Ramírez et al., 2022; Granovsk et al., 1994). Clustering of GalNAc residues also negatively influence their elongation, although elongated clustered *O*-glycans are sometimes found *in vivo* (*e.g.*, on mucins (Brockhausen et al., 2009; Granovsk et al., 1994)). In addition, most tissues express C1GalT and C2GnTs, resulting in a high abundance of core 1 and 2 *O*-glycans. In contrast, the expression of C3GnT, and thus the synthesis of cores 3 and 4, is more restricted, mainly in the mucins of gastrointestinal tissue (Gill et al., 2011). Finally, physical separation of the enzymes in the Golgi is one mechanism that allows the cell to orchestrate the activity of all these enzymes. Indeed, early acting enzymes (*e.g.*, C1GalT1 and C2GnTs) are enriched in cis- and medial-Golgi, while late acting enzymes (*e.g.*, β4GalT and sialyltransferases) are located in trans-Golgi (Gill et al., 2011).

2.2 Molecular functions of *O*-glycans

Glycosylation is involved in many physiological processes. It is generally accepted that glycosylation impacts the solubility and *in vivo* stability of proteins. *O*-glycans could participate in these functions, for instance by providing additional sialic acids

that are directly correlated with extended serum half-life of EPO (Castilho et al., 2012).

O-glycosylation has been implicated in protection against proteases. Indeed, densely *O*-glycosylated domains are often observed on secreted proteins. These *O*-glycans patches play a role in protecting proteins against proteases. This is abundantly demonstrated by mucins, which are very resistant to proteases and serve to protect cells against pathogens in mucosal surfaces (Daniel et al., 2020). This could also be the case for the extended hinge regions of certain antibodies which show *O*-glycosylation clusters (de Haan et al., 2020; Plomp et al., 2015). For example, *O*-glycosylation protects the hinge region of mouse IgG2b against proteases (H. Kim et al., 1994) and neuraminidase treatment enhanced the susceptibility of IgA1 to Streptococci proteases (Reinholdt et al., 1990). Likewise, this could explain why recombinant IgA1 lacking *O*-glycans have a short lifespan in serum (Göritzer et al., 2017) and the degradation observed for IgG3 in *N. benthamiana* leaves (Kallolimath et al., 2020).

However, isolated glycosites could also protect from and regulate processing by host proteases. This is nicely exemplified by the specific *O*-glycosylation of FGF23 by GalNAcT3, which prevents its cleavage by proprotein convertase (Kato et al., 2006). Similar results have been observed for other proteins such as angiopoietin-like protein 3 and G-CSF (Ramírez-Alanis et al., 2018; Schjoldager et al., 2010). Interestingly, a glycoproteomic study revealed that several of the GalNAcT3-specific sites are located in close proximity to protease cleavage sites (Schjoldager et al., 2015).

O-glycosylation could also play a role in controlling protein secretion. Indeed, KO of DmGalNAcT9 led to aberrantly glycosylated protein cargoes that affected the morphology and maturation of secretory vesicles (Ji et al., 2018). Mutation of EPO *O*-glycosylation site resulted in poor secretion of the protein (Dubé et al., 1988).

O-glycans may also have direct structural functions. Indeed, *O*-glycans and especially the negatively charged sialic acids, can impart steric and electrostatic repulsion that could stabilise the extended conformation of the protein backbone to which they are attached, such as IgA1 hinge region or unfolded mucin domains. Furthermore, the α linkage of GalNAc to the peptide confers unique structural features to the protein backbone (Hang & Bertozzi, 2005). In addition, in mucin, *O*-glycans act as water keeper allowing the formation of hydrogels on mucosal surfaces.

O-glycans also provide specific structures for molecular interactions and signal perceptions. For example, neutrophils are recruited to the site of inflammation by *O*-glycan binding to L- and P-selectins (Ellies et al., 1998a; Hang & Bertozzi, 2005), and

extended core 2 *O*-glycans regulate T-cell activation and susceptibility to galectin-1-mediated cell death (Cheng & Radhakrishnan, 2011; Nguyen et al., 2001; Tian & Ten Hagen, 2009). These interactions also mediate cell adhesion (Schwientek et al., 2000). *O*-glycans are used to modulate the interaction between proteins. For example, the human group-specific component protein is converted to a potent macrophage activating factor by the removal of Gal and NeuAc from its sialylated core 1 glycan (Nabeshima et al., 2020).

These functions highlight how cells can use the tight regulation of *O*-glycosylation mediated by the large GalNAcT family and downstream enzymes to regulate the fate and functions of glycoproteins.

The implications of *O*-glycans in cell physiology are well illustrated by their dysregulation in cancers. Different mechanisms in cancer cells lead to the formation of short, truncated *O*-glycans (core 1 and single GalNAc have been termed Tumor (T) and Tumor new (Tn) associated antigens) that are heavily sialylated (Burchell et al., 2018; Poiroux et al., 2017; Zlocowski et al., 2019). This has many consequences in cell physiology, with implications for cancer cell signalling and adhesion. For example, it modulates the immune response by binding to macrophage galactose-specific lectins and sialic acid-binding immunoglobulin-like lectins that induce T-cell death and produce signals that negatively regulate the immune system, respectively (Burchell et al., 2018). In addition, the abundance of *O*-glycans in several epithelial cancers is directly related to metastasis and invasiveness, resulting in poor clinical outcome (Radhakrishnan et al., 2014; Zlocowski et al., 2019). Consequently, *O*-glycans are good clinical markers for cancers and evaluation of aggressiveness (*e.g.*, switch from core 3-4 to core 1 glycans in gastrointestinal cancers) (Burchell et al., 2018; Kudelka et al., 2015). However, this also provides opportunities to use these antigens for therapeutic purposes (Pinho & Reis, 2015). Indeed, these cancer-specific antigens have raised the possibility to use them as vaccines as patients naturally produce anti-T and anti-Tn antibodies (Zlocowski et al., 2019). One promising strategy was to fuse the STn antigen to the keyhole limpet hemocyanin carrier protein (Julien et al., 2009). However, this drug, named Theratope, failed to prove its efficacy in phase III clinical trial. This failure put an end to the hype on Tn-based cancer vaccines. Some authors have suggested that differences in Tn profile between different patients, that was not monitored during the trial, could explain the inability to prove efficacy and that the field requires further investigation (Buskas et al., 2009; Dalziel et al., 2014; Julien et al., 2012; Zlocowski et al., 2019). Other groups have focused on targeting these glycoepitopes for passive immunisation. For instance, the

firm Glycotope uses a proprietary cell technology to produce cancer antigens and antibodies targeting glycopeptides epitopes for treatment (Kehler et al., 2022).

The complexity of *O*-glycosylation has made it a very difficult modification to study. Indeed, the lack of consensus sequence for initiation, the large variety of structures and the site clustering have hampered the classical analytical methods. As a result, the *O*-glycoproteome is not yet well understood both in terms of its extent and its functionality. In recent years, several tools developed specifically for the study of *O*-glycosylation have been investigated.

The SimpleCell technology was developed to reduce the high heterogeneity of *O*-glycans (Steentoft et al., 2011). When COSMC is knockout using Zinc Finger nuclease approach, the formation of core 1 and 2 is abolished, leading to a *O*-glycoproteome mainly composed of (sialylated-)GalNAc. This greatly simplifies downstream analysis of *O*-glycans, such as lectin-purification enhanced MS. Using this method, Steentoft et al. (2011) analysed the *O*-glycoproteome of many cell lines derived from different human tissues, allowing identification of new *O*-glycoproteins and new sites on known *O*-glycoproteins. In addition, the use of new bacterial derived specific *O*-glycoproteases circumvented the difficulties associated with the low efficiency of traditional proteases (e. g., trypsin) on highly *O*-glycosylated peptides, allowing the development of new *O*-glycans specific MS pipelines (Mulagapati et al., 2017; Taleb et al., 2022; S. Yang et al., 2018; W. Yang et al., 2018).

2.3 *O*-glycoengineering of pharmaceutical systems

Due to the implications of *O*-glycans in modulating the activity and stability of glycoproteins, it is important that pharmaceutical *O*-glycosylated proteins have the appropriate *O*-glycosylation profile for their purpose. Thus, heterologous production systems should be able to perform tailored *O*-glycosylation of therapeutics. In addition, these systems must also achieve a reproducible homogeneity as required by regulatory agencies (P. Zhang et al., 2013). At present, industrial production systems do not meet these requirements and produce pharmaceuticals whose *O*-glycosylation profile is neither human nor optimised for enhanced activity of the pharmaceuticals (Strasser, 2013).

In addition, there is also a fundamental interest in generating glycoproteins with defined *O*-glycosylation sites and structure. Indeed, this could provide scientists with research and diagnostic tools to elucidate the function of particular *O*-sites or structure in the complex and heterogenous context of the native protein.

Engineering of human *O*-glycosylation has mainly focused on the fundamental study of the *O*-glycoproteome. As mentioned above, the first major step was the generation of the SimpleCell technology in 12 human cell lines, which considerably reduced the heterogeneity of *O*-glycans (Steentoft et al., 2011, 2013). These studies were followed by the KO of several GalNAcTs (T1, 2, 3, 7, 10 or 11 (Narimatsu, Joshi, Schjoldager, et al., 2019)) or their overexpression (OE) (T1, 2 or 3 (Schjoldager et al., 2015)) in HEK293 SimpleCells. The authors continued this work by KO or OE of core forming enzymes, elongating galactosyltransferases and sialyltransferases. In total, this group has generated more than 100 glycoengineered cell lines (also including other glycosylation types) to study interactions of glycans, for example with influenza HA, or to study the glycosylation of mucin reporters (Konstantinidi et al., 2022; Narimatsu, Joshi, Nason, et al., 2019; Nason et al., 2021). This work has made possible to generate relatively homogeneous *O*-glycosylation profiles and to considerably broaden the knowledge of *O*-sites in the human proteome but were not used with the intention to produce pharmaceutical products.

In the industry, the standard CHO cells express only a subset of four GalNAcT isoenzymes (T2, 7, 11 and 19) and predominantly possess sialylated core 1 *O*-glycans (Bennett et al., 2012; Z. Yang, Bennett, et al., 2012). Only a few studies report the engineering *O*-glycosylation in CHO cells. The SimpleCell technology was applied to CHO cells and used for expression of EPO that displayed the expected truncated *O*-glycan (Z. Yang et al., 2014). Another group used transient overexpression of core forming GTs (C2GnT, C3GnT) with downstream enzymes (β3GnT or blood group related FucTs, GalT and GalNAcT) to generate specific *O*-glycans on P-selectin glycoprotein ligand-1/mouse IgG2b Fc. They reported successful generation of the desired structures albeit with varying efficiencies and high heterogeneity (Lindberg et al., 2013; J. Liu et al., 2015). Finally, overexpression of ST6GalNAcT1 or C3GnT in a *cosmc* KO CHO cell line allowed the generation of sialylTn or core 3 *O*-glycans on Etanercept (Biel et al., 2022). This FDA approved drug has 3 *N*- and 13 *O*-glycosylation sites. The glycovariants had different *in vitro* binding affinities and *in vivo* potencies. Notably, the authors report that *O*-glycoengineering had no impact on the *N*-glycosylation profile of the protein.

The *O*-glycosylation machinery, consisting of hGalNAcT1, 2 or 3 and DmC1GaLT1 with *Bacillus subtilis* UDP-Gal/GalNAc 4-epimerase and human UDP-Gal/GalNAc transporter, was also introduced into *S. cervisiae* and found to be functional (Amano et al., 2008). However, in yeast, GalNAc-*O*-glycosylation compete with *O*-mannosylation for Ser and Thr acceptors. While this PTM can be reduced by gene

KO or with chemical inhibitors, a complete KO of *O*-mannosyltransferases is lethal (Strahl-Bolsinger et al., 1999).

Following the remarkable advances in engineering human-like *N*-glycosylation in plant systems (Castilho et al., 2010; Strasser, 2023), several groups have undertaken the task of *de novo* introduction of GalNAc-*O*-glycosylation in plant cells. This also has the advantage of reducing undesirable heterogeneity as there is no competition with endogenous enzymes since this pathway is completely absent in plants. In addition, this could also provide a tool to study the effect of specific glycoforms on protein properties.

Initial studies tested the ability of plants to generate *O*-glycans through stable integration of the *GALNT2* coding sequence in *N. benthamiana* (Daskalova et al., 2010). The localisation in the Golgi apparatus was verified by fusion with a fluorescent protein. Modification of endogenous proteins with GalNAc was observed, suggesting that GalNAcT2 is sufficient for introducing *O*-glycosylation in plant cells. For efficient glycosylation, the presence of activated donor nucleotides, as well as the substrate and enzyme, must be tightly regulated. While GalNAc has never been demonstrated in plant protein glycans, there is evidence of accumulation of UDP-GalNAc in plant tissues and polysaccharides (Alonso et al., 2010). In addition, the UDP-GlcNAc transporter ROCK1 from Arabidopsis was shown to have UDP-GalNAc transport activity in the ER/Golgi (Niemann et al., 2015). Although these corroborate the initial findings of Daskalova et al. (2010), overexpression of heat labile toxin subunit B of *E. coli* with MUC1 (LTB-MUC1 which has 5 *O*-sites) using the MagnICON viral system led to poor glycosylation of this substrate. Complete *O*-glycosylation of this substrate was only obtained when a UDP-GlcNAc/GalNAc C4-epimerase from *Yersinia enterocolitica* and a UDP-GalNAc transporter from *Caenorhabditis elegans* were cointegrated with GalNAcT2 (Daskalova et al., 2010).

Later studies by another group have confirmed that GalNAcT2 is not sufficient for efficient glycosylation of MUC1-YFP in transient expression in *N. benthamiana* and stably transformed BY-2 cells and Arabidopsis plants (Z. Yang, Bennett, et al., 2012; Z. Yang, Drew, et al., 2012). However, they disagree on the need for a UDP-GalNAc transporter and show that GalNAcT2 and a *Pseudomonas aeruginosa* UDP-GlcNAc/GalNAc C4-epimerase were sufficient for glycosylation of MUC1-YFP substrate (5 sites). On the contrary, a following study shown that transient expression of GalNAcT2 only was sufficient for efficient glycosylation of EPO (1 *O*-site) in *N. benthamiana* (Castilho et al., 2012). The discrepancies between these studies are puzzling. It could be partly explained by differences in enzyme expression levels and

the number of *O*-sites on the reporter proteins. In addition, the different expression systems used, transient or stable expression of the glycosylation machinery and different construct design, make the comparison more difficult. Of note, the use of MagnICON led to the necrosis of the leaves, which could induce stress on cells and affect glycosylation.

MS analysis revealed that 3 sites out of 5 were occupied when MUC1-YFP was expressed with GalNAcT2 in *N. benthamiana,* which is in accordance with GalNAcT2 activity on this substrate (Hassan et al., 2000; Z. Yang, Drew, et al., 2012). Co-expression of GalNAcT4 resulted in the complete glycosylation of the 5 sites (Z. Yang, Drew, et al., 2012). The ability of plants to produce pharmaceuticals with complex *O*-glycans was further developed by generation of sialylated core 1 *O*-glycans. Co-expression of GalNAcT2, DmC1GalT1 and sialyltransferases ST3GalI and ST6GalNAcIII/IV along with CMP-NeuAc biosynthetic pathway led to the efficient production of mono- and disialylated core 1 *O*-glycans on EPO (Castilho et al., 2012). Other proteins were tested for the *O*-glycosylation capacity in *N. benthamiana* leaves including MUC16 (approximately 30 sites) and interferon α 2b (1 site) (Z. Yang, Drew, et al., 2012), G-CSF (1 site)(Ramírez-Alanis et al., 2018) and IgA1 (3-5 sites) (Dicker et al., 2016). In all these experiments, the authors reported no obvious phenotypes for the plants transformed with the *O*-glycosylation machinery.

3 Plant Hydroxyproline *O*-glycosylation

The plant cell wall is a very important organelle for plant development and cell integrity. It withstands the high turgor pressure of plant cells and confers rigidity while, at the same time, being sufficiently extensible to allow growth (of sometimes fast-growing cell such as root hairs). It is also the place for signalling between cells and sensing environmental cues. It is therefore not surprising that in order to achieve and regulate all of these functions, 10 to 15% of the genome of *Arabidopsis thaliana* is devoted to the building, maintenance, and signalling functions of the cell wall (Decker et al., 2014).

The cell wall is mainly composed of polysaccharides, with a minor proportion of proteins (5-10%) (Y. Zhang & Showalter, 2020). The Cell Wall Proteins (CWP) are dedicated to cell wall support, structure, signalling, and interactions with other components (Silva et al., 2020). Among the CWPs, the superfamily of Hydroxyproline Rich GlycoProteins (HRGPs) is particularly abundant. Interestingly, the cell wall of the unicellular algae *Chlamydomonas reinhardtii* is almost entirely made of around 30 extensin-like HRGPs (Keskiaho et al., 2007; Saito et al., 2014). HRGPs are implicated in virtually every aspect of plant growth and development, from cellular processes such as mitosis and cell expansion to organ-specific and systemic mechanisms such as reproduction or responses to biotic and abiotic stresses (Petersen et al., 2021).

HRGPs have characteristics of intrinsically disordered proteins (IDPs): their amino acid composition is biased towards polar and charged residues and Pro (Johnson et al., 2017; Silva et al., 2020). They do not conform the structure-function relationship usually accepted for proteins. The absence of structural constraints has allowed IDPs to evolve rapidly, leading to little sequence conservation, with the exception of specific motifs, high indel events and domain swapping (Johnson et al., 2017). The IDPs are ideally suited to functions that require transient molecular recognition, and have been associated with stress response, signalling and molecular recognition pathways in plants (Pietrosemoli et al., 2013). HRGPs are commonly divided in three classes: arabinogalactan proteins (AGP), extensins (EXT) and proline-rich proteins (PRP). This classification is based on several conserved motifs that give rise to differences in PTMs. However, many HRGPs mix motifs from the different subclasses (hybrid HRGPs), or other functional domains form other protein families (chimeric HRGPs). A first genome-wide identification of HRGPs in Arabidopsis was based on the biased amino acid composition (>50% Pro, Ala, Ser, Thr) and motif search for $SerPro_n$ (X. Liu et al., 2020; Showalter et al., 2010). A total of 166 HRGPs,

including 85 AGPs, 59 EXTs, 18 PRPs and 4 AGP/EXT hybrids, were identified in Arabidopsis using this method. Johnson et al. built another bioinformatic pipeline based on amino acid bias and motif counting to classify HRGPS into 24 subgroups (excluding chimeric HRGPs) (Johnson et al., 2017).

Another hallmark of HRGPs is their extensive post-translational modifications. Their Pro rich sequons are targets for plant prolyl-4-hydroxylases, and hydroxyproline is the site of subsequent glycosylation. *O*-glycans can account for up to 95% of the molecular weight of AGPs, whereas EXTs are more moderately glycosylated. PRPs are more varied, from heavily to non-glycosylated. The glycosylation of HRGPs can be partly predicted using the Hyp contiguity hypothesis (N. Zhang, Wright, Wang, et al., 2019). This is based on the assumption that the HRGP glycosylation motifs are small functional units within the disordered proteins.

At the molecular level, the exact functions of HRGPs remain elusive. Their role as structural components, both in strengthening and plasticising the cell wall, has been largely documented (Hijazi, Velasquez, et al., 2014; Petersen et al., 2021). They form crosslinks with other cell wall components and take part in the remodelling the cell wall polymer network (Hijazi, Roujol, et al., 2014). They could also serve as signalling molecules and are able to bind extracellular calcium, which could be important for Ca^{2+} signalling (Y. Zhang & Showalter, 2020). They can also interact with pathogenic and non-pathogenic microorganisms (Nguema-Ona et al., 2014; Showalter, 1993). Due to the predominance of *O*-glycans on HRGPs, the oligosaccharides are thought to be very important for these functions, although understanding at the molecular level is still poor. The first step of plant *O*-glycosylation is the hydroxylation of proline.

3.1 Proline hydroxylation

Proline hydroxylation is a widespread PTM. The third and fourth carbons can be the site of the hydroxylation which leads to the formation of a chiral centre (in addition to the chiral centre already found on C2). Thus, 3- and 4-hydroxyproline both possess four stereoisomers that are found in nature. In animals, trans-4-hydroxy-L-proline is the most abundant and mainly found in collagen and structural proteins, but this modification also serves as a regulation mechanism for the activity of some proteins such as the hypoxia inducible factor transcription factors (HIFs). Trans-3-hydroxy-L-proline is also found in collagen but to a much lesser extent. Trans-4-hydroxy-L-proline is also present in plants mainly in HRGPs, but also in several other proteins as described below. The other isoforms are mainly found in bacteria and fungi, where

various roles such as signal transmission, protein regulation and antibiotic activity, have been described (Hu et al., 2022; B. Wang et al., 2023).

In plants, Prolyl-4-Hydroxylases catalyse hydroxylation of prolines in HRGPs and probably in many other peptides, such as secreted peptide hormones. However, P4Hs are found in all kingdoms from bacteria to animals. The most extensively studied enzymes are those found in humans, where proline hydroxylation is the most common PTM (Gorres & Raines, 2010).

The prolyl-4-hydroxylases (P4Hs) belong to the non-heme iron (II) 2-Oxoglutarate-Dependent Oxygenases (2ODOs), which is the second largest enzyme superfamily in plants after the cytochrome P450 monooxygenases. The enzymes in this family are widely distributed in nature and catalyse diverse types of reactions, including hydroxylation, in a variety of highly specialised pathways (Hagel & Facchini, 2018). These specialisations have led to an overall low sequence conservation. However, the catalytic mechanism and binding of 2-oxoglutarate (2OG) (and the involved residues) are remarkably well conserved and occupy a similar position in a characteristic β-barrel jelly roll fold composed of eight beta sheets (Figure I.3.1).

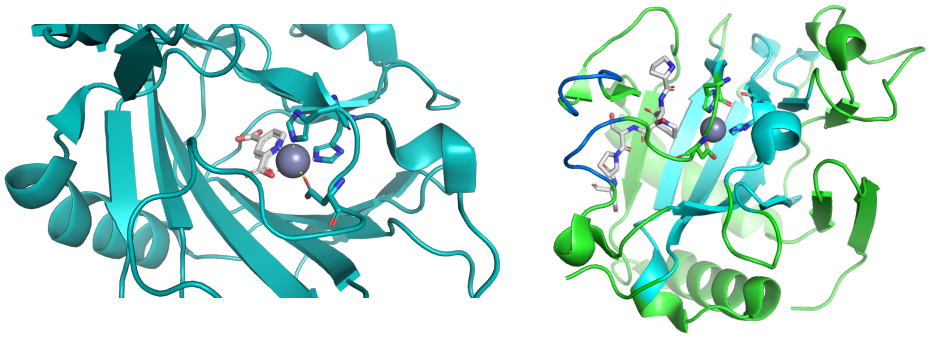


Figure I.3.1: Crystal structures of *C. reinhartdii* P4H1. The active site of CrP4H1 complexed with Zn^{2+} and pyridine 2,4-dicarboxylate, which replace Fe^{2+} and 2-oxoglutarate respectively (PDB: 2JIG (Koski et al., 2007)) (A). The three catalytic residues of the $HXDX_nH$ motif are represented in teal sticks and pyridine 2,4-dicarboxylate in white. These residues coordinate octahedrally the zinc atom (dark grey). CrP4H1 complexed with Zn^{2+} and $(SP)_5$ peptide (PDB: 3GZE (Koski et al., 2009)) (B). The three catalytic residues of the $HXDX_nH$ motif are shown as sticks. The residues implicated in the jellyroll motif are coloured in cyan and the two loops containing substrate binding residues in blue. The $(SP)_5$ peptide is shown in white.

P4Hs catalyse the irreversible formation of (2S,4R)-4-hydroxyproline (Hyp). Like other 2ODOs, P4Hs use a high-valent iron-oxo intermediate to initiate this reaction

(Proshlyakov et al., 2004). 2ODOs require molecular oxygen O_2 and 2OG as co-substrates, and iron (II) as cofactor, and further release CO_2 and succinate. The conserved active site is composed of the HisXaaAspXaa$_n$His motif that coordinates the iron (II) (Figure I.3.1) (Gorres & Raines, 2010; Hagel & Facchini, 2018).

The catalytic reaction, proposed by (Martinez & Hausinger, 2015), is as follows (Figure I.3.2): the enzyme in the resting state coordinates octahedrally the Fe^{2+} cofactor with the three key residues of HisXaaAspXaa$_n$His and three water molecules within the active site. 2OG then replaces two water molecules and binds Fe^{2+} in a bidentate configuration. The substrate, a polypeptide containing Pro for P4Hs, binds in the active site and displaces the last water molecule. O_2 then enters the active site and is activated by formation of a superoxo Fe^{3+} intermediate, which attacks and decarboxylates 2OG to CO_2 and succinate. CO_2 exits the catalytic pocket leaving a reactive Fe^{4+}-oxo intermediate, which creates a radical on the proline by attacking the hydrogen atom on the C4. Hydroxyproline is then generated by hydroxyl radical rebound.

The addition of oxygen to proline has a profound influence on the chemical and structural properties of this amino acid. Stretches of proline residues adopt a particular helical conformation, termed polyproline II helix, often observed in plant HRGPs, which is stabilised by conversion to hydroxyproline (Gorres & Raines, 2010). In addition, this provides a reactive oxygen atom that is the site for the plant-specific Hyp-*O*-glycosylation.

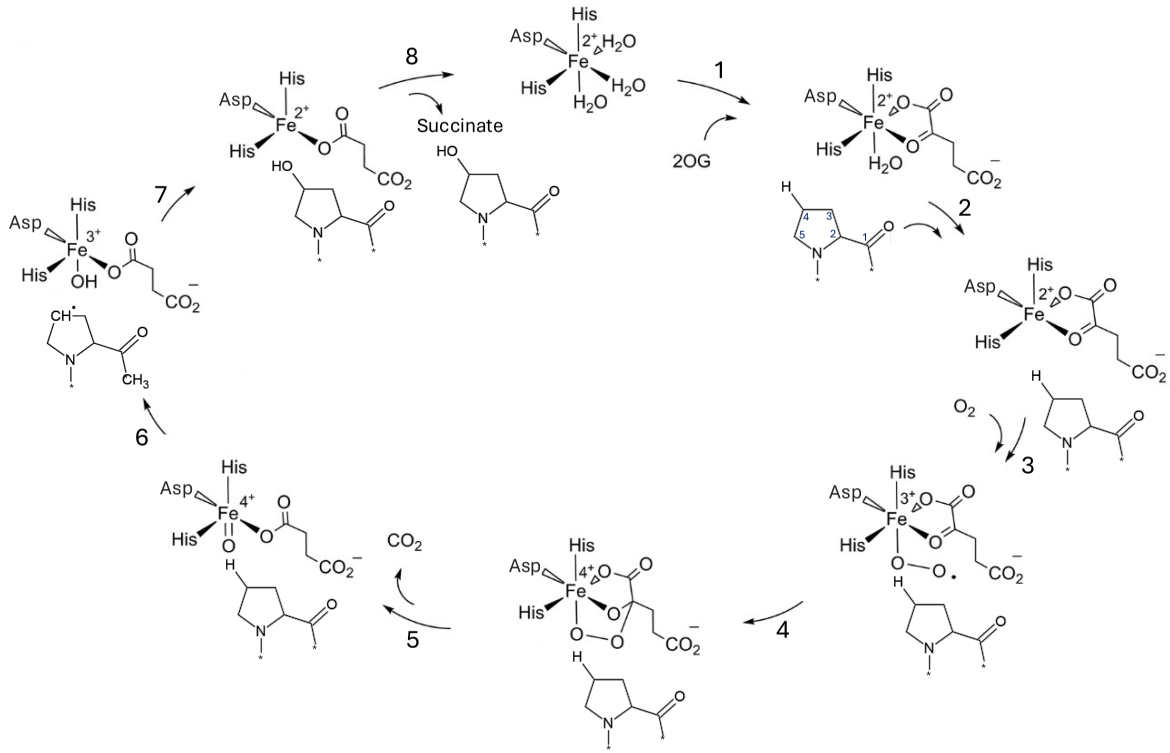


Figure I.3.2: Catalytic mechanism of hydroxylation by 2ODOs proposed by Martinez & Hausinger (2015)**.** Starting from the central upper state: 2OG, the substrate and O_2 bind in the active site successively (steps 1, 2 and 3). Decarboxylation of 2OG will create a reactive iron-oxo species (steps 4 and 5) that attacks the hydrogen of the substrate leaving a radical (step 6). The hydroxylated product is generated by radical rebound (step 7). The product and succinate leave the catalytic site and the enzyme is ready for the next catalysis (step 8).

3.1.1 P4Hs in humans

Three types of P4Hs have been described in humans. The first type, responsible for the high abundance of Hyp in humans, are the collagen CP4Hs that form $\alpha_2\beta_2$ tetramers. The α subunit is solely responsible for the P4H activity (catalysis and substrate binding) while the β-subunit functions independently as protein disulphide isomerase. The enzyme localises in the ER where it mainly hydroxylates the XaaProGly triplets of collagen (with a preference for Pro in the Xaa position), with an increase affinity for longer repetition of this triplet. Hydroxyproline residues are very important for the formation and stability of the collagen triple helix but are never glycosylated. The enzymes can bind polyproline II helix but are unable to modify it (Gorres & Raines, 2010).

The second type is involved in oxygen sensing by hydroxylation of the HIFs that maintain oxygen homeostasis (Konkina et al., 2021). They are similar to the α subunit of CP4H, except for their localisation in the cytosol. The P4H Domain Proteins (PHD) can sense the cytosolic concentration of its co-substrate, O_2. Hydroxylation of a single proline residue in LeuXaaXaaLeuAlaPro sequences of HIFs serves as a signal for their proteasomal degradation (Tiainen et al., 2005).

Finally, there is also evidence of the existence of a P4H with a transmembrane domain that localises in the ER. Its functions and substrates are currently unknown, but mutants were associated with neurological syndromes (Myllykoski et al., 2021).

3.1.2 P4Hs in plants

P4Hs have been identified and characterised in many plant species. They belong to a multigene family. Several members have been identified in the genome of *A. thaliana* (13 (Velasquez et al., 2011)), *Solanum lycopersicum* (13 (Kutyrieva-Nowak et al., 2024)), *N. benthamiana* (11 (Mócsai et al., 2021)), *C. reinhardtii* (10 (Keskiaho et al., 2007), *P. patens* (6 (Parsons et al., 2013) and *Dianthus caryophyllus* (2 (Vlad et al., 2010)) (Figure I.3.3). All P4Hs localise in the ER and/or the Golgi apparatus (Mócsai et al., 2021; Parsons et al., 2013; Velasquez, Ricardi, et al., 2015; Yuasa et al., 2005). AtP4H2 relocalises from the ER to the Golgi upon formation of heterodimers with P4H5 and 13, both of which contain a ArgLysXaaArgLys motif at the N-terminus, probably responsible for their localisation (Velasquez, Ricardi, et al., 2015). In green algae, all P4H members are soluble while higher plants have both soluble and membrane proteins (Hieta & Myllyharju, 2002; Keskiaho et al., 2007; Parsons et al., 2013; Yuasa et al., 2005). Studies on a NtP4H9 cytoplasmic tail and stalk domain revealed several targeting determinants that could allow recycling between the ER and the Golgi

(Yuasa et al., 2005). Purified recombinant AtP4H1 and 2, CrP4H1 and NbP4H1, 4, 9 and 10 were used in *in vitro* experiments (Hieta & Myllyharju, 2002; Keskiaho et al., 2007; Mócsai et al., 2021; Tiainen et al., 2005). This confirmed the requirement for the conserved catalytic residues and for cofactors/co-substrates for activity. Several plant P4Hs also possess a C-terminal cysteine-rich ShK toxin domain, the function of which is unknown (Fragkostefanakis et al., 2014; Tiainen et al., 2005). This domain has no influence on the P4H activity of the enzyme *in vitro* (Keskiaho et al., 2007). ShK domain is known as a toxin homology domain and has also been found in different proteins with various function such as proteases or peroxidases, for instance in *C. elegans* (Dauplais et al., 1997).

Mutations of any residue from the HisXaaAspXaa$_n$His motif (His180, Asp182 and His260) to Ala or Glu completely inactivate AtP4H1 (Hieta & Myllyharju, 2002). Another conserved histidine, probably involved in binding the C-1 carboxyl group of 2OG to the Fe atom and decarboxylation, is sometimes replaced by an arginine in plant P4Hs (Arg278 in AtP4H1). This substitution also exists in Drosophila and *Paramecium bursaria* Chlorella virus 1 P4Hs (Hieta & Myllyharju, 2002). In addition, mutation to alanine of residues implicated in 2OG or substrate binding (Lys270, Ser272 and Arg278) also inactivates the enzyme or strongly reduces its activity (Hieta & Myllyharju, 2002). Based on crystallographic structural data of CrP4H1, Koski et al. (2007) identified Tyr134 and Tyr168 as other 2OG binding residues. Their mutation to phenylalanine strongly reduced the catalytic activity and affinity for 2OG. Nine putative peptide binding residues were identified (Arg93, Ser95, Trp99, Glu127, Gln130, Tyr140, Arg161, Trp243, His245), which strongly or totally decreased activity when mutated. Using another structure complexed with a (SerPro)$_5$ peptide, Koski et al. (2009) showed the implication of several residues in two loops between the β sheets of the jelly roll motif, in the binding of the peptide substrate (Figure I.3.1 B). Mutation of these residues also strongly reduced the activity.

Proline hydroxylation in plants is mainly related to the creation of new *O*-glycosylation sites on proteins, that are important for their functions in the cell wall and signalling pathways. Accordingly, they are coregulated, along with the appropriate glycosyltransferases, during developmental stages (e.g., during fruit ripening (Kutyrieva-Nowak et al., 2024) and root development (Velasquez, Ricardi, et al., 2015)) or in response to stresses (e.g., during hypoxia (Konkina et al., 2021; Vlad et al., 2007), wounding (Salvá & Jamet, 2001; Vlad et al., 2007) and salt stress (Azariadis et al., 2023)). In addition, P4Hs could also be involved in iron sensing and deficiency responses (Vigani et al., 2013).

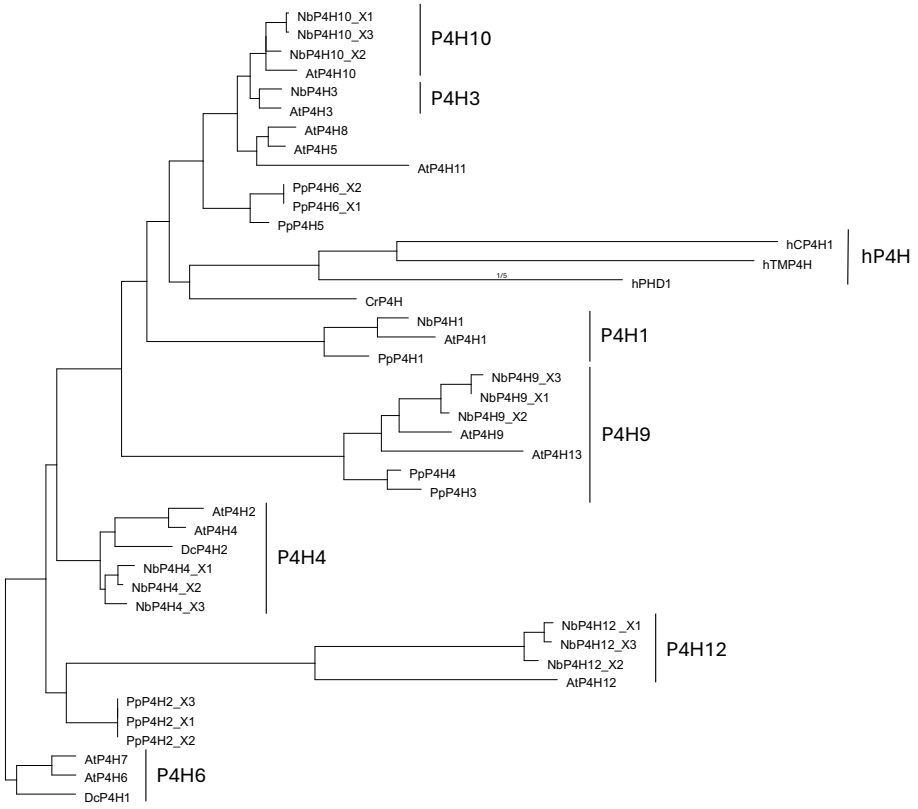


Figure I.3.3: Phylogenic tree of plant and human P4Hs. PhyML tree inference of MAAFT aligned plant and human P4H amino acid sequences. Clusters with different species were grouped and named according to *A. thaliana* and *N. benthamiana* homologs. The human PHD1 branch was shortened (1/5) for representation.

3.1.3 Plant proline hydroxylation code

Plant P4Hs can bind polyproline and hydroxylate this substrate, which is a competitive inhibitor of CP4H (Hieta & Myllyharju, 2002). Initially, it was believed that the polyproline II conformation was the determinant for recognition by plant P4Hs, but other peptides could be also modified. A search for a consensus sequence for Pro hydroxylation thus began and led to several observations.

The first Pro hydroxylation code was described by Shpak et al. (1999) and was intimately linked to the glycosylation status and clustering of Pro residues. It was termed Proline Contiguity Hypothesis and stated that Pro in contiguous stretches $SerPro_n$ found in EXTs are modified and glycosylated with short arabinose chains,

while clustered non-contiguous Pro in $(XaaPro)_n$ found in AGPs are hydroxylated and glycosylated with arabinogalactan glycans.

However, hydroxylation also occurs on isolated Pro residues, suggesting that other consensus motifs must exist. Using site directed mutagenesis of residues around the single hydroxyproline residue of sporamin, the sequence [AVSTG]-Pro-[AVST]-[GAVPSTC]-[APSED] (where Pro is the modification site) was found to be required for hydroxylation of this peptide in BY-2 cells (Shimizu et al., 2005). This study also showed that the more stringent motif [AVSG]-Pro-[AVST]-[GAVPSTC]-[APS] is necessary for correct arabinogalactosylation of the protein.

Canut et al. (2016) described the Extended Pro hydroxylation code that includes HRGPs, allergens, systemins and other hydroxylated plant protein sequence data that have been collected in 25 plant species. Briefly, Pro residues can be hydroxylated when they are located after Ala, Gln, Hyp, Pro, Ser, Thr, and Val residues, whereas Pro residues following the other amino acids could not. However, the authors observed many exceptions to this rule, which they could not correlate with any of variables they tested. While it is efficient at predicting Hyp in classical AGP and EXT, it fails for other HRGPs and other proteins with low Pro content, for which the rules established for sporamin also failed (Duruflé et al., 2017). In addition, the code does not consider tissue- or species-specific motifs. To date, no general consensus sequence for Pro hydroxylation has been identified, one of the difficulties being that plant P4H families are big and of misunderstood specificity.

3.1.4 Specificities of plant P4Hs

The size of the P4H family in plants suggests a high functional redundancy between the genes. However, little work has been done to identify the molecular targets of each P4H. Nevertheless, several evidence point to the specificity of P4Hs.

AtP4H1 and 2 show clear preferences to different substrates *in vitro* (Hieta & Myllyharju, 2002; Tiainen et al., 2005). While both enzymes hydroxylated efficiently polyproline substrates, AtP4H1 can hydroxylate, albeit inefficiently, collagen-like and HIF-like peptides, whereas AtP4H2 cannot. In addition, AtP4H1 prefers the second Pro in collagen-like peptide $(ProProGly)_n$ unlike CrP4H1 which preferentially hydroxylates the first Pro (Hieta & Myllyharju, 2002; Keskiaho et al., 2007). AtP4H1 and 2 and CrP4H1 are able to hydroxylate AGP- and EXT-like peptides derived from their own species. Two P4Hs from *D. caryophyllus* were also characterized and showed similar activity towards polyproline substrates but not towards EXT and AGP substrates *in vitro* (Vlad et al., 2010).

In Arabidopsis roots, AtP4H5 is required for the correct hydroxylation of EXT $SerPro_4$ motifs in which it modifies Pro1, 2 and 3 sequentially, but never Pro4 (Velasquez, Ricardi, et al., 2015). AtP4H2 and 13, also required for complete hydroxylation of $SerPro_4$ in roots, are thought to complete the hydroxylation of the last proline residue. This is corroborated by the fact that AtP4H2 only hydroxylates the last Pro in Pro_3 motif (Tiainen et al., 2005). The somewhat low efficiency observed could be due to a requirement of Hyp in the preceding amino acids. Accordingly, AtP4H5 could serve as an initiating enzyme on EXT sequences. Indeed, *p4h5* mutant displays the same phenotype as the triple *p4h2,5,13* mutant, which is more severe than the single *p4h2* or *p4h13* mutants (Velasquez, Ricardi, et al., 2015). Consistently, AtP4H5 had a higher expression than the other two partners in roots. In addition, AtP4H5 preferentially hydroxylates EXT substrates *in vitro*, even in presence of four-fold higher concentration of AGPs.

P4H mutants display various phenotypes in plants. For example, in *S. lycopersicum*, silencing of *P4H1* resulted in smaller leaves, while silencing of *P4H7* or *P4H9* produced bigger leaves (Fragkostefanakis et al., 2014). *C. reinhardtii* has a defective cell wall when *P4H1* is silenced (Keskiaho et al., 2007). This indicates that the other P4Hs are not able to fully compensate the absence of some isoenzymes that may have distinct substrate specificities. In contrast, silencing of several *P4H* subfamilies in *N. benthamiana* has no obvious phenotype (Mócsai et al., 2021), but the aim of this study was not plant physiology, but the production of recombinant proteins. Thus, subtle or conditional phenotypes at the cellular level may have been missed. In general, *P4Hs* show differential expression (Fragkostefanakis et al., 2014; Mócsai et al., 2021), which further suggests that organ-specific targets could be affected by a specific subset of P4Hs.

In *N. benthamiana* P4Hs have similar activity but show weak substrate and site preferences. Using four different peptides with multiple Pro in an *in vitro* assay, Mócsai et al. (2021) showed that NbP4H1, 4, 9 and 10 could modify all peptides (except for NbP4H4 and EPO peptide), but with different efficiencies. However, they showed clear preference regarding the site of hydroxylation. RNAi co-silencing of *NbP4H1* and *NbP4H10* led to the greatest reduction of Hyp but not complete abolition; surprisingly, this also reduced the number of pentoses attached to the remaining Hyp (Mócsai et al., 2021). CRISPR/Cas9 KO of the four *P4H4* genes that are strongly expressed in leaves led to a reduction of the number of Hyp on an IgA1 hinge peptide but, unexpectedly, the number of unmodified peptides was also reduced (Uetz et al., 2022). The authors speculated that this may be due to a rebalancing by other P4Hs enzymes. In *N. tabacum* NtP4H9 was unable to

hydroxylate sporamin *in vitro* (Yuasa et al., 2005), indicating that other P4Hs are responsible for this modification in BY-2 cells (Shimizu et al., 2005).

The expression level of each P4H isoform could also affect how hydroxylation is regulated. In a *P4H1* KO *P. patens* line, overexpression of PpP4H1 was able to restore proline hydroxylation of EPO and increased the percentage of modified Hyp. It also led to the conversion of an additional proline that was not modified in the WT (Parsons et al., 2013). In *N. benthamiana*, overexpression of four different P4Hs with IgA1 resulted in increased levels of hydroxylation in the hinge region. The highest levels were obtained when two P4Hs from different families were co-expressed (P4H1 and P4H10) (Mócsai et al., 2021). Lower Hyp formation was also reported in transient systems compared to stable production on PMPs (Parsons et al., 2013; Z. Yang, Bennett, et al., 2012). The production of large amounts of proteins in transient expression could allow the proteins to avoid hydroxylation of less favoured sites.

3.2 *O*-glycosylation

Hydroxyproline-*O*-glycosylation is a PTM specific of the plant kingdom. It was first associated with HRGPs, but many other targets have also been described. Like other glycosylation types, Hyp-*O*-glycosylation is thought to impact solubility, thermal stability, conformation/folding and proteolytic resistance (Hijazi, Velasquez, et al., 2014; Strasser et al., 2021). Indeed, although they are disordered and exposing their backbone to proteases, AGPs are very resistant to degradation probably because of glycosylation. In addition, *O*-glycosylation is often absolutely necessary for the activity of HRGPs and functional roles of the *O*-glycans as molecular signal/ligands or crosslinker of cell wall component have therefore been proposed.

There is also an interplay between Hyp-*O*-glycosylation and secretion. Indeed, several studies have reported the very efficient secretion of highly Hyp-*O*-glycosylated proteins (Xu et al., 2007, 2008). Similarly, the use of P4H inhibitor in soybean led to the disappearance of a major PRP, PRP2, from the cell wall (Schmidt et al., 1991). In Arabidopsis, the proline rich domain of Fascilin-like AGP 4 (FLA4) is not required for activity, but its removal led to accumulation of the protein in intracellular compartments and lower overall abundance (Seifert, 2020). Based on the observation of partially glycosylated protein accumulation in intracellular compartments, it has been postulated that the first steps of hydroxylation and glycosylation occur very rapidly and are followed by a rate limiting step (N. Zhang, Wright, Wang, et al., 2019). An unknown retention mechanism would prevent partially glycosylated proteins from being secreted. Mutants completely lacking the Hyp-*O*-glycosylation machinery would lose these retention signals and allow secretion of non-glycosylated proteins (Beuder et al., 2020). As some glycosyltransferases are already present in the ER,

glycosylation could protect disordered domains from degradation pathways (Basu et al., 2013; Oka et al., 2009).

As with Hyp hydroxylation, initiation of glycosylation must be tightly regulated. Indeed, the amino acids surrounding the Hyp define the type of glycosylation it undergoes. The initiating GTs must detect this in order to modify the peptide accordingly. Primary amino acid sequence and Hyp site clustering appear to be the major discriminant for the addition of galactose or arabinose by GTs (Shpak et al., 1999). Since many Hyps are not glycosylated, there must also be a selectivity of the initiating enzymes to control this process. Indeed, several studies have described proteins with clustered hydroxyprolines that were non glycosylated, such as tobacco chitinase (Sticher et al., 1992). In data mining of MS experiments, Duruflé et al. (2017) studied five hydroxyproline-rich peptides (including a PRP) from Arabidopsis that are non-glycosylated. Their conclusions were that the Extended Pro hydroxylation code was ineffective in predicting these hydroxylation sites, probably suggesting that proline hydroxylation was more widespread, although its function outside the context of glycosylation is unclear.

The intrinsic features of HRGPs, their large glycans moieties and the apparently high functional redundancy both in the HRGP superfamily and within glycosyltransferases have made them notably difficult to study. Indeed, single KOs of HRGPs and related GTs have generally no effect (Basu, Tian, et al., 2015; Y. Zhang & Showalter, 2020). In addition, glycosylation is a modification that is particularly difficult to study. Despite this, small change in the glycosylation status of HRGPs could be linked to important phenotypical changes, although often without the exact protein involved (if any) identified (MacAlister et al., 2016; Y. Zhang, Held, et al., 2021). A tool that has widely used to isolate, identify, and characterise *in vivo* HRGPs is the β-glucosyl Yariv reagent, which is able to bind and precipitate the β-galactose present on AGPs. Carbohydrate reactive antibodies and to a lesser extent lectins (due to the sometimes-poor characterisation/specificity of their exact substrate) have also been used as probes for the study of HRGPs (Seifert, 2020). Other tools include monosaccharide analysis, NMR, and MS. Some studies also elegantly used glycosylhydrolases (GH) to selectively modify and analyse glycans (Tryfona et al., 2010, 2012). Glycan analysis of HRGPs has also further been hampered by the difficulties to separate these proteins and to isolate them from the carbohydrate matrix of the cell wall. Because of these difficulties, some authors postulate that the *O*-glycome is largely underappreciated (Petersen et al., 2021).

The use of synthetic HRGPs glycomodules as tools to study Hyp-*O*-glycosylation has allowed to hypothesise rules regarding the specificity of the glycosylation types:

arabinogalactan (AGP motif) and arabinose (EXT motif). Various glycomodules used in different plant species (*A. thaliana*, *L. lycopersicum, N. tabacum, N. benthamiana* and *C. reinhardtii*) and different culture systems (cell suspensions, hairy roots and whole plants) are listed in Table I.3.1. Some of them, such as $(SerPro)_{32}$, show remarkable efficiency and specificity. Stably transformed Arabidopsis revealed a cell-type specific accumulation of the different glycomodules when expressed with *p35S* (Estévez et al., 2006). Comparison of the results between Arabidopsis and tobacco revealed important differences: some constructs, for example $(AlaPro)_{51}$ and $(SerProPro)_{24}$, could not be expressed at all in Arabidopsis, hydroxylation efficiencies varied, and monosaccharide composition differed. This suggested that there are some cell-type- and species-specific mechanisms for the hydroxylation and glycosylation of the glycomodules and that glycomodules that are not recognised are targeted for degradation. In a subsequent study, (Xu et al., 2008) expressed different glycomodule-GFP constructs in Arabidopsis cell culture, and the results were in accordance with BY-2 cells data, with the exception of monosaccharide composition, indicating that most of the differences are related to tissue specific mechanisms.

Actually, HRGPs are part of a continuum of lowly to highly glycosylated proteins that mix different *O*-glycosylation motifs with other protein domains, resulting in a very diverse and heterogenous superfamily. As mentioned above, the classical separation of HRGPs into AGPs and EXTs is actually more complex as many proteins mix and match different glycomodules and protein domains. However, for the sake of clarity, the glycosylation process will be presented based on the classical AGP/EXT classification. Additionally, PRPs and other types of peptides that do not necessarily fall into this classification will be also discussed.

Table I.3.1: Overview of the use of AGP and EXT glycomodules. Various protein carriers have been fused to glycomodules to study Hyp *O*-glycosylation code and increase yields of recombinant proteins. The percentage of hydroxylated proline on each substrate in BY-2 cells is given in the "Hydroxylation efficiency" column, No Data (ND) is indicated when it was not calculated. The percentage of Hyp modified with the different glycosylation types in BY-2 cells is given in the "Glycosylation type" column. The analysis of the glycosylation with the asterisk was not conducted but assumed based on the contiguity hypothesis and protein molecular mass increase. The sequence of GAGP is SPSPTPTPPGPHSPPPTL and YK is SPPPPSPSPPPPYYYK.

HRGP motif	Glycomodule	Hydroxylation efficiency in BY-2 cells	Glycosylation in BY-2 cells	Carrier protein	Expression system
AGP motif	$(SerPro)_{32}$	100%	Arabinogalactan (100%)	GFP	BY-2 (Shpak et al., 1999; N. Zhang, Dolan, et al., 2016), tobacco hairy roots (N. Zhang, Wright, Wang, et al., 2019), Arabidopsis plants and suspension cells (Estévez et al., 2006; Xu et al., 2008)
	$(SerPro)_5$	ND	Arabinogalactan*	GFP	*N. Benthamiana* (Dolan et al., 2014)
	$(SerPro)_{10}$	100%	Arabinogalactan (100%)	hGH, Venus	BY-2 (Xu et al., 2010), *C. reinhardtii* (Ramos-Martinez et al., 2017)
	$(SerPro)_{20}$	ND	Arabinogalactan*	GFP, EPO, INF2α, SCF, Venus	BY-2 (Karki et al., 2022; X. Wang et al., 2021; Xu et al., 2007; N. Zhang, Gonzalez, et al., 2016), C. reinhardtii (Ramos-Martinez et al., 2017)
	$(ValPro)_{11}$	80%	Arabinogalactan (60%), also arabinose and unmodified	GFP	BY-2 (Tan et al., 2003), Arabidopsis (Estévez et al., 2006)
	$(AlaPro)_{51}$	100%	Arabinogalactan (100%)	GFP	BY-2 (Tan et al., 2003), Arabidopsis (Estévez et al., 2006)
	$(AlaPro)_{20}$	ND	Arabinogalactan*	AAT	BY-2 (N. Zhang, Wright, Caraway, et al., 2019)

Table I.3.1: Continued

HRGP motif	Glycomodule	Hydroxylation efficiency in BY-2 cells	Glycosylation in BY-2 cells	Carrier protein	Expression system
AGP motif	$(ThrPro)_{101}$	70%	Arabinogalactan (50%), also arabinose and unmodified	GFP	BY-2 (Tan et al., 2003), Arabidopsis (Estévez et al., 2006)
	3xGAGP	ND	Unmodified (10%), arabinose (30%), arabinogalactan (60%)	GFP	BY-2 (Shpak et al., 1999), Arabidopsis (Estévez et al., 2006)
	SlAGP1	90%	Unmodified (10%), arabinose (35%), arabinogalactan (55%)	GFP	BY-2 (Zhao et al., 2002), tomato plants and suspension cells (Gao et al., 1999; W. Sun et al., 2004), Arabidopsis suspension cells (Estévez et al., 2006; Xu et al., 2008)
EXT motif	$(SerProPro)_{24}$	90%	Arabinose (50-70%)	GFP	BY-2 (Shpak et al., 2001), Arabidopsis plants and suspension cells (Estévez et al., 2006; Xu et al., 2008)
	$(SerProProPro)_{15}$	75%	Arabinose (75%), arabinogalactant (25%)	GFP	BY-2 (Shpak et al., 2001), Arabidopsis (Estévez et al., 2006)
	$(SerProProProPro)_{18}$	100%	Arabinose (95%)	GFP	BY2 (Shpak et al., 2001), tobacco hairy roots (N. Zhang, Wright, Wang, et al., 2019), Arabidopsis plants and suspension cells (Estévez et al., 2006; Xu et al., 2008)
	$(YK)_{20}$	100%	Arabinose (100%)	GFP	BY-2 (Held et al., 2004), Arabidopsis (Estévez et al., 2006)

3.2.1 AGP type

AGPs are a part of a very diverse class of proteins with little sequence identity. However, they are ubiquitous in plants and conserved through evolution, highlighting their important role. They are mostly found in the extracellular space and at the membrane, but also in intracellular multivesicular bodies (Silva et al., 2020). They are characterised by a high abundance of Pro, Ala, Ser and Thr residues and the repetition of Ala-Pro, Ser-Pro, Thr-Pro, Pro-Gly and Pro-Val dipeptides (Silva et al., 2020; Strasser et al., 2021). Another characteristic of AGPs is the importance of the complex type II arabinogalactans glycans, which can account for up to 95% of the molecular mass of the proteins. Half of classical AGPs are predicted to have a GPI anchor and many GPI anchored proteins carry AGP modules in Arabidopsis (Seifert, 2020; Showalter & Basu, 2016).

AGPs are involved in all aspects of plant development and growth. Their many functions have recently been reviewed and there are many examples of their essential role in plant physiological processes (Leszczuk et al., 2023; Seifert, 2020; Silva et al., 2020; Strasser et al., 2021). It is believed that glycans fully account for the functions of classical AGPs and that the protein backbone serves only as a scaffold (Seifert, 2020; Shpak et al., 1999). Glycosylation of AGPs begins with hydroxylation of XaaPro dipeptides motifs. Five AtP4Hs were co-expressed with several AGP genes (Konkina et al., 2021). Next, the Hyps are glycosylated with type II arabinogalactan glycans. Type II arabinogalactan consists of a backbone of β(1,3)-galactose with side chains β(1,6)-galactose (Figure I.3.4). The galactose residues are then decorated with Ara, Rha, Fuc, Xyl and (O-methyl)-GlcA. Glycosylation occurs by multistep addition of single saccharide residues and is initiated by addition of a first Gal on Hyp, catalysed by eight distinct Hyp-galactosyltransferases. There are two groups: GalT2-6, which have a Galectin domain in addition to the GalT domain, and HPGT1-3, which only have the GalT domain (Figure I.3.4 B) (Basu et al., 2013; Basu, Tian, et al., 2015; Basu, Wang, et al., 2015; Ogawa-Ohnishi & Matsubayashi, 2015; Oka et al., 2009). Their activity was confirmed by heterologous expression and *in vitro* characterisation, revealing that they indeed act specifically on alternating but not consecutive Hyps. The enzymes are located in the early Golgi with some members also localised in the ER, suggesting that galactosylation begins in the ER and continues in the Golgi (Basu et al., 2013; Oka et al., 2009). Higher order mutant from the GalT2-6 group resulted in a strong reduction in AGPs and abnormal growth phenotype of Arabidopsis plants (Basu, Wang, et al., 2015; Y. Zhang, Held, et al., 2021). Interestingly, recombinant GalT2 and 5 from *P. pastoris* have significantly lower activity (Basu et al., 2013; Basu, Wang, et al., 2015) than galactosyltransferase

activity isolated from ER microsomal fractions (Basu et al., 2013; Basu, Wang, et al., 2015) or after HPGT transient expression in *N. benthamiana* (Basu, Tian, et al., 2015; Oka et al., 2009). This could be due to differences in the *in vitro* assays and presence of several GalTs in the microsomal fractions. An interesting hypothesis is that, by analogy with mammalian GalNAcTs, some GalTs would have initiating activities (maybe HPGT in the ER) and that other enzymes such as GalT2 and 5 could recognise galactosylated substrates via their Galectin domain and act as finishing enzymes. Further studies are required to confirm this possibility.

Glycosyltransferases with activities for elongation of the β(1,3)-galactose backbone (At1g77810 and AtKSN4 (Y. Qu et al., 2008; Suzuki et al., 2017)), branching and elongation of β(1,6)-galactose side chains (AtGALT29A and AtGALT31A (Dilokpimol et al., 2014; Geshi et al., 2013)), transfer of α(1,2)-fucose (AtFUT4 and AtFUT6 (Tryfona et al., 2014; Wu et al., 2010)), β(1,2)-Glucuronic acid (AtGLCAT14A-E (Knoch et al., 2013; Y. Zhang et al., 2020)) and β-arabinose (AtRAY1 (Gille et al., 2013)), have been identified (Figure I.3.4). Some glycosyltransferases remain to be identified including α-arabinosyl-, xylosyl- and rhamnosyltransefrases. Some authors have also reported Ser galactosylation on some AGPs (Saito et al., 2014; N. Zhang, Wright, Wang, et al., 2019) which could be analogous to that found in EXTs. Most of these enzymes are located in the Golgi apparatus.

The presence of particular monosaccharides can be important for the general functions of AGPs. For example, the Arabidopsis *mur1* mutant, which is deficient in fucose biosynthesis displays abnormal root phenotype (Van Hengel & Roberts, 2002). Detection of fucosylated AGPs revealed that they are precisely localised at the plasma membrane of certain cell-types such root caps and elongated cells (Van Hengel & Roberts, 2002). A particular attention was paid to GlcA. This monosaccharide confers a negative charge to AGPs and provides a potential binding site for Ca^{2+} in a pH dependent manner, which was confirmed by simultaneous KO of 3 GLCATs in Arabidopsis (Y. Zhang & Showalter, 2020). The exact function of calcium binding capacity of AGPs is unclear. In seaweed, high GlcA content and thus calcium binding, could ensure salt tolerance (Strasser et al., 2021). In addition, GlcA is often methylated, which can have a high impact on the functionality of the residue, especially for binding of polar species (Temple et al., 2019). Two enzymes with this activity have been identified (AtAGM1-2).

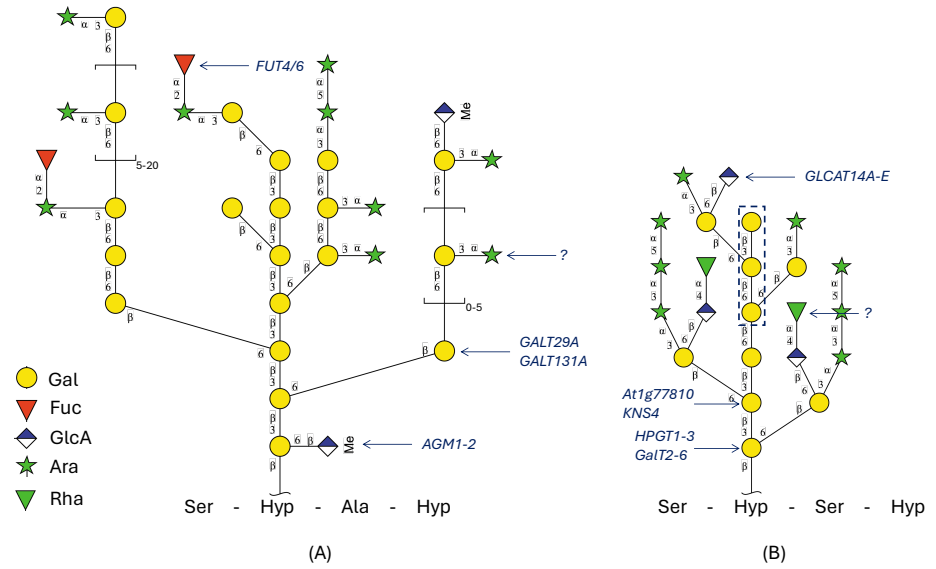


Figure I.3.4: Structural differences of type-II arabinogalactan glycans. Schematic representation of model AGP glycans representative of Arabidopsis leaves AGPs (A) and of the glycomodules $(SP)_{32}$ and $(AP)_{51}$ expressed in BY-2 cells (B) (Tan et al., 2010; Tryfona et al., 2012). Enzymes responsible for the addition of each modification are represented with arrows. The repetitive trigalactosyl unit described by Tan et al. (2010) in BY-2 AGPs is highlighted with dashes.

There is a great variability in AGP glycans between species and tissues or even between different developmental stages of the same cells. For instance, AGPs from BY-2 cells lack fucose but contain Rha, which is absent in Arabidopsis (Tan et al., 2010; Xu et al., 2008). *N. tabacum* AGPs also generally have a higher GlcA content. Apart from monosaccharide composition, there is also structural variability (Figure I.3.4). The structure of AGPs from $(AlaPro)_{51}$ and $(SerPro)_{32}$ glycomodules produced in BY-2 cells was elucidated by NMR spectroscopy (Tan et al., 2004, 2010). It revealed that the backbone is composed of repeat β(1,3)-trigalactosyl units that are linked together by β(1,6) linkage. β(1,6) monogalactosyl side chains can be linked to the first and second galactoses and are heavily substituted with Ara, GlcA and Rha where some microheterogeneity has been observed. This repeating units could be the building blocks of much larger glycans (up to 150 residues) (Tan et al., 2010). This is very different from the structure observed for the AGPs from Arabidopsis leaves, radish roots or wheat flour (Tryfona et al., 2010, 2012). Combination of GHs, MS, and carbohydrate analysis revealed a β(1,3)-galactose backbone mostly decorated with long β(1,6)-galactose side chains (between 5 and 20 residues) and sometimes directly substituted with β(1,6)-(Me)GlcA. The sides chains are themselves

substituted with (Me)GlcA, Ara and Fuc (Tryfona et al., 2012). This variability is well-illustrated by the different profiles observed for glycomodules by western blotting (see references in Table 3.1).

Final processing of AGPs within the cell wall is carried out by GHs that can act on AGP glycans (β-galactosyl- and arabinosylhydrolase activities have been described (Silva et al., 2020)), and phospholipases that can cleave GPI anchor (Showalter & Basu, 2016; Silva et al., 2020). During tomato ripening, there is a reduction of the AGP epitopes in the cell wall. There is also a remodelling of the glycans (reduction of the molecular weight) and redistribution of AGP localisation, possibly via internalisation. SlP4H3 expression is linked with this process and its silencing delayed abscission of fruits (Kutyrieva-Nowak et al., 2024; Perrakis et al., 2019). AGP31 coexists in various glycosylated forms, and it is not known if this results from incomplete glycosylation or GH activity. In addition, the less glycosylated forms of AGP31 also undergo proteolytic processing of some domains (Hijazi et al., 2012).

Many roles have been proposed for AGP glycans. Firstly, they might participate in the cell wall architecture and act as structural components. Indeed, AGPs can interact with other cell wall components via non-covalent and covalent linkages. A group of AGP/PRP hybrids has been identified in several species including Arabidopsis and tobacco (Hijazi et al., 2012; Hijazi, Velasquez, et al., 2014), which play a central role in the cell wall organisation. In fact, their multi-domain organization, containing a short N-terminal AGP domain, a histidine (His)-stretch, a PRP domain and a C-terminal PAC (Proline-rich protein and AGP, containing Cys), results in a complex supra-molecular scaffold that interacts with different components (Hijazi, Roujol, et al., 2014; Tan et al., 2013). For example, AGP1 from *Daucus carrota* was shown to bind the pectin fraction in a calcium-dependant manner (Baldwin et al., 2001). This is supported by AFM measurements of the interaction between $(SerPro)_{51}$ and $(SerPro_4)_{18}$ glycomodules that is dependent on pH and Ca^{2+} and B^{3+} concentrations (Tan et al., 2018). In addition, AtAGP31 can interact with itself to form aggregates in solution and deglycosylation of the tobacco homologue NtTTS prevents this aggregation phenomenon (Hijazi, Roujol, et al., 2014). Covalent linkages between the AGP Rha and Rhamnogalacturonan I or Ara and arabinoxylan, have also been described (Tan et al., 2013).

The calcium binding activity of AGPs glycans could also be involved in intracellular Ca^{2+} signalling or serve as a calcium reserve in the plasma membrane (Lopez-Hernandez et al., 2020). Glycans could also directly serve as molecular signals and be implicated in receptor/ligand binding. The plasma membrane location and GPI anchoring of AGPs suggest their involvement in signal perception and transduction.

Indeed, cell signalling is altered by Yariv reagent that binds to AGPs (Shimizu et al., 2005). In addition, the release of the glycans via GHs or the whole AGP via phospholipases could be a way to diffuse these signals under specific conditions. Also, AGPs are involved in interactions with root microorganisms and alteration of AGPs reduces the attachment of symbiotic rhizobium (Nguema-Ona et al., 2014).

3.2.2 Extensin type

Extensins are modular highly repetitive proteins. They are characterised by the $SerPro_{2-5}$ motif, Tyr crosslinking motifs and abundant basic amino acids (Hijazi, Velasquez, et al., 2014; Showalter & Basu, 2016). They constitute a large family of classical and chimeric EXT-like proteins that share high sequence homology. As for other Hyp *O*-glycosylated proteins, their destination is the apoplast where they are involved in the architecture of the cell wall. They participate in cell expansion and growth, particularly of tip-growing cells like pollen tube and root hairs (Beuder et al., 2020). EXTs are thought to guide the correct deposition of wall polysaccharides (mainly pectin) during growth and mitosis (Chen et al., 2015). They also play a role in cell wall repair and response to mechanical stresses and pathological and non-pathological microbial interactions (Merkouropoulos & Shirsat, 2003; Salvá & Jamet, 2001).

EXTs provide a scaffold for cell wall components that organise and reinforce the cell wall. AtEXT3 is able to self-assemble into a dendritic network that can serve as a template for components for cell plate formation (Cannon et al., 2008). Silencing of AtP4H5 produces cell wall fibrils that are more parallel to the plasma membrane (Velasquez, Ricardi, et al., 2015). However, apart from a near-lethal mutant *ext3*, EXT mutants show subtle phenotypes, probably due to high functional redundancy (Hijazi, Velasquez, et al., 2014). However, roots seem to be more sensitive to EXT variations and mutants often show aberrant root morphologies (Marzol et al., 2018).

EXTs undergo many PTMs in the secretory pathway and apoplast. The $SerPro_n$ motifs are hydroxylated and then glycosylated with short arabinose chains. The serine residue is also monogalactosylated (Figure I.3.5). In the cell wall, type III peroxidases catalyse the formation of intra- and intermolecular tyrosine cross-linking (Held et al., 2004). The crosslinking motifs are Tyr-Xaa-Tyr or Val-Tyr-Lys, the former being able to cross link with itself and, in total, crosslinks can involve 4 Tyr residues (Chen et al., 2015). This leads to the insolubilisation of EXT and the formation of a three-dimensional network in the cell wall that interacts with other components (Hijazi, Velasquez, et al., 2014). For example, basic EXTs can form by acid-base interaction with the acidic pectins (Cannon et al., 2008; Chen et al., 2015; Nuñez et al., 2009).

To date it is not known if some EXT GHs contribute to the remodelling of EXT glycans in the cell wall.

HPAT1-3 transfer the first β-arabinose onto consecutive Hyp (Ogawa-Ohnishi et al., 2013). This is followed by the stepwise addition of a second and a third β(1,2)-arabinose by RRA1-3 and XEG113, respectively (Egelund et al., 2007; Gille et al., 2009). The fourth α(1,3)-Ara is transferred by ExaD (Møller et al., 2017) (Figure I.3.5). An additional fifth arabinose residue can also be added but the linkage and enzyme involved are still unknown. Ser α-monogalactosylation in the $SerPro_n$ sequons is mediated by SGT1 (Saito et al., 2014). This glycosyltransferase is located in the ER membrane and possibly in the Golgi where it recognises Hyp in EXT sequences. Of note, this is the only known glycosyltransferase with a type 1 membrane topology that has no homology outside the plant kingdom (Saito et al., 2014).

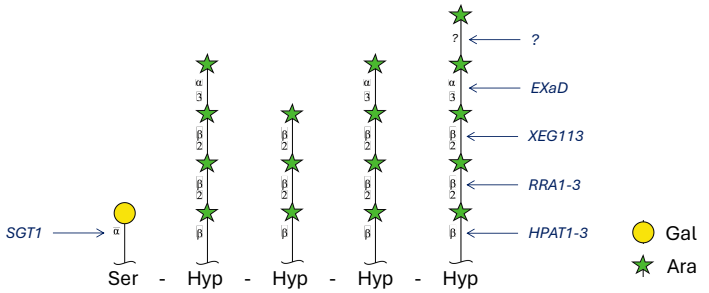


Figure I.3.5: Glycosylation of EXT SP_4 sequons. Schematic representation of average EXT glycosylation. Enzymes responsible for the addition of each modification are represented with arrows.

EXT glycosylation is important for cell expansion, particularly in root hair and pollen tube (Beuder et al., 2020; Dardelle et al., 2010; Velasquez et al., 2011; Velasquez, Ricardi, et al., 2015). Indeed, *hpat* and *sgt1* mutants show disrupted root hair elongation (MacAlister et al., 2016; Saito et al., 2014). In fact, both Hyp-arabinosylation and Ser-galactosylation are required and have additive effects (Velasquez, Marzol, et al., 2015). Overexpression of *AtP4H2*, *5* or *13* resulted in increased root hair density and length consistent with the involvement of these P4Hs in modification of EXT substrates (Velasquez et al., 2011). EXTs are also implicated in many other cell types and *hpat* mutants show effects at the whole plant level. HPAT 1 and 3 are redundantly required for full male fertility (MacAlister et al., 2016). It should be noted that *hpat* mutants in Arabidopsis and *P. patens* have very different phenotypes (MacAlister et al., 2016).

The formation of the EXT network is promoted by recognition and alignment of the hydrophilic glycosylated domains and the hydrophobic Tyr-rich domains. In fact, EXT can self-assemble in a network even in absence of cross-linking (Cannon et al., 2008). Release of H_2O_2 during stress could promote crosslinking in an ultra-rapid defence response (Showalter, 1993). This reinforcement of the cell wall could limit the spread of pathogenic bacteria, as reported in Arabidopsis overexpressing *EXT1* (Nguema-Ona et al., 2014).

EXTs adopt a rod-shaped structure with a triple polyproline II helix. This structure could be stabilised by the glycans that wrap around the backbone (Showalter & Basu, 2016; Velasquez, Marzol, et al., 2015). Glycans, in particular the fourth Ara residue, were also shown to increase crosslinking rate *in vitro* (Beuder et al., 2020; Chen et al., 2015). In addition, glycosylation could provide curvature and relaxation of EXTs structure, and this will reduce the rigidity and allow growth (Velasquez, Marzol, et al., 2015). Consistently, secretion of unglycosylated EXTs could be toxic as mutants in glycosylation show arrest of growth (Beuder et al., 2020; Petersen et al., 2021; Velasquez, Marzol, et al., 2015). Due to the importance of proper EXTs glycosylation for their function, glycosylation defects are sensed and controlled in the secretory pathway by specific papain-type cysteine endopeptidases (Helm et al., 2008).

3.2.3 Other glycosylation types

PRPs – PRPs are the most varied and less well-known type of HRGPs. They are characterised by Pro_2 or Pro_3 motifs within larger repeating units such as LysLysProCysProPro and ProProXaaYaaLys/Thr (Xaa and Yaa can be Val, Tyr, His, or Glu) (Hijazi et al., 2012; Hijazi, Velasquez, et al., 2014; Showalter et al., 2010). Little is known about the function of PRPs, but several members such as AtPRP4 have structural roles in the CW (Duruflé et al., 2017). Some PRPs are nodulin proteins involved in interaction with soil bacteria (Showalter, 1993). Like other HRGPs, PRPs can be modified with Hyp and glycans. However, contrary to most HRGPs, the Pro rich domains are weakly hydroxylated and the PRPs are usually non- or weakly glycosylated (Hijazi et al., 2012). For example, only the second proline of diPro repeats is hydroxylated in a soybean PRP (Showalter, 1993). Some PRP, such as AGP31 PRP domain, however, are heavily glycosylated with large glycans most likely composed of Gal and Ara and EXT-like short Ara chains (Hijazi et al., 2012). The biosynthesis of PRP glycans in not known, but monosaccharide analysis revealed the presence of Gal, Ara, Xyl, Fuc, Glc, Man and GlcA (Hijazi et al., 2012; Hijazi, Roujol, et al., 2014).

Allergens – The analysis of the major allergen from *Artemisia vulgaris* Art v 1 revealed unusual glycosylation types (Leonard et al., 2005). The protein has a N-terminal defensin-like domain and a C-terminal hydroxyproline-rich domain, which contains both contiguous and non-contiguous sites. Consecutive Hyp residues are modified with single β-arabinose instead of chains as in EXTs. An arabinogalactan composed of β(1,6)-galactose highly substituted with α-arabinose chains was also observed and termed type-III arabinogalactan. Similar glycosylation pattern was then also found on *Ambrosia artemisiifolia* allergen Amb a 4 with differences in the type-III arabinogalactan, which contained a single galactose with α-arabinose chains capped by single β-arabinoses (Léonard et al., 2010).

While it is generally accepted that glycan specific IgE are not responsible for clinical symptoms of allergic patients (Altmann, 2007), the glycans of Art v 1 are important determinants for allergic reaction in a number of patients (Leonard et al., 2005). In ELISA, monoarabinose glycans, but not type-III arabinogalactan, were found to be necessary for the binding of antibodies from patients (Leonard et al., 2005). Comparison of binding affinities with Ara h 2, the main allergen from peanuts, produced in *E. coli* or *N. benthamiana* revealed a 1000-fold higher affinity with the plant-produced antigen (Üzülmez et al., 2021), further suggesting an important role of Hyp and its glycosylation in IgE binding. It should be noted that the allergen produced in *N. benthamiana* contained 7 Hyps compared to 3 in the natural protein.

Secreted peptides – Small peptide hormones are also targets of Pro hydroxylation and glycosylation. CLAVATA 3 and associated peptides (CLV3 and CLEs), C-terminal encoded peptides (CEPs) and peptides containing sulphated tyrosine (PSY) are all hydroxylated in GlyHypAspPro, AspSerHypGlyVal, and ValHypHypSer, respectively (Petersen et al., 2021). CLE-RS2 and RGFs peptides also contain hydroxyprolines (K. H. Lee et al., 2020).

These Hyp are further glycosylated with arabinose chains of three residues with the same linkage as those found in EXTs, suggesting that the same GTs are involved in this process. However, a fourth arabinose is never observed. This raised the questions of the ability of the HPAT-initiating enzymes to differentiate these non-contiguous substrates from AGPs and of ExAD to discriminate EXTs from the peptide substrates. Glycosylation of these peptides is necessary for their full activity (K. H. Lee et al., 2020; Petersen et al., 2021; Taylor et al., 2012). Because of their involvement in development, the reduced activity of secreted peptide hormones could also explain some of the phenotypes observed in *hpat* mutants (Beuder et al., 2020).

Other *O*-glycosylated peptides are the Hydroxyproline rich Systemin peptides (HypSys), which have been identified in Solanaceae and related species (tomato, tobacco, petunia, nightshade and sweet potato) (Pearce et al., 2009; Ryan & Pearce, 2003; Taylor et al., 2012). These 18 amino acids long peptides are derived from a single peptide precursor and have up to 5 Hyp and 9 pentoses. They have a role in inducing defence mechanisms. The activity of unglycosylated forms is greatly reduced (Pearce et al., 2009; Taylor et al., 2012).

Solanaceous lectins – Another type of HRGPs, described only in Solanaceae, is the lectins (Fragkostefanakis et al., 2014; Showalter, 1993). Potato tuber lectin is the most characterised member and consists of a GlyCys rich domain and a Ser-Pro rich domain, related to EXT. This domain is highly glycosylated with arabinoses for up to 50% of the molecular mass. They are expressed in response to wounding and viral infection and are able to agglutinate GlcNAc (Showalter, 1993).

Ser/Thr O-glycosylation – There are also reports of plant-type *O*-glycosylation on Ser and Thr residues. Arabinogalactan glycans have been reported on Ser/Thr in lyceum (Qin et al., 2000). Furthermore, recombinant GM-CSF in rice suspension cells is arabinosylated on Ser/Thr (J. Kim et al., 2016). Interestingly, the authors observed galactosylated arabinose like those found in microalgae (Mathieu-Rivet et al., 2020). It is not known whether these glycans are biologically relevant. However, they could maybe result from a misinterpretation of Ser-galactosylation and Hyp-arabinosylation as GM-CSF harbours SerHyp dipeptides.

3.3 Implication of plant *O*-glycosylation in PMF

3.3.1 Downside

In the context of PMF, non-native glycosylation has been a source of concern as it can impact the biological activity of pharmaceuticals and confers undesirable properties to the products. Consequently, many successful attempts have been made to engineer plants to eliminate non-human glycans, particularly core α(1,3)-fucose and β(1,2)-xylose from *N*-glycans (for instance: (Jansing et al., 2019; Mercx et al., 2017; Strasser et al., 2008)). *In silico* analysis of the human proteome revealed that 30% of the human proteins carry a recognition sequence for plant P4Hs (Gomord et al., 2010). In addition, several examples of non-human *O*-glycosylation on pharmaceuticals have been reported when produced in plants, including EPO (Castilho et al., 2012; Weise et al., 2007), IgA1 (Karnoup et al., 2005), MUC1 peptide (Pinkhasov et al., 2011), α-iduronidase and GM-CSF (J. Kim et al., 2016) (Figure I.3.6). It is also possible that other proteins have been overlooked. Yet, until recently,

little work had focused on elimination of plant specific *O*-glycans (Mócsai et al., 2021; Parsons et al., 2013; Uetz et al., 2024).



Figure I.3.6: Hydroxylation of plant produced pharmaceutical. Tryptic peptides isolated from several plant produced proteins (Cardon et al., 2019; Karnoup et al., 2005; Parsons et al., 2013; Pinkhasov et al., 2011; Weise et al., 2007). The hydroxylated prolines are represented in bold.

One of the main concerns of non-human glycosylation is the potential immunogenicity of these structures. However, after more than 25 years of PMF research, there is still no clear evidence supporting this. In fact, the Influenza VLPs vaccine from Medicago transiently produced in WT *N. benthamiana* harbours plant-typical N-glycans, which have no adverse effect in clinical trials (Ward et al., 2014). Despite this, analysing the effects plant *O*-glycosylated biopharmaceuticals in humans would be a labour- and cost-intensive process that could pose additional difficulties for the approval of PMPs. Immunogenicity of plant *O*-glycans in mammal models has been demonstrated in several studies and anti-AGP antibodies mainly target carbohydrate-containing epitopes (Gomord et al., 2010). For these reasons, engineering plant systems to produce pharmaceuticals devoid of their plant aberrant *O*-glycans is necessary.

Elimination of plant specific *O*-glycans has already been studied in several systems and have mainly focused on disrupting Pro hydroxylation to abolish glycosylation. Human EPO produced in *P. patens* and transiently in *N. benthamiana* is hydroxylated within the SerProPro motif but not glycosylated (Castilho et al., 2012; Parsons et al., 2013; Weise et al., 2007). In *P. patens*, *P4H1* KO was sufficient to completely eliminate Hyp from EPO, whereas *P4H2-6* KO had no effect (Parsons et al., 2013). Several reports have shown the presence of Hyp and arabinoses on the hinge region of IgA1 produced in maize (Karnoup et al., 2005) and in transiently *N. benthamiana* (Dicker et al., 2016; Uetz et al., 2022). Actually, the hinge region shares sequence identity with

a maize EXT (Karnoup et al., 2005). In *N. benthamiana*, CRISPR/Cas9 KO of the four genes in the *P4H4* subfamily led to an unexpected decrease of non-hydroxylated Pro residues in the hinge of transiently expressed IgA1 and an increase in arabinosylation of Hyp, although the pattern of hydroxylation was modified (Uetz et al., 2022). The authors concluded that P4H4 did indeed modify the Pro of the hinge region but that a rebalancing by other P4Hs was responsible for an increase in hydroxylation. On the other hand, in a later work, they showed an important reduction from 90 to less than 50% of hydroxylated peptides and of arabinosylation by knocking out two of the three *P4H10* genes (Uetz et al., 2024). It is important to note that in all these experiments, no obvious abnormal growth phenotype was observed. This is consistent with studies performed in *C. reinhardtii*, where *p4h1* and *cw-1* mutants showed important defects in cell wall architecture, but no impact on growth kinetics (Keskiaho et al., 2007).

Alternative strategies for elimination of plant specific *O*-glycans could be based on the use of inhibitors and medium optimisation. Indeed, the use of several P4H inhibitors (2,2-dipyridyl, an iron chelator; 3-4 dehydroxyproline, pyridine-2,4-dicarboxylate and pyridine-2,5-dicarboxylate, which are analogues to Hyp and 2 OG, respectively) has resulted in a reduction of AGPs and an alteration of plant morphology reminiscent of *P4H* or *GT* mutants (Fragkostefanakis et al., 2014; Ludovic et al., 2015; Moriguchi et al., 2011; Parsons et al., 2013; Tiainen et al., 2005). However, the effect is not complete (Moriguchi et al., 2011) and may have other impacts on the plant physiology. Moreover, N. Zhang, Dolan, et al. (2016) have shown that medium composition, and in particular the nitrogen supply, has a major impact on the secretion and glycosylation of GFP-fused glycomodules. This phenomenon is still unexplored but could contribute to reducing undesirable *O*-glycosylation of PMPs in the same way that culture media are used to modulate glycosylation in CHO cells (Brühlmann et al., 2017).

3.3.2 Upside

On the contrary, Hyp-*O*-glycosylation could also be beneficial for PMF. Indeed, the *O*-glycans can enhance the secretion and protect the proteins from proteolysis, which can highly impact yields. Indeed, plant production systems are known for lower yields compared to traditional systems and high protease activity (Santos et al., 2016). In addition, this can also improve the *in vivo* stability of the proteins and present an alternative to polyethylene glycol conjugation to increase serum half-life (Gomord et al., 2010). Indeed, hGM-CSF produced in rice cell suspension culture is *O*-arabinosylated, which reduces its activity but improves its stability compared with the same protein produced in yeast (J. Kim et al., 2016).

With this in mind, the AGP and EXT glycomodules were used to boost the yield of recombinant proteins. This proprietary technology was used on several proteins: GFP, EPO, CSF, α-antitrypsin (AAT) and interferon α2, and in several production systems: stably transformed *N. tabacum* BY-2 cells and hairy root culture, and transient expression in *N. benthamiana* (Karki et al., 2022; Xu et al., 2007; N. Zhang, Dolan, et al., 2016; N. Zhang, Wright, Caraway, et al., 2019). The results are quite striking as the yields of INFα2 can be increased by a factor of 1500 in BY-2 cells (N. Zhang, Dolan, et al., 2016). $(AlaPro)_{20}$-AAT was biologically active in inhibition assays (N. Zhang, Wright, Caraway, et al., 2019) while INFα2-$(SerPro)_{20}$ half-life in mice serum was greatly extended without showing immunogenicity (Xu et al., 2007). $(SerPro)_{20}$-EPO was able to stimulate differentiation into red blood cells *ex vivo* (Karki et al., 2022). The use of red blood cell differentiated with Hyp-*O*-glycosylated EPO *ex vivo* is an interesting alternative to circumvent the immunogenic potential of Hyp-*O*-glycosylation for clinical applications.

Objectives

Nicotiana tabacum BY-2 cell suspension culture is an alternative to the traditional systems to produce therapeutic glycoproteins. The glycosylation of these products is very important for their properties and is defined as a critical quality attribute in the industry. Glycoengineering is a well implemented strategy to control glycosylation and improve the quality of therapeutic proteins and their production processes. GalNAc-*O*-glycosylation is a type of glycosylation typical of metazoans that has broad functions in protein biology. This modification is not well understood and is heterogeneous, which complicates its systematic control in therapeutic production pipelines. In contrast, hydroxyproline-*O*-glycosylation is a typical plant modification with a potential immunogenic problem. This hampers the development of plant cells as a therapeutic production system. It has been repeatedly demonstrated that BY-2 cells are amenable to *N*-glycoengineering and could address all these issues.

The objective of this work is therefore to engineer BY-2 cells to produce therapeutic glycoproteins with a homogeneous and humanised *O*-glycosylation profile. To achieve this, two steps are required: (1) the *de novo* introduction of human GalNAc-*O*-glycosylation and (2) the complete elimination of plant Hyp-*O*-glycosylation on recombinant proteins. These strategies and their combination in BY-2 cells have largely been unexplored but are needed tofurther improve the quality of the products and potentially increase the interest in the BY-2 platform.

Firstly, GalNAc-*O*-glycosylation will be introduced by expression of one of the fifteen human initiating enzyme, GalNAcT2, to generate the Tn-antigen. This *O*-glycan will then be elongated by co- expression of C1GalT1 to form the core 1 structure. In addition, the effect of expressing an epimerase to boost the availability of UDP-GalNAc will be evaluated. Therapeutically relevant *O*-glycoproteins (Immunoglobulin A1, erythropoietin and the Spike from SARS-CoV2) will be expressed in the generated cell lines to test their capacity for correct *O*-glycosylation. Lectin blotting will be used to check the presence of the expected O-glycans on endogenous and recombinant pharmaceutical proteins. Mass spectrometry analysis will confirm O-glycans structures on purified recombinant *O*-glycoproteins.

Secondly, Pro hydroxylation will be addressed by suppressing P4H activities using CRISPR/Cas9 edition of the corresponding genes. The P4H genes will have first to be identified in *N. tabacum*. The tobacco genome will be searched for genes homologous to known plant P4Hs from *Physcomitrella patens* and *Arabidopsis thaliana*. Then, the different P4H subgroups identified will be individually targeted by CRISPR/Cas9 editing to uncover their specificities. Reporter cell lines based on the fusion of GFP with two different Hyp-glycomodules will be generated to enable rapid phenotyping by western blotting. In these reporter cell lines, inactivation of

relevant P4H genes will be visualised by a reduced mobility shift of glycosylated GFP. The genome edition will also be analysed by sequencing in interesting cell lines.

The *O*-glycoengineered cell lines generated in this work will provide more versatility to BY-2 cells for the production of improved therapeutic proteins. In addition to industrial applications, these cell lines may also provide insights and tools for the study of *O*-glycosylation pathways, which are not fully characterised to date.

Results and Discussion

1 Introducing GalNAc-*O*-glycosylation in BY-2 cells

We introduced GalNAc-*O*-glycosylation in BY-2 cells by heterologous expression of GalNAcT2 and C1GalT1 for initiation and elongation of the *O*-glycan and tested the effect of an UDP-GlcNAc/UDP-GalNAc epimerase. The human GalNAcT2 is well characterised. It has a broad substrate specificity and is responsible for initiating *O*-glycosylation on a wide range of substrates and was successfully expressed in plants by different groups (Castilho et al., 2012; Daskalova et al., 2010; Z. Yang, Drew, et al., 2012). C1GalT1 is the enzyme responsible for elongating single GalNAc *O*-glycans to the core 1 structure. In humans, this enzyme requires a specific chaperone. The C1GalT enzyme of *D. melanogaster* was selected because it does not require a chaperone to be fully active. Finally, the WbpP epimerase from *P. aeruginosa* has already been used in plants in the GalNAc-*O*-glycosylation context, but it is not clear whether this enzyme increases the availability of UDP-GalNAc and leads to more efficient *O*-glycosylation. In the first experiments, endogenous BY-2 proteins were used to evaluate the *O*-glycosylation.

1.1 Genetic constructs and BY-2 cell transformation

ORFs coding for human GalNAcT2, *D. melanogaster* C1GalT1 and *P. aeruginosa* WbpP (UDP-GlcNAc C4-epimerase) were synthetised and cloned into Golden Gate entry plasmids (Engler et al., 2014). Each coding sequence was then cloned under the control of the Cauliflower mosaic virus *35S* promotor (*p35S*) and the *A. tumefaciens* Nopaline Synthase terminator (*tNOS*). Additionally, a sequence encoding the 3xFLAG tag was fused at 5' of the *WbpP* sequence. The three constructs were assembled into binary vectors containing the left and right borders (LB and RB) for *A. tumefaciens* mediated plant transformation. In total, four different plasmids were generated (Figure II.1.1). They encoded the GalNAcT2 alone (GalNAc), GalNAcT2 + WbpP (GalNAcE), GalNAcT2 + C1GalT1 (Core1) and GalNAcT2 + C1GalT1 + WbpP (Core1E). Additionally, all plasmids contained a 5' *nptII* kanamycin resistance cassette for selection of the transgenic BY-2 cells and a 3' tobacco *SAR* genetic insulator, which should reduce positional and silencing effects (Diamos & Mason, 2018). The *XylT*/*FucT* KO (ΔXF) BY-2 cell line was transformed with all the constructs and transformants were obtained. The ΔXF line is a *N*-glycoengineered cell line previously generated in our laboratory by CRISPR/Cas9 KO of the four *FucT* and the two *XylT* genes, responsible for the addition of the non-human core α(1,3)-fucose and β(1-2)-xylose residues, respectively. The KO was confirmed by

genomic DNA sequencing and expression of IgG, which showed the absence of xylose and core α(1,3)-fucose in *N*-glycans (Mercx et al., 2017).

1.2 Glycosylation analysis on endogenous proteins

Glycosylation of the endogenous proteins was analysed in the different *O*-glycoengineered BY-2 cell lines. Total cellular soluble proteins (TSP) were extracted and analysed by lectin blotting to confirm the presence of the expected *O*-glycans. Two lectins were used: *Helix pomatia* lectin (HPL) and peanut (*Arachis hypogaea*) agglutinin (PNL). HPL binds to terminal α- and β-GalNAc residues while PNL has activity toward unsubstituted core 1. Both lectins are coupled with biotin to visualise the blots using Streptavidin-HRP.

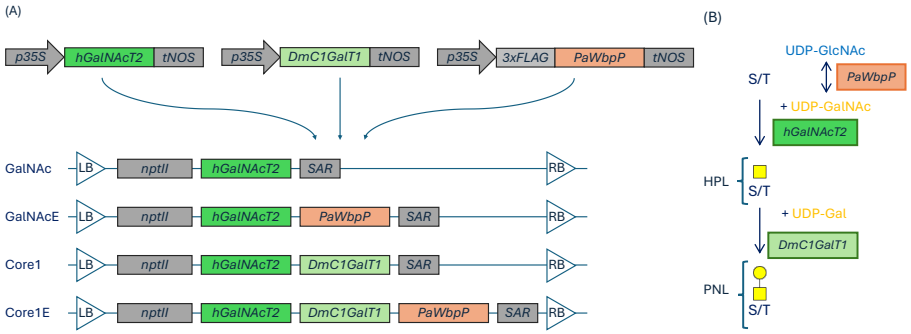


Figure II.1.1: Constructions of the *O*-glycosylation machinery. (A) Schematic representation of the generated constructs. The coding sequences from human *GalNAcT2*, *D. melanogaster C1GalT1* and FLAG-tagged *P. aeruginosa WbpP UDP-GlcNAc C4-epimerase* were cloned under the control of *p35S* and *tNOS* in Level 1 Golden Gate vectors. The expression cassettes were then assembled to form the four different constructs in binary plasmids. The T-DNA also contained the *nptII* kanamycin resistance gene and a *SAR* genetic insulator and were flanked by LB and RB for *A. tumefaciens* mediated transformation. (B) Schematic representation of the introduced pathway. The expressed enzymes are highlighted. The glycans recognized by the two lectins used in protein blotting are shown by brackets.

First, the initiation of GalNAc-*O*-glycosylation pathway was introduced in ΔXF BY-2 cells by expression of GalNAcT2 with or without WbpP epimerase. Figure II.1.2 A and B shows HPL-probed blots of GalNAc and GalNAcE cell lines. An increase in HPL signals in all the lanes of GalNAc (panel A) and GalNAcE (panel B) lines was observed compared to the ΔXF line (first lane in each panel). This result suggests that expression of GalNAcT2 alone in BY-2 cells enables initiation of GalNAc-*O*-glycosylation. In both blots, a weak signal was observed in ΔXF BY-2, which could be due to the fact that HPL is not totally specific and show some activity with terminal α- and β-GlcNAc (Sanchez et al., 2006) that may be present on endogenous glycoproteins. Alternatively, some endogenous proteins could be biotinylated and

recognised by Streptavidin-HRP. However, we did not test the reactivity of the blots with Strep-HRP alone. The effect of the epimerase expression was difficult to evaluate from these blots. Indeed, compared to the GalNAc line (panel B, lane 3), there was an increase in signal intensity for several cell lines (*e.g.*, GalNAcE-4), whereas the signals decreased in others (*e.g.*, GalNAcE-7). These variations in signal intensity suggested a different capacity of the cell lines for efficient *O*-glycosylation. This could be due to the expression level of the glycosyltransferase. A western blot with the microsomal fraction (MF) of the cell lines was probed with anti-GalNAcT2 antibodies (Figure II.1.2). It showed that the expression of GalNAcT2 is mostly constant among GalNAc and GalNAcE cell lines (the expected molecular mass is 64.7 kDa). Interestingly, GalNAcE-5 that showed no expression of GalNAcT2, had a HPL signal comparable to ΔXF. Another feature that could impact efficient glycosylation was the availability of UDP-GalNAc, which should be increased by the expression of the cytosolic WbpP UDP-GlcNAc C4-epimerase. Therefore, a western blot with TSP was probed with an anti-FLAG antibody (Figure II.1.2, expected MW = 40.6 kDa). It showed that the expression of the epimerase was variable among the cell lines (upper band around 40 kDa, just above a non-specific band). However, its exact effect was difficult to quantify. For instance, GalNAcE-7 and -4 had a similar expression of GalNAcT2 but HPL signal was stronger in GalNAcE-4 although the expression of the epimerase was weaker in this cell line.

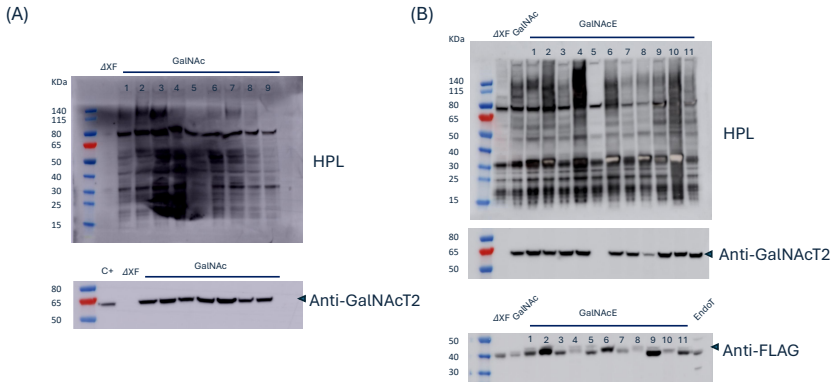


Figure II.1.2: Initiation of GalNAc-*O*-glycosylation in BY-2 cells. GalNAcT2 was expressed with or without a UDP-GlcNAc C4-Epimerase in ΔXF BY-2 cells. (A) HPL and anti-GalNAcT2 blots of GalNAc cell lines with the ΔXF line as negative control. (B) HPL, anti-GalNAcT2 and anti-FLAG blots of GalNAcE cell lines with GalNAc and ΔXF lines as positive and negative controls. Twenty-five µg of TSP were loaded in each lane for the HPL, and anti-FLAG blots and 25 µg of MF were used for anti-GalNAcT2 blots. GalNAcT2 expressed in *N. benthamiana* leaves and a cell line expressing FLAG-tagged EndoT were used as positive controls. A Coomassie-stained gel related to the lectin blot, showing the loading of each lane, is given in Supplementary figure S1 A.

Next, elongation of the *O*-glycans into the core 1 structure was achieved by co-expression of GalNAcT2 and DmC1GalT1 with or without the WbpP epimerase. Lectin blots of TSPs from these cell lines were probed with HPL or PNL (Figure II.1.3). In lines expressing C1GalT1 (Core1 and Core1E), there was a clear overall decrease in HPL signals compared to the GalNAc control, indicating a conversion of GalNAc-*O*-glycans to core 1. The residual signals were probably due to unmodified GalNAc or to low HPL reactivity toward substituted GalNAc (Sanchez et al., 2006). When probed with PNL, increased signals, albeit highly variable, were observed when C1GalT1 was expressed (unspecific signals saturated the detector and the membrane of this blot was cut to allow a longer peroxidase exposure; the complete membrane is shown in Supplementary figure S1 C). Western blots against GalNAcT2 and FLAG-WbpP are also shown in Figure II.1.3. As observed previously, GalNAcT2 expression was relatively constant between the cell lines, with the exception of the line Core1E-7, whereas epimerase expression was more variable. The effect of the WbpP epimerase expression on UDP-GalNAc substrate pool is again difficult to evaluate.

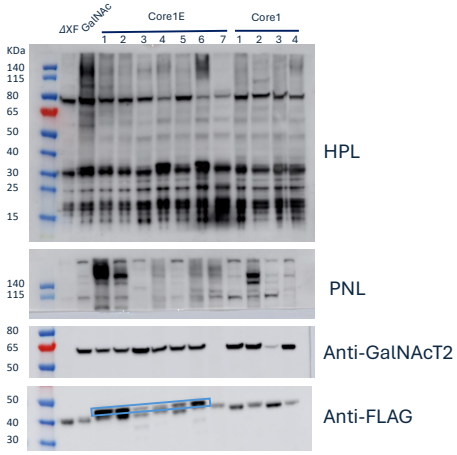


Figure II.1.3: Formation of core 1 *O*-glycans in BY-2 cells. GalNAcT2 and C1GalT1 were co-expressed with or without a UDP-GlcNAc C4-Epimerase in ΔXF BY-2 cells. TSPs from Core1 and Core1E lines were used for HPL, PNL and anti-FLAG blots. MFs from Core1 and Core1E lines were used for anti-GalNAcT2 blot. Twenty-five µg of TSP or MF were loaded in each lane ΔXF and GalNAc cell lines were used as controls in all blots. The band corresponding to the epimerase is highlighted with a blue box. A Coomassie-stained gel related to the lectin blot, showing the loading of each lane, is given in Supplementary figure S1 B.

The results of GalNAcT2 expression alone clearly showed that the expression of this enzyme is sufficient to initiate GalNAc-*O*-glycosylation on endogenous proteins in BY-2 cells. This is in agreement with the results of Castilho et al. (2012) who demonstrated efficient *O*-glycosylation of EPO-Fc in *N. benthamiana* leaves by co-expression with human GalNAcT2 independently of the presence of *Y. enterocolitica* UDP-GlcNAc 4-epimerase and *C. elegans* UDP-GlcNAc/UDP-GalNAc transporter. However, this is at odds with previous data from Z. Yang, Drew, et al. (2012) who showed that co-expression of PaWbpP epimerase was required to glycosylate MUC1-YFP after transient expression in *N. benthamiana* leaves, or from those of Daskalova et al. (2010) who generated transgenic GalNAcT2 *N. benthamiana* plants where both an epimerase and UDP-GalNAc transporter were needed. Z. Yang, Bennett, et al. (2012) also included the epimerase in their stably transformed BY-2 cells and *A. thaliana* plants. The reason for this discrepancy is not clear. However, none of these studies reported on the expression level of GalNAcT2. Although our data showed that the expression of an UDP-GalNAc epimerase is not required to initiate *O*-glycosylation in BY-2 cells, the potential of the epimerase to boost the UDP-GalNAc substrate pool and promote a more efficient *O*-glycosylation is unclear and further experiments are required to clearly decipher this point.

Finally, the results from the lectin blots of Core1 and Core1E cell lines indicate that GalNAc-*O*-glycans can be elongated by expression of C1GalT1, as evidenced by decreased reactivity with HPL and increased PNL signals. We used the same approach as Amano et al. (2008) and Castilho et al. (2012) which consisted of expressing C1GalT1 from *D. melanogaster*. Indeed, this protein does not require the specific COSMC chaperone as the human C1GalT homolog. The loss of the *N*-glycosylation site in hC1GalT1 during evolution may be responsible for the need of a specific chaperone for correct folding in ER (González-Ramírez et al., 2022; Ju & Cummings, 2002). COSMC is closely related to hC1GalT1 and is thought to originate from a gene duplication event prior to its divergence as a chaperone without GT activity (Taujale et al., 2020). When co-expressed in *N. benthamiana*, COSMC and hC1GalT1 lead to a poor elongation of the *O*-glycans compared to expression of DmC1GalT1 alone (Castilho et al., 2012). In accordance, our results show that expression of both hGalNAcT2 and DmC1GalT1 in BY-2 cells leads to the formation of core 1 glycan structures on endogenous proteins.

The effect of introducing the *O*-glycosylation machinery on the fitness of BY-2 cells was roughly assessed using a growth assay. The fresh weight of three independent cell lines obtained for each construct was monitored at different timepoints and in several independent experiments (five for the GalNAc lines and three for the others)

(Figure II.1.4). Compared to ΔXF BY-2 cells, the glycoengineered cell lines grew similarly with no statistical difference (Student's t-test, $p<0.05$). The cells were also observed under microscope and did not show any aberrant shape or size (data not shown).

Our results confirmed that BY-2 cells can be stably engineered to produce glycoproteins with GalNAc-*O*-glycans without obvious deleterious phenotypes. They also demonstrated that BY-2 cells constitute an interesting platform for the production of therapeutic *O*-glycoproteins. Therefore, three relevant reporter proteins will be expressed in the generated cell lines to demonstrate their ability to produce GalNAc-*O*-glycoproteins. This will also enable us to better characterise the new glycoengineered BY-2 cell lines and their *O*-glycosylation profile.

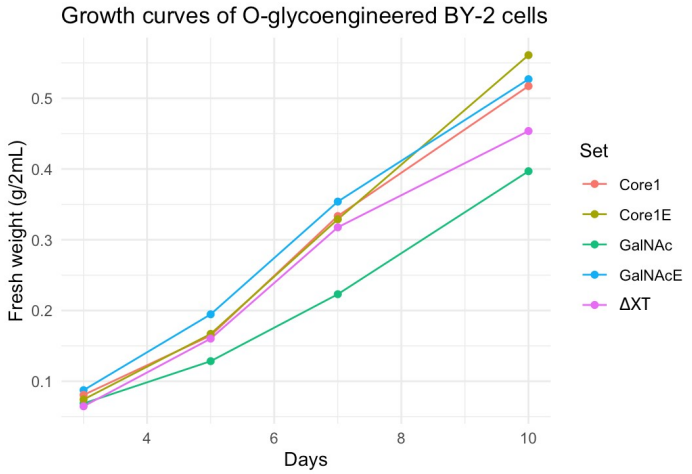


Figure II.1.4: Growth curves of *O*-glycoengineered BY-2 cells. The fresh weight of three independent cell lines transformed with each construct was harvested at 3, 5, 7 and 10 days and the mean for each construct was compared to ΔXF BY-2 cells.

2 GalNAc-*O*-glycosylation of therapeutic proteins

To confirm the results obtained with endogenous BY-2 cell proteins and to better characterise the ability of glycoengineered BY-2 cells to perform correct *O*-glycosylation of therapeutic proteins, several reporter proteins were chosen: human IgA1, EPO and SARS-CoV-2 Spike glycoprotein. EPO and Spike are commercial products used as an erythropoietic drug and a vaccine immunogen, respectively, in while IgA1 is being studied for its unique properties that open up new therapeutic prospects due to their different binding capabilities and effector functions. There are many examples of production of these proteins in the different plant expression platforms (for instance, IgA1 (Göritzer et al., 2017; Karnoup et al., 2005; Ma et al., 1994), EPO (Pires et al., 2012; Weise et al., 2007), Spike (Maharjan & Choe, 2021; Rebelo et al., 2022; Y.-J. Shin et al., 2021)). However, they were not expressed in a system competent for GalNAc-*O*-glycosylation, with the exception of a few reports (Castilho et al., 2012; Dicker et al., 2016).

2.1 Immunoglobulin A1

IgAs are the most synthesized antibodies in humans (with IgA1 accounting for 85% of all IgAs) (Novak et al., 2012). In serum, they are present in a short-lived monomeric form. However, they are mainly found in mucosal surfaces in the form of secretory IgAs. Interactions between the C-terminus tails pieces of the heavy chains enable dimerisation, which is further stabilised by the J chain. Dimeric IgAs interact with the secretory component to form the secretory IgA.

IgA1 is highly glycosylated and characterised by its long hinge region (19 amino acids versus 5 in IgG1), which is GalNAc-*O*-glycosylated *in vivo* with three to six (Thr225, Thr228, Ser232 and Ser230, Thr233, Thr236) sialylated core 1 *O*-glycans in serum and complex elongated *O*-glycans in secretions (Mattu et al., 1998; Novak et al., 2012; Royle et al., 2003). GalNAcT2 is mainly responsible for this modification *in vivo* but GalNAcT1 and T11 may also contribute (Iwasaki et al., 2003; Wandall et al., 2007). *O*-glycosylation is important to confer rigidity and extended conformation to the hinge region (Narimatsu et al., 2010; Royle et al., 2003) and is also proposed to play a role in receptor interaction (Herr et al., 2003; Rouwendal et al., 2016). It also confers enhanced resistance to proteolysis (Reinholdt et al., 1990). Aberrantly glycosylated hinge is prone to aggregation and is responsible for several diseases. For example, hypogalactosylation is responsible for the formation of aggregates that are recognised by autoantibodies and deposited in kidneys leading to renal failure in IgA1 nephropathy (Novak et al., 2012; Xue et al., 2013). IgA1 also contains two *N*-glycosylation sites: N263 located in the CH2 and N459 in the tail piece. Unlike N297

in IgGs, N263 faces outward and is not involved in antibody stabilisation and receptor interaction. N459 is located near the dimer interface and potentially interacts with the J chain, suggesting a role in dimer formation.

IgA1s have several advantages that make them interesting therapeutic candidates. IgA1s are highly effective at recruiting immune cells and delivering potent cytotoxic mechanism (de Sousa-Pereira & Woof, 2019). They can also be used in combination with other Ig subclasses as their receptors differ and their effector functions are therefore different and can be complementary. The long hinge confers enhanced avidity, and the secretory form has higher agglutination capacity and stability. IgA1s are the prime candidate for mucosal applications given their natural prevalence at those sites.

In addition, there are several reports of IgA1 production in plants, with or without the GalNAc-*O*-glycosylation machinery, allowing comparison with our results.

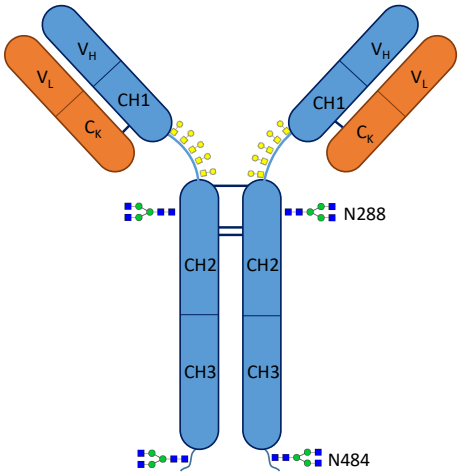


Figure II.2.1 Schematic representation of IgA1 produced in glycoengineered BY-2 cells. The heavy and light chains are represented in blue and orange, respectively. The different domains are indicated (V = variable, C_K = constant kappa and CH = constant heavy). The disulfide bridges are represented by blue lines. *N*-glycosylation sites are indicated, and the hinge region has been arbitrarily represented with five core 1 *O*-glycans.

2.1.1 Genetic constructs

The cDNAs encoding the heavy and light chains (HC and LC) of a chimeric human IgA1 with the variable region from Trastuzumab were received in pEAQ from Prof. Richard Strasser's lab (Dicker et al., 2016). The encoded antibody is composed of the

α1-HC and κ-LC with the HER2-binding variable region from Trastuzumab IgG1. The barley α-amylase signal peptide was fused to the N-terminus of both chains. Both sequences were PCR amplified for cloning into Golden Gate plasmids under the control of *PMA4* promoter and the *N. tabacum Extensin* terminator (*tExt*). *pPMA4* is a strong constitutive promoter from the plasma membrane *ATPase 4* gene of *N. plumbaginifolia,* engineered with two enhancers from *p35S* and a mutated 5'UTR (De Muynck et al., 2009). They were then combined with *GalNAcT2* and *C1GalT1* CDS in binary plasmids along with the *nptII* kanamycin resistance gene. In total, three constructs were generated: They encoded IgA1 alone (IgA), IgA1 + GalNAcT2 (IgAG) and IgA1 + GalNAcT2 + C1GalT1 (IgAC) (Figure II.2.2).

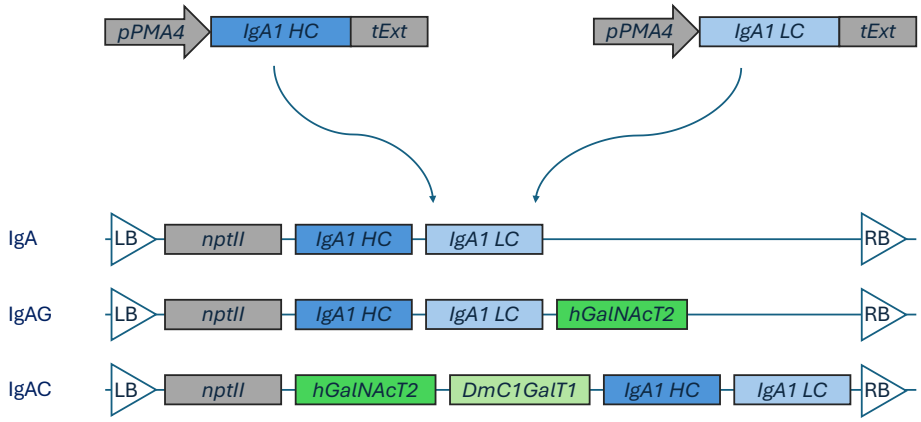


Figure II.2.2: IgA1 *O*-glycovariants constructs. Coding sequences from chimeric human Trastuzumab IgA1 HC and LC were cloned under the control of *pPMA4* and *tExt* in Level 1 Golden Gate vectors. The expression cassettes were then assembled to form the three different binary plasmids with the *O*-glycosyltransferase expression cassettes previously obtained. The plasmids also contain the *nptII* kanamycin resistance gene and are flanked by LB and RB for *A. tumefaciens* mediated transformation.

2.1.2 Production and purification of IgA1 in ΔXF BY-2 cells

IgA1 production in BY-2 cells was tested with or without the *O*-glycosylation machinery and a purification protocol based on Protein A affinity was set up. The purified IgA1 proteins were then used for downstream glycosylation analysis. First, lectin probed blots and enzymatic deglycosylation were performed to roughly characterise the structures of the *N*- and *O*-glycans . Mass spectrometry analysis was then conducted to confirm the results. All three IgA1 constructs were used to transform ΔXF BY-2 cells. Stable transgenic cell lines were obtained and screened by western blot to identify the best producing lines. In addition, the effect of the

medium composition on the accumulation of IgA1 in the extracellular medium was tested. The amount of IgA1 in the EM of BY-2 cells cultivated in MS or D11b medium were compared (Figure II.2.3 A). Lanes 1 and 2 were loaded with EM from a MS culture at 7 and 10 days, respectively, while lanes 3 and 4 were loaded with EM from a D11b culture at the same days. The western blots clearly showed that D11b allowed better accumulation of higher molecular mass IgA1, which supposedly has better integrity (the expected molecular mass of the full IgA1 is 148.5 kDa + 4 *N*-glycosylation sites). This is expected as D11b is a medium that was developed to improve antibody yields in BY-2 cells (Vasilev et al., 2013).

EM from cultures in D11b was used to purify the different IgA glycovariants by affinity on Protein A agarose (ProtA). A western blot and a Coomassie stained gel of the purification are shown in Figure II.2.3 A and B. We observed that IgA1 could be specifically purified using ProtA. Indeed, high amount of IgA1 was recovered in the first elution fractions (E1-E3) and appear to be quite pure as each Coomassie-stained band was also detected by anti-Fab antibodies. However, many bands were observed indicating that the protein is present in several forms. ProtA is routinely used for IgG affinity purification due to its strong and specific binding. However, the binding of IgA is weak. This was well demonstrated in our experiments as a significant signal was still observed in the flow through. To confirm that IgA1 could still be recovered, a follow-up purification was carried out using the flow through as input. A smaller but still high amount of IgA1 could be recovered after a second cycle of purification. The band profile was different between the elutions, which could indicate that the different forms of IgA1 have a different affinity for ProtA (Supplementary figure S2).

The ProtA purified fractions from each glycovariant were analysed by size exclusion chromatography (SEC) on a 25 mL Superdex G200 column. The elution profile is shown in Figure II.2.4 A. The three glycovariants displayed similar profiles and eluted mainly as a broad monodisperse peak between fractions 14 and 17. This peak most likely corresponded to the intact IgA antibody. Two other small peaks were also visible. The first, at fraction 12, was probably the dimeric IgA1s while the second, after fraction 20, was probably composed of degradation fragments. The main peak was slightly shifted among the samples, probably due to differences in glycosylation of the hinge region.

Next, each SEC fraction was analysed by SDS-PAGE with Coomassie staining (Figure II.2.4 B, C and E). A faint band was observed above 200 kDa in fractions corresponding to the first peak (fractions 11-13), again suggesting the presence of dimeric IgA1. The second peak was composed of an intense band above 150 kDa corresponding to the intact antibody and many smaller bands (fractions 14-19).

These smaller bands probably originated from degradation or disassembly of the antibody during sample preparation for the SDS-PAGE. Indeed, non-boiled samples only appeared as bands above 100 kDa (Figure II.2.4, panels D and F and Supplementary figure S3).

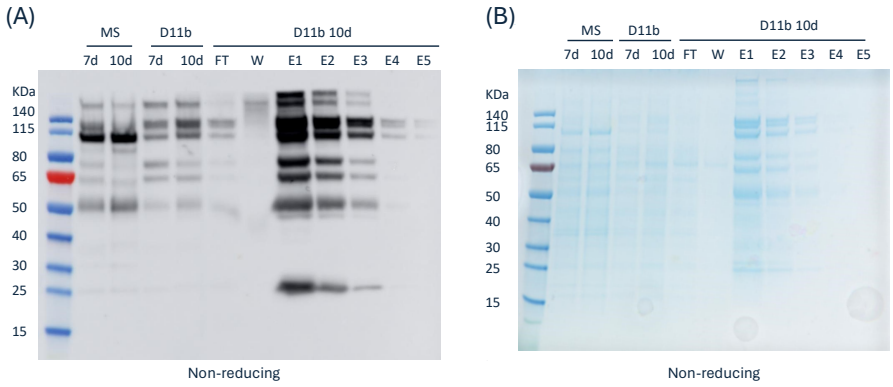


Figure II.2.3: Production and purification of IgA1 in BY-2 cells. IgA1 was purified from BY-2 cells culture medium using ProtA after 10 days in MS or D11b medium. The ProtA agarose beads were incubated overnight with the clarified, buffered extracellular medium. The beads were washed in 10 mL Tris-Cl pH 8 (W) and IgA1 was eluted at pH 3 in glycine 35 mM in five 500 µL fractions(E1-E5) . EM from MS and D11b cultures and purification fractions were analysed using (A) anti-Fab western blot and (B) Coomassie stained SDS-PAGE. The productivity of IgA1 in MS and D11b is compared, and a 50 mL D11b culture was purified. Each lane was loaded with 35 µL of each fraction.

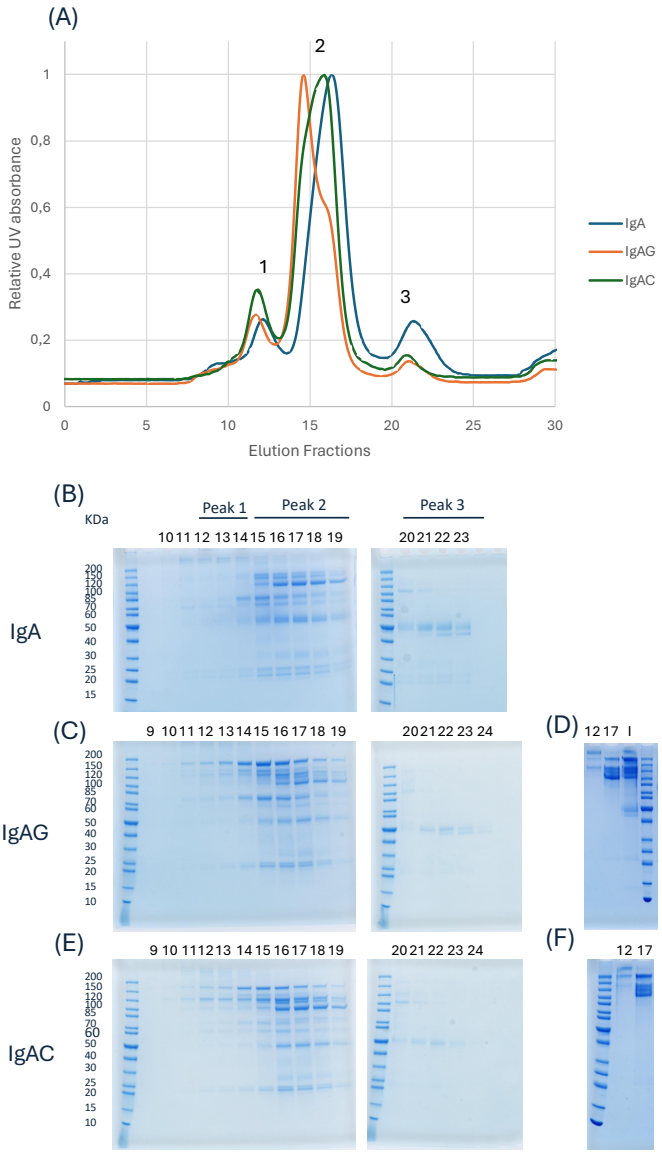


Figure II.2.4: Size exclusion chromatography of IgA1 *O*-glycovariants. The first ProteinA purified fraction E1 of IgA1 (200 µg) was ran on a 25 mL Superdex G200 column. (A) Elution profile of each glycovariant. (B, C and E) SDS-PAGE analysis of SEC fractions. Fractions of 0.5 mL were collected and analysed by SDS-PAGE and Coomassie blue staining after boiling of the samples (40 µL loaded in each lane). Fractions 11-13 corresponded to the first peak, 14-19 to the second and 20-23 to the third. (D and F) Samples from the peaks 1 and 2 of IgAG and IgAC and from the input (I) were analysed by SDS-PAGE and Coomassie blue staining without boiling during the sample preparation.

2.1.3 Analysis of glycosylation profile by lectin blotting

To assess the glycan identities of each *O*-glycovariant, blots of purified IgA1 were probed using two different lectins that bind specific *O*-glycans. Figure II.2.5 shows three blots, probed with anti-α1-HC antibody, HPL (specific to α- and β-GalNAc) and PNL (specific to unsubstituted core 1). The first blot (A) shows that all lanes were loaded with a similar amount purified IgA1. The second blot (B) shows that GalNAcT2 was required for detection of the antibody with HPL. Similarly, GalNAT2 and C1GalT1 were needed for detection using PNL as observed in the third blot (C). These results clearly indicated that IgA1 could be *O*-glycosylated with GalNAc-*O*-glycans in BY-2 cells when the appropriate enzymes were expressed. In addition, both GalNAc and core 1 *O*-glycans could coexist on IgA1 expressed with both GalNAcT2 and C1GalT1 as HPL and PNL signals were observed for IgAC. Alternatively, the signals could be due to a small reactivity of HPL toward core 1 glycans (Sanchez et al., 2006). The same bands were detected with the lectins and the anti-α1-HC antibody, suggesting that *O*-glycans are found on the HC. However, the band around 35 kDa was only weakly detected by the antibody but gave a strong signal with both lectins. This band therefore contained the hinge region but had lost many epitopes bound by the antibody, which is not surprising as it lost half of its molecular mass.

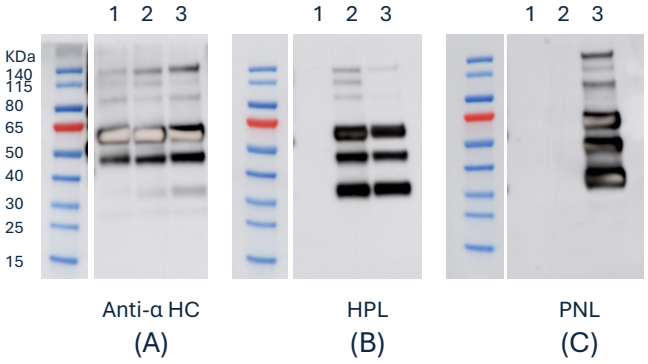


Figure II.2.5: *O*-glycosylation profile of purified IgA1 analysed by western blot and lectin blots. Blots of purified IgA (1), IgAG (2), IgAC (3) in reducing conditions. Blot (A) is probed using anti-α-HC antibody, (B) using HPL lectin and (C) using PNL lectin. Each lane was loaded with 1,25 µg of purified IgA1.

2.1.4 Analysis of the glycosylation profile by *in vitro* deglycosylation

IgA1 carries two *N*-glycosylation sites on each Fc domain, one in the CH2 and one in the tail piece. To characterise the *N*-glycans present on IgA1 produced in ΔXF BY-2 cells, de-*N*-glycosylation was carried out using either EndoH or PNGaseF. EndoH cleaves high-mannose *N*-glycans while PNGaseF can cleave all glycan types except those containing core α(1,3)-fucose. Figure II.2.6 A shows a western blot of IgA and IgAG treated with the glycosidases. A small molecular mass shift, which could correspond to the release of two *N*-glycans (approximately 5 kDa), was observed for the PNGaseF-treated samples but not for EndoH-treated ones. These results indicate two features of IgA1 *N*-glycans: (1) the *N*-glycans are hybrid or complex as they cannot be released by EndoH and (2) they do not contain core α(1,3)-fucose as they are cleaved by PNGaseF, which is expected in ΔXF BY-2 cells. Moreover, there was no visible difference between the ΔXF and *O*-glycoengineered cell lines, suggesting that GalNAcT2 expression had not impacted *N*-glycosylation.

To confirm the identity of core 1 *O*-glycans in IgAC, *in vitro* de-*O*-glycosylation was carried out using an *O*-glycosidase that specifically cleaves this glycan from glycoproteins. As shown on Figure II.2.6 B, there was a reduction of the PLN signal after 4 h of treatment with *O*-glycosidase compared with the control sample, indicating the removal of core 1 *O*-glycans from the IgA. In addition, this also confirmed that the PNL signals were probably due to core 1 *O*-glycans. A weak signal was still visible after treatment with 80,000 units of *O*-glycosidase, suggesting that the clustering of *O*-glycans in the hinge region reduced or abolished the activity of the enzyme at certain glycosites. These results further confirmed that co-expression of GalNAcT2 and C1GalT1 enabled BY-2 cells to produce IgA1 with core 1 *O*-glycans.

2.1.5 Analysis of the glycosylation profile by mass spectrometry

IgA1 purified proteins were analysed by MS to confirm the data obtained by lectin blots and to allow more precise identification of the *N*- and *O*-glycosylation status of the antibody. MS analysis of IgA1 glycopeptides was performed by the BOKU Core Facility Mass Spectrometry (Vienna, Austria). Precise determination of the glycans identity and the number of sites occupied was carried out by LC-MS/MS after trypsin and GluC digestion. Firstly, glycopeptides containing the first and second sites of *N*-glycosylation (AN*LTCTLTGLR in the CH2 and LAGKPTHVN*VSVVMAE in the tail piece) were analysed.

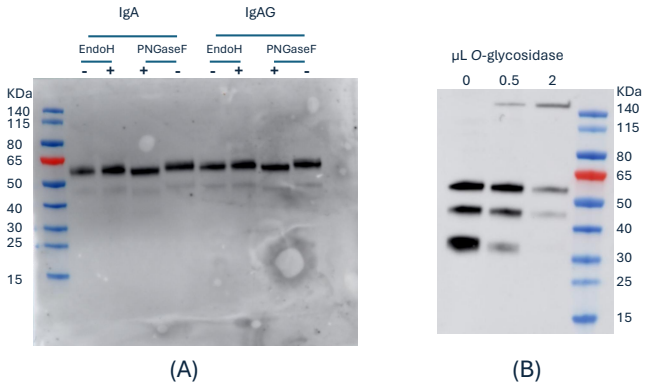


Figure II.2.6: *in vitro* deglycosylation of IgA1. (A) Purified IgA and IgAG were treated with EndoH and PNGaseF. Purified proteins (125 ng) were treated with 0.1 µL of EndoH or PNGaseF (500 u/µL). The samples were loaded on gel and the blot was probed with anti-α-HC. (B) Purified IgAC was treated with *O*-glycosidase. Purified proteins (100 ng) were treated with 0, 0.5 or 2 µL of *O*-glycosidase (40000 u/µL) and analysed by PNL-probed blotting. The proteins were denatured by 10 minutes of boiling in 0.5% SDS and 40mM DTT prior to enzymatic reaction.

Figure II.2.7 shows the mass spectra of glycopeptides corresponding to the CH2 site (N263) of the three different IgA1 *O*-glycovariants. The overall profile was similar between the three variants, with GnGn and GnM being the most abundant glycans (50-75%, Supplementary table S1). This was in agreement with the *in vitro* de-*N*-glycosylation assay, which revealed that IgA1 carries mostly hybrid or complex *N*-glycans. The other *N*-glycans detected were mostly oligomannoses (Man4 to Man8) and truncated *N*-glycans (MM). However, there were several differences in abundance of some glycans between the IgA1 glycovariants. These differences were most likely due to variations across samples or technical variability as the analysis was done only once. They could also be the result of the *O*-glycosylation engineering although this is unlikely since the glycosylation sites are not adjacent and this has never been reported in the literature. Analysis of IgA1 *N*-glycosylation profile in other cell lines will be necessary to confirm this. The spectra also showed a small amount of MU (± 5%) and single Gn (<3%) glycans, which was probably due to fragmentation of other *N*-glycans. Notably, N263 was fully occupied.

Glycosylation analysis of the tail piece site gave results of poorer quality (Supplementary table S1). Indeed, the signal intensities were low and led to poor quantification. For IgA and IgAG, the glycan composition corresponds to GnGn, GnM and oligomannose structures, but it could not be determined for IgAC. This

glycosylation site was located at the C-terminus of the protein in a disordered region and was found to be cleaved for an IgA1 expressed in *N. benthamiana* (Dicker et al., 2016). It was probably also partly cleaved in our samples as it was not found in the peptide mapping. This was problematic as the tail piece and its glycosylation are implicated in the formation of dimeric and secretory IgA (de Sousa-Pereira & Woof, 2019; Göritzer et al., 2020a). On the contrary, some applications benefit from monomeric IgA1. For instance, for cancer immunotherapy, some studies mutated or removed the tail piece to obtain only the monomeric form of IgA1 (Rouwendal et al., 2016). Underglycosylation of Igs produced in *N. benthamiana* leaves is often observed and can be overcome by expression of the single unit oligosaccharyltransferase (OST) from *Leishmania spp* in *N. benthamiana* (Beihammer et al., 2023; Castilho et al., 2018). While this is not the case for N263 that exhibits high occupancy, the expression of *Leishmania major* OST led to efficient glycosylation of N459, which promoted dimerization.

Next, the glycosylation status of the hinge region was analysed (peptide: HYTNPSQ DVTVPCPVPSTPPTPSPSTPPTPSPSCCHPR). The PTM search included HexNAc and Hex for GalNAc-*O*-glycosylation as well as Pro hydroxylation and Pent as the hinge region was shown to be modified by pentoses (most likely arabinoses) on Hyps when expressed in plants (Göritzer et al., 2017; Karnoup et al., 2005). The mass spectra of the glycopeptides corresponding to the hinge region of the three *O*-glycovariants are shown in Figure II.2.8. Panel A shows the superimposed spectra of the three MS experiments. A clear difference in the distribution of the masses of the hinge peptide is observed for the *O*-glycovariants. Panel B presents the spectrum of IgA1 produced in ΔXF BY-2 cells (IgA). The hinge region was heavily modified with plant-specific hydroxyproline glycosylation. Indeed, more than 90% of the identified peptides were modified with a least one Hyp and 80% were glycosylated with up to 18 pentoses (Supplementary table S2). On average, 40% of the 14 Pro residues found in the glycopeptide were glycosylated. The most abundant glycoforms had 5-6 Hyps with 8-13 pentoses attached. A small amount of hexoses was also detected, which could correspond to plant Ser-galactosylation (Supplementary table S2). The mass spectrum of the hinge from IgAG is then shown in panel C (Supplementary table S3). Mass shifts corresponding to HexNAc addition confirmed that expression of GalNAcT2 initiated GalNAc-*O*-glycosylation on the hinge region of IgA1. While the amount of unmodified hinge was similar, the variety of glycopeptides was more heterogenous. Hydroxyproline and GalNAc-*O*-glycosylation coexisted but with a reduction in the extent of plant-specific glycosylation (1-3 Hyps with 0-3 pentoses) compared to IgA. Finally, panel D shows the results of IgAC hinge region (Supplementary table S4). A strong reduction of Pro hydroxylation and a nearly

complete disappearance of pentoses was observed. All the identified peptides were modified. Additionally, Hex were detected that likely correspond to the galactose residue of core 1 *O*-glycans. The formation of the core 1 seems to be very efficient, with most of the GalNAc being modified by this residue (0.93 Hex/HexNAc residue ratio). Three to five sites were occupied in accordance with the glycosylation of IgA1 observed in humans (Mestecky et al., 2008).

Figure II.2.9 A summarises the abundance of IgA1 hinge peptides modified with different numbers of Pro (0-8) in each sample. This graph clearly shows that GalNAc-*O*-glycosylation inhibited Pro hydroxylation of the IgA1 hinge region, and that the strongest reduction was achieved with core 1 *O*-glycans with almost 50% of peptides with no Hyp for IgAC. Furthermore, the elongation of GalNAc-*O*-glycans seemed to promote the *O*-glycosylation (Figure II.2.9 B). Indeed, the proportion of unmodified peptide was decreased (from 15% to 0%) when comparing IgAG to IgAC. This result was surprising as elongation of GalNAc-*O*-glycans was thought to inhibit initiation of neighbouring sites (Perrine et al., 2009). However, only one experiment was performed and thus quantification results should be interpreted with caution. Analysis of IgA1 glycovariants produced in different cell lines would reduce the variance and provide a clearer picture.

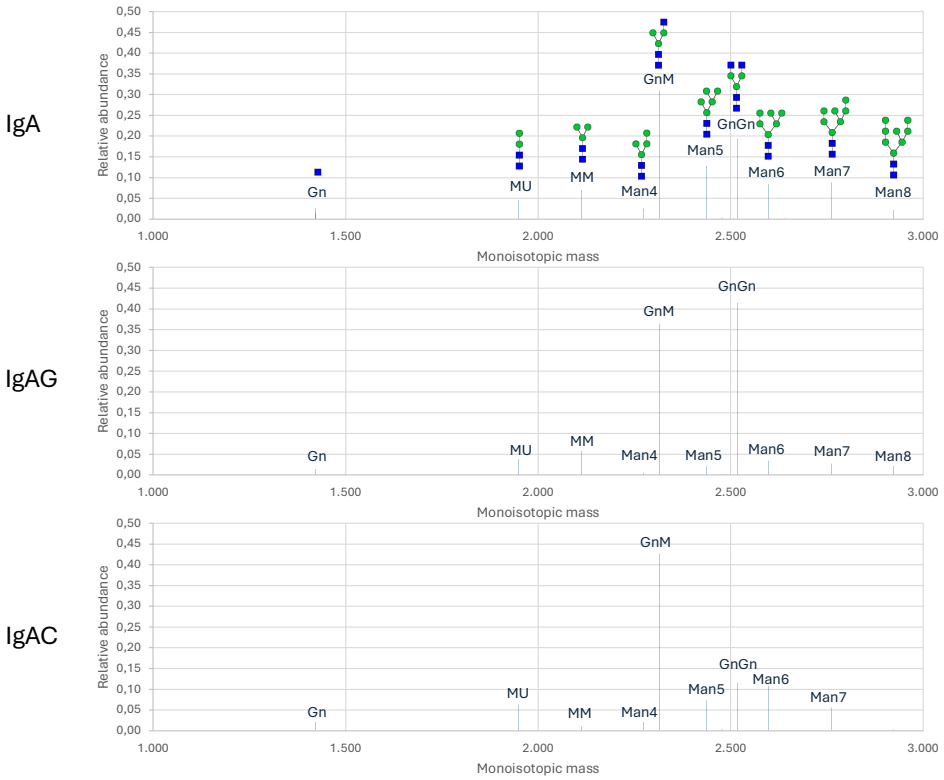


Figure II.2.7 *N*-glycans composition of the CH2 site of IgA1 *O*-glycovariants analysed by Orbitrap MS/MS. GluC/tryptic peptides containing the N263 site (AN*LTCTLTGLR) identified for each IgA1 *O*-glycovariants were reported in the spectra. The given spectra were cleaned to show only peaks corresponding to IgA1 glycopeptides.

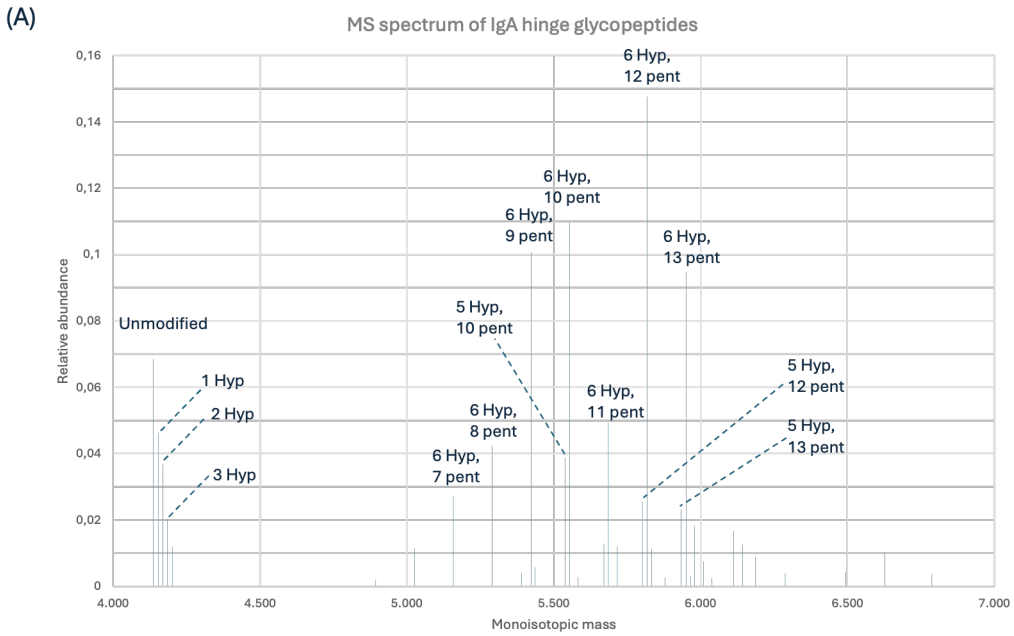


Figure II.2.8: *O*-glycans analysis of the hinge region of IgA1 *O*-glycovariants from BY-2 cells by Orbitrap MS/MS. Glycopeptides corresponding to the GluC/tryptic peptides of the hinge region (HYTNPSQDVTVPCPVPSTPPTPSPSTPPTPSPSCCHPR) identified were reported in the MS spectra. The peptides with a relative abundance above 2% were annotated. Spectra of (A) IgA, (B) IgAG and (C) IgAC showing the identified PTMs.

(B)

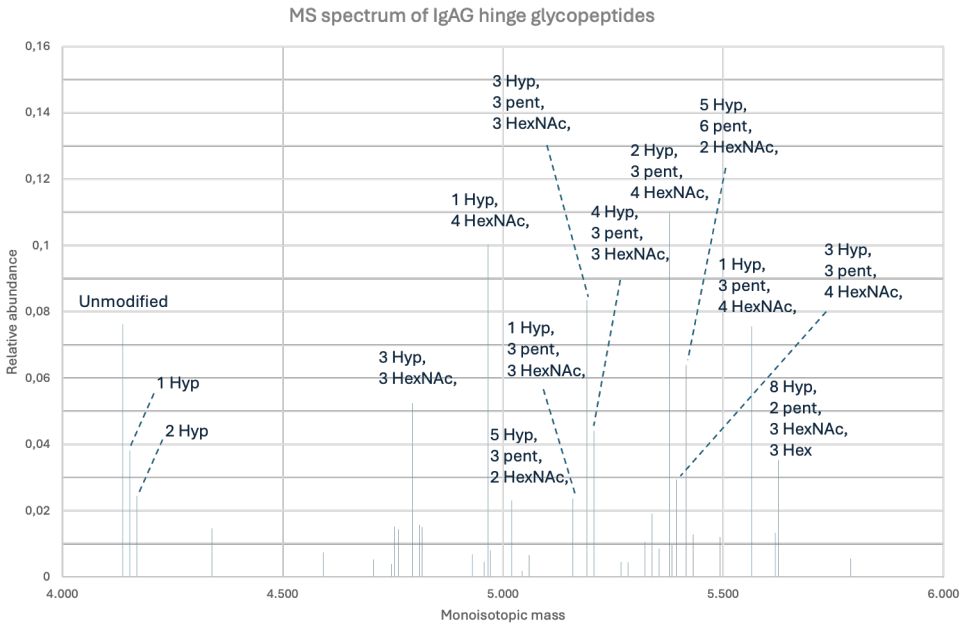


Figure II.2.8: Continued

(C)

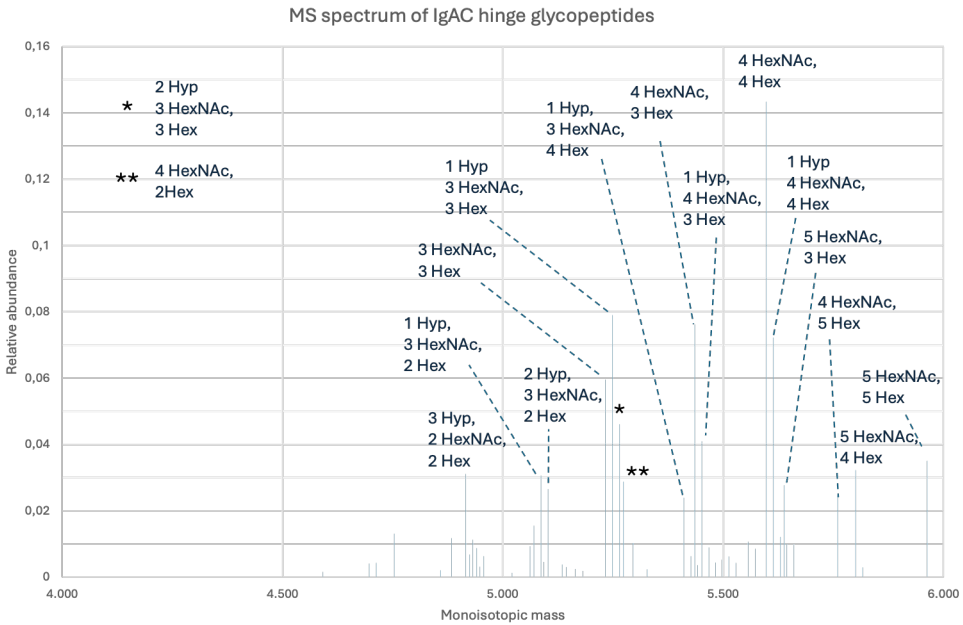


Figure II.2.8: Continued

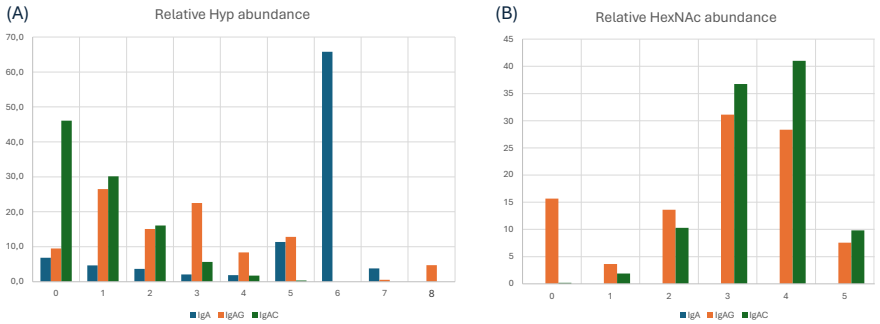


Figure II.2.9: Abundance of *O*-glycosylated hinge peptides. Sum of the relative abundance of the number of (A) Hyp and (B) HexNAc residues in the hinge region of each IgA1 glycovariants measured by MS.

2.2 EPO

EPO is a 165 amino acids glycoprotein that is a hormone, a cytokine and a growth factor. It is the main regulator of erythropoiesis but also has anti-apoptosis and anti-inflammatory properties that protect tissues (Chateauvieux et al., 2011). Many tissues produce EPO at the periphery of injury sites following hypoxia, trauma, or inflammation. Recent studies have also suggested an effect of EPO on immune cells (Peng et al., 2020). Its action is mainly regulated by interaction with the homodimeric EPO receptor (EPOR) to trigger proliferation and differentiation of erythrocytes. However, EPOR can also form heterodimeric complexes with other partners such as the β common receptor that regulates the other function of EPO (Suresh et al., 2020).

This is why EPO, its mutants and derivatives are very important therapeutics. In fact, EPO is a leading drug for the treatment of anaemia (Walsh & Walsh, 2022), with potential in the treatment of cancer and neurological diseases (Peng et al., 2020). The human EPO glycoprotein contains 3 *N*- and 1 *O*-glycosylation sites (Asn24, Asn38, Asn83 and Ser126). The *N*-sites are modified with sialylated complex type *N*-glycans. Although they have a negative impact on the binding to EPOR, the effect of sialic acids on the serum half-life of EPO increases the potency *in vivo* (Darling et al., 2002). Darbepoetin α is an EPO-based drug with two additional *N*-glycosylation sites that has a threefold longer half-life (Chateauvieux et al., 2011). Little is known about the biological relevance of the *O*-site, but it is important for the proper secretion of the protein (Dubé et al., 1988). It is occupied with a sialylated core 1 *O*-glycan, which could also contribute to the longer half-life. In CHO, the site occupancy for *O*-glycosylation is 50% (Taschwer et al., 2012).

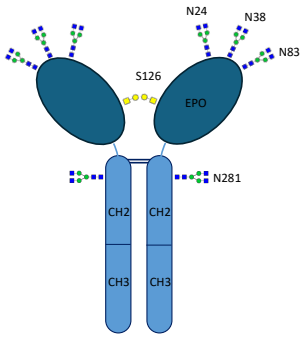


Figure II.2.10. Schematic representation of the EPO-Fc fusion protein. The Fc part (hinge, CH2 and CH3) is represented in light blue, while EPO is in dark blue. Interchain disulfide bridges are represented by dark blue line. N-glycosylation sites are indicated with complex type structures without core fucose and xylose, and the O-glycosylation site with core 1 O-glycan.

2.2.1 Genetic constructs and BY-2 cell transformation

To facilitate the purification of EPO, the protein was fused at its C-terminus with the Fc part of human IgG2 (hinge-CH2-CH3), which contains a conserved *N*-site in the CH2 region (Magy et al., 2014). The EPO coding sequence we used did not contain the sequence encoding the endogenous signal peptide and has a mutation leading to G140R substitution. The expression cassette was constructed by assembling the sequence of EPO-Fc with the sequence encoding the signal peptide of the *Medicago sativa* protein disulfide isomerase (PDI). The composite ORF was under the control of *pPMA4* and the terminator of *Saccharomyces cerevisiae* regulatory subunit for the plasma membrane H^+-ATPase (*tPMP1*). This terminator was shown to interact with ER components and is recruited to the ER even in the absence of a coding sequence (Loya et al., 2008). We therefore thought that it could lead to an efficient translation of secretory proteins. The EPO-Fc expression cassette was then combined with the cassettes encoding the GalNAc-*O*-glycosylation enzymes and *hptI* hygromycin resistance cassette to form four different constructs: EPO-Fc (EPO), C1-GalT1 + EPO-Fc (EPO-C), WbpP + EPO-Fc (EPO-E), WbpP + C1GalT1 + EPO-Fc (EPO-CE) (Figure III.2.11). All four constructs were used to transform GalNAc BY-2 cells (obtained in 1.2), while EPO-Fc was also used to transform ΔXF BY-cells.

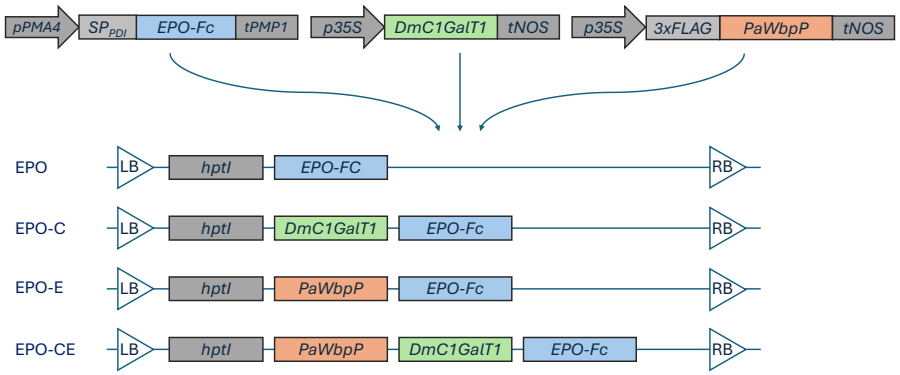


Figure II.2.11: EPO-Fc construction with GalNAc-*O*-glycosylation enzymes. Coding sequence from EPO and IgG2 Fc were fused with the sequence encoding the PDI signal peptide under the control of *pPMA4* and *tPMP1* in Level 1 Golden Gate vectors. The expression cassettes were then assembled to form the four different constructs in binary plasmids with C1GalT1 and WbpP expression cassettes previously obtained. The transgenes also contained the *hptI* hygromycin resistance gene for selection in GalNAcT2 BY-2 and were flanked by LB and RB for *A. tumefaciens* mediated transformation.

2.2.2 Accumulation and purification of EPO-Fc in the EM of in BY-2 cells

The production and accumulation of EPO-Fc in the EM of the generated BY-2 cells was assessed and elites cell lines were screened by western blot using an anti-Fc antibody. Figure II.2.12 A shows the screening performed on EM of ΔXF cell lines. In reducing conditions, EPO-Fc accumulated as two bands: a lower band around 35 kDa probably corresponding to the cleaved EPO-Fc protein, and an upper band around 60 kDa corresponding to the intact fusion protein (expected molecular mass: 44.2 kDa + 4 *N*-glycosylation sites). Accumulation of EPO-Fc in the EM was then compared between cultures in MS or D11b medium after 7 and 10 days by western blot using anti-Fc (Figure II.2.12 B). The upper band was detected in D11b EM, suggesting a better accumulation of the intact proteins in the D11b medium. The intensity of the lower bands was similar between the two media, although two close bands were observed in MS media and the band from the D11b medium was slightly bigger, suggesting a different proteolytic cleavage.

ProtA was used to purify EPO-Fc from the EM of ΔXF BY-2 cells grown in D11b through affinity with the Fc tag. Figure II.2.13 shows a western blot with anti-Fc probed and a Coomassie stained gel of the different fractions. We observed that that EPO-Fc could be efficiently purified using ProtA. The elution fractions 1 and 2 contained the two major bands of 35 kDa and 60 kDa, but also two additional bands with smaller masses, which were also recognised by the anti-Fc.

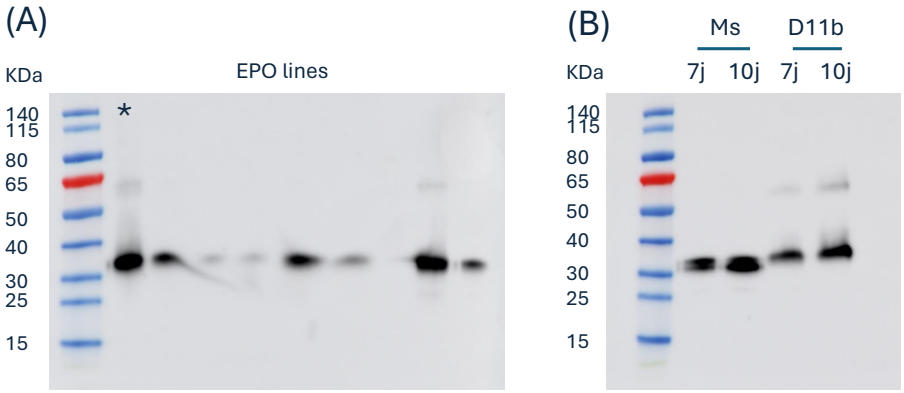


Figure II.2.12: Production of EPO-Fc in BY-2 cells. (A) Western blotting of EM from EPO producing ΔXF BY-2 cell lines probed with anti-Fc antibodies. (B) Comparison of the accumulation of EPO-Fc in MS and D11b media from the cell line denoted by the asterisk in (A). Each lane was loaded with 60 µL of EM. Both blots are in reducing conditions.

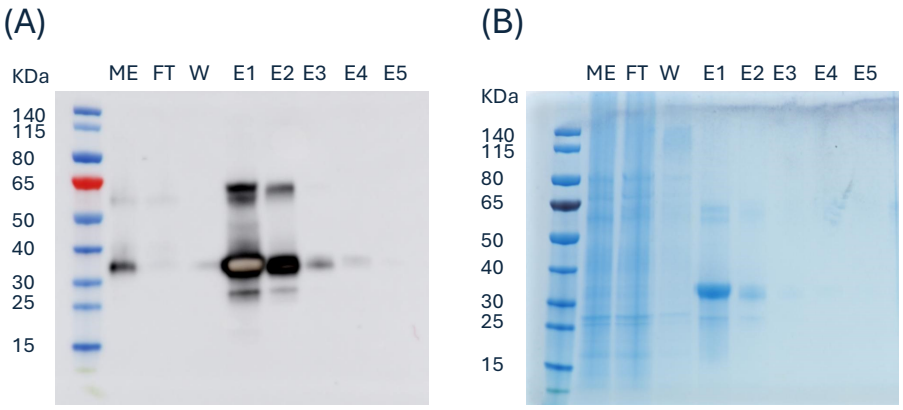


Figure II.2.13: Purification of EPO-Fc from BY-2 cells EM. The EM from a 50 mL culture was clarified and incubated with 200 µL of ProtA beads overnight at 4°C. The beads were then washed with 0.1 M Tris-Cl pH 8, and EPO-Fc eluted with pH 3 in glycine 100 mM. Thirty-five µL of each fraction was analysed by reducing (A) anti-Fc western blot and (B) Coomassie stained gel.

2.2.3 Analysis of EPO-Fc *N*-glycosylation profile by *in vitro* deglycosylation

Enzymatic *in vitro* deglycosylation was carried out on the EM to roughly assess the glycosylation status of EPO. The *N*-glycosylation of EPO-Fc was analysed by *in vitro* deglycosylation with EndoH and PNGaseF. Figure II.2.14 shows the deglycosylation profile of the 35 kDa EPO-Fc band produced in ΔXF BY-cells. A small shift, corresponding to the loss of one *N*-glycan was observed after treatment with PNGaseF but not with EndoH, indicating that *N*-glycans were of the complex type without core fucose as expected for a ΔXF cell line. This small shift also suggests that the lower EPO-Fc band contained only the Fc *N*-glycosylation site. The higher band could not be analysed in this assay due to its low abundance in the EM.

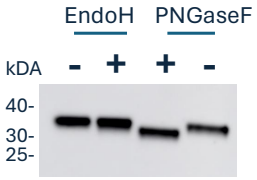


Figure II.2.14: *In vitro* deglycosylation of EPO-Fc produced in ΔXF BY- cells. Extracellular medium of EPO-Fc producing BY-2 cells was treated with EndoH and PNGaseF. EM (36 µL) was treated with 0.1 µL of EndoH or PNGaseF (500 u/µL) and analysed by anti-Fc western blotting. The proteins were denatured by 10 minutes of boiling in 0.5% SDS and 40 mM DTT prior to enzymatic reaction.

2.2.4 Generation of *O*-glycosylated EPO-Fc and analysis by lectin blotting

Next the *O*-glycosylation status of EPO-Fc expressed with the *O*-glycosylation enzymes was investigated. The genetic constructs encoding the EPO-Fc and *O*-glycosylation enzymes were used to transform GalNAc BY-2 cells and their *O*-glycosylation capacity was analysed by western (anti-GalNAcT2 and anti-FLAG) and HPL lectin blotting (Supplementary figure S4). Anti-GalNAcT2, anti-FLAG and HPL showed similar results compared to the GalNAc and GalNAcE cell lines (Figure II.1.2) indicating that the generated cell lines were able to perform GalNAc-*O*-glycosylation. Protein A purified EPO-Fc from these cell lines was analysed by lectin blotting to detect modification with GalNAc-*O*-glycans. The flowthrough that contains endogenous BY-2 proteins was used as a positive control of *O*-glycosylation. Figure II.2.5 shows blots probed with HPL (A) or anti-Fc antibody (B) in reducing (+) and non-reducing conditions (-). Compared to the flowthrough used as positive control (reduced), HPL signal could not be detected for EPO-Fc purified from GalNAc BY-2 cells, although EPO-Fc was present as evidenced by the anti-Fc blot. Similar results were obtained when co-expressing EPO-Fc and C1GalT1 in a GalNAcT2 cell line and probing the blot with PNL (data not shown). The reasons for these results were not clear, especially as the expression of GalNAcT2 allowed production of GalNAc-*O*-glycosylated EPO in *N. benthamiana* (Castilho et al., 2012). As a significant proportion of the protein was cleaved, it was possible that the *O*-glycosylation site was absent on the majority of the purified proteins. A more precise determination of the identity of both EPO-Fc bands recovered after purification was required to answer this question.

2.2.5 Characterization of protein A purified EPO-Fc

To understand why *O*-glycosylated EPO-Fc could not be detected by lectins when the protein was produced in GalNAc BY-2 cells, the precise identity of the purified protein was determined. The two bands observed in the E1 elution fraction were separated by SDS-PAGE and analysed by QTOF MS after trypsin digestion. Figure II.2.16 A shows the sequence coverage corresponding to the peptides identified. Both bands have a full-length C-terminal Fc including the hinge region. Peptides corresponding to the EPO and Fc regions were identified in the upper 60 kDa band. In contrast, no peptide corresponding to EPO was identified in the lower 35 kDa band, suggesting that EPO-Fc was cleaved in the 35 kDa band, with most of the EPO being removed. However, a deglycosylation step prior the trypsin digestion would have facilitated the identification of some glycosylated peptides.

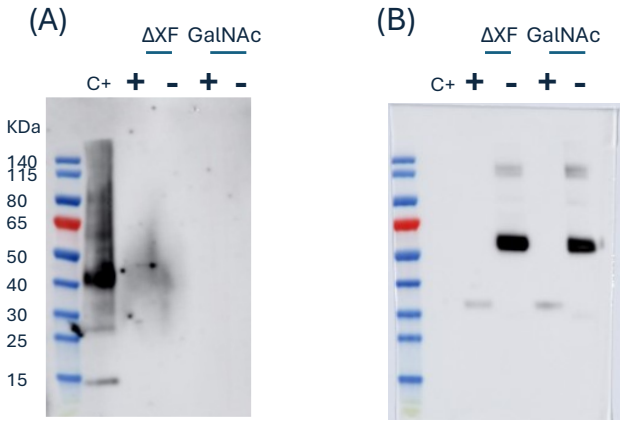


Figure II.2.15: Analysis of EPO-Fc GalNAc-*O*-glycosylation by lectin blotting. EPO-Fc purified from a GalNAc BY-2 cell line or from a ΔXF BY-2 cell line using ProtA were analysed by blotting probed with (A) HPL or (B) anti-Fc antibodies under reducing (+) and non-reducing (-) conditions. The flowthrough from the purification of EPO-Fc from the GalNAc BY-2 cell line was used as a positive control (C+) for the lectin. Each lane was loaded with 50 µL of sample.

In addition, the peptides were screened for hydroxyproline modification. A single hydroxyproline conversion was observed in EPO region at Pro122, which has already been observed on EPO produced in *N. benthamiana* and *P. patens* (Castilho et al., 2011; Weise et al., 2007). In the Fc region, 14 hydroxyprolines were identified, five of which are located in the Pro rich hinge region, which could be the site for potential subsequent glycosylation. The prolines were 5 to 65% hydroxylated in the identified peptides. To our knowledge, this was never described for IgGs produced in plants.

Next, to further characterise the purified EPO-Fc, and particularly its oligomerisation state, the first elution fraction (E1) was analysed by SEC. The panel A of Figure II.2.17 A shows the SEC elution profile. Purified EPO-Fc eluted as a monodispersed peak centred around fraction 9 (Peak 1). The fractions were then analysed by western blot using anti-Fc (Figure II.2.17 B and C) and anti-EPO (D and E) antibodies. The blots were performed under reducing and non-reducing conditions to evaluate the molecular mass of the protein fragment (which provides information on the cleavage site) and the dimerisation state of the Fc portion. Fractions 9-11, corresponding to the peak 1, were composed mainly of cleaved EPO-Fc dimers as shown by the comparison between the anti-Fc western blots in both conditions. However, this peak also contained some full-length EPO-Fc, but its abundance was very low. Interestingly, the anti-EPO antibody showed the presence of some EPO epitopes on the cleaved EPO-Fc 35kDa band. Since the polyclonal anti-EPO antibody was raised

against the C-terminal part of EPO (amino acids 93 to 193), this is compatible with a cleavage occurring upstream of the fusion with Fc, in the EPO domain, and leaving a small portion of EPO attached to the Fc domain. This was in agreement with the observed molecular mass of this 35 kDa band, which was slightly higher than the expected molecular mass of a free Fc. However, no peptide corresponding to EPO was identified in the MS analysis of the 35 kDa band. Peptides corresponding to the last thirty amino acids of EPO could neither be detected, even in the full-length 60 kDa band, suggesting that the experimental MS procedure was not appropriate for identifying this sequence. The low sensitivity of the anti-EPO antibody used probably prevented the detection of the full-length EPO-Fc in the 60 kDa band.

This was confirmed by western blot analysis using different antibodies (Figure II.2.16 B). The same blot was probed with a monoclonal antibody raised against the N-terminus of EPO (left panel), the previously used polyclonal anti-EPO antibody that targets the C-terminal part of EPO (centre panel) and the anti-Fc antibody (right panel). This confirmed that the 60 kDa band corresponded to the full-length fusion protein, although it was not detected with the C-terminal antibody. This also supported that the lower band contained a N-terminus cleaved EPO fused to Fc.

The instability of the Fc fusion proteins when produced in plants was previously reported. The cleavage often occurred in the vicinity of the fusion site and was thought to be due to the extended structures of the hinge or linker region, as well as physical constraints caused by dimerization (Gattinger et al., 2021). The Fc fusion was useful for easily purifying EPO, which was required for downstream analysis of glycosylation. However, cleavage of a significant part of EPO in the purified EPO-Fc sample probably resulted in the loss of the Ser126 *O*-glycosylation site, making this substrate unsuitable for assessing the *O*-glycosylation capacity of the generated BY-2 cell lines.

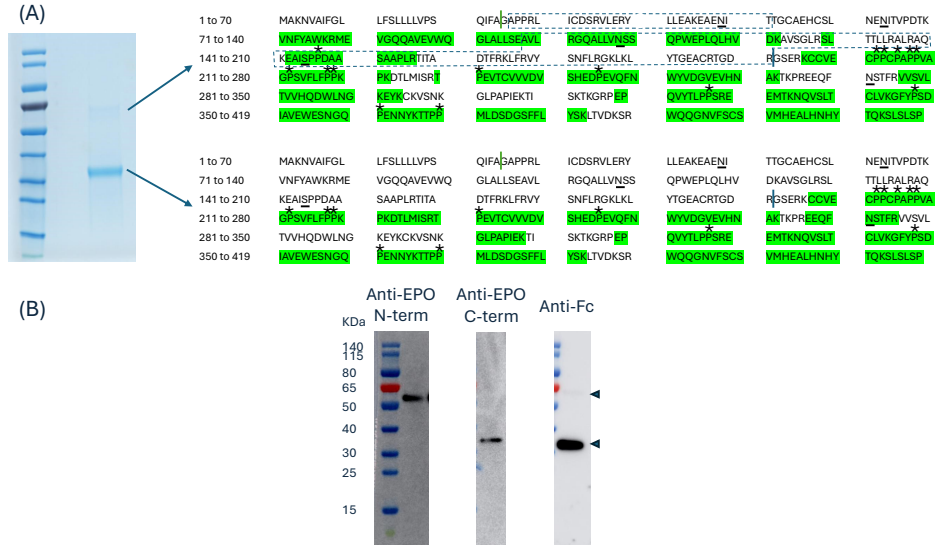


Figure II.2.16: Analysis of EPO-Fc size variants. (A) MS analysis. The EPO-Fc of different sizes were separated by SDS-PAGE, the bands were isolated and digested in gel with trypsin. The peptides were analysed by QTOF MS in DDA or HDMSE modes. The identified peptides are coloured in green. The identified Hyp residues are highlighted with asterisk and the junction between EPO and Fc is marked by a straight blue line. The predicted end of the signal peptide is marked by a straight green line and endogenous glycosylation sites underlined. (B) Western blot analysis of purified EPO-Fc using different antibodies: a monoclonal anti-EPO raised against the N-terminus of the protein (left blot), a polyclonal anti-EPO raised against the C-terminus (centre) and the anti-Fc antibody (right). The same amount of protein was loaded for each blot and reducing conditions were used. The epitopes recognised by the N-terminal (monoclonal) and C-terminal (polyclonal) antibodies are shown in boxes on the sequence coverage.

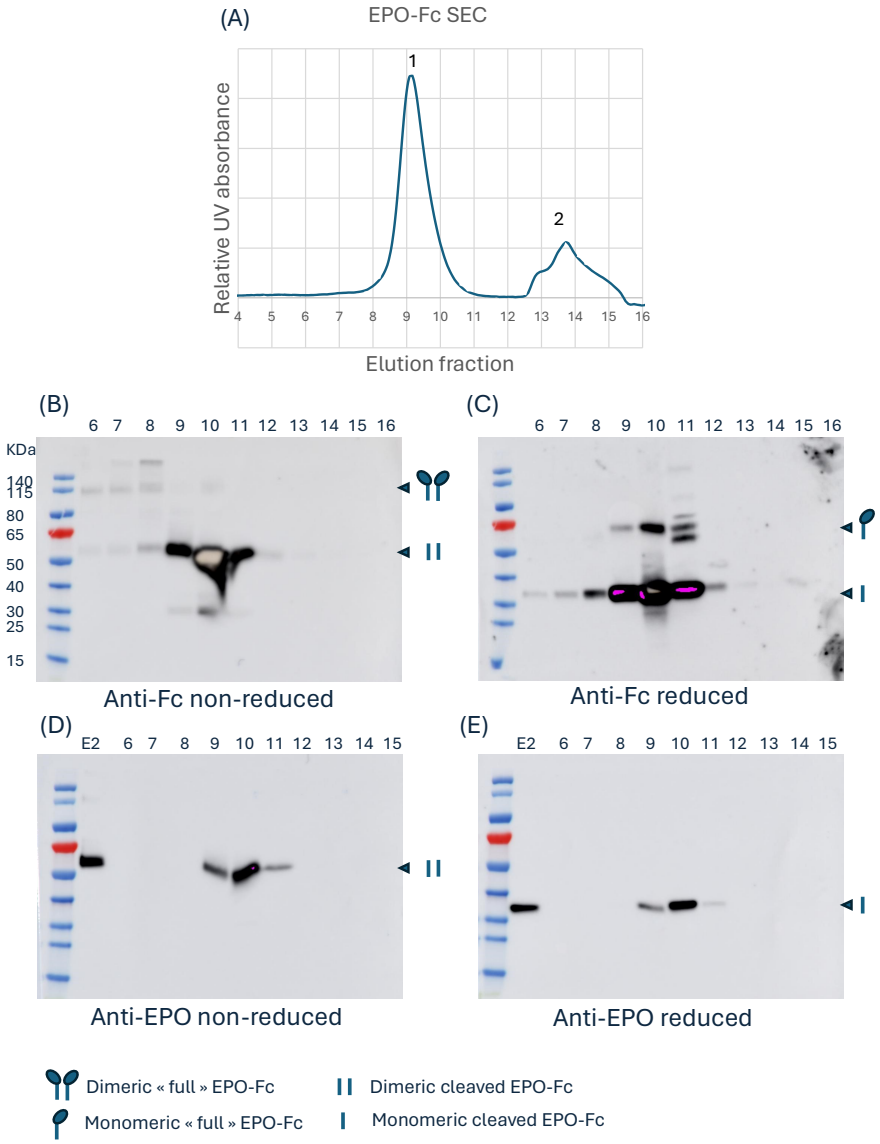


Figure II.2.17: SEC analysis of purified EPO-Fc. The first ProteinA purified fraction (500 µL) was run on a 25 mL Superdex G200 colum. (A) Elution profile of EPO-Fc. (B) Western blotting in non-reducing conditions using anti-Fc antibodies, (C) in reducing conditions using anti-Fc antibodies, (D) non-reducing conditions using anti-EPO antibodies, (E) reducing conditions using anti-EPO antibodies. Elution fractions of 1 mL were collected during SEC and analysed. Fractions 9 to 11 corresponded to the first peak. For anti-EPO probed blots, a control with the second elution fraction (E2) was also added. Each lane was loaded with 50 µL of sample.

2.3 SARS-CoV-2 Spike

The use of recombinant pathogen proteins as vaccine immunogens is a well-established vaccination strategy. The efficacy of these vaccines depends on the properties of the proteins, particularly their glycosylation. Numerous reports have demonstrated that many viral surface glycoproteins are GalNAc-*O*-glycosylated, and the involvement of *O*-glycosylation in viral biology is being increasingly studied (Ming et al., 2024).

In the context of the SARS-CoV-2 pandemic, the Spike glycoprotein (S) proved to be an immunogenic target for vaccine strategies. This 1267 amino acid protein assembles as homotrimers and contains a polybasic furin cleavage site. After processing, it is composed of two subunits, S1 responsible for receptor recognition and S2 for membrane fusion and viral entry. Like many other viral glycoproteins, S is highly *N*-glycosylated (22 sites) and the glycans of the S1 subunit have an impact on infectivity (Shajahan et al., 2020; Watanabe et al., 2020). The Spike is also an *O*-glycoprotein (Bagdonaite et al., 2021; Campos et al., 2022; Sanda et al., 2021; Y. Zhang, Zhao, et al., 2021). However, the degree of *O*-glycosylation is the subject of debate. Indeed, the number of sites and their occupancy vary greatly from one study to another. This is probably due to the different variants produced and expression systems used. Nevertheless, the blocking of *O*-glycosylation reduces infectivity *in vitro* (Q. Yang et al., 2020), highlighting the importance of this modification. Several sites often described in literature are located in the receptor binding domain (RBD) and near the furin cleavage site, which could affect protein binding and processing.

2.3.1 Genetic constructs and expression in BY-2 cells

A S1 DNA sequence encoding the amino acids 1-685 was codon optimised for *A. thaliana* and synthesised. The encoded protein contained the endogenous viral signal peptide and the furin cleavage site. A 10-histidine tag with a $(G_4S)_2$ linker was also inserted at the C-terminal of the furin cleavage site. The expression cassettes were generated in pAUX3131 under the control of *pPMA4* with different terminators: t*NOS*, *tExt* and the t*PMP1*. Each expression cassette was then cloned into a pPZP plasmid containing the *hptI* hygromycin resistance gene, a cassette encoding the fluorescent *mCherry*, a 3' *SA*R and the LB and RB borders (Figure II.2.18). WT and ΔXF BY-2 cells were transformed using *A. tumefaciens* and several transformed lines were screened. Figure II.2.19 A and B shows western blot of TSP and EM samples, respectively, from S1 cell lines using an anti-RBD antibody. In TSP, a band corresponding to the intact S1 was found just below 115 kDa (expected molecular mass: 78.7 kDa + 13 *N*-glycosylation sites: 32.5 kDa) whereas no intact S1 protein

accumulated in the EM. Indeed, in panel B, S1 was detected as a strong band between 25 and 30 kDa. Comparison of the signal intensity between many different cell lines also showed that the use of *tExt* resulted in slightly higher overall accumulation (data not shown).

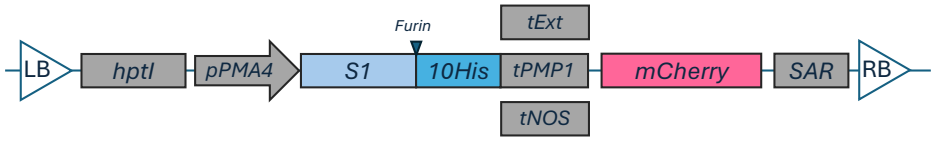


Figure II.2.18: Genetic constructions for expression of S1. The S1 coding sequence containing the endogenous signal sequence was fused to a 10His tag under the control of *pPMA4* and three different terminators: *tExt*, *tPMP1* or *tNOS*. The final plasmids also contained a *hptI* hygromycin resistance gene, a mCherry fluorescent reporter expression cassette and a *SAR* genetic insulator. The Furin cleavage site is indicated by an arrow.

2.3.2 Nickel affinity purification of S1 from intracellular protein extracts

Purification of the His-tagged S1 from TSP was performed by affinity on Ni-NTA resin. Western blots and a Coomassie stained gel of the purification fractions are shown in Figure II.2.19 C and D. In panel C, the blots probed with the anti-RBD antibody and HisProbe show that His-tagged S1 was recovered in the elution fractions E2-E4. The Coomassie stained gel shows that many proteins were removed in the washing fractions compared to the TSP, but contaminants remained. Additional washing using 80 mM of imidazole concentration resulted in a little loss of S1 (data not shown).

2.3.3 Analysis of the *N*-glycosylation profile by *in vitro* deglycosylation

Characterisation of the *N*-glycosylation profile of S1 was carried out by *in vitro* deglycosylation using PNGaseF and EndoH on cellular extracts and purified proteins. Figure II.2.20 A shows the western blot profile of deglycosylated S1 in TSP of WT and ΔXF BY-2 cells. The profile of the WT TSP sample was the same for EndoH and PNGaseF treatments and consisted of a band slightly below the untreated protein (1) and a lower band (2) of similar intensities. This corresponds to a population of S1 decorated with oligomannoses or complex GnGnXF *N*-glycans. In ΔXF TSP sample, the profile for EndoH was the same as the WT, while PNGaseF produced only the lower band (2), as the complex *N*-glycans did not contain any core α(1,3)-fucose. The profile of purified S1 produced in ΔXF BY-2 cells and *in vitro*

deglycosylated consisted of only the lower band (2) with both EndoH and PNGaseF treatments, indicating that the purified S1 protein harboured only unprocessed high mannose glycans. Our hypothesis is therefore that the His tag was cleaved early in the secretory pathway before the processing of the oligomannoses *N*-glycans. Consequently, the His-tagged purified S1 contains unprocessed glycans. TSP sample additionally contains S1 glycovariants processed with complex *N*-glycans, but without the His tag, meaning that secreted S1 cannot be easily purified using the His tag. As GalNAc-*O*-glycosylation occurs in the Golgi apparatus, this complicates the analysis of the glycosylation profile using available tools (lectin blots and MS). For this reason, and due to the high degradation of S1 in the EM, this protein was not studied further.

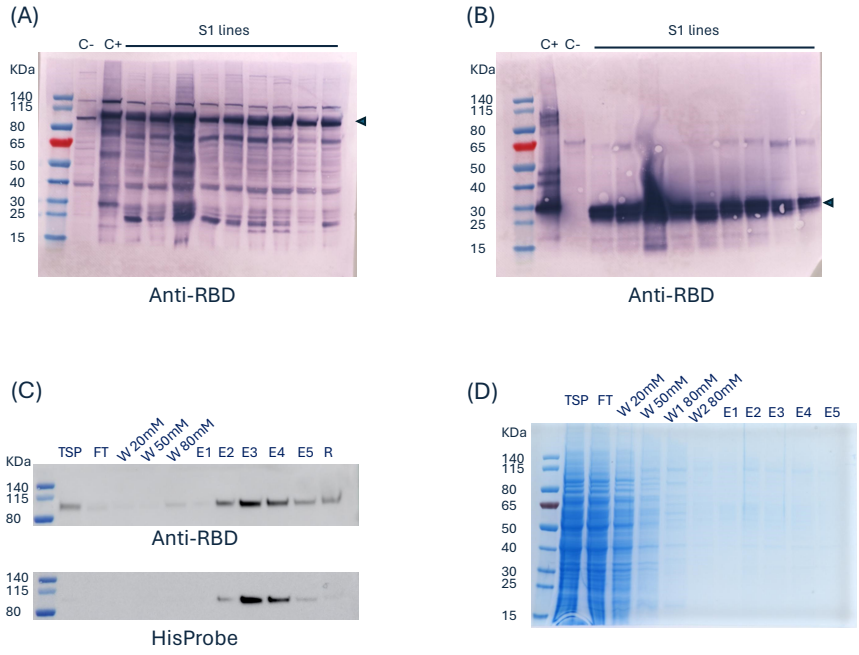


Figure II.2.19: Production and purification of S1 in BY-2 cells. Screening of the productivity of S1 cell lines in (A) TSP (30 µg) and (B) EM (40 µL) analysed by anti-RBD western blot. EM and TSP from WT BY-2 cells were used as negative control (C-) and TSP from *N. benthmiana* transiently expressing S1 as a positive control (C+). Arrows indicate intact and degraded S1 in TSP and EM, respectively. Purification of S1 was done using affinity on Ni-NTA beads. The beads (100 µL) were incubated with TSP (1 mL) overnight and washed (W) with increasing concentration of imidazole (20 to 50 mM) before elution (E) in 300 mM of imidazole. Residual S1 was remove from the beads by boiling (R). The fractions were analysed by (C) anti-RBD and HisProbe western blots and (D) Coomassie stained SDS-PAGE. The following volumes were loaded in the lanes: 25 µL TSP/FT; 50 µL W, 30 µL E.

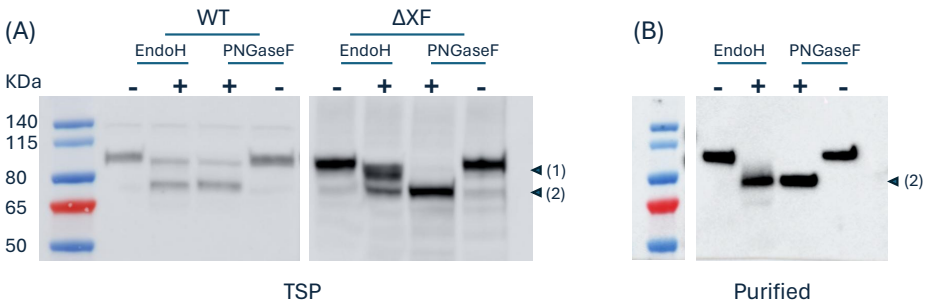


Figure II.2.20: *In vitro* deglycosylation of S1 from TSP or purified fractions. (A) TSPs from WT and ΔXF BY-2 cells and (B) purified S1 were treated with EndoH and PNGaseF. TSP (40 µL) or Nickel affinity purified proteins (25 µL) were treated with 0.1 µL of EndoH or PNGaseF (500 u/µL). The samples were loaded on gel and the blot was probed with anti-RBD antibodies. The proteins were denatured by 10 minutes of boiling in 0.5% SDS and 40mM DTT prior to enzymatic reaction.

2.4 Discussion

BY-2 cells have been used to produce many therapeutic glycoproteins (Santos et al., 2016). Here, we chose to express three *O*-glycoproteins to serve as reporters for the *O*-glycoengineered cell lines: IgA1, EPO and S1. These proteins carry both *N*- and *O*-glycans. By treatment with EndoH and PNGaseF glycosidases, we were able to show that these proteins were predominantly glycosylated with core fucose-free complex type glycans when produced in ΔXF BY-2 cells. MS analysis of IgA1 glycopeptides confirmed this and revealed that they did not contain any xylose. These results confirmed the ability of the engineered ΔXF cell line to produced glycoproteins with a humanised complex glycosylation profile.

We were able to show that the expression of GalNAcT2 and C1GalT1 was sufficient to obtain efficient glycosylation with GalNAc and core 1 glycans in the IgA1 hinge region, in accordance with what was described after transient expression in *N. benthamiana* leaves (Castilho et al., 2012; Dicker et al., 2016). Indeed, more than 85% of the hinge peptides contained at least three *O*-glycans when both enzymes were expressed (Figure II.2.8 B).

Expression of GalNAcT2 alone in BY-2 cells enables the generation of a homogeneous GalNAc glycosylation without interference of other processing enzymes. This *O*-glycan, also termed Tn antigen, is generally not found in glycoproteins from healthy tissues. However, truncated *O*-glycans are often present in diseases (Magalhães et al., 2021), for example epithelial cancer cells generally display Tn glycans and their sialylated version at the cell surface (Kudelka et al., 2015; Pinho & Reis, 2015). Thus, a cell line capable of producing human glycoproteins with homogeneous Tn *O*-glycans may provide tools for research and therapeutic applications. Here we chose to use GalNAcT2 as the initiating enzyme for GalNAc-*O*-glycosylation as it has a broad substrate specificity and is mainly responsible for modifying IgA1 hinge. However, GalNAcTs belong to a family of fifteen active members, each with its own specificity and activity towards unmodified or previously modified glycopeptides. Careful selection of the GalNAcT isoform or combination of isoforms to be expressed would enable the activity on the desired glycoprotein to be fine-tuned. Alternatively, BY-2 cell lines expressing a panel of GalNAcTs would provide a platform for *in vivo* characterisation of their specificities.

Next, co-expression of C1GalT1 with GalNAcT2 generated the core 1 *O*-glycans on BY-2 produced proteins. This reaction was highly efficient, converting most of the glycans in the hinge region of IgA1 (Figures II.2.7 and II.2.8 B). In addition, it seemed

to improve the initiation of *O*-glycosylation, which was unexpected. The core 1 structure is one of the most abundant in humans (Gill et al., 2011) and serve as the basis for core 2 and its sialylated counterparts. Further work could include the introduction of the sialic acid pathway in the *O*-glycoengineered BY-2 cell lines. This pathway, composed of seven enzymes in addition to the glycosyltransferases, has already been successfully introduced in plants and shown to be functional (Castilho et al., 2010). However, generation of asialo-glycans can also be advantageous, for example asialo-EPO is non-erythropoietic but can be used as a neuroprotective drug (Kittur et al., 2023).

The presence of UDP-GalNAc in the plant secretory pathway is a subject of debate in the literature. It was suggested that GalNAc is part of extracellular carbohydrate-protein complexes (Hori, 1978; Wold & Hillestad, 1976). However, the expression of GalNAcT2 alone did not result in efficient glycosylation in *N. benthamiana,* and expression of a transporter and cytosolic epimerase were needed to achieve this (Daskalova et al., 2010). Since then, however, the transport of UDP-GlcNAc and UDP-GalNAc to the Golgi by ROCK1 was demonstrated in Arabidopsis (Niemann et al., 2015) and several reports from another group demonstrated GalNAc-*O*-glycosylation in several plant species without a transporter but with an epimerase (Z. Yang, Bennett, et al., 2012; Z. Yang, Drew, et al., 2012). Nevertheless, our results and those of the Strasser's group (Castilho et al., 2012; Dicker et al., 2016), clearly demonstrated that GalNAcT2 alone is sufficient to initiate efficient *O*-glycosylation.

However, the UDP-GalNAc pool could still be a limiting factor for complete *O*-glycosylation. Comparison of the results obtained with or without expression of the WbpP epimerase did not provide a straightforward answer to this question (Figure II.1.2) and thus a cell line expressing GalNAcT2 was transformed with a genetic construct expressing EPO with or without WbpP. However, this did not lead to a significant increase in the HPL signals on blots with endogenous proteins (Supplementary figure S4). In addition, the effect on the *O*-glycosylation of EPO could not be assessed as the major EPO-Fc degraded protein could not be detected with the lectins. EPO was probably also not the most appropriate reporter for the *O*-glycosylation efficiency and *O*-glycoproteins that are more highly *O*-glycosylated, such as mucins, might be more interesting for this purpose.

MS characterisation of IgA1 hinge region also revealed the presence of plant typical Hyp-*O*-glycosylation, identified as pentoses. The Pro of the hinge region were hydroxylated and then glycosylated, most probably, with short arabinose chains of up to five residues. This phenomenon was first observed on IgA1 produced in maize where there is homology between the IgA1 hinge and maize EXTs (Karnoup et al.,

2005). However, no such homology was found in the genome of *N. tabacum*. Nevertheless, the expression of GalNAc-*O*-glycosylation enzymes strongly reduced the extent of this aberrant Pro modification (Figure II.2.8 A). This suggests that GalNAcT2 and C1GalT1 act before P4Hs on the hinge region. Confocal imaging pointed to a localisation of human GalNAcT2 in the Golgi in *N. benthamiana* and BY-2 cells (Daskalova et al., 2010; Z. Yang, Drew, et al., 2012) as it is the case in human cells. On the other hand, P4Hs are localised both in the ER and the Golgi. Our results suggest that ER-resident P4Hs are probably not responsible for the modification of IgA1 and that some late acting P4Hs would recognise and modify the hinge region of IgA1. In *N. benthamiana*, KO of two of the three genes of the P4H10 group led to a strong reduction of this modification (Uetz et al., 2024), demonstrating the involvement of this family in Pro hydroxylation of the hinge region.

Two other *O*-glycosylation reporters were tested: EPO and the S1 glycoprotein. Both proteins were prone to degradation during their production and secretion in plant cells. This is a well-known drawback of PMF. S1 was processed early in the secretory pathway resulting in the loss of the His-tag needed for purification and then extensively degraded in the EM. Similarly, EPO-Fc was found in the EM mainly as a degraded protein in which most of the EPO domain was cleaved. This protein was the major component of the purified sample and had probably lost the single Ser126 *O*-glycosylation site present in EPO. A larger protein corresponding to a more intact EPO-Fc was also purified, but lectin blots did not reveal the presence of *O*-glycans. The reason of this is unclear as EPO-Fc was glycosylated by GalNAcT2 in *N. benthamiana* leaves (Castilho et al., 2012). However, the *O*-glycosylation site is located next to a Hyp residue that could be glycosylated in BY-2 cells but not in *N. benthamiana*. Interestingly, many Hyp were also detected in the Fc region by MS. To the best of our knowledge, this was never described for IgG produced in plants. Further work is needed to evaluate if this is due to the EPO-Fc construct or whether all plant produced IgGs are potentially hydroxylated and glycosylated.

3 Proline hydroxylation in BY-2 cells

The presence of Hyp modifications and plant-specific *O*-glycans can be detrimental for therapeutic proteins. To enable the production of better pharmaceuticals in BY-2 cells, this modification must be eliminated. To this end, a CRISPR/Cas9 KO screening of the P4Hs was undertaken to find out the enzymes critical for this modification. The *P4H* genes were identified in *N. tabacum* by homology with plant P4Hs and classified into different subgroups. The expression of each P4H relative to the *O*-glycosylation pathway was assessed using a gene co-expression network. Genes from the different subgroups were then targeted together in CRISPR/Cas9 KO experiments to uncover their function. To facilitate the phenotypical characterisation of the P4H mutants, a reporter system based on the expression of two reporter glycomodules fused to GFP was also set-up.

3.1 Identification and analysis of P4Hs in *Nicotiana tabacum*

The genes coding for P4Hs activity in *N. tabacum* were not known, with the exception of one NtP4H (Yuasa et al., 2005). To identify these *P4H* genes, a BLAST search was performed on cDNAs and genome scaffolds in the Solgenomics Database (genome and cDNA deposited by Edwards et al. (2017)) using PpP4H1 and AtP4H1 sequences as well as the consensus sequence from the CP4H domain smart00702. This identified 21 coding sequences in tobacco, including NtP4H from Yuasa et al. (2005). The 21 sequences were then compared by BLAST to RNA-seq data from BY-2 cells obtained in our laboratory (Ahmed et al., 2020) to obtain an overview of the expression of different P4Hs in our system. In addition, differences between our data and that of Edwards et al. (2017), possibly coming from the genetic diversity of the tobacco population or sequencing/annotation errors, could be corrected. Possible splice variants could also be identified. The sequences were also compared with *N. sylvestris* and *N. tomentosiformis* proteins to assess their ancestry. These data are presented in Table II.3.1.

First, a phylogenic analysis of the NtP4Hs was carried out using the NGPhylogeny OneClick pipeline (based on PhyML) (Lemoine et al., 2019). This analysis revealed that NtP4Hs are grouped into six homology groups: P4H1, P4H3/10, P4H4, P4H6 and P4H12 (Figure II.3.1). The groups were named based on homology with Arabidopsis *P4Hs*. *NtP4Hs* were named based on their ancestry with *N. tomentosiformis* (A) and *N. sylvestris* (B). The tobacco P4H sequences were then compared with those of other plant species: *C. reinhaardtii*, *P. patens*, *A. thaliana*, *D. carnation* and *N. benthamiana* (Figure II.3.2). Close homologs are found in the other

species, but differences can be noted between the *Nicotiana* relatives and the other species. Not all Arabidopsis genes have homologs in tobacco and P4H12 have diverged. An identity matrix of the NtP4H sequences was computed (Supplementary figure S5). Overall, it shows high conservation between members of each group (identity > 70%). In addition, several members show low overall identity with other P4Hs (identity <30%, P4H6_A3 and both P4H12). In plants, P4Hs were found as soluble or transmembrane proteins from the ER to the Golgi apparatus. We used SignalP 6.0, TMHMM and WolfPSort to evaluate the presence of a signal peptide and a transmembrane domain, and to predict their subcellular localisation, respectively. Most sequences were predicted to have at least one SP or TM, and to localise in the secretory pathway (Table II.3.1). Some did not, which could be explained by errors in the prediction algorithm or in the sequence annotation. Furthermore, it cannot be ruled out that some P4Hs are active in other cellular compartments with functions that have not yet been described in plants.

A more detailed analysis of the sequences revealed other information. First, the $HXDX_nH$ motif of the active site is conserved in all but four genes (Table II.3.1). The sequence P4H6_A3 corresponds to a short, truncated proteins and is probably a pseudogene. In P4H10_B1, the last histidine is replaced by a serine. The other two are members of the P4H12 group in which HXD is mutated to NXG. Substitutions at these positions are also observed in the closest homologs in Arabidopsis (KXD) and *N. benthamiana* (NXA). Other substitutions in the 2OG binding were observed in these last three genes such K180M/A, suggesting that they are most likely non-active P4Hs. Substitutions in the 2OG binding residues (Ser182 or His187) were also observed in other groups (Table II.3.1). For instance, P4H1 and 3/10 have the H187R mutation that was observed in the functional AtP4H1 enzyme (Hieta & Myllyharju, 2002). In addition, the P4H9 group, which contains the previously characterised NtP4H (Yuasa et al., 2005), has a S182V mutation. These two mutations certainly affect the affinity for 2OG but do not inactivate the enzymes as the activity of P4Hs with these substitutions has been demonstrated. However, P4H6_A2 had large deletions in the 2OG binding domain and are thus most likely unable to bind this substrate. All genes from the P4H4, P4H6 (except the truncated P4H6_A3) and P4H12 groups contain a C-terminal Cys-rich ShK domain of unknown function.

Structure predictions were made from the P4H sequences using AlphaFold (Jumper et al., 2021). With the exception of the truncated proteins, all structures displayed the characteristic beta-fold jellyroll of P4Hs. Figure II.3.3 A shows a structural alignment of the prediction of the P4H1_B (without its TM) and CrP4H1 crystal structure

(PDB: 2JIJ). A very good conservation of the position of alpha helixes (green) and beta strands (red) was observed. Figure II.3.3 B shows a sequence alignment of both proteins. The eight beta strands forming the jellyroll fold are highlighted with roman numerals. Again, the structural properties are conserved between the two enzymes. In addition, the residues involved in the binding of the iron cofactor and 2OG are conserved (arrows and red triangles, respectively). The residues identified by Koski et al. (2009) for peptide substrate binding are also shown. Here, some variations are observed that could lead to different specificities. Substitution in these amino acid residues is also observed between the NtP4H subgroups (Supplementary figures S6).

Table II.3.1: *in silico* analysis of NtP4Hs. NtP4Hs could be classed into six homology groups and their ancestry traced back to *N. tomentosiformis* and *N. sylvestris* by phylogenic analysis. Sequence analysis compared to AtP4H1 revealed mutations in the iron and 2OG binding sites which may inactivate the enzymes. In addition, predictions of the signal peptide, transmembrane domain and subcellular localisation suggested their presence in the secretory pathway. Finally, comparison with BY-2 cell RNA-seq showed that not all genes are expressed in BY-2.

Phylogeny				**Conserved amino acids**			**Localisation**			**RNA-seq**
Gene id	Name	Group	Ancestry	$HXDX_nH$	Substrate binding	ShK Domain	Signal peptide	TM domain	Secretory pathway	Expression in BY-2 cells (FPKM)
0012519g0010.1	P4H1_A	P4H1	*N. tomentosiformis*	Yes	H187R	No	No	Yes	Yes	/
0008084g0010.1	P4H1_B	P4H1	*N. sylvestris*	Yes	H187R	No	No	Yes	Yes	19.28
0002295g0050.1	P4H4_A1	P4H4	*N. tomentosiformis*	Yes	Yes	Yes	Yes	No	Yes	92.75
0009936g0050.1	P4H4_B1	P4H4	*N. sylvestris*	Yes	Yes	Yes	No	No	No	/
0000059g0270.1	P4H4_A2	P4H4	*N. tomentosiformis*	Yes	Yes	Yes	Yes	No	No	/
0006860g0030.1	P4H4_B2	P4H4	*N. sylvestris*	Yes	Yes	Yes	Yes	No	No	28.19
0001727g0170.1	P4H6_A1	P4H6	*N. tomentosiformis*	Yes	Yes	Yes	Yes	No	Yes	36.6
0003431g0060.1	P4H6_B1	P4H6	*N. sylvestris*	Yes	Yes	Yes	Yes	No	Yes	10.36
0002265g0030.1	P4H6_A2	P4H6	*N. tomentosiformis*	Yes	Deletion	Yes	No	No	No	/
0001239g0130.1	P4H6_A3	P4H6	*N. tomentosiformis*	No	Truncated	No	No	No	Yes	/
0001772g0060.1*	P4H9_A1	P4H9	*N. tomentosiformis*	Yes	S182V	No	No	Yes	Yes	14.04

Table II.3.1: Continued

Phylogeny				Conserved amino acids			Localisation			RNA-seq
Gene id	Name	Group	Ancestry	$HXDX_nH$	Substrate binding	ShK Domain	Signal peptide	TM domain	Secretory pathway	Expression in BY-2 cells (FPKM)
0004400g0030.1	P4H9_B1	P4H9	*N. sylvestris*	Yes	S182V	No	No	Yes	Yes	/
0003878g0090.1	P4H9_A2	P4H9	*N. tomentosiformis*	Yes	S182V	No	No	No	Yes	11.11
0007941g0030.1	P4H9_B2	P4H9	*N. sylvestris*	Yes	S182V	No	No	No	Yes	/
0001685g0040.1	P4H10_A1	P4H3/10	*N. tomentosiformis*	Yes	H187R	No	No	Yes	Yes	/
0001561g0090.1	P4H10_B1	P4H3/10	*N. sylvestris*	No	K180A, W186Y, H187Y	No	No	Yes	Yes	/
0002105g0070.1	P4H10_A2	P4H3/10	*N. sylvestris*	Yes	H187R	No	No	Yes	Yes	18.46
0006892g0020.1	P4H10_B2	P4H3/10	*N. tomentosiformis*	Yes	H187R	No	No	Yes	Yes	/
0003030g0090.1	P4H3	P4H3/10	*N. sylvestris*	Yes	H187R	No	No	Yes	Yes	17.99
0000084g0070.1	P4H12_A	P4H12	*N. tomentosiformis*	No	Y81S, T123I, K180M, S182Y, T184Y, W186F, H187Y	Yes	Yes	No	Yes	1.15
0004439g0140.1	P4H12_B	P4H12	*N. sylvestris*	No	Y81T, T123I, K180M, S182Y, T184Y, W186F, H187Y	Yes	Yes	No	Yes	4.01

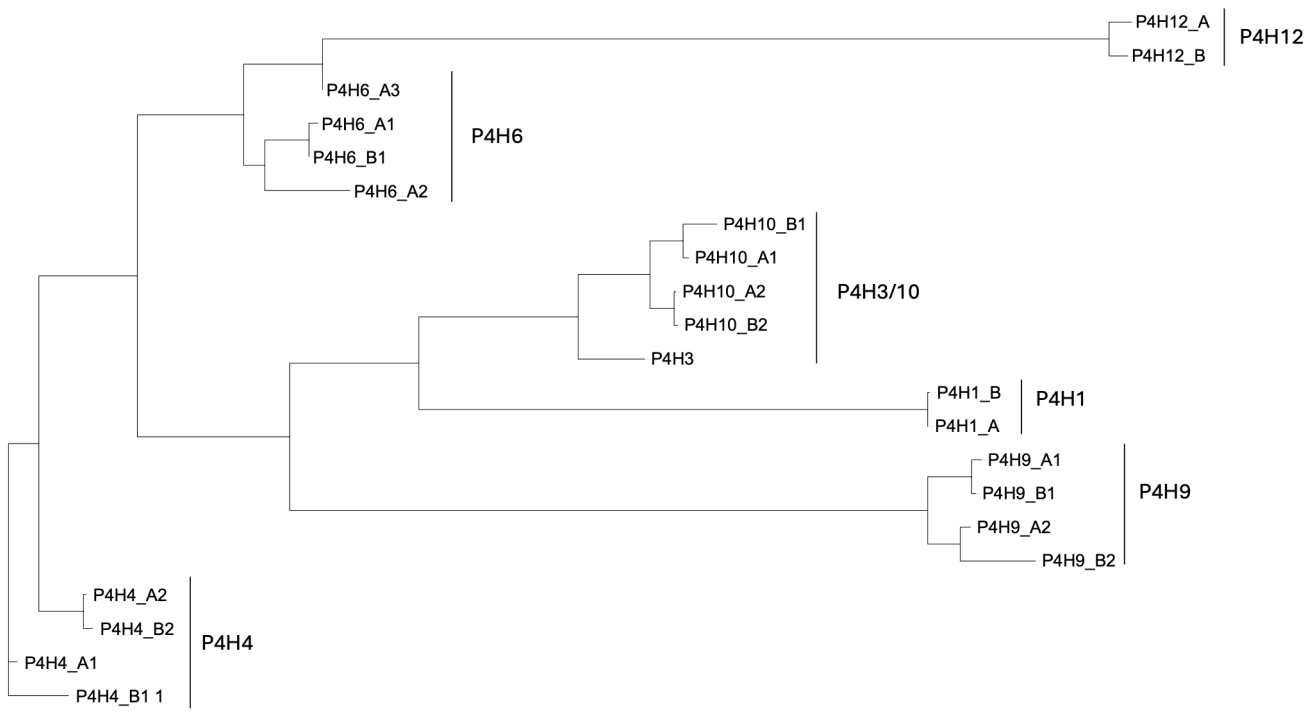


Figure II.3.1: Phylogenic tree of NtP4Hs. The P4H proteins identified in *N. tabacum* were aligned using MAAFT and phylogeny was inferred with PhyML. NtP4Hs are clustered in six homology groups and show high homology between pair of homeologs.



Figure II.3.2: Phylogenic tree of plant P4Hs. The P4H proteins identified in *N. tabacum* and other plant species were aligned using MAAFT and phylogeny was inferred with PhyML. The different homology groups are shown.

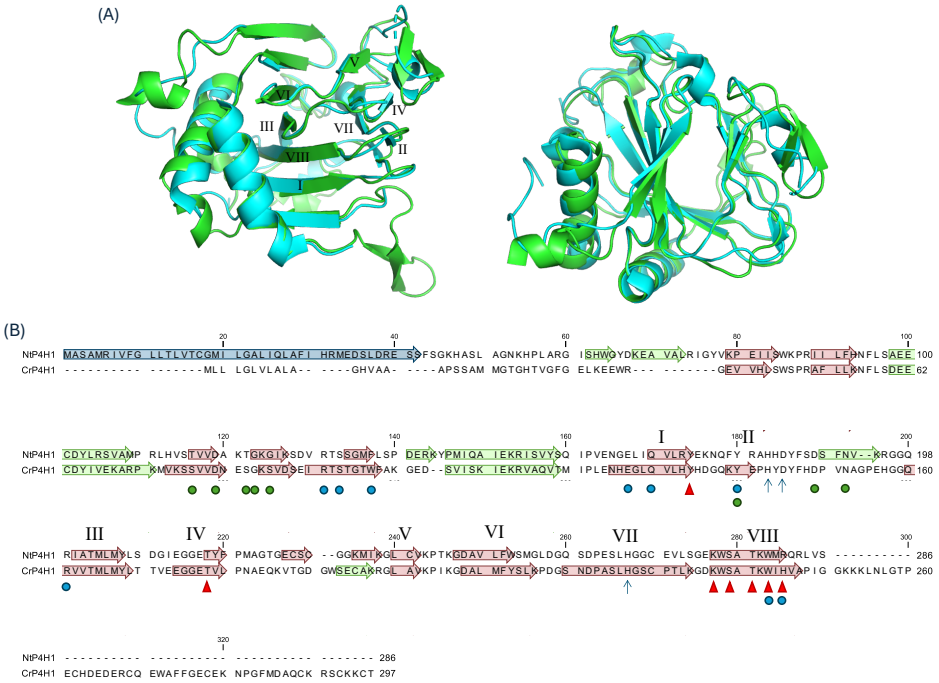


Figure II.3.3: Alignments of NtP4H1_B and CrP4H1. (A) Structural alignment. Structural data was inferred using AlphaFold for NtP4H1_B (green, without the TM) and taken from the CrP4H1 crystal structure 2JIJ (cyan) (Koski et al., 2007). The beta strands involved in the jellyroll fold are denoted by roman numerals. (B) Sequences alignment. The transmembrane domain is shown in the sequences by a blue arrow, alpha helixes by green arrows and beta strands by red arrows. The residues implicated in the binding of iron and 2OG are highlighted by arrows and red triangles, respectively. The peptides binding residues identified are shown with blue (Koski et al., 2007) and green (Koski et al., 2009) dots.

3.2 Co-expression analysis of *NtP4Hs*

Expression data from Edwards et al. (2017) were assembled into a gene expression network using the BioNERO R package (Almeida-Silva & Venancio, 2022). Briefly, the RNA-seq analysis was conducted on roots, whole shoot and shoot apex tissues of 8-week-old *N. tabacum* cv. K326 plants (accession number GSE95717). The samples were harvested at four time points (ZT 0, 6, 12 and 18 h) and in triplicate. The data was cured according to the package documentation before network inference. The output is a network composed of different modules (denoted by different colours) in which genes are linked by their expression profile. In total, 58 co-expression modules were identified in the RNA-seq data. Figure II.3.4 shows a plot of the number of genes identified in each module. At the top of the graph, modules with few genes are found, which probably indicates their implication in specialised pathways or in cellular response to specific signals. For example, the module darkred contains many genes associated with phospholipase D activity (6/9 genes). In contrast, the modules at the bottom contain many genes that are most likely implicated in normal/broad cellular functions.

The position of *P4H* genes in this network and the genes to which they are directly correlated (neighbours) may be helpful for obtaining information about their involvement in the EXT and AGP glycosylation pathways and their importance in cell physiology. Thus, the network was searched for *P4H*s. Table II.3.2 show the modules in which *P4H* genes are found. Two *P4H6* genes are not present as they were not expressed and, therefore, removed from the network (*P4H6_A2* and *A3*). Most of the *P4H* genes were found in modules with more than 1000 genes with 9 and 3 *P4H*s found in the two largest modules, bisque and chocolate2, respectively. Only one gene (*P4H6_A1*) was found in a module with few genes (plum4, 50 genes). This suggests that *P4Hs* are involved in important general processes of the cell. Furthermore, four *P4H* genes were detected as housekeeping genes (*P4H4_A1, P4H10_A2, P4H10_B1 and P4H10_B2*), *i. e.*, genes broadly expressed with low variation across the samples. Gene ontology (GO) and InterPro (IPR) term enrichment analysis was performed on the different modules. It confirms the broad functions of the bisque and chocolate2 modules in basic cellular processes, which are enriched with many terms (such as GO:0015629 actin cytoskeleton and GO:06413 translation initiation in bisque and GO:0006260 DNA replication and GO:0006397 mRNA processing in chocolate2). Interestingly, the module chocolate2 was enriched in IPR terms corresponding to glycosyltransferase families GT8 (IPR002495), GT31 (IPR002659) (which contains the arabinogalactan GalTs) and GT48 (IPR003440). Notably, some of the smaller modules which contained only one *P4H* gene were enriched in terms associated with the plant CW (aquamarine2 1/15 terms,

blanchealmond 6/19 and mediumpurple3 2/12). The plum4 module contains only one term corresponding to heat shock protein 70 family.

Next, a search for the network neighbours of each *P4H* was carried out. Firstly, some *P4H* genes were direct neighbours (expression correlation > 0.7) (Figure II.3.5), indicating that these genes are co-expressed and may therefore act together in the same pathway. In addition, at least three groups (*P4H4, 9* and *3/10*) were expressed together. Then, to uncover the co-expression of *P4H*s with *Hyp GT*s, a search for genes with high homology to known *A. thaliana AGP* and *EXT* related *GT*s was conducted in the neighbours of *P4H*s (Supplementary table S5). Five *P4H*s were exclusively linked to genes involved in the synthesis of AGPs glycans. Only one gene (*P4H10_B2*) was connected with both *EXT* and *AGP GT*s, while none were exclusively connected with *EXT* related genes. The remaining 14 genes were not co-expressed with any of the *GT*s used in the search. It cannot be ruled out that some *GTs* were not included in the search due to their low identity with *A. thaliana* homologs and because some enzymes were not discovered, especially in tobacco. In addition, the search could have included *HRGP* genes that may also be co-expressed with specific *P4H*s.

Based on this analysis, most *P4H* genes have a broad expression linked to both EXT and AGP biosynthetic pathways and show a low overall correlation with the *GTs* for these substrates. We did not find P4Hs co-expressed exclusively with *EXT GTs* in contrast to Arabidopsis where three P4Hs (AtP4H5, 2 and 13) were shown to mainly modify EXTs *in vivo* (Velasquez, Ricardi, et al., 2015), but these *AtP4Hs* have no close homologs in *N. tabacum*. However, some P4Hs are exclusively related to AGP specific *GTs*, which could mean that these P4Hs are implicated in the hydroxylation of AGP substrates. Nevertheless, these findings cannot be correlated with the different P4Hs subgroups. Therefore, further experiments are required to clarify the involvement of P4Hs in specific modification of EXT and AGP substrates.

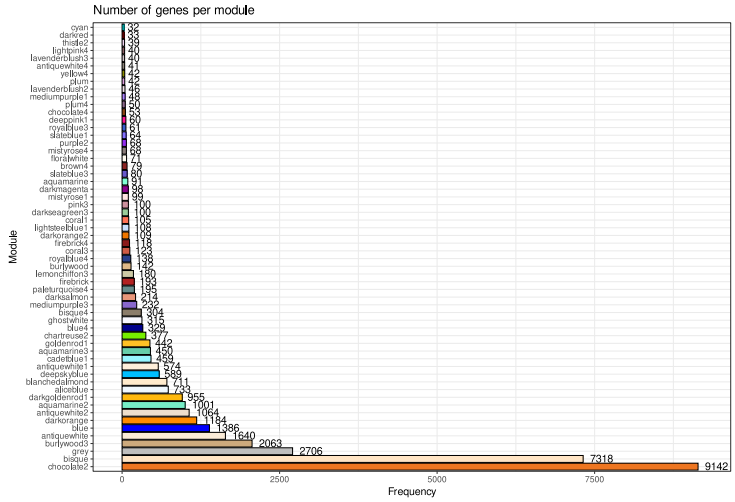


Figure II.3.4: Number of genes per modules. The number of genes identified in each module of the co-expression network was plotted and classed increasingly.

Table II.3.2: Co-expression modules of *NtP4Hs*. The co-expression network was searched for *P4Hs* genes and their position in each module. The number of genes in the identified modules is indicated.

Modules	**Number of genes**	***NtP4Hs***
bisque	7318	*P4H4_B1, P4H4_A2, P4H9_A1, P4H9_B1, P4H9_B2, P4H10_A1, P4H10_B1 P4H12_A, P4H12_B*
chocolate2	9142	*P4H10_A2, P4H10_B2*
burlywood3	2063	*P4H1_B*
darkorange	1184	*P4H1_A, P4H9_A2*
aquamarine2	1001	*P4H4_B2*
blanchedalmond	711	*P4H6_B1*
bisque4	304	*P4H4_A1*
mediumpurple3	232	*P4H3*
plum4	50	*P4H6_A1*

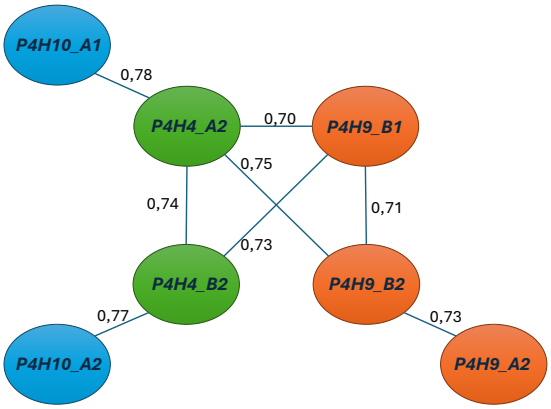


Figure II.3.5: Co-expressed *NtP4Hs*. Several *NtP4Hs* genes were correlated in the co-expression network. The coefficient of correlation between each gene is shown on the lines. P4H4 are coloured in green, P4H9 is orange and P4H3/10 in blue.

4 Silencing of P4H activity in BY-2 cells

To deepen our understanding of EXT and AGP proline hydroxylation and to develop BY-2 cells as a platform for the production of pharmaceutical glycoproteins without the potential immunogenic plant-specific *O*-glycans, the inactivation by CRISPR/Cas9of three groups of P4Hs (1, 4 and 10) were tested. Elimination of P4H activity should abolish the ability of BY-2 to generate Hyp-*O*-glycans. However, verification of the KO and the assessment of the phenotype variation can be difficult and time-consuming. Therefore, to simplify phenotypic characterisation of the KO cell lines, reporter cell lines expressing reporter substrates based on AGP and EXT glycomodules were first generated.

4.1 Generation of reporter cell lines

Reporter cell lines were generated to allow simple phenotypic characterisation for studying the elimination of P4H activity in BY-2 cells on AGP and EXT-like substrates. They were based on a secreted GFP fused in N-terminus to $(SP)_{32}$ or $(SP_4)_{18}$ glycomodules (Shpak et al., 1999). These glycomodules can drive very efficient and specific (relative to AGP and EXT types) Hyp-*O*-glycosylation on the numerous Pro residues they carry. When expressed in BY-2 cells, they significantly increase the secretion and molecular mass of the GFP and further analysis revealed that the Pro residues within the glycomodules were fully hydroxylated and glycosylated (see Table I.3.I and reference therein). The *p35S* was used to drive high expression and the *N. benthamiana* calreticulin signal peptide to target the reporter protein to the secretory pathway. Genetic constructs containing a selection marker (kanamycin or hygromycin) and the *p35S-SP$_{Cal}$-[(SP)$_{32}$ or (SP$_4$)$_{18}$]-GFP-tNOS* expression cassette were made (Figure II.4.1). A control construct expressing secreted GFP without a glycomodule tag was also generated. WT BY-2 cells were transformed and transgenic calli were selected. The expression and glycosylation status of each GFP variant was assessed by western blot using anti-GFP antibodies. Figure II.4.2 shows a blot comparing the profile of the three GFP constructs in the EM of BY-2 cells. The secreted GFP was visible in the EM at 27 kDa. For the two glycomodule constructs, this band was still barely visible, but most of the signal was detected at much higher molecular masses. Glycosylation of the glycomodules was responsible for this shift since unglycosylated $(SP)_{32}$-GFP and $(SP_4)_{18}$-GFP were expected at 32.8 kDa and 35 kDa respectively. Interestingly, both glycomodules showed a different profile: $(SP_4)_{18}$-GFP was secreted as a single band between 80 and 115 kDa, whereas $(SP)_{32}$-GFP had a much broader profile observed as a smear between 80 and over 140 kDa. This highlights the differences of glycosylation between the two glycomodules. Indeed, the glycosylation structure of AGP is larger and more heterogeneous than

the glycosylation structure of EXT, which led to the smear profile of $(SP)_{32}$-GFP compared to $(SP_4)_{18}$-GFP. Comparison of EM proteins and TSP showed the presence of intermediate glycosylated GFP products in the TSP (Supplementary figure S7).

These reporter cell lines are very useful to study proline hydroxylation and *O*-glycosylation in BY-2 cells. Indeed, a defect in P4Hs activity should perturb the glycosylation process of the GFP glycomodules and lead to a large shift in the band corresponding to the GFP reporter proteins or impact their secretion. This would reveal the ability of specific P4Hs to target AGP- or EXT-like substrates.

An initial test of the reporter cell lines was carried out by chemically inhibiting P4Hs with 2,2-dipyridyl, an iron chelator, that can be used to inhibit proline hydroxylation in plants. This chemical has already been used in tobacco seedlings and BY-2 cells (Moriguchi et al., 2011; Shimizu et al., 2005). Thus, 0.1 mM of 2,2-dipyridyl was added to 3 day-old BY-2 cell cultures and the accumulation of glycosylated $(SP)_{32}$-GFP was monitored after 3, 5 and 7 days. The culture became bright pink, indicating the formation of complex between iron and 2,2-dipyridyl. However, glycosylated $(SP)_{32}$-GFP was still detected both in the EM and TSP (Supplementary figure S7) while a severe growth defect was observed. This is in agreement with the results from Moriguchi et al. (2011) who showed only a slight reduction in hydroxylation but with growth defects upon treatments with up to 1 mM. This highlights that 2,2-dipyridyl probably affected all the iron-dependent cellular processes, resulting in growth defects. We decided to use the CRISPR/Cas9 strategy to inactivate *P4H*s in BY-2 cells.

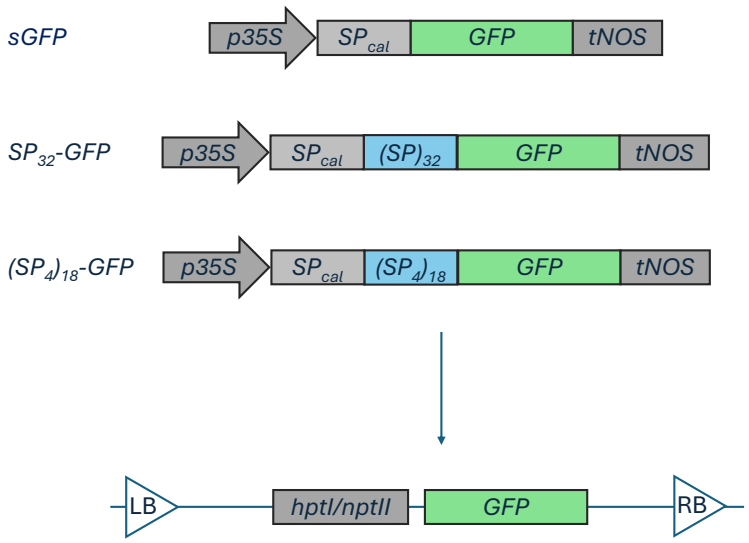


Figure II.4.1: Constructions for the GFP reporter proteins. Sequences coding for the $(SP)_{32}$ or $(SP_4)_{18}$ glycomodules were fused in N-terminal of GFP coding sequence with the signal peptide sequence from *N. benthamiana* calreticulin. The expression cassettes were under the control of *p35S* and *tNOS*. A control construct without the glycomodules was also made. Binary plasmids with the constructs also contain hygromycin *hptI* or kanamycin *nptII* resistance gene and LB and RB for *A. tumefaciens* mediated transformation.

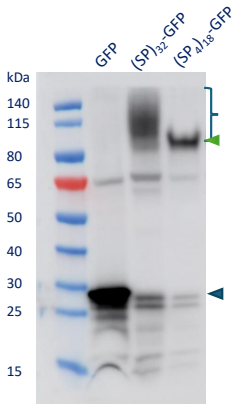


Figure II.4.2: Western blot profile of the GFP reporters. The EM from cell lines producing secreted GFP, $(SP)_{32}$-GFP or $(SP_4)_{18}$-GFP were analysed by anti-GFP western blot. Each lane was loaded with 50 µL of EM. The blue arrow indicates untagged GFP, the green arrow glycosylated $(SP_4)_{18}$-GFP and the bracket glycosylated $(SP)_{32}$-GFP.

4.2 Design of CRISPR/Cas9 guide RNAs and vector construction

CRISPR/Cas9 is a powerful strategy to specifically knockout selected genes. It relies on the pairing of the guide RNA (gRNA) sequence with the sequence of the targeted gene. The guide sequence therefore controls the specificity and efficiency of the genome edition. The cDNA sequences of the identified *P4H*s were used to design gRNAs for CRISPR/Cas9 KO. The guides were designed using CRISPOR (Concordet & Haeussler, 2018) and manually verified. The best guides were selected to match the following criteria: their position at the first part of the CDS, minimisation of possible off target, appropriate GC content (40-60%) and their secondary structure, according to Liang et al. (2016). In addition, guides were selected to target simultaneously the close homeologs. Three guides were designed for each gene in the P4H4 and P4H1 groups (Figure II.4.3 A and B). To simplify the KO analysis and because of the high edition efficiency in BY-2 cells observed in our lab, only two gRNAs were made for the genes of the P4H3/10 (Figure II.4.3 C) and other groups.

Constructs containing an array of gRNAs and the *Cas9* cDNA for editing several genes were made with the Golden Gate CRISPR/Cas toolkit (Hahn et al., 2020). This toolkit allows the generation a polycistronic tRNA-gRNA under the control of the *A. thaliana U6* promotor into Golden Gate modules (Figure II.4.4). The endogenous tRNA processing machinery is expected to release the gRNAs. A tRNA-gRNA construct was generated for each of the P4H group containing gRNAs targeting all the genes within the group. The multiplexes contained six gRNAs for P4H1 and P4H4 (two pairs of genes with three gRNAs) and P4H6 and P4H10 (three pairs with two gRNAs), four gRNAs for P4H9 (two pairs with two gRNAs) and two gRNAs for P4H12 (one pair of genes).These constructs were then assembled in binary plasmids with *hptI* or *nptII* resistance gene and the *Cas9* expression cassette containing an *A. thaliana* codon optimised *Cas9* fused to the simian virus 40 nuclear localisation signal under the control of *p35s* and *tNOS* (Figure II.4.4). The constructs were transformed into *A. tumefaciens*. $(SP)_{32}$- and $(SP_4)_{18}$-GFP reporter BY-2 cell lines were transformed with the constructs to obtain KO of the P4H1, 4 and 3/10 groups.

However, the transformation could not be carried out for the three other groups due to problems with the *A. tumefaciens* strains. Indeed, digestion of the plasmid DNA extracted from the different *A. tumefaciens* strains did not result in the expected profile on agarose gel electrophoresis. The LBA4404 strain used in our laboratory contains the pAL4404 plasmid and a constitutive virulence plasmid pBBRvirGN54D (Van Der Fits et al., 2000), as well as the binary plasmid containing the gRNA constructs.

Following the unexpected digestion results, *Agrobacterium* plasmid DNA was used to transform *E. coli* and selection was performed on spectinomycin (binary plasmid) or gentamycin (*VirG* plasmid). Both plasmids were digested by EcoRI (Figure II.4.5). While the binary plasmid migrated as expected, the pBBRvirGN54D plasmid appeared much higher than anticipated (>10 kb instead of 6kb). Sequencing of this plasmid revealed the presence of the Tn904 transposon in the *VirG* gene. We have previously described the insertion of this transposon in binary plasmids, which probably prevented expression of potentially toxic constructs in *A. tumefaciens* (Navarre et al., 2023). This would indicate that replication of the T-DNA or translation of its proteins present some toxicity for *Agrobacterium*. Why this occurred in only three of the constructs and never in the binary plasmid is unclear. To circumvent this, particle bombardment could be used to deliver the transgenes in BY-2 cells.

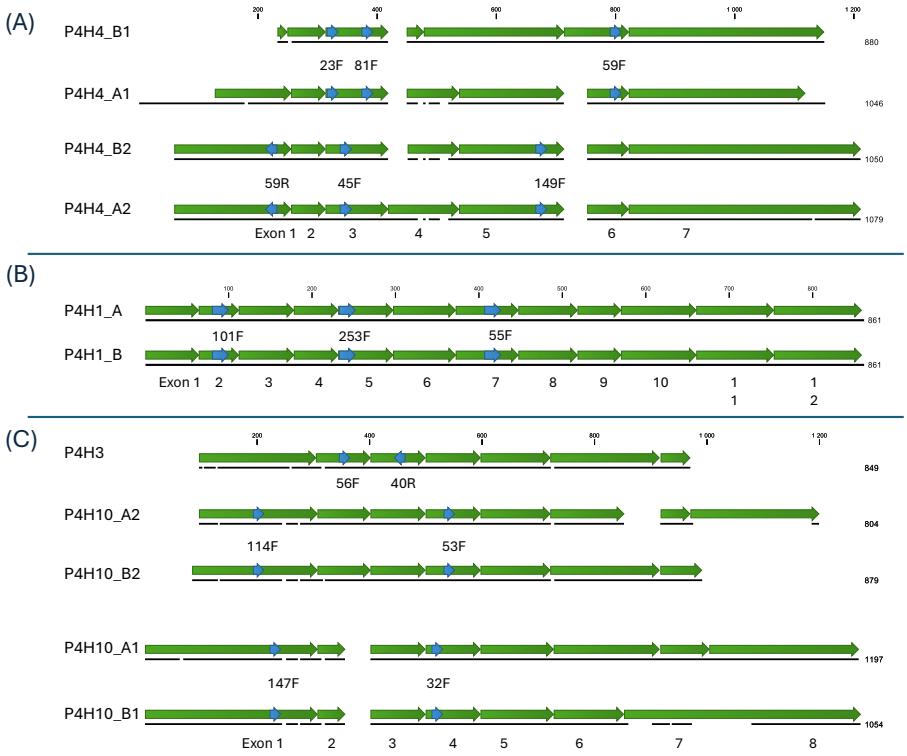


Figure II.4.3: Guide RNAs targeting *P4Hs* groups. cDNA sequences from (A) *P4H4*, (B) *P4H1* and (C) *P4H3/10* were aligned within each group and gRNAs targeting each gene were indicated. Exons are shown in green and gRNAs in blue. The sequences of the guides are given in Supplementary table S6.

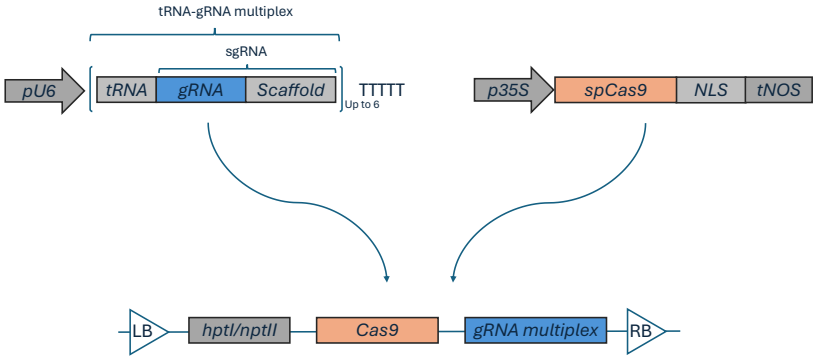


Figure II.4.4: Guide RNAs constructs. Each gRNA was cloned into a construct containing the scaffold and interspaced by tRNAs and assembled into a multiplex under the control of *AtU6* promoter. The multiplex is cloned in binary plasmids with an Arabidopsis codon-optimised *SpCas9* with a nuclear targeting signal under the control of *p35S* and *tNOS* and with a resistance gene and LB and RB for BY-2 cells transformation.

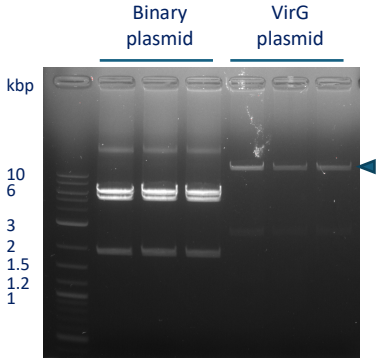


Figure II.4.5: Recombination of the VirG plasmid. EcoRI digestion of *A. tumefaciens* plasmid DNA. The expected size of the VirG plasmids is 6 kB. The binary plasmid had the expected profile.

4.3 Knock out of *P4H4*

The $(SP)_{32}$-GFP and $(SP_4)_{18}$-GFP cell lines were transformed with the multiplex construct containing the six guides targeting the four genes in the *P4H4* group. The cell lines were screened by western blotting of EM samples using an anti-GFP antibody to uncover potential difference in molecular mass of the glycosylated reporter (Figure II.4.6). Compared to the parent cell lines, the profile of $(SP_4)_{18}$-GFP and $(SP)_{32}$-GFP in the KO cell lines showed no significant shift of the bands, suggesting that the elimination of P4H4 had no impact on the Pro hydroxylation of the reporter. To ensure that the genes were correctly knocked out, the four *P4H4* genes were amplified from genomic DNA and compared to the WT sequence using ICE (Conant et al., 2022). Genomic DNA from the BY-2 cell lines was extracted and each gene was amplified individually with specific primers The PCR products were run on agarose gels (Figure II.4.6 B) to verify correct amplification. Most amplifications showed a single band at the correct size. Sequencing of the amplicons was required to confirm occurrence of indel events. Interestingly, some PCR products also showed bands at lower molecular mass, which can indicate a large deletion event between two gRNAs. (for instance, cell line $(SP_4)_{18}$ #1 for P4H4_A2/B2 guides 1-2). The PCR products were sequenced and analysed by ICE.

However, sequencing of P4H4_B1 amplicon resulted in the identification of P4H4_A1 despite the use of several pairs of primers. This prompted us to investigate this gene in the genome of *N. tabacum*. The same PCR was performed on genomic DNA from *N. tabacum* plants. The gene P4H4_B1 was specifically amplified from *N. tabacum* DNA (Supplementary figure S8). The sequencing chromatogram showed an almost equal signal for both P4H4_B1 and A1, indicating that these homeologs were both amplified with the same pair of primers from genomic DNA. This was not the case when using BY-2 genomic DNA, as only P4H4_A1 was amplified, strongly suggesting that the gene P4H4_B1 was modified in BY-2 cells.

The genome editing results based on the sequencing for the other three *P4H4* genes were generated by ICE analysis. Indeed, as genome editing events can occur independently in different cells, the genomic DNA extracted and therefore the PCR products are composed of several different sequences. Sequencing of these products results in overlapping chromatograms that can be separated into individual sequences by alignment with a WT sequence using software (such as ICE). The software also computes the relative abundance of each sequence and calculates editing scores. The indel percentage score (% indel) indicates the proportion of the sequences in which an editing event occurred. The KO score specifies the percentage of the sequences

in which the indels probably inactivate the gene (for instance ± 3 indels do not contribute to this score whereas ± 1 or 2 indels contribute as they give rise to frame shifts. The computed scores for the six ΔP4H4 cell lines analysed are given in Table II.4.1. The editing efficiency between the different cell lines greatly varied, with some cell lines being completely inactivated (*e.g.* $(SP_4)_{18}$ #1) while others not (*e.g.* $(SP_4)_{18}$ #3). The analysis also confirmed the large deletion events observed in some cell lines However, this was not related to the profile observed in western blotting as the P4H4 inactivated and not inactivated cell lines had similar profiles compared to the parental cell line. This suggests that the *P4H4* family is either involved in the hydroxylation of Pro in sequons different from those found in the two glycomodules tested, or that they are involved in those modifications but that other enzymes also carry them out and can compensate for the absence of P4H4s.

Table II.4.1: Analysis of *P4H4* KO. PCR amplicons from ΔP4H4 cell lines were sequenced and analysed by ICE. The percentage of indel and KO score are indicated. The KO score is computed based on the indel percentage that leads to a gene inactivation (*e.g.* a frame shift).

Gene/guide	Score	$(SP_4)_{18}$ 1	$(SP_4)_{18}$ 3	$(SP_4)_{18}$ 8	$(SP)_{32}$ 18	$(SP)_{32}$ 21	$(SP)_{32}$ 22
P4H4_B2 g1-2	%indel	100	0	100	81	100	15
	KO	100	0	99	81	51	11
P4H4_B2 g3	%indel	93	0	96	95	68	99
	KO	90	0	28	72	68	98
P4H4_A2 g1-2	%indel	100	0	100	96	95	49
	KO	100	0	100	96	95	0
P4H4_A2 g3	%indel	54	0	92	100	100	100
	KO	20	0	48	48	100	0
P4H4_A1 g1-2	%indel	100	0	100	100	100	100
	KO	96	0	97	52	100	28
P4H4_A1 g3	%indel	84	19	97	96	96	97
	KO	80	19	97	96	96	97

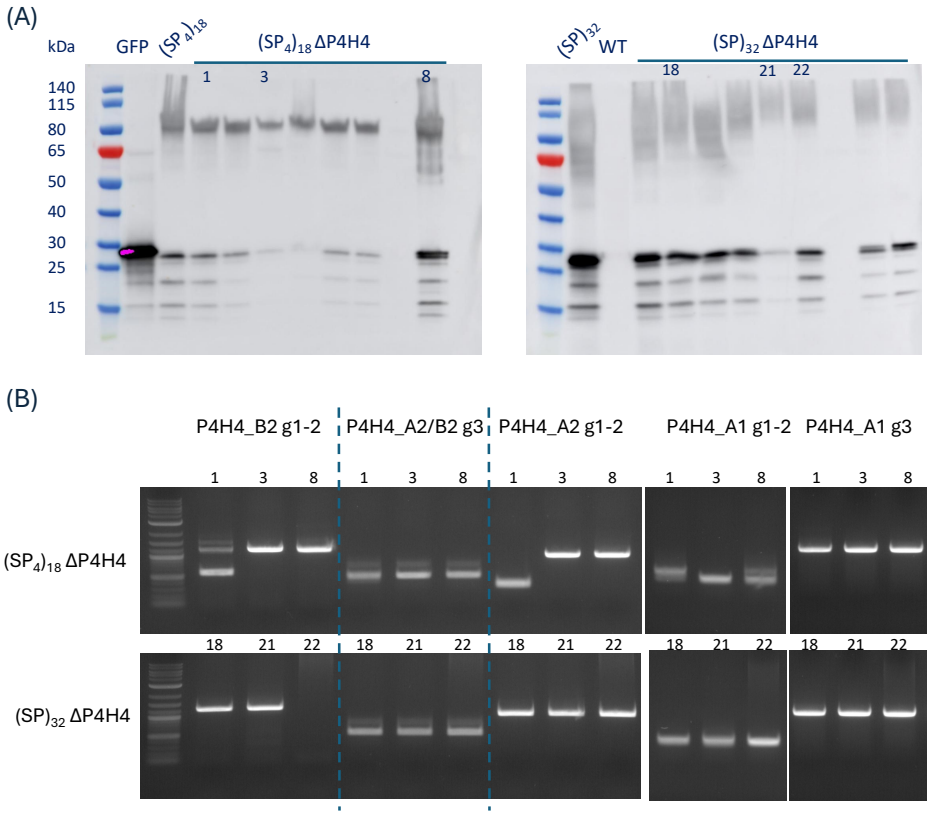


Figure II.4.6: Analysis of P4H4 KO. (A) Analysis of the ΔP4H4 KO in $(SP_4)_{18}$-GFP (left panel) and $(SP)_{32}$-GFP (right panel) by anti-GFP western blot. Secreted GFP and the parental cell line were used as controls. Each lane was loaded with 40 µL of EM. The numbered cell lines were used for further analysis. (B) PCR amplification of the *P4H4* genes from genomic DNA from three $(SP)_{32}$ KO and three $(SP_4)_{18}$ KO cell lines. Several amplifications were made to target the zones containing the gRNA sites in each gene.

4.4 Knock out of *P4H1* and *P4H3/10*

Similarly to P4H4, anti-GFP western blot analysis was conducted for the reporter cell lines transformed with either the *P4H1* or *P4H3/10* guides. For ΔP4H1, cell lines were obtained from $(SP)_{32}$-GFP and $(SP_4)_{18}$-GFP, but only $(SP_4)_{18}$-GFP ΔP4H3/10 were obtained. Figure II.4.9 shows western blots using the ME and TSP from those cell lines. No GFP signal was observed in the inactivated cell lines and in the parental cell line compared with the signal from the ΔP4H4 line. It is possible that the reporter *GFP* gene was silenced in the parental cell line, making the phenotypical characterisation of the KO with the glycosylated reporter impossible. Therefore, the characterisation was not pursued, and the KO experiments are inconclusive.

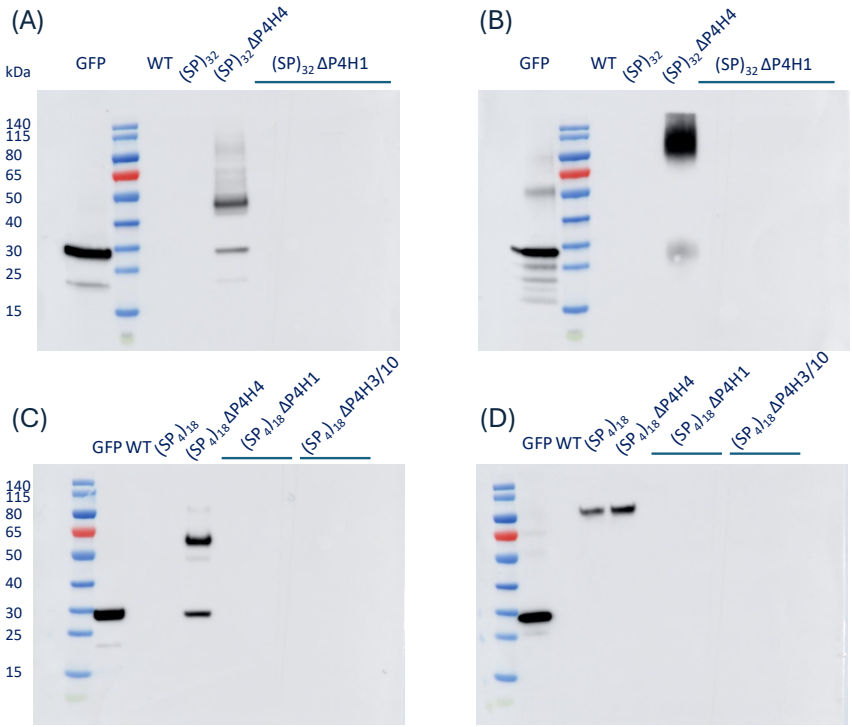


Figure II.4.7: Analysis of P4H1 and P4H3/10 KO by western blotting. (A) TSP and (B) EM from three $(SP)_{32}$-GFP ΔP4H1 cell lines were analysed by western blots. (C) TSP and (D) EM from three $(SP_4)_{18}$-GFP ΔP4H1 and $(SP_4)_{18}$-GFP ΔP4H3/10 cell lines were analysed by western blots. Samples from secreted GFP, the parental cell lines and a ΔP4H4 cell line were used as control. Each lane was loaded with 40 µg of TSP or 55 µL of EM (except for GFP and $(SP_4)_{18}$-GFP ΔP4H4 where only 5 µL was used).

4.5 Discussion

Glycosylation of hydroxyproline is a typical PTM in plant cells. It results in clusters of arabinose chains or large complex arabinogalactan polysaccharides. Their potential immunogenicity is debatable: some Hyp-*O*-glycan determinants have been shown to be part of allergic reaction (Leonard et al., 2005; Léonard et al., 2010), while INFα2-$(SerPro)_{20}$ has not been found to be immunogenic in mice (Xu et al., 2007). However, the presence of these glycans can affect protein properties.

The sequences recognised for typical plant *O*-glycosylation are rich in Pro, Ser and Thr, residues also found in the GalNAc-*O*-glycosylation sequons. This may lead to potential competition between the two pathways and hamper the humanisation of *O*-glycosylation in plant cells. Yet, our results on the hinge region of IgA1 showed that GalNac-*O*-glycosylation of Ser/Thr strongly reduces neighbouring Hyp-*O*-glycosylation (Figure II.2.8 A). On the other hand, it cannot be ruled out that close Hyp-*O*-glycosylation prevents optimal/complete GalNac-*O*-glycosylation. In addition, Hyp residues and potentially linked *O*-glycans can be found in other regions that are not GalNAc-*O*-glycosylated such as in EPO or in the Fc region of IgG2.

The extent of hydroxyproline glycosylation is higher in transgenic BY-2 cell lines than in widely used transient expression system in *N. benthamiana* leaves. Indeed, our data showed that the hinge region of IgA1 was modified with 5-6 Hyps and pentoses whereas modification with pentoses was reduced in IgA1 produced in *N. benthamiana* (Dicker et al., 2016; Göritzer et al., 2017; Uetz et al., 2024). It has been already suggested that transient system was less prone to aberrant Hyp-*O*-glycosylation (Parsons et al., 2013; Z. Yang, Bennett, et al., 2012).

We set up a CRISPR/Cas9 KO screen of the different *P4H* groups to identify the P4H enzymes involved in the glycosylation of AGPs and EXTs, and potentially a member that is required to start the pathways as it was demonstrated for AtP4H5 (Velasquez, Ricardi, et al., 2015). Firstly, the *N. tabacum P4H* genes were identified and the sequences analysed. In total 21 genes were found and grouped into six subfamilies. Most of the proteins were predicted to be located in the secretory pathway and to be functional P4Hs on the basis of the catalytic and 2OG binding residues. A gene co-expression network of *N. tabacum* cv. K326 was then deduced to uncover potential co-expression of *P4H*s and *AGP*- or EXT-related genes, (Velasquez, Marzol, et al., 2015) but no correlation was found.

To facilitate the phenotypic characterisation of the BY-2 *P4H* knockout cell lines, we developed a GFP reporter system based on the fusion of glycomodules as described by Shpak et al. (1999). We used the two hyperglycosylated repeat glycomodules $(SP)_{32}$

and $(SP_4)_{18}$ which were shown to drive a highly efficient and specific glycosylation, with all Pro residues being hydroxylated and modified either with arabinogalactans or arabinose in BY-2 cells (Table I.3.1). This also allows for distinguishing between the potential specificity of the P4Hs families towards EXT $(SP_4)_{18}$ or AGP $(SP)_{32}$ substrates. As shown in Figure II.4.1, the reporter system performed very well and the difference in the profile of GFP fused to the two glycomodules highlighted the variation in complexity and heterogeneity between arabinose and arabinogalactan glycans. In the KO cell lines, the involvement of a specific P4H subfamily in the hydroxylation of one glycomodule would be marked by a shift in the GFP electrophoretic mobility, suggesting that some prolines were not hydroxylated and thus not glycosylated. Alternatively, this could also lead to a decrease in secretion of the underglycosylated GFP. Indeed, it is thought that some retention mechanisms prevent secretion of partly glycosylated proteins as shown by the accumulation of small size intermediates intracellularly and the presence of only highly glycosylated species in the extracellular medium (Karki et al., 2023; N. Zhang, Dolan, et al., 2016) (Figure S6). Several GalT and GlcAT were found located in the Golgi but also in unknown small organelles positive for Exocyst (EXO70A2) mediated secretion (Poulsen et al., 2014). EXO70A2 is also involved in the secretion of EXT cargoes and *exo70a2/hpat1-3* mutant showed increased secretion of non-glycosylated products compared to the *hpat1-3* mutant, possibly involving EXO70A2 in the retention mechanism of intermediates (Beuder et al., 2020).

We obtained KO cell lines for three groups of P4Hs: P4H4, P4H1 and P4H3/10. Several P4Hs from these groups possess conserved catalytic residues and are expressed in BY-2 cells. The PpP4H1 homolog was shown to be solely responsible for EPO hydroxylation when produced recombinantly in this species (Parsons et al., 2013). In addition, CRISPR/Cas9 KO of *P4H4* or *P4H10* in *N. benthamiana* led to a slight or strong decrease, respectively, of Pro hydroxylation of transiently expressed IgA1 hinge region (Uetz et al., 2022, 2024). In our experiments, the GFP reporters did not accumulate in the EM or in the TSP of the *ΔP4H1* and *ΔP4H3/10* cell lines (Figure II.4.7). However, when the parental cell line was analysed, the same result was obtained suggesting a silencing of the GFP reporter prior to the transformation with the Cas9/gRNAs constructs, making the experiment inconclusive. For the *ΔP4H4* cell lines, the analysis of the glycomodule profiles by western blot showed no significant differences with the parental cell lines. The inactivation of *P4H1* genes was verified by genomic DNA sequencing, which revealed editing in several cell lines. This suggests that either the genes in the P4H4 group are not involved in Pro hydroxylation in the $(SP)_n$ and SP_n repeats or that P4H4s can hydroxylate the glycomodules, but that other P4Hs may compensate for their absence. Importantly,

in all KO cell lines analysed, no obvious growth defect was observed and several ΔP4H4 cell lines were examined by microscopy and did not show aberrant morphology.

Further work is required to complete the screening of P4H groups. New reporter cell lines should be screened to find a cell line with a high and stable expression of the glycomodule reporters. These new cell lines will be transformed with the remaining constructs using the biolistic method to bypass the problems encountered with the *Agrobacterium* strain.

Alternatively, several P4H inhibitors can be used to reduce or abolish Pro hydroxylation. We have tested the chemical 2,2-dypiridyl, which is an iron chelator. However, there was no visible effect on the glycosylation of the reporter glycomodules but a severe growth defect, prohibiting the use of this compound to circumvent proline hydroxylation during protein production. Several other inhibitors such as 2OG analogues, proline analogues and 3,4-dihydroxybenzoic acid were used in Arabidopsis, carrot, BY-2 cells, soybean and carnation and showed an effect on *O*-glycosylated proteins and phenotypes typical of aberrant *O*-glycosylation such as disappearance of PRP from the culture medium (Schmidt et al., 1991) and aberrant cell wall morphology (Fragkostefanakis et al., 2014; Vlad et al., 2010). Interestingly, cultured soybean cells did not show aberrant morphology after inhibition (Schmidt et al., 1991).

General Discussion and Perspectives

Plant cell suspensions such as BY-2 have emerged as a powerful alternative platform for the production of recombinant therapeutic proteins. A generally recognised advantage of eukaryotic systems is their ability to perform complex PTMs but there are differences between the glycosylation capacity of plant and human cells. While most of the *N*-glycosylation pathway is similar between plants and mammals, the final processing steps differ. A remarkable property of plants is their ability to produce a simpler and more homogeneous *N*-glycosylation profile than mammalian cells. The proteins produced in plants are decorated with several non-human glycans. Many groups have developed strategies to address these differences: the removal of xylose and core α(1,3)-fucose, and the introduction of the pathway for terminal sialylation of complex *N*-glycans (Strasser, 2023). Plants, and particularly BY-2 cells, are well tolerant to glycoengineering and able to produce homogeneous humanised *N*-glycans without adverse phenotypes.

However, the capacity of plant cells for *O*-glycosylation is intrinsically different from that of humans. In fact, GalNAc-*O*-glycosylation is completely absent from plants. This modification is not fully understood, but there is increasing evidence that it plays important roles in protein bioactivity (Thompson & Wakarchuk, 2022). In addition, truncated GalNAc-*O*-glycans are often found at the surface of cancer cells and have been proposed as glycoepitopes for therapy (Julien et al., 2012; Kudelka et al., 2015). Due to its complexity and the lack of knowledge, *O*-glycosylation is difficult to control and an expression platform capable of producing therapeutics with a homogeneous and defined *O*-glycosylation profile is needed. Glycoengineering of plant cells offers a great opportunity because the complete absence of the GalNAc-*O*-glycosylation pathway enables expression of selected human glycosyltransferases without endogenous competition for protein substrates. Therefore, to broaden the potential applications and increase the performance of BY-2 cells as a platform for therapeutic protein production, we have engineered BY-2 cell lines to produce single GalNAc and core 1 *O*-glycans.

Collectively, our results demonstrated that the expression of hGalNAcT2 in BY-2 cells initiates GalNAc *O*-glycans on endogenous proteins and on the hinge region of IgA1. This glycan is termed the Tn antigen and its sialylated and non-sialylated forms are enriched at the surface of cancer cells (Burchell et al., 2018; Poiroux et al., 2017; Zlocowski et al., 2019). It has therefore been proposed that proteins carrying these *O*-glycans could be used as cancer vaccines or to produce anti-cancer antibodies for passive immunization (Kehler et al., 2022; Pinho & Reis, 2015). The expression of cancer associated proteins in the GalNAc engineered BY-2 cells could therefore be used for this purpose. The GalNAc-*O*-glycan was also elongated by co-expression of DmC1GalT1 to yield core 1 *O*-glycans. The glycosylation was very efficient on IgA1, which was mostly glycosylated with three to five core 1 *O*-glycans. This structure is the most abundant *O*-glycan in humans and is the basis for the core 2. The core 2

structure and its processed forms are important ligands for the development of the immune system (Cheng & Radhakrishnan, 2011; Ellies et al., 1998b; Nguyen et al., 2001). Engineering BY-2 cells to produce this structure by the expression of C2GnT and other enzymes such as β4GalTs and FucTIII for the formation of the Lewis X epitopes could be the next step in the development of BY-2 cells to produce tailored *O*-glycoproteins. Alternatively, the sialylated core 1 is an abundant structure found on many proteins such as IgA1 and could also be engineered in BY-2 cells. This would require the expression of the sialic acid synthesis pathway that is absent in plant cells (composed of seven enzymes) and the appropriate sialyltransferases. Engineering of the sialic acid pathway in BY-2 cells has been already demonstrated (Misaki et al., 2006; Paccalet et al., 2007), as has the generation of sialylated core 1 *O*-glycans in *N. benthamiana* (Castilho et al., 2012) and should thus be readily achievable in BY-2 cells.

The two other reporter proteins tested, EPO-Fc and the SARS-CoV2 Spike subunit 1, were subject to high proteolysis when produced in the secretory pathway of BY-2 cells. In the case of S1, no intact protein could be recovered in the EM making this substrate unsuitable for characterising the *O*-glycosylation capacity of the engineered BY-2 cells. For EPO-Fc, most of the proteins seem to be cleaved in the EPO domain just upstream of the Fc fusion. When expressed in GalNAc BY-2 cells, we were unable to detect HPL signal on purified EPO-Fc proteins. This result is surprising because GalNAcT2 is known to modify EPO (Castilho et al., 2012), but the *O*-site was probably cleaved during the proteolytic processing of the protein and the amount of intact EPO-FC protein was therefore too small to be detected by lectin blot.

The presence of sufficient amount of UDP-GalNAc in plant cells is controversial. Our data showed that BY-2 cells can efficiently glycosylate the hinge region of IgA1 without the need to boost this substrate pool, which is in agreement with other studies in *N. benthamiana* (Castilho et al., 2012; Dicker et al., 2016). Some authors have suggested that a UDP-GlcNAc/GalNAc epimerase is capable to promoting more efficient glycosylation (Daskalova et al., 2010). Although it is not necessary for IgA1, it could be for more heavily glycosylated substrates such as the mucin domain proteins (*e.g.* MUC16 has thirty *O*-sites). UDP-GalNAc has been detected in plants, but its function is unknown. Analysis of plant nucleotide sugars present in the secretory pathway (for example by carbohydrate electrophoresis (Barnes et al., 2016)) and studies of GalNAc incorporation in plant polysaccharides could contribute to uncovering its function and help determine the requirement to boost this substrate pool to achieve efficient *O*-glycosylation.

When produced in BY-2 cells (and other plant systems (Castilho et al., 2012; Göritzer et al., 2017; Karnoup et al., 2005; Weise et al., 2007)), EPO-Fc and IgA1 were modified with the plant-typical Hyp-*O*-glycosylation. Due to its potential immunogenicity and adverse impact on protein properties, this modification could

be a major obstacle to the development and approval of plant-made pharmaceuticals. To learn more about this modification in BY-2 cells, we studied the *P4H* genes responsible for proline hydroxylation. The *P4H* genes were largely unknown in *N. tabacum* and we identified 21 genes in the genomic databases, which clustered into six homology groups. Sequence analysis revealed that most were probably functional P4Hs. We constructed a co-expression network from RNA-seq data and analysed it for neighbour *NtP4H* genes. Several genes (*P4H4_A1/A2/B2*, *P4H9_B1* and *P4H10_A2/B2*) were correlated with AGP glycosyltransferases but no direct correlation was found for EXT glycosyltransferases.

We have initiated a CRISPR/Cas9 KO strategy of the *P4H* genes in BY-2 cell lines expressing two GFP reporters using EXT or AGP glycomodules. To date, our results demonstrated that the P4H4 group is either not implicated in hydroxylation of the glycomodules or that other P4H groups can compensate its absence. This corroborates results from Uetz et al. (2022) who suggested that P4H4 silencing led to a rebalancing by other P4Hs that could also hydroxylate the IgA1 hinge region in *N. benthamiana*. Other results obtained in plants showed important reduction of Hyps by KO of P4H1 and P4H10 in *P. patens* and *N. benthamiana* highlighting the importance of exploring all the P4H subgroups (Parsons et al., 2013; Uetz et al., 2024). Binary plasmids containing the gRNAs to target the other five groups have been generated. Transformation of BY-2 cells with these constructs into cell lines highly expressing glycomodule-GFP could allow us to study the involvement of each family in the Hyp-*O*-gycosylation process. Alternatively, the expression of therapeutic proteins known to be targeted by plant P4Hs (Mócsai et al., 2021) and the use of other analytical techniques such as Hyp quantification (Stephen C. Fry, 1993) could help refine the results.

On the other hand, co-expression of IgA1 with the GalNAc-*O*-glycosylation machinery showed a strong reduction in Hyp abundance in the hinge region (Figure II.2.9). Thus, complete elimination of P4H activity could not be an absolute requirement, at least for some proteins. Indeed, in the case of IgA1 the GalNAc- and Hyp-*O*-glycosylation sites are directly adjacent and can therefore influence each other.

In addition, Hyp-*O*-glycosylation has the potential to protect proteins and increase their secretion. This has led to the development of the glycomodules as tags to increase the yields of therapeutic proteins (Table I.3.1). While the $(SP)_{20}$ tag has not shown immunogenicity problems in mice (Xu et al., 2007), the use of this tag family in conjunction with cleavable linkers in BY-2 cells could be a reliable approach to improve the low yield and degradation observed for many proteins. However, this is not compatible with the complete elimination of P4H activity as the glycans are thought to be required for the glycomodule effect on secretion. It remains to be

demonstrated whether some P4H families only target specific glycomodules or whether the hydroxylation process is not regulated at this level.

In conclusion, this work enabled the development of the BY-2 cell platform towards the humanisation of PMP *O*-glycosylation. It has paved the way for further engineering and production of tailored *O*-glycans on recombinant proteins that can be used both for therapeutic applications and as reagents for research into the function of these poorly characterised modifications.

Supplementary Figures and Tables

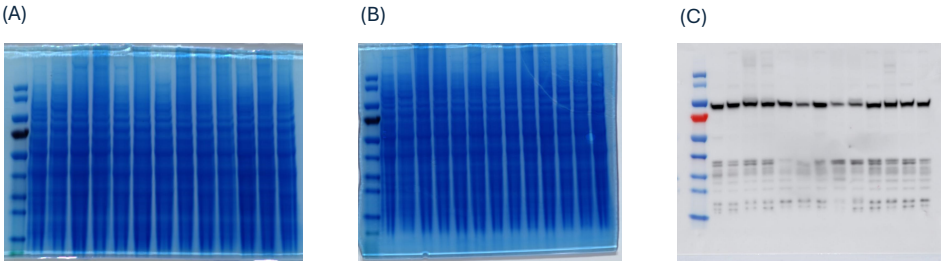


Figure S1: Analysis of *O*-glycosylation on endogenous BY-2 cells proteins. (A and B) Coomassie stained gel corresponding to the HPL blots of Figures II.1.2 (GalNAcE) and II.1.3. (C) Full PNL lectin blot from Figure II.1.3.

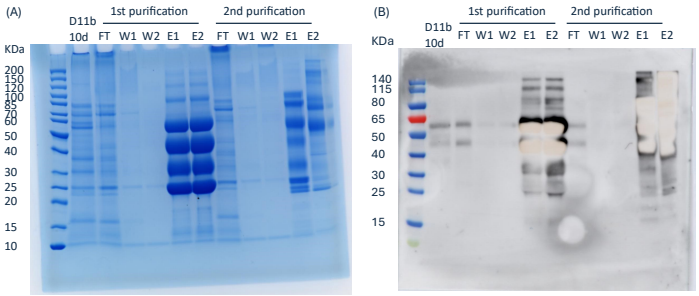


Figure S2: Follow-up purification of IgA1 from the flowthrough. IgA1 was purified from the EM of IgA1 BY-2 cells cultures after ten days in D11b (1st purification). The flowthrough was used as an input for a second purification. The fractions from the two purifications were analysed by SDS-PAGE under reducing conditions: (A) Coomassie stained gel and (B) anti-α HC western blot. Each lane was loaded with 35 µL of sample.

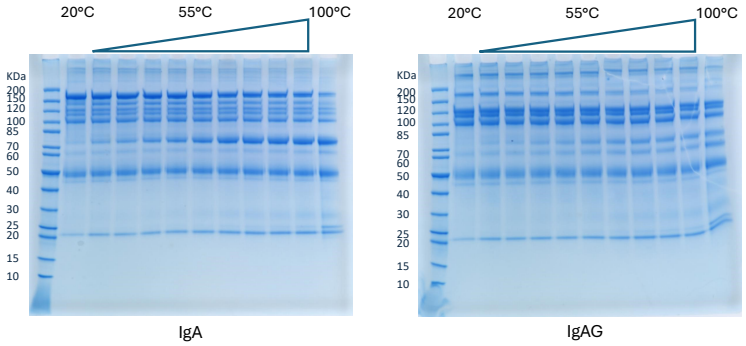


Figure S3: The SDS-PAGE profile of IgA1 depends on the sample preparation. IgA1 samples were solubilized in Laemmli buffer by incubation at 20°C for 20', 55°C for increasing time and then at 20°C until 20' or boiled for 10' before incubation at 20°C for 10'. The band profile depends on the thermal processing prior to PAGE analysis and varies between *O*-glycovariants.

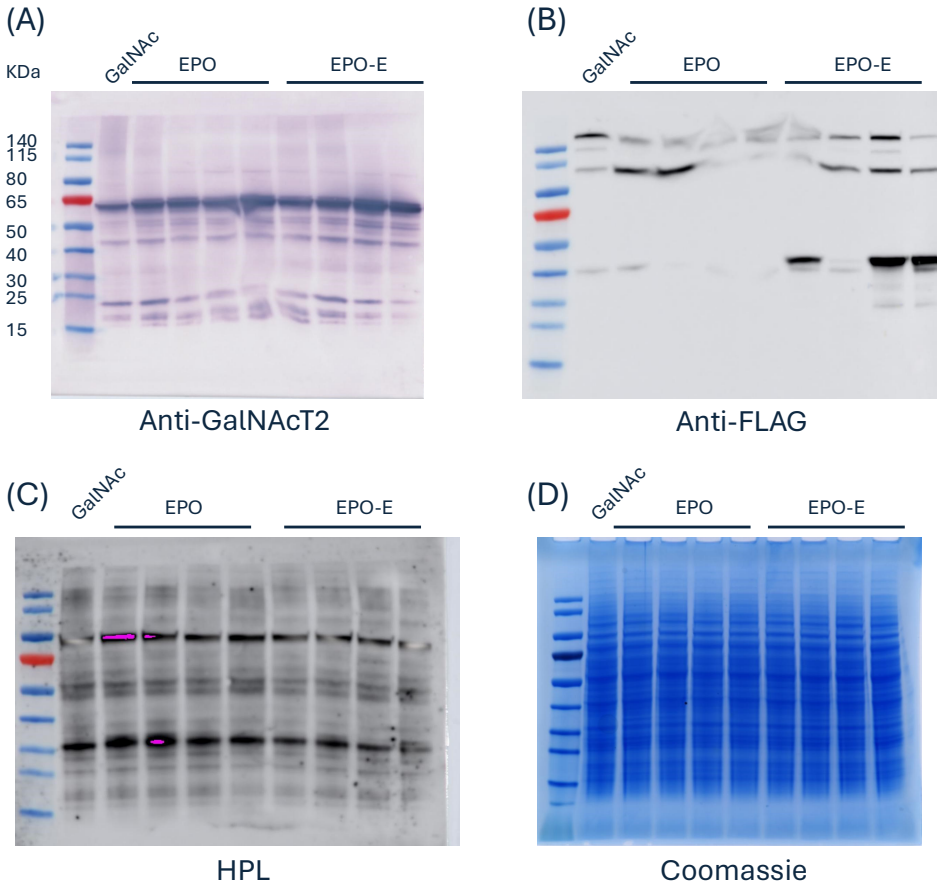


Figure S4: Analysis of EPO-Fc/GalNAc BY-2 cell lines. GalNAc BY-2 cell lines were transformed with EPO and EPO-E and analysed by SDS-PAGE. Blots were probed with (A) anti-GalNAcT2, (B) anti-FLAG and (C) HPL and (D) the gel was stained with Coomassie blue. Each lane was loaded with 25 µg of MF (GalNAcT2) or TSP (FLAG, HPL and Coomassie).

	P4H1		P4H4				P4H6				P4H9				P4H3/10					P4H12	
	P4H1_A	P4H1_B	P4H4_A1	P4H4_B1	P4H4_A2	P4H4_B2	P4H6_A1	P4H6_B1	P4H6_A2	P4H6_A3	P4H9_A1	P4H9_A2	P4H9_B1	P4H9_B2	P4H3	P4H10_A1	P4H10_B1	P4H10_A2	P4H10_B2	P4H12_A	P4H12_B
P4H1_A	100,0	99,1	37,8	30,0	37,5	38,9	35,5	37,0	34,7	15,0	36,2	37,3	36,4	36,5	46,2	44,1	39,0	47,9	46,9	31,3	29,3
P4H1_B	99,1	100,0	38,3	27,7	37,2	38,7	27,7	28,1	35,1	19,3	34,2	33,3	33,6	34,4	40,6	37,6	37,6	41,0	39,8	25,3	23,7
P4H4_A1	37,8	38,3	100,0	74,2	80,5	82,8	58,8	58,4	44,6	21,3	40,0	39,6	35,8	38,6	40,2	40,1	36,0	41,1	42,5	39,3	38,9
P4H4_B1	30,0	27,7	74,2	100,0	64,4	64,7	48,6	48,4	33,7	55,9	33,4	32,8	29,2	32,7	34,1	34,9	31,3	34,6	34,5	30,8	30,5
P4H4_A2	37,5	37,2	80,5	64,4	100,0	94,1	58,1	58,6	43,9	67,1	35,1	38,2	34,8	37,5	40,1	39,8	34,9	39,2	38,9	38,3	37,0
P4H4_B2	38,9	38,7	82,8	64,7	94,1	100,0	59,4	60,0	45,2	67,1	35,4	38,6	35,0	37,9	41,1	40,6	35,5	40,2	39,8	38,9	37,6
P4H6_A1	35,5	27,7	58,8	48,6	58,1	59,4	100,0	97,2	52,3	24,5	36,4	37,1	36,0	36,9	42,8	41,5	37,9	41,7	40,8	39,0	39,0
P4H6_B1	37,0	28,1	58,4	48,4	58,6	60,0	97,2	100,0	52,9	25,1	36,7	37,3	35,4	38,7	45,2	41,7	38,5	42,3	40,8	39,9	39,6
P4H6_A2	34,7	35,1	44,6	33,7	43,9	45,2	52,3	52,9	100,0	17,1	30,6	29,9	29,6	29,9	37,5	37,6	37,9	37,3	38,4	26,4	28,3
P4H6_A3	15,0	19,3	21,3	55,9	67,1	67,1	24,5	25,1	17,1	100,0	18,0	14,9	15,1	19,4	16,3	16,7	10,5	15,7	15,9	19,5	18,7
P4H9_A1	36,2	34,2	40,0	33,4	35,1	35,4	36,4	36,7	30,6	18,0	100,0	82,5	94,2	75,8	40,7	37,4	33,4	40,0	39,3	32,4	31,2
P4H9_B1	36,4	33,6	35,8	29,2	34,8	35,0	36,0	35,4	29,6	15,1	94,2	81,1	100,0	75,5	39,7	36,7	32,1	38,9	38,6	32,4	31,5
P4H9_A2	37,3	33,3	39,6	32,8	38,2	38,6	37,1	37,3	29,9	14,9	82,5	100,0	81,1	85,5	39,5	36,5	30,7	38,5	39,6	28,2	27,6
P4H9_B2	36,5	34,4	38,6	32,7	37,5	37,9	36,9	38,7	29,9	19,4	75,8	85,5	75,5	100,0	39,0	36,8	31,8	38,8	38,1	30,6	29,0
P4H3	46,2	40,6	40,2	34,1	40,1	41,1	42,8	45,2	37,5	16,3	40,7	39,5	39,7	39,0	100,0	69,2	58,0	73,8	72,9	27,6	27,0
P4H10_A1	44,1	37,6	40,1	34,9	39,8	40,6	41,5	41,7	37,6	16,7	37,4	36,5	36,7	36,8	69,2	100,0	85,8	81,8	80,4	25,2	25,5
P4H10_B1	39,0	37,6	36,0	31,3	34,9	35,5	37,9	38,5	37,9	10,5	33,4	30,7	32,1	31,8	58,0	85,8	100,0	72,2	70,3	22,7	22,5
P4H10_A2	47,9	41,0	41,1	34,6	39,2	40,2	41,7	42,3	37,3	15,7	40,0	38,5	38,9	38,8	73,8	81,8	72,2	100,0	96,9	28,5	27,9
P4H10_B2	46,9	39,8	42,5	34,5	38,9	39,8	40,8	40,8	38,4	15,9	39,3	39,6	38,6	38,1	72,9	80,4	70,3	96,9	100,0	28,5	27,4
P4H12_A	31,3	25,3	39,3	30,8	38,3	38,9	39,0	39,9	26,4	19,5	32,4	28,2	32,4	30,6	27,6	25,2	22,7	28,5	28,5	100,0	94,1
P4H12_B	29,3	23,7	38,9	30,5	37,0	37,6	39,0	39,6	28,3	18,7	31,2	27,6	31,5	29,0	27,0	25,5	22,5	27,9	27,4	94,1	100,0

Figure S5: Identity matrix of NtP4Hs. The identity for each pair of P4Hs was computed as follow: 100 * (identical positions) / (aligned positions + internal gap positions). The cells are color coded depending on the percentage: green > 60%; 30% < Yellow < 60%; red < 30%.

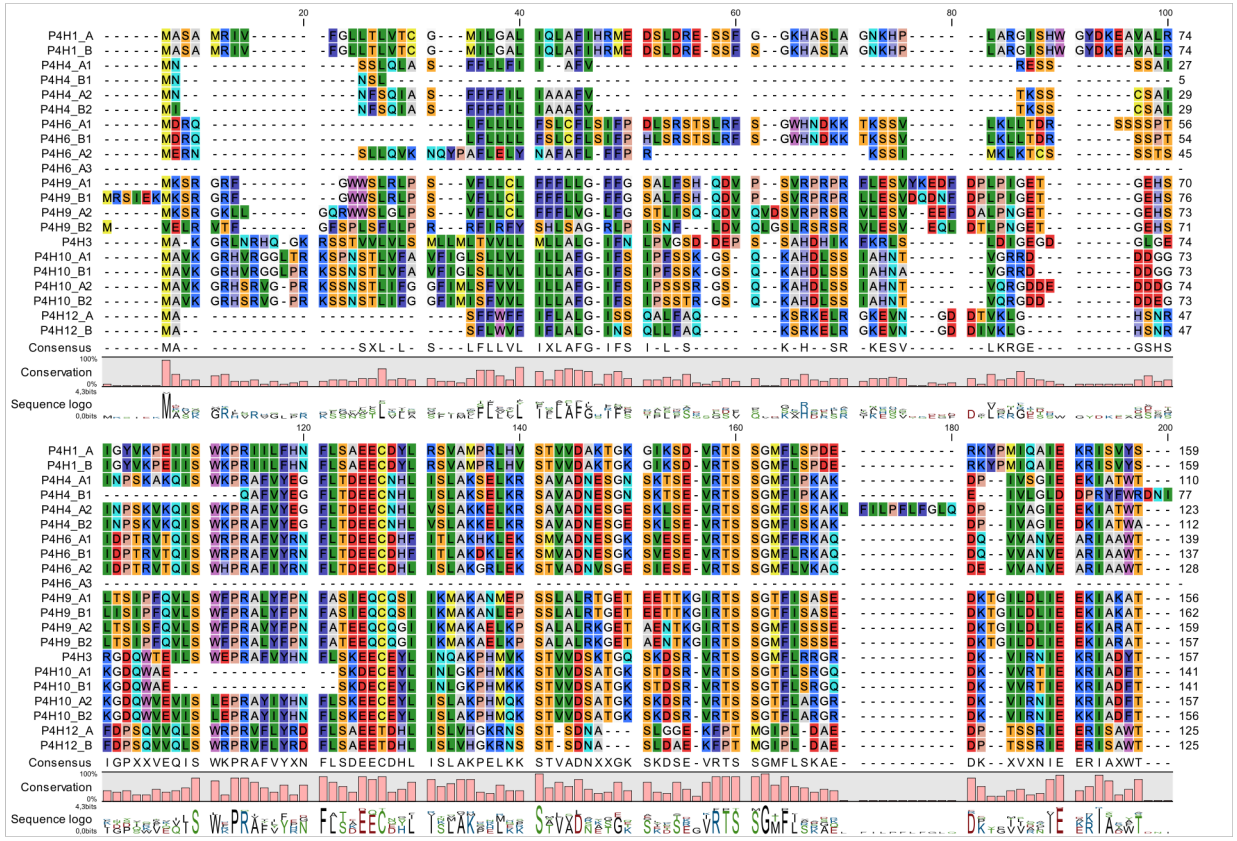


Figure S6: Alignment of NtP4Hs. The amino acid sequences of identified NtP4Hs were aligned with MAAFT.

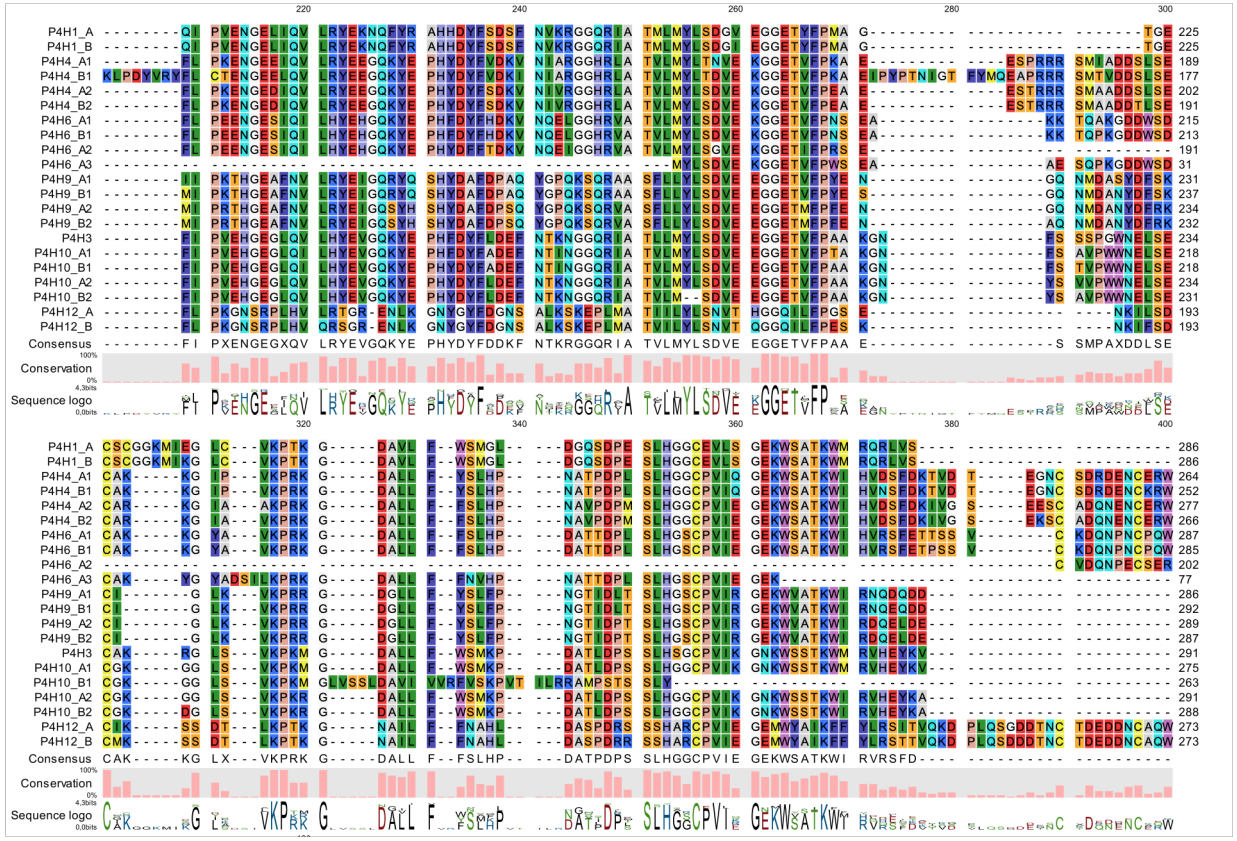


Figure S6: Continued

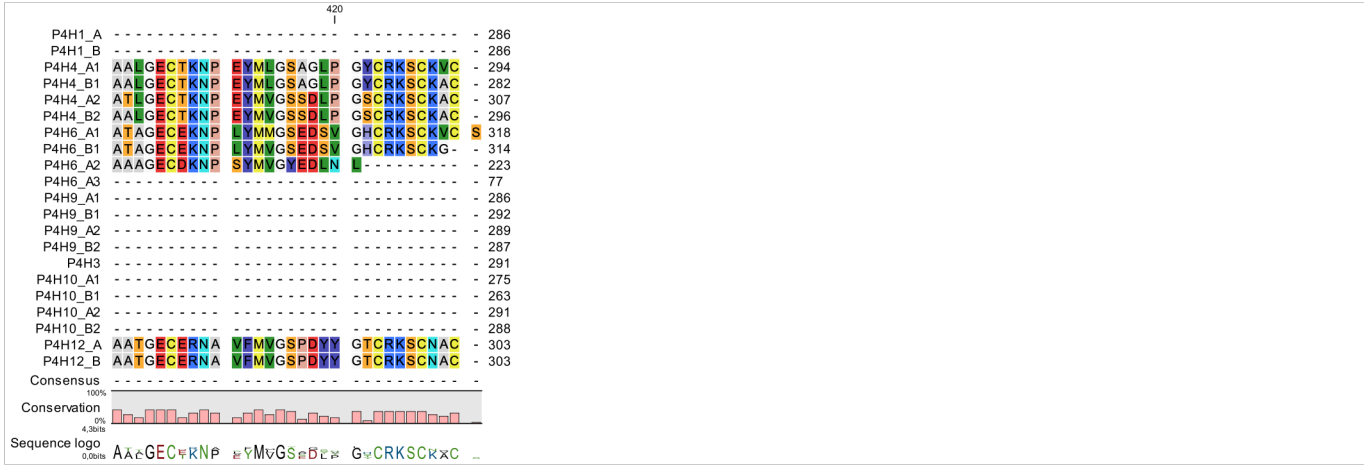


Figure S6: Continued

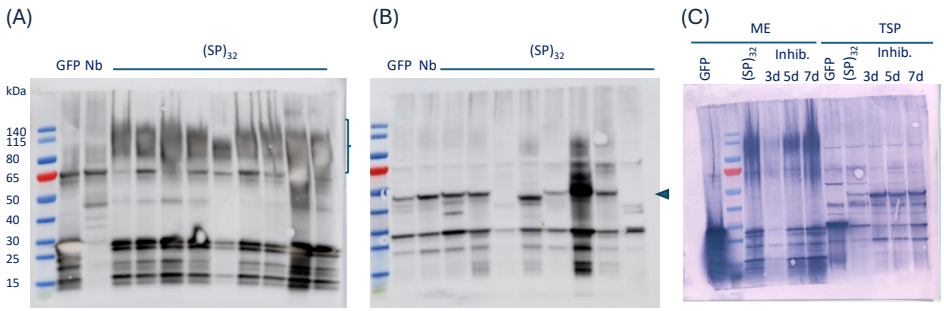


Figure S7: Analysis of $(SP)_{32}$-GFP lines and chemical inhibition. (A) EM (40 µL) and (B) TSP (25 µg) from $(SP)_{32}$-GFP cell lines were analysed by anti-GFP western blotting. The blot is also loaded with secreted GFP from BY-2 cells and $(SP)_{32}$-GFP from *N. benthamiana*. (C) Anti-GFP western analysis of chemical inhibition of P4Hs in BY-2 cells. P4Hs were inhibited by addition of 0.1 mM of 2,2-dypiridyl in the culture medium of $(SP)_{32}$ BY-2 cells. TSP and EM was collected after 3, 5 and 7 days. 40 µL of EM and 25 µg of TSP was loaded in each well.

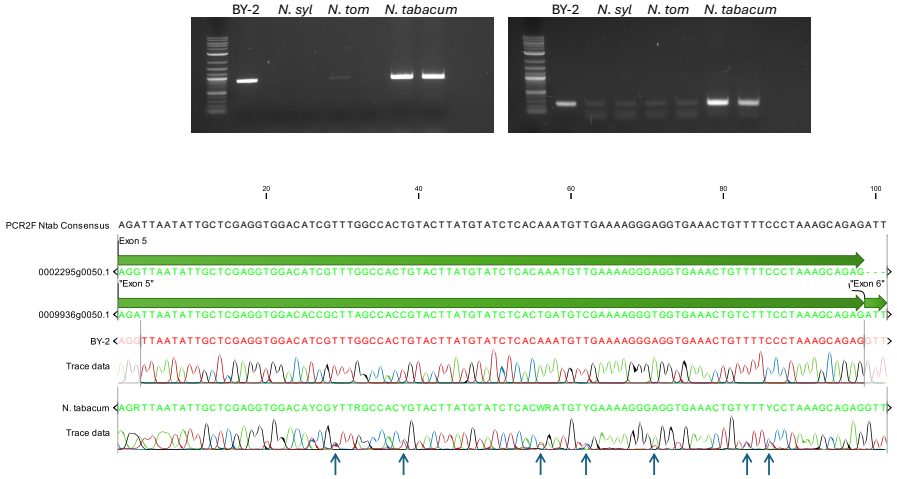


Figure S8: Analysis of the P4H4_A1/B1 pair in tobacco relatives. The pair of genes was PCR amplified from genomic DNA from BY-2 cells and *N. sylvestris*, *N. tomentosiformis* and *N. tabacum* plants in a single reaction. The PCR product from BY-2 cells and *N. tabacum* plant were sequenced. In BY-2 cells, only P4H4_A1 is amplified while the sequencing contains nearly equal signal from each gene in tobacco plants (differences shown by the arrows).

Table S1: Abundance of each glycoforms in the two *N*-glycosylation sites of IgA1 *O*-glycvariants by MS.

Glycoform	**relative ratios [%]**		
***N*-site1 (ANLTCTLTGLR)**	IgA	IgAG	IgAC
N1@single-GlcNAc_(Gn)	2,6	1,4	2,1
H2N2@MU	4,6	3,8	6,3
H3N2@MM	7,1	5,8	10,5
H3N3@GnM	31,0	36,4	42,7
H3N4@GnGn	19,4	41,5	11,6
H4N2@Man4	2,5	0,5	2,1
H5N2@Man5	12,8	2,1	7,4
H6N2@Man6	8,5	3,4	10,8
H7N2@Man7	8,8	2,8	5,6
H8N2@Man8	2,2	2,2	0,4
H4N3@Man4Gn	0,3	0,2	0,5
H5N3@Man5Gn	0,2	-	-
Sum	100	100	100
***N*-site2 (LAGKPTHVNVSVVMAE)**	IgA	IgAG	IgAC
Not glycosylated	11,4	6,7	-
H3N3@GnM	10,8	13,1	-
H3N4@GnGn	18,0	69,4	-
H6N2@Man6	13,1	-	-
H7N2@Man7	34,5	10,7	-
H8N2@Man8	12,2	-	-
Sum	100	100	-

Table S2: Abundance of each glycoforms in the hinge of IgA by MS.

Sorted by Mass		**Sorted by relative abundance**	
Glycoform	Relative ratio %	Glycoform	Relative ratio %
0xHyp0xPent0xHexNAc0xHex	6,8	6xHyp12xPent0xHexNAc0xHex	14,8
1xHyp0xPent0xHexNAc0xHex	4,6	6xHyp10xPent0xHexNAc0xHex	11,0
2xHyp0xPent0xHexNAc0xHex	3,7	6xHyp9xPent0xHexNAc0xHex	10,1
3xHyp0xPent0xHexNAc0xHex	2,0	6xHyp13xPent0xHexNAc0xHex	9,5
4xHyp0xPent0xHexNAc0xHex	1,2	0xHyp0xPent0xHexNAc0xHex	6,8
6xHyp5xPent0xHexNAc0xHex	0,2	6xHyp11xPent0xHexNAc0xHex	4,9
6xHyp6xPent0xHexNAc0xHex	1,2	1xHyp0xPent0xHexNAc0xHex	4,6
6xHyp7xPent0xHexNAc0xHex	2,7	6xHyp8xPent0xHexNAc0xHex	4,2
6xHyp8xPent0xHexNAc0xHex	4,2	5xHyp10xPent0xHexNAc0xHex	3,9
4xHyp9xPent0xHexNAc0xHex	0,4	2xHyp0xPent0xHexNAc0xHex	3,7
6xHyp9xPent0xHexNAc0xHex	10,1	6xHyp7xPent0xHexNAc0xHex	2,7
7xHyp5xPent1xHexNAc2xHex	0,6	5xHyp12xPent0xHexNAc0xHex	2,6
5xHyp10xPent0xHexNAc0xHex	3,9	5xHyp13xPent0xHexNAc0xHex	2,3
6xHyp10xPent0xHexNAc0xHex	11,0	3xHyp0xPent0xHexNAc0xHex	2,0
6xHyp5xPent1xHexNAc3xHex	0,3	6xHyp12xPent0xHexNAc1xHex	1,8
5xHyp11xPent0xHexNAc0xHex	1,3	6xHyp13xPent0xHexNAc1xHex	1,7
6xHyp11xPent0xHexNAc0xHex	4,9	5xHyp11xPent0xHexNAc0xHex	1,3
6xHyp10xPent0xHexNAc1xHex	1,2	6xHyp12xPent0xHexNAc2xHex	1,3
5xHyp12xPent0xHexNAc0xHex	2,6	6xHyp10xPent0xHexNAc1xHex	1,2
6xHyp12xPent0xHexNAc0xHex	14,8	4xHyp0xPent0xHexNAc0xHex	1,2
7xHyp8xPent1xHexNAc2xHex	1,1	6xHyp6xPent0xHexNAc0xHex	1,2
6xHyp10xPent0xHexNAc2xHex	0,3	7xHyp8xPent1xHexNAc2xHex	1,1
5xHyp13xPent0xHexNAc0xHex	2,3	7xHyp18xPent0xHexNAc0xHex	1,0
6xHyp13xPent0xHexNAc0xHex	9,5	5xHyp10xPent0xHexNAc4xHex	0,9
7xHyp13xPent0xHexNAc0xHex	0,3	6xHyp7xPent1xHexNAc4xHex	0,8
6xHyp12xPent0xHexNAc1xHex	1,8	7xHyp5xPent1xHexNAc2xHex	0,6
6xHyp7xPent1xHexNAc4xHex	0,8	7xHyp17xPent0xHexNAc0xHex	0,4
4xHyp9xPent0xHexNAc4xHex	0,3	4xHyp9xPent0xHexNAc0xHex	0,4
6xHyp13xPent0xHexNAc1xHex	1,7	5xHyp12xPent0xHexNAc3xHex	0,4
6xHyp12xPent0xHexNAc2xHex	1,3	7xHyp14xPent1xHexNAc3xHex	0,4
5xHyp10xPent0xHexNAc4xHex	0,9	7xHyp13xPent0xHexNAc0xHex	0,3
5xHyp12xPent0xHexNAc3xHex	0,4	6xHyp5xPent1xHexNAc3xHex	0,3
7xHyp17xPent0xHexNAc0xHex	0,4	6xHyp10xPent0xHexNAc2xHex	0,3
7xHyp18xPent0xHexNAc0xHex	1,0	4xHyp9xPent0xHexNAc4xHex	0,3
7xHyp14xPent1xHexNAc3xHex	0,4	6xHyp5xPent0xHexNAc0xHex	0,2

Table S3: Abundance of each glycoforms in the hinge of IgAG by MS.

Sorted by Mass		**Sorted by relative abundance**	
Glycoform	Relative ratio %	Glycoform	Relative ratio %
0xHyp0xPent0xHexNAc0xHex	7,6	2xHyp3xPent4xHexNAc0xHex	11,0
1xHyp0xPent0xHexNAc0xHex	3,8	1xHyp0xPent4xHexNAc0xHex	10,0
2xHyp0xPent0xHexNAc0xHex	2,4	3xHyp3xPent3xHexNAc0xHex	8,3
0xHyp0xPent1xHexNAc0xHex	1,5	0xHyp0xPent0xHexNAc0xHex	7,6
3xHyp0xPent2xHexNAc0xHex	0,7	1xHyp3xPent5xHexNAc0xHex	7,6
7xHyp1xPent0xHexNAc2xHex	0,5	5xHyp6xPent2xHexNAc0xHex	6,4
0xHyp0xPent3xHexNAc0xHex	0,4	3xHyp0xPent3xHexNAc0xHex	5,2
3xHyp0xPent2xHexNAc1xHex	1,5	4xHyp3xPent3xHexNAc0xHex	4,4
1xHyp0xPent3xHexNAc0xHex	1,4	1xHyp0xPent0xHexNAc0xHex	3,8
3xHyp0xPent3xHexNAc0xHex	5,2	8xHyp2xPent3xHexNAc3xHex	3,5
4xHyp0xPent3xHexNAc0xHex	1,6	3xHyp3xPent4xHexNAc0xHex	2,9
5xHyp3xPent1xHexNAc0xHex	1,5	2xHyp0xPent0xHexNAc0xHex	2,4
2xHyp3xPent1xHexNAc1xHex	0,7	1xHyp3xPent3xHexNAc0xHex	2,4
3xHyp0xPent3xHexNAc1xHex	0,5	5xHyp3xPent2xHexNAc0xHex	2,3
1xHyp0xPent4xHexNAc0xHex	10,0	4xHyp0xPent4xHexNAc2xHex	1,9
2xHyp3xPent2xHexNAc0xHex	0,8	4xHyp0xPent3xHexNAc0xHex	1,6
5xHyp3xPent2xHexNAc0xHex	2,3	3xHyp0xPent2xHexNAc1xHex	1,5
2xHyp2xPent3xHexNAc0xHex	0,2	5xHyp3xPent1xHexNAc0xHex	1,5
3xHyp2xPent3xHexNAc0xHex	0,7	0xHyp0xPent1xHexNAc0xHex	1,5
1xHyp3xPent3xHexNAc0xHex	2,4	1xHyp0xPent3xHexNAc0xHex	1,4
3xHyp3xPent3xHexNAc0xHex	8,3	5xHyp6xPent3xHexNAc0xHex	1,3
4xHyp3xPent3xHexNAc0xHex	4,4	1xHyp6xPent0xHexNAc3xHex	1,3
4xHyp5xPent2xHexNAc0xHex	0,4	8xHyp1xPent3xHexNAc3xHex	1,2
5xHyp5xPent2xHexNAc0xHex	0,4	3xHyp0xPent4xHexNAc2xHex	1,1
3xHyp0xPent4xHexNAc2xHex	1,1	3xHyp6xPent2xHexNAc0xHex	1,0
4xHyp0xPent4xHexNAc2xHex	1,9	5xHyp0xPent4xHexNAc2xHex	0,9
5xHyp0xPent4xHexNAc2xHex	0,9	2xHyp3xPent2xHexNAc0xHex	0,8
2xHyp3xPent4xHexNAc0xHex	11,0	3xHyp0xPent2xHexNAc0xHex	0,7
3xHyp6xPent2xHexNAc0xHex	1,0	2xHyp3xPent1xHexNAc1xHex	0,7
3xHyp3xPent4xHexNAc0xHex	2,9	3xHyp2xPent3xHexNAc0xHex	0,7
5xHyp6xPent2xHexNAc0xHex	6,4	3xHyp6xPent4xHexNAc0xHex	0,6
1xHyp6xPent0xHexNAc3xHex	1,3	7xHyp1xPent0xHexNAc2xHex	0,5
8xHyp1xPent3xHexNAc3xHex	1,2	3xHyp0xPent3xHexNAc1xHex	0,5
1xHyp3xPent5xHexNAc0xHex	7,6	4xHyp5xPent2xHexNAc0xHex	0,4
5xHyp6xPent3xHexNAc0xHex	1,3	5xHyp5xPent2xHexNAc0xHex	0,4
8xHyp2xPent3xHexNAc3xHex	3,5	0xHyp0xPent3xHexNAc0xHex	0,4
3xHyp6xPent4xHexNAc0xHex	0,6	2xHyp2xPent3xHexNAc0xHex	0,2

Table S4: Abundance of each glycoforms in the hinge of IgAC by MS.

Sorted by Mass		**Sorted by relative abundance**	
Glycoform	Relative ratio %	Glycoform	Relative ratio %
3xHyp0xPent2xHexNAc0xHex	0,2	0xHyp0xPent4xHexNAc4xHex	14,3
2xHyp0xPent1xHexNAc2xHex	0,4	1xHyp0xPent3xHexNAc3xHex	7,9
3xHyp0xPent1xHexNAc2xHex	0,4	0xHyp0xPent4xHexNAc3xHex	7,6
3xHyp0xPent2xHexNAc1xHex	1,3	1xHyp0xPent4xHexNAc4xHex	7,2
2xHyp0xPent1xHexNAc3xHex	0,2	0xHyp0xPent3xHexNAc3xHex	6,0
1xHyp0xPent2xHexNAc2xHex	1,2	2xHyp0xPent3xHexNAc3xHex	4,6
3xHyp0xPent2xHexNAc2xHex	3,1	1xHyp0xPent4xHexNAc3xHex	4,1
1xHyp0xPent3xHexNAc1xHex	0,7	0xHyp0xPent5xHexNAc5xHex	3,5
4xHyp0xPent2xHexNAc2xHex	1,1	0xHyp0xPent5xHexNAc4xHex	3,2
2xHyp0xPent3xHexNAc1xHex	0,9	3xHyp0xPent2xHexNAc2xHex	3,1
5xHyp0xPent2xHexNAc2xHex	0,3	1xHyp0xPent3xHexNAc2xHex	3,1
3xHyp0xPent3xHexNAc1xHex	0,6	0xHyp0xPent4xHexNAc2xHex	2,9
2xHyp0xPent1xHexNAc4xHex	0,1	0xHyp0xPent5xHexNAc3xHex	2,8
2xHyp0xPent2xHexNAc3xHex	0,9	2xHyp0xPent3xHexNAc2xHex	2,7
0xHyp0xPent3xHexNAc2xHex	1,6	0xHyp0xPent4xHexNAc5xHex	2,5
1xHyp0xPent3xHexNAc2xHex	3,1	1xHyp0xPent3xHexNAc4xHex	2,4
2xHyp3xPent1xHexNAc2xHex	0,5	0xHyp0xPent3xHexNAc2xHex	1,6
2xHyp0xPent3xHexNAc2xHex	2,7	3xHyp0xPent2xHexNAc1xHex	1,3
4xHyp0xPent3xHexNAc2xHex	0,4	2xHyp0xPent4xHexNAc4xHex	1,2
2xHyp0xPent4xHexNAc1xHex	0,3	1xHyp0xPent2xHexNAc2xHex	1,2
2xHyp6xPent1xHexNAc0xHex	0,2	4xHyp0xPent2xHexNAc2xHex	1,1
0xHyp3xPent0xHexNAc4xHex	0,2	0xHyp0xPent3xHexNAc5xHex	1,1
0xHyp0xPent3xHexNAc3xHex	6,0	2xHyp3xPent2xHexNAc2xHex	1,0
1xHyp0xPent3xHexNAc3xHex	7,9	1xHyp3xPent3xHexNAc3xHex	1,0
2xHyp0xPent3xHexNAc3xHex	4,6	2xHyp3xPent3xHexNAc3xHex	1,0
0xHyp0xPent4xHexNAc2xHex	2,9	2xHyp0xPent2xHexNAc3xHex	0,9
2xHyp3xPent2xHexNAc2xHex	1,0	2xHyp0xPent4xHexNAc3xHex	0,9
4xHyp3xPent2xHexNAc2xHex	0,2	2xHyp0xPent3xHexNAc1xHex	0,9
1xHyp0xPent3xHexNAc4xHex	2,4	1xHyp0xPent3xHexNAc5xHex	0,9
2xHyp0xPent3xHexNAc4xHex	0,6	1xHyp0xPent3xHexNAc1xHex	0,7
0xHyp0xPent4xHexNAc3xHex	7,6	3xHyp0xPent3xHexNAc1xHex	0,6
1xHyp3xPent2xHexNAc3xHex	0,4	2xHyp0xPent3xHexNAc4xHex	0,6
1xHyp0xPent4xHexNAc3xHex	4,1	1xHyp2xPent3xHexNAc3xHex	0,6
2xHyp0xPent4xHexNAc3xHex	0,9	0xHyp6xPent2xHexNAc1xHex	0,5
1xHyp3xPent3xHexNAc2xHex	0,4	2xHyp3xPent1xHexNAc2xHex	0,5
0xHyp6xPent2xHexNAc1xHex	0,5	1xHyp3xPent3xHexNAc2xHex	0,4
1xHyp2xPent3xHexNAc3xHex	0,6	2xHyp2xPent3xHexNAc3xHex	0,4
2xHyp2xPent3xHexNAc3xHex	0,4	3xHyp0xPent1xHexNAc2xHex	0,4
0xHyp0xPent3xHexNAc5xHex	1,1	2xHyp0xPent1xHexNAc2xHex	0,4
1xHyp0xPent3xHexNAc5xHex	0,9	4xHyp0xPent3xHexNAc2xHex	0,4
0xHyp0xPent4xHexNAc4xHex	14,3	1xHyp3xPent2xHexNAc3xHex	0,4
1xHyp0xPent4xHexNAc4xHex	7,2	5xHyp0xPent2xHexNAc2xHex	0,3

Table S4: Continued.

Sorted by Mass		**Sorted by relative abundance**	
Glycoform	Relative ratio %	Glycoform	Relative ratio %
2xHyp0xPent4xHexNAc4xHex	1,2	2xHyp0xPent4xHexNAc1xHex	0,3
0xHyp0xPent5xHexNAc3xHex	2,8	1xHyp0xPent5xHexNAc4xHex	0,3
1xHyp3xPent3xHexNAc3xHex	1,0	2xHyp6xPent1xHexNAc0xHex	0,2
2xHyp3xPent3xHexNAc3xHex	1,0	4xHyp3xPent2xHexNAc2xHex	0,2
0xHyp0xPent4xHexNAc5xHex	2,5	2xHyp0xPent1xHexNAc3xHex	0,2
0xHyp0xPent5xHexNAc4xHex	3,2	0xHyp3xPent0xHexNAc4xHex	0,2
1xHyp0xPent5xHexNAc4xHex	0,3	3xHyp0xPent2xHexNAc0xHex	0,2
0xHyp0xPent5xHexNAc5xHex	3,5	2xHyp0xPent1xHexNAc4xHex	0,1

Table S5: Co-expression of AGP and EXT related *GTs* with *P4Hs* in the generated network. For each gene, the number of identified neighbours and of EXT and AGP *GTs* neighbours is indicated as well as their implication in the biosynthetic pathways.

Gene	**Neighbours**	***GT* neighbours**	**Biosynthetic pathway**
P4H4_A1	423	2	AGP
P4H4_B2	553	12	AGP
P4H4_A2	1091	3	AGP
P4H9_B1	1843	1	AGP
P4H10_B2	41	3	AGP/EXT
P4H10_A2	446	9	AGP

Table S6: Sequences of the gRNAs used for inactivation of P4H1, 4 and 3/10.

Name	Sequence
P4H4	
23F	AAAATCGGAGCTGAAGAGAT
81F	AGTGAGGTTAGGACTAGCTC
59F	CTTGTCTGAATGTGCAAAGA
59R	GACTTTAGAGGGATTGATGA
45F	GCTGTGGCAGATAATGAGTC
149F	CTCAGTGATGTCGAGAAAGG
P4H1	
101F	AGCATTTATTCATAGGATGG
253F	TGAAGCCTGAAATTATTAGC
55F	TGTTTTTGAGCCCTGATGAG
P4H3/10	
56F	AATGGACGGAGATTCTATCT
40R	ACTGTTGACTTCACCATGTG
114F	GTCCTCATTCTTCTCGCTTT
53F	GAGGACGTGATAAAGTGATC
147F	TCGATTCCTTTCAGTTCTAA
32F	GTTCTGGAACATTTCTTTCC

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), 30 g/L sucrose, 0.2 g/L KH_2PO_4, 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 blots 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. benthamiana 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}).

2 BY-2 cell culture growth kinetics

Three independent transgenic cell lines from each cell type were inoculated at known concentration 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 (Millipore, 475855) and the cell pack fresh weight was measured after centrifugation for 5 min at 2,700*g*.

3 Binary plasmid construction

Unless stated otherwise, the constructions were made using the Golden Gate plant toolkit (Engler et al., 2014) following the protocol described by (Marillonnet & Grützner, 2020). The following table presents the primers used for cloning:

Table M1: Primers used for genetic constructions.

Primer	Sequence
IgA HC F	ACGACCGAAGACTTAATGGCTAACAAGCACCTTAG
IgA HC R	TGAAGACTTAAGCTAGTAGCAGGTTCCATCCACTT
IgA LC F	ACGACCGAAGACTTAATGGCTAACAAGCACCTTAG
IgA LC R	TGAAGACTTAAGCTACTAACACTCTCCCCTGTTGA
IgG2 Fc 1R	ACGACCGAAGACATGTCCCTTGTTTGACACTTTAC
IgG2 Fc 1F	ACGACCGAAGACATTTCGGAAAGAAAATGTTGCGTTG

IgG2 Fc 2R	AGCATTGAAGACATAAGCTATGGTGATAAACTGAGTGATT
IgG2 Fc 2F	ACGACCGAAGACATGGACTCCCTGCTCCAATTGAAAA
tExt F	GGTCTCGGCTTAAAGCAGAATGCTGAGCTAA
tExt R	GGTCTCGAGCGCCGTCATAACTGTAGAAATG
GFP F	GAAGACAGAGGTATGGTGAGCAAGGGCGAGGA
GFP R	GAAGACATAAGCTTACTTGTACTGCTCGTCCA

The cloning of the level 0 plasmids pICH41258-SP_{PDI}, pICH41295-pPMA4, pGEMT-tPMP1 and pICH41331-SAR and the level 1 pICH47732-nptII and pICH47732-hptI are described in Navarre et al. (2023). The generated binary plasmids are listed in Table M2.

3.1 GalNAc-*O*-glycosylation plasmids

ORFs encoding human GalNAcT2 (NP_004472.1), *D. melanogaster* C1GalT1 (NP_609258.1) and *P. aeruginosa* WbpP GlcNAc C4-epimerase (AAF23998.1) were designed for entry into level 0 CDS1 Golden Gate plasmid (pICH41308): AATG and GCTT fusion sites were added with BsaI restriction sites in 5' and 3', respectively. The sequences were synthesised by Genewiz (GalNAcT2) and Genscript (WbpP and C1GalT1) and cloned into pICH41308. Next, expression cassettes were constructed into level 1 plasmids under the control of *p35S* (pICH51288) and *tNOS* (pICH41421). Additionally, a N-terminal FLAG tag (DYKDHDGDYKDHDLDA AAADYKDDDD) (pICSL30005) was fused to the WbpP construct. Different level 1 plasmids were used to define correct position of the expression cassettes in the final constructs: pICH47742 (position 2, GalNAcT2 or C1GalT1 or WbpP), pICH47751 (position 3, C1GalT1 or WbpP) and pICH47761 (position 4, WbpP). In addition, the neomycin phosphotransferase II expression cassette from the kit (pICSL70004) was transferred in the position 1 plasmid pICH47732. Finally, the *N. tabacum Rb7* scaffold attachment region (SAR) genetic insulator was PCR amplified from pPZP-gS-SAR (see 3.4) to add 5'-AATG and GCTT-3' fusion sites with BsaI restriction sites. It was then cloned into the gene level 0 plasmid (pICH41331) and then into level 1 position 3, 4 and 5 (pICH47751, pICH47761 and pICH47772). The expression cassettes were then assembled into the level M pAGM8031 plasmid to generate: pAGM8031-nptII-GalNAcT2-SAR, pAGM8031-nptII-GalNAcT2-C1GalT1-SAR, pAGM8031-nptII-GalNAcT2-C1GalT1-WbpP-SAR and pAGM8031-nptII-GalNAcT2-WbpP-SAR.

3.2 IgA1 plasmids

Sequences coding for IgA1 HC and LC were received from Strasser's lab in pEAQ (Göritzer et al., 2017) and PCR amplified to add 5'-AATG and GCTT-3' and BpiI restriction sites for cloning into the level 0 CDS1 plasmid pICH41308. The *N. tabacum Ext* terminator was PCR amplified to add 5'-GCTT and CGCT-3' and BsaI restriction sites from pAUX3131-pPMA4-S1-tExt (see 3.4) and cloned into pICH41276. The ORFs were then assembled with *pPMA4* and *tExt* in different level 1 plasmids: pICH47742 and pICH47761 for HC; pICH47751 and pICH47772 for LC. The HC and LC expression cassettes were then assembled with the *nptII*, *GalNAcT2* and *C1GalT1* expression cassettes into the level M pAGM8031 plasmid to generate: pAGM8031-nptII-HC-LC, pAGM8031-nptII-HC-LC-GalNAcT2 and pAGM8031-nptII-GalNAcT2-C1GalT1-HC-LC.

3.3 EPO plasmids

A sequence from human EPO (NM_000799.1) without the signal peptide (starting at G27) was synthesized (IDT). The sequence was flanked by 5'-AGGT and TTCG-3' and BpiI restriction site for cloning into the CDS2 ns plasmid pAGM1299. To keep the sequence in frame a GS linker was added at the 3' end (GGT TCG). The Fc fragment of LoBm2 IgG2 (E235 to K462) (Magy et al., 2014) was PCR amplified for cloning into Golden Gate module. The sequence contained a BpiI site that was removed (GGTCTC to GGACTC, keeping the GGN encoded Gly) according to the Golden Gate protocol (Marillonnet & Grützner, 2020). The amplified Fc fragment contained 5'-TTCG and GCTT-3' and BpiI restriction site and was cloned in the CT module pAGM1301. The ORFs of EPO and Fc were assembled with *pPMA4*, the SP from PDI and *tPMP1* in level 1 plasmids at position 2, 3 and 4 (pICH47742, pICH47751 and pICH47761). The expression cassettes were cloned with *nptII*, *C1GalT1* and *WbpP* to form pAGM8031-hptI-EPOFc, pAGM8031-hptI-C1GalT1-EPOFc, pAGM8031-hptI-WbpP-EPOFc and pAGM8031-hptI-WbpP-C1GalT1-EPOFc.

3.4 Spike glycoprotein plasmids

The S1 expression cassettes were made using the pPZP-RCS2 binary vectors (Goderis et al., 2002). An Arabidopsis codon optimised sequence of S1 (amino acid 1-685) with a C-terminal $(G_4S)_2$ linker followed by a 10His tag and flanked by 5'-KpnI and 3'-SacI was synthesized (Genewiz). KpnI/SacI digestion was used to replace gB by S1 in pAUX3131-pPMA4-gB-tNOS (Smargiasso et al., 2019) or pAUX3131-pPMA4-gB-tExt. The 0.7 kb terminator of *PMP1* was PCR amplified

from pRS315-PMP1 (Navarre et al., 1992) to add 5'-SacI and 3'-AgeI and the PCR product was used to replace *tNOS*, forming pAUX3131-pPMA4-S1-tPMP1. The S1 expression cassettes were inserted into pPZP-hptI-mCherry I-SceI cloning site and a *N. tabacum Rb7* SAR was cloned in the Pi-PspI site.

3.5 Reporter GFP plasmids

The reporter and control GFP plasmids were based on *Aequorea victoria* sequence (AAA27722.1). First, the GFP sequence from the CDS1 module pICH41531 was PCR amplified to add 5'-AGGT and GCTT-3' fusion sites and BsaI restriction site for cloning into the CDS2 plasmid pICH41264. Then the expression cassette for secreted GFP was built by assembly of *p35S* (pICH51288), the *N. benthamiana calreticulin* signal peptide (SPcal, pICH37326), pICH41264-GFP and *tNOS* (pICH41421) in pICH47742. Next, $(SP)_{32}$ and $(SP_4)_{18}$ glycomodule sequences were synthesised with the 5'-AGGT and TTCG-3' fusion site for CDS2 ns modules and cloned into pUC57 (Genscript). The glycomodules were assembled with *p35S*, SP_{cal}, a C-terminal GFP (pICSL50008) and *tNOS* in pICH47742 to form the expression cassettes. The three expression cassettes were then cloned in position 2 of the binary plasmid pAGM8031 with *hptI* or *nptII* expression cassette in position 1.

3.6 CRISPR/Cas9 plasmids

To generate the CRISPR/Cas9 constructs, the cloning toolkit designed by Hahn et al. (2020) was used. Each guide RNA was synthetised as a pair of oligonucleotides with Golden Gate compatible ends (forward 5'-GAAGACAATGCA-[gRNA sequence]- GTTTTTGTCTTC-3'; reverse 5'-GAAGACAAAAAC-[gRNA sequence]- TGCATTGTCTTC-3'). The matching oligonucleotides were annealed by heating 10' at 100°C and allowing cooling to room temperature. The annealed gRNAs were then cloned into individual plasmids by Golden Gate reaction (guide 1 pFH49, guide 2 pAK002, guide 3 pAK003, guide 4 pAK004, guide 5 pAK005 and guide 6 pAK006). The guides were then assembled into pICH47751 (position 3) with the *Arabidopsis thaliana* U6 promoter (pFH34) and the appropriate linker for the desired number of guides. The final plasmids were generated by assembly of the guides construct with the *nptII* resistance (position 1) and Cas9 (pFH52, p35S-spCas9-tNOS, position 2) in pAGM8031.

Table M2: Generated binary plasmids

Name	Backbone	Resistance	Constructs
GalNAc	pAGM8031	*nptII*	p35S-GalNAcT2-tNOS
GalNAcE	pAGM8031	*nptII*	p35S-GalNAcT2-tNOS_p35S-3xFLAG-WbpP-tNOS
Core1	pAGM8031	*nptII*	p35S-GalNAcT2-tNOS_ p35S-C1GalT1-tNOS
Core1E	pAGM8031	*nptII*	p35S-GalNAcT2-tNOS_ p35S-C1GalT1-tNOS_p35S-3xFLAG-WbpP-tNOS
IgA	pAGM8031	*nptII*	pPMA4-IgA1 HC-tExt_pPMA4-IgA1 Lc-tExt
IgAG	pAGM8031	*nptII*	pPMA4-IgA1 HC-tExt_pPMA4-IgA1 Lc-tExt_ p35S-GalNAcT2-tNOS
IgAC	pAGM8031	*nptII*	p35S-GalNAcT2-tNOS_ p35S-C1GalT1-tNOS_ pPMA4-IgA1 HC-tExt_pPMA4-IgA1 Lc-tExt
EPO	pAGM8031	*hptI*	pPMA4-SP_{PDI}-EPO-Fc-tPMP1
EPO-E	pAGM8031	*hptI*	p35S-3xFLAG-WbpP-tNOS_ pPMA4-SP_{PDI}-EPO-Fc-tPMP1
EPO-C	pAGM8031	*hptI*	p35S-C1GalT1-tNOS_ pPMA4-SP_{PDI}-EPO-Fc-tPMP1
EPO-CE	pAGM8031	*hptI*	p35S-3xFLAG-WbpP-tNOS_p35S-C1GalT1-tNOS_ pPMA4-SP_{PDI}-EPO-Fc-tPMP1
gS-tExt	pPZP	*hptI*	pPMA4-S1-10His-tEXT_p35S-mCherry-tNOS_SAR
gS-tNOS	pPZP	*hptI*	pPMA4-S1-10His-tNOS_p35S-mCherry-tNOS_SAR
gS-tPMP1	pPZP	*hptI*	pPMA4-S1-10His-tPMP1_p35S-mCherry-tNOS_SAR
$(SP)_{32}$	pAGM8031	*hptI*/*nptII*	p35S-SP_{Cal}-$(SP)_{32}$-GFP-tNOS
$(SP_4)_{18}$	pAGM8031	*hptI*/*nptII*	p35S-SP_{Cal}-$(SP_4)_{18}$-GFP-tNOS
ΔP4H1	pAGM8031	*hptII*	p35S-Cas9-tNOS_pU6-(6 gRNAs P4H1)
ΔP4H4	pAGM8031	*nptII*	p35S-Cas9-tNOS_pU6-(6 gRNAs P4H4)
ΔP4H10	pAGM8031	*nptII*	p35S-Cas9-tNOS_pU6-(6 gRNAs P4H10)

4 *A. tumefaciens* transformation

Electrocompetent *A. tumefaciens* strain LBA4404 VirG (Van Der Fits et al., 2000) were transformed using 1 µL *E. coli* DNA minipreparation. The conditions for the electroporation were 25 µF, 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 $MgSO_4$) for 4 hours at 28°C and plated on 2xYT agar plates supplemented with 40 µg/mL gentamycin, 20 µg/mL rifampicin, and 50 µg/mL spectinomycin. Colonies were visible after three days of incubation at 28°C.

5 *Agrobacterium*-mediated plant cell transformation

A. tumefaciens LBA4404 VirG-mediated stable transformation of BY-2 cells was carried out as described by Navarre & Chaumont (2022) with *nptII* or *hptI* resistance marker. 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 Genomic DNA extraction

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

7 Characterization of *P4H* homeologs and analysis of genome modifications

Genomic DNA was amplified by PCR using Q5 High-Fidelity DNA Polymerase (New England Biolabs, M0491) with the primers listed in Table M3. The PCR products were separated by gel electrophoresis on 1% agarose gel, purified (innuPREP, Innuscreen) and sequenced (Microsynth AG). Data analysis of Sanger sequencing for edited cell lines was carried out using Synthego ICE (Conant et al., 2022).

Table M3: Primers used for amplification of *P4H* gDNA.

Primer	Sequence
P4H4_B2 g1-2 F	TTTGACAACGATAAAAAGTGTTACAAGCAC
P4H4_B2 g1-2 R	GACAAATTTATCTTCTGACCCTGGCATGAT
P4H4_A2/B2 g3 F	TTTTGGCTGTGGTTTCTTGAAATCAAATCT
P4H4_A2/B2 g3 R	AAATTTAGGAAACAATATATTCAGGGGGAC
P4H4_A2 g1-2 F	CACGTCCTCCCAACTATAA
P4H4_A2 g1-2 R	ACACGATTTAGCAACACAAC
P4HA_A1 g1-2 F	AGCATTTGTATATGAAGGATTTCTTACGGGA
P4HA_A1 g1-2 R	ATCCAGAAATTCCAAACACTAGCACAATGA
P4HA_A1 g3 F	CGATATGAGGAGGGGCAGAAATATGAGCC
P4HA_A1 g3 R	AGCATATACTCTGGATTCTTGGTGCATTCC

8 Protein extraction

Two mL of the indicated BY-2 culture were filtered on four layers of Miracloth. The filtrate was then centrifuged 10 min at 7,200 *g* at 4°C. The supernatant corresponds to the extracellular medium (EM). Filtered BY-2 cells were used to harvest cellular total soluble proteins (TSP) and microsomal fractions (MF). 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-Cl, 2 mM Na_2EDTA, pH 8.0) supplemented with 1 mM phenylmethylsulfonylfluoride (PMSF), and protease inhibitor cocktail (leupeptin, aprotinin, antipain, pepstain, and chymostain, 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) 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. To separate TSP from the MF, the remaining sample was centrifuged for 15 min at 130,000 *g* at 4°C. The supernatant, corresponding to TSP, was retrieved and the pellet, corresponding to the MF, was resuspended in 50 µL resuspension buffer (3 mM KH_2PO_4, 330 mM sucrose, 3 mM KCl, pH 7.8 (KOH)). The protein concentrations in the TSP and MF were quantified by Bradford assay.

9 SDS-PAGE, western and lectin blotting analysis of proteins

Indicated amount of culture medium (µL), TSP or MF (µg) were analysed by SDS-PAGE (reducing, if not stated otherwise) after denaturation (5 min at 100°C for EM and TSP; 15 min at 56°C for MF). Gels were either stained with Coomassie Brilliant Blue G-250 (SERVA, 17524) or transferred onto a PVDF membrane (Bio-Rad Laboratories) for western blotting. The PVDF membrane was saturated in TBST + 3% milk powder for western blotting or in TBST + 2% BSA for lectin blotting before incubation with:

- Rabbit polyclonal antibodies against SARS-CoV-2 RBD (Sino Biological, 40592-T62, 1:2,000)
- Rabbit polyclonal antibodies against hEPO (C-terminal) (Invitrogen, PA5-120937, 1:2,000)
- Mouse monoclonal IgG2A antibody against hEPO (N-terminal) (R&D Systems, MAB2871, 1:3000)
- Goat polyclonal antibodies against human IgG Fab coupled to HRP (Sigma, A0293, 1:1,000)
- Goat polyclonal antibodies against human IgG Fc coupled to HRP (Sigma, A0170, 1:1,000)
- Goat polyclonal antibodies against α heavy chain coupled to HRP (Invitrogen, 31417, 1:25,000)
- Rabbit polyclonal antibodies against GFP (generated by (Duby et al., 2001), 1:2000 and Merck, SAB4301138, 1:5,000)
- Rabbit polyclonal antibodies against GalNAcT2 (Sino Biological, 13764-T62, 1:2,000)
- Mouse monoclonal antibody against FLAG coupled to HRP (Sigma-Aldrich, A8592, 1:1,000)

- Biotinylated *Helix pomatia* lectin (Sigma-Aldrich, L6512, 1:2,500)
- Biotinylated peanut agglutinin (Vector Laboratories, 1:2,500)
- HisProbe coupled to HRP (Thermo Fisher Scientific, 15165, 1:1,000)

followed by incubation with:

- anti-rabbit HRP polyclonal antibodies (SynAbs, LO-RG-1-HRP, 1:10,000,)
- anti-rabbit alkaline phosphatase (AP) polyclonal antibodies (Sigma-Aldrich, A3687, 1:10,000)
- Streptavidin-HRP (Roche, 11089153001, 1:5,000)

Western blots using HRP-conjugated antibodies (or Streptavidin-HRP) were revealed with BM Chemiluminescence Blotting Substrate (Roche, 11500694001). Western blots using AP-conjugated antibodies were revealed with BM Purple AP substrate, precipitating (Roche, 11442074001). Chemiluminescence was detected using the Amersham Imager 600 (GE Healthcare).

10 IgA and EPO-Fc Protein purification

IgA and EPO-Fc purification was adapted from the protocol described by Mercx *et al.* (2017). 50-250 mL of 10-days old BY-2 suspension culture in D11b medium was filtered on three layers of Miracloth and the filtrate was centrifuged (13,900 g, 30 min). The supernatant was recovered, supplemented with 10% 1 M Tris-Cl pH 8 and incubated for 16 h at 4°C with 400 µL of Pierce R Protein A Plus Agarose (Thermoscientific # 22812) previously washed three times with 4 mL of 0.1 M Tris-Cl pH 8. The sample was centrifuged at ~50 g for 5 min and the supernatant was discarded. The pelleted beads were washed twice with 10 mL of 0.1 M Tris-Cl pH 8 and transferred into a 10 mL Poly-Prep chromatography column (Bio-Rad #7311550). IgA1 were eluted with 5 × 500 µL 0.1 M glycine pH 3 and immediately buffered with 10% 1 M Tris-Cl pH 8.

11 Mass spectrometry analysis IgA1 glycopeptides

The IgA1 *O*-glycovariants were analysed in BOKU Core Facility Mass Spectrometry. The samples were digested in-solution. The proteins were S-alkylated with iodoacetamide and digested with Trypsin (Promega) over night. The peptide samples were then digested with GluC (Promega) for 4h.

The digested samples were loaded on a nanoEase C18 column (nanoEase M/Z HSS T3 Column, 100Å, 1.8 µm, 300 µm X 150 mm, Waters) using 0.1 % formic acid as the aqueous solvent. A gradient from 1% B (B: 80% Acetonitrile, 0.1% FA) to 40% B in 30 min was applied, followed by a 5 min gradient from 40% B to 95% B that facilitates elution of large peptides, at a flow rate of 6 µL/min. Detection was performed with an Orbitap MS (Exploris 480, Thermo) equipped with the standard H-ESI source in positive ion, DDA mode (= switching to MSMS mode for eluting

peaks). MS-scans were recorded (range: 350-3200 m/z) and the 8 highest peaks were selected for fragmentation. Instrument calibration was performed using Pierce FlexMix Calibration Solution (Thermo Scientific).

The possible glycopeptides were identified as sets of peaks consisting of the peptide moiety and the attached N-glycan varying in the number of *N*-acetylhexosamine (HexNAc) units, hexose (Hex), pentose (Pent) and deoxyhexose (dHex) residues, as well as additional hydroxyproline modifications for the *O*-glycosylation site. The theoretical masses of these glycopeptides were determined with a spread sheet using the monoisotopic masses for amino acids and monosaccharides.

Manual glycopeptide searches were made using FreeStyle 1.8 (Thermo). For the relative quantification of the different glycoforms the peak intensities of the deconvoluted spectra were compared (using an in-house software tool (Freestyle_parser_v0.3.R)).

12 Mass spectrometry analysis of EPO-Fc

EPO-Fc samples were analysed by QTOF mass spectrometry using both data dependent acquisition (DDA) and Ion Mobility Separation High-Definition Enhanced MS (HDMSE) modes.

Coomassie Blue stained protein bands were manually excised. Gels plugs were transferred with 200 µl of HPLC water to 0.5 mL polypropylene Protein LoBind Eppendorf tubes. Water was removed and gel plugs were incubated under shaking at 20°C for 5 min with 200 µl of a solution containing 50 mM ammonium carbonate (pH 8.0) in 50% (v/v) acetonitrile. This solution was removed, and gel plugs were incubated under shaking at 20°C for 5 min with 200 µl of 10% (v/v) acetonitrile. Acetonitrile was removed and gels plugs were dried under vacuum with a Savant Speed Vac Concentrator. 20 µl of a digestion buffer containing 100 mM TEAB (triethylammonium bicarbonate (pH 8.0) and 0.5 µg of trypsin (Promega) were added to each tube to rehydrate gel plugs. Proteolysis was allowed to continue overnight at 37°C and stopped by adding 10 µl of 1% (v/v) TFA. Supernatants of each tube were recovered and transferred in new 0.5 mL polypropylene Protein LoBind Eppendorf tubes. 50 µl of peptide extraction solution containing 50% (v/v) acetonitrile and 0.1% (v/v) TFA were added to the gel plugs. After incubation for 5 min, extracts were combined with the first. A second extraction with 50 µl of 100% (v/v) acetonitrile was done for 5 min. Extracts were combined with the first and dried under vacuum with Savant Speed Vac Concentrator. Before peptide separation, the sample was dissolved in 20 µl of 0.1% (v/v) formic acid and 2% (v/v) acetonitrile (ACN).

Peptide mixture was separated by reverse phase chromatography on a NanoACQUITY UPLC MClass system (Waters) working with MassLynx V4.1 (Waters) software. 5 µL were injected on a trap C18, 100Å 5 µm, 180 µm x 20 mm column (Waters) and desalted using isocratic conditions with at a flow rate of 8 µL/min using a 95% formic acid and 5% (v/v) ACN buffer for 5 min. Protein mixture was subjected to reverse phase chromatography on a C18, 100Å 1.8 µm, 75 µm x 150 mm column (Waters) PepMap for 35 min at 35°C at a flow rate of 300 nL/min using a two part linear gradient from 5% (v/v) ACN, 0.1% formic acid to 40% (v/v)) ACN, 0.1% formic acid for 15 min and from 40% (v/v) ACN, 0.1% formic acid to 85% (v/v)) ACN, 0.1 % formic acid for 10 min. The column was re-equilibrated at initial conditions after washing 30 min at 85% (v/v)) ACN, 0.1% formic acid at a flow rate of 300 nL/min. For online LC-MS analysis, the nanoUPLC was coupled to the mass spectrometer through a nano-electrospray ionization (nanoESI) source emitter.

DDA analysis were performed on an SYNAPT G2-Si high definition mass spectrometer (Waters) equipped with a NanoLockSpray dual electrospray ion source (Waters).Precut fused silica PicoTipR Emitters for nanoelectrospray, outer diameters : 360 µm; inner diameter: 20 µm; 10 µm tip; 2.5" length (Waters) were used for samples and Precut fused silica TicoTipR Emitters for nanoelectrospray, outer diameters : 360 µm; inner diameter: 20 µm; 2.5" length (Waters) were used for the lock mass.solution. The eluent was sprayed at a spray voltage of 2.8 kV with a sampling cone voltage of 25 V and a source offset of 30 V. The source temperature was set to 80°C. The cone gas flow was 20 liters/h with a nano flow gas pressure of 0.4 bar and the purge gas was turned off. The SYNAPT G2Si instrument was operated in DDA (data-dependent mode), automatically switching between MS and MS2. Full scan MS and MS2 spectra (m/z 50- 1500) were acquired directly after injection to 35 min in resolution mode (20,000 resolution FWHM at m/z 400) with a scan time of 0.1 sec. Tandem mass spectra of up to 10 precursors were generated in the trapping region of the ion mobility cell by using a collision energy ramp from 17/19 V (low mass, start/end) to up to 65/75 V (high mass,start/end). Charged ions (+2, +3, +4) are selected to be submitted to the MSMS fragmentation over the m/z range from 50 to 2,000 with a scan time of 0.25 sec. For the post-acquisition lock mass correction of the data in the MS method, the doubly charged monoisotopic ion of [Glu1]-fibrinopeptide B was used at 100 fmol/µl using the reference sprayer of the nanoESI source with a frequency of 30 s at 0.5µl/min into the mass spectrometer.

HDMSE analyses were performed on an SYNAPT G2-Si high-definition mass spectrometer (Waters) equipped with a NanoLockSpray dual electrospray ion source

(Waters). Precut fused silica PicoTipR Emitters for nanoelectrospray, outer diameters: 360 μm; inner diameter: 20 μm; 10 μm tip; 2.5" length (Waters) were used for samples and Precut fused silica TicoTipR Emitters for nanoelectrospray, outer diameters: 360 μm; inner diameter: 20 μm; 2.5" length (Waters) were used for the lockmass solution. The eluent was sprayed at a spray voltage of 2.4 kV with a sampling cone voltage of 25 V and a source offset of 30 V. The source temperature was set to 80°C. HDMSE method in resolution mode was used to collect data from 2 min after injection to 35 min. This method acquires MSE in positive and resolution mode over the m/z range from 50 to 2,000 with a scan time of 1 sec. with a collision energy ramp starting from ion mobility bin 20 (20 eV) to 110 (45 eV). The collision energy in the transfer cell for low-energy MS mode was set to 4 eV. For the post-acquisition lock mass correction of the data in the MS method, the doubly charged monoisotopic ion of [Glu1]-fibrinopeptide B was used at 100 fmol/μl using the reference sprayer of the nanoESI source with a frequency of 30 s at 0.5 μl/min into the mass spectrometer.

HDMSE and DDA data were processed with UNIFI (Waters) software using database containing protein sequence of interest. Propionamide as the fixed cysteine modification, Methionine oxidation, disulfide bridge and Hydroxyl Proline as variable modifications. Trypsin as the digestion enzyme were selected and one miss cleavage allowed.

13 *In vitro* deglycosylation assay

Indicated amount of TSP and purified proteins (μg) or EM (μL) were denatured by heating for 10 min at 100°C in the manufacturer's glycoprotein denaturing buffer, to ensure optimal glycosite accessibility. Samples were then incubated for 1 h at 37°C with 50 U of EndoH or PNGaseF or up to 80,000 U of *O*-glycosidase (New England Biolabs, P0703, P0704 and P0733) using the manufacturer's recommended conditions. Control samples were run in parallel but without the glycosidase. Samples were then visualized by SDS-PAGE followed by western blotting.

14 Representation of *N*-glycan structures

Schematic *N*-glycan structures were generated with GlycoWorkbench (Ceroni et al., 2008).

References

Adolf, G. R., Kalsner, I., Ahorn, H., Maurer-Fogy, I., & Cantell, K. (1991). Natural human interferon-α2 is O-glycosylated. *Biochemical Journal*, *276*(2), 511–518. https://doi.org/10.1042/bj2760511

Ahmed, J., Mercx, S., Boutry, M., & Chaumont, F. (2020). Evolutionary and Predictive Functional Insights into the Aquaporin Gene Family in the Allotetraploid Plant Nicotiana tabacum. *International Journal of Molecular Sciences*, *21*(13), 4743. https://doi.org/10.3390/ijms21134743

Almeida-Silva, F., & Venancio, T. M. (2022). BioNERO: an all-in-one R/Bioconductor package for comprehensive and easy biological network reconstruction. *Functional & Integrative Genomics*, *22*(1), 131–136. https://doi.org/10.1007/s10142-021-00821-9

Almon, E., Khoury, T., Drori, A., Gingis-Velitski, S., Alon, S., Chertkoff, R., Mushkat, M., Shaaltiel, Y., & Ilan, Y. (2017). An oral administration of a recombinant anti-TNF fusion protein is biologically active in the gut promoting regulatory T cells: Results of a phase I clinical trial using a novel oral anti-TNF alpha-based therapy. *Journal of Immunological Methods*, *446*, 21–29. https://doi.org/10.1016/j.jim.2017.03.023

Alonso, A. P., Piasecki, R. J., Wang, Y., LaClair, R. W., & Shachar-Hill, Y. (2010). Quantifying the Labeling and the Levels of Plant Cell Wall Precursors Using Ion Chromatography Tandem Mass Spectrometry . *Plant Physiology*, *153*(3), 915–924. https://doi.org/10.1104/pp.110.155713

Altmann, F. (2007). The Role of Protein Glycosylation in Allergy. *International Archives of Allergy and Immunology*, *142*(2), 99–115. https://doi.org/10.1159/000096114

Amano, K., Chiba, Y., Kasahara, Y., Kato, Y., Kaneko, M. K., Kuno, A., Ito, H., Kobayashi, K., Hirabayashi, J., Jigami, Y., & Narimatsu, H. (2008). Engineering of mucin-type human glycoproteins in yeast cells. *Proceedings of the National Academy of Sciences*, *105*(9), 3232–3237. https://doi.org/10.1073/pnas.0710412105

Atsmon, J., Brill-Almon, E., Nadri-Shay, C., Chertkoff, R., Alon, S., Shaikevich, D., Volokhov, I., Haim, K. Y., Bartfeld, D., Shulman, A., Ruderfer, I., Ben-Moshe, T., Shilovitzky, O., Soreq, H., & Shaaltiel, Y. (2015). Preclinical and first-in-human evaluation of PRX-105, a PEGylated, plant-derived, recombinant human acetylcholinesterase-R. *Toxicology and Applied Pharmacology*, *287*(3), 202–209. https://doi.org/10.1016/j.taap.2015.06.004

Azariadis, A., Vouligeas, F., Salame, E., Kouhen, M., Rizou, M., Blazakis, K., Sotiriou, P., Ezzat, L., Mekkaoui, K., Monzer, A., Krokida, A., Adamakis, I. D., Dandachi, F., Shalha, B., Kostelenos, G., Figgou, E., Giannoutsou, E., & Kalaitzis, P. (2023). Response of Prolyl 4 Hydroxylases, Arabinogalactan Proteins and Homogalacturonans in Four Olive Cultivars under Long-Term Salinity Stress in Relation to Physiological and Morphological Changes. *Cells*, *12*(11). https://doi.org/10.3390/cells12111466

Bagdonaite, I., Thompson, A. J., Wang, X., Søgaard, M., Fougeroux, C., Frank, M., Diedrich, J. K., Yates, J. R., Salanti, A., Vakhrushev, S. Y., Paulson, J. C., & Wandall, H. H. (2021). Site-Specific O-Glycosylation Analysis of SARS-CoV-2 Spike Protein Produced in Insect and Human Cells. *Viruses*, *13*(4), 551. https://doi.org/10.3390/v13040551

Baldwin, T. C., Domingo, C., Schindler, T., Seetharaman, G., Stacey, N., & Roberts, K. (2001). DcAGP1, a secreted arabinogalactan protein, is related to a family of basic proline-rich proteins. *Plant Molecular Biology*, *45*, 421–435. https://doi.org/https://doi.org/10.1023/A:1010637426934

Barnes, J., Tian, L., Loftis, J., Hiznay, J., Comhair, S., Lauer, M., & Dweik, R. (2016). Isolation and analysis of sugar nucleotides using solid phase extraction and fluorophore assisted carbohydrate electrophoresis. *MethodsX*, *3*, 251–260. https://doi.org/10.1016/j.mex.2016.03.010

Basu, D., Liang, Y., Liu, X., Himmeldirk, K., Faik, A., Kieliszewski, M., Held, M., & Showalter, A. M. (2013). Functional Identification of a Hydroxyproline-O-galactosyltransferase Specific for Arabinogalactan Protein Biosynthesis in Arabidopsis. *Journal of Biological Chemistry*, *288*(14), 10132–10143. https://doi.org/10.1074/jbc.M112.432609

Basu, D., Tian, L., Wang, W., Bobbs, S., Herock, H., Travers, A., & Showalter, A. M. (2015). A small multigene hydroxyproline-O-galactosyltransferase family functions in arabinogalactan-protein glycosylation, growth and development in Arabidopsis. *BMC Plant Biology*, *15*(1), 295. https://doi.org/10.1186/s12870-015-0670-7

Basu, D., Wang, W., Ma, S., DeBrosse, T., Poirier, E., Emch, K., Soukup, E., Tian, L., & Showalter, A. M. (2015). Two Hydroxyproline Galactosyltransferases, GALT5 and GALT2, Function in Arabinogalactan-Protein Glycosylation, Growth and Development in Arabidopsis. *PLOS ONE*, *10*(5), e0125624. https://doi.org/10.1371/journal.pone.0125624

Beihammer, G., König-Beihammer, J., Kogelmann, B., Ruocco, V., Grünwald-Gruber, C., D'Aoust, M.-A., Lavoie, P.-O., Saxena, P., Gach, J. S., Steinkellner, H., & Strasser, R. (2023). An oligosaccharyltransferase from Leishmania donovani increases the N-glycan occupancy on plant-produced IgG1. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1233666

Bennett, E. P., Hassan, H., Mandel, U., Hollingsworth, M. A., Akisawa, N., Ikematsu, Y., Merkx, G., van Kessel, A. G., Olofsson, S., & Clausen, H. (1999). Cloning and Characterization of a Close Homologue of Human UDP-N-acetyl-α-d-galactosamine:Polypeptide N-Acetylgalactosaminyltransferase-T3, Designated GalNAc-T6. *Journal of Biological Chemistry*, *274*(36), 25362–25370. https://doi.org/10.1074/jbc.274.36.25362

Bennett, E. P., Mandel, U., Clausen, H., Gerken, T. A., Fritz, T. A., & Tabak, L. A. (2012). Control of mucin-type O-glycosylation: A classification of the polypeptide GalNAc-transferase gene family. *Glycobiology*, *22*(6), 736–756. https://doi.org/10.1093/glycob/cwr182

Beuder, S., Dorchak, A., Bhide, A., Moeller, S. R., Petersen, B. L., & MacAlister, C. A. (2020). Exocyst mutants suppress pollen tube growth and cell wall structural defects of *hydroxyproline O-arabinosyltransferase* mutants. *The Plant Journal*, *103*(4), 1399–1419. https://doi.org/10.1111/tpj.14808

Biel, T. G., Faison, T., Matthews, A. M., Zou, G., Ortega-Rodriguez, U., Pegues, M. A., Azer, N., Gomez, F., Johnson, S., Rogstad, S., Chen, K., Xie, H., Agarabi, C., Rao, V. A., & Ju, T. (2022). An etanercept O-glycovariant with enhanced potency. *Molecular Therapy - Methods & Clinical Development*, *25*, 124–135. https://doi.org/10.1016/j.omtm.2022.03.002

Brockhausen, I., Benn, M., Bhat, S., Marone, S., Riley, J. G., Montoya-Peleaz, P., Vlahakis, J. Z., Paulsen, H., Schutzbach, J. S., & Szarek, W. A. (2006). UDP-Gal: GlcNAc-R β1,4-galactosyltransferase—a target enzyme for drug design. Acceptor specificity and inhibition of the enzyme. *Glycoconjugate Journal*, *23*(7–8), 525–541. https://doi.org/10.1007/s10719-006-7153-x

Brockhausen, I., Dowler, T., & Paulsen, H. (2009). Site directed processing: Role of amino acid sequences and glycosylation of acceptor glycopeptides in the assembly of extended mucin type O-glycan core 2. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1790*(10), 1244–1257. https://doi.org/10.1016/j.bbagen.2009.05.020

Brühlmann, D., Sokolov, M., Butté, A., Sauer, M., Hemberger, J., Souquet, J., Broly, H., & Jordan, M. (2017). Parallel experimental design and multivariate analysis provides efficient screening of cell culture media supplements to improve biosimilar product quality. *Biotechnology and Bioengineering*, *114*(7), 1448–1458. https://doi.org/10.1002/bit.26269

Büll, C., Joshi, H. J., Clausen, H., & Narimatsu, Y. (2020). Cell-Based Glycan Arrays—A Practical Guide to Dissect the Human Glycome. *STAR Protocols*, *1*(1), 100017. https://doi.org/10.1016/j.xpro.2020.100017

Burchell, J. M., Beatson, R., Graham, R., Taylor-Papadimitriou, J., & Tajadura-Ortega, V. (2018). O-linked mucin-type glycosylation in breast cancer. *Biochemical Society Transactions*, *46*(4), 779–788. https://doi.org/10.1042/BST20170483

Buskas, T., Thompson, P., & Boons, G.-J. (2009). Immunotherapy for cancer: synthetic carbohydrate-based vaccines. *Chemical Communications*, *36*, 5335. https://doi.org/10.1039/b908664c

Bustamante, J. J., Gonzalez, L., Carroll, C. A., Weintraub, S. T., Aguilar, R. M., Muñoz, J., Martinez, A. O., & Haro, L. S. (2009). O-Glycosylated 24 kDa human growth hormone has a mucin-like biantennary disialylated tetrasaccharide attached at Thr-60. *PROTEOMICS*, *9*(13), 3474–3488. https://doi.org/10.1002/pmic.200800989

Campos, D., Girgis, M., & Sanda, M. (2022). Site-specific glycosylation of SARS-CoV-2: Big challenges in mass spectrometry analysis. *PROTEOMICS*, *22*(15–16), e2100322. https://doi.org/10.1002/pmic.202100322

Cannon, M. C., Terneus, K., Hall, Q., Tan, L., Wang, Y., Wegenhart, B. L., Chen, L., Lamport, D. T. A., Chen, Y., & Kieliszewski, M. J. (2008). Self-assembly of the plant cell wall requires an extensin scaffold. *Proceedings of the National Academy of Sciences*, *105*(6), 2226–2231. https://doi.org/10.1073/pnas.0711980105

Canut, H., Albenne, C., & Jamet, E. (2016). Post-translational modifications of plant cell wall proteins and peptides: A survey from a proteomics point of view. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1864*(8), 983–990. https://doi.org/10.1016/j.bbapap.2016.02.022

Cardon, F., Pallisse, R., Bardor, M., Caron, A., Vanier, J., Ele Ekouna, J. P., Lerouge, P., Boitel-Conti, M., & Guillet, M. (2019). *Brassica rapa* hairy root based

expression system leads to the production of highly homogenous and reproducible profiles of recombinant human alpha-L-iduronidase. *Plant Biotechnology Journal*, *17*(2), 505–516. https://doi.org/10.1111/pbi.12994

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*(6), 813–823. https://doi.org/10.1093/glycob/cwr009

Castilho, A., Neumann, L., Daskalova, S., Mason, H. S., Steinkellner, H., Altmann, F., & Strasser, R. (2012). Engineering of Sialylated Mucin-type O-Glycosylation in Plants. *Journal of Biological Chemistry*, *287*(43), 36518–36526. https://doi.org/10.1074/jbc.M112.402685

Castilho, A., Strasser, R., Stadlmann, J., Grass, J., Jez, J., Gattinger, P., Kunert, R., Quendler, H., Pabst, M., Leonard, R., Altmann, F., & Steinkellner, H. (2010). In Planta Protein Sialylation through Overexpression of the Respective Mammalian Pathway. *Journal of Biological Chemistry*, *285*(21), 15923–15930. https://doi.org/10.1074/jbc.M109.088401

Ceroni, A., Maass, K., Geyer, H., Geyer, R., Dell, A., & Haslam, S. M. (2008). GlycoWorkbench: A Tool for the Computer-Assisted Annotation of Mass Spectra of Glycans. *Journal of Proteome Research*, *7*(4), 1650–1659. https://doi.org/10.1021/pr7008252

Chateauvieux, S., Grigorakaki, C., Morceau, F., Dicato, M., & Diederich, M. (2011). Erythropoietin, erythropoiesis and beyond. *Biochemical Pharmacology*, *82*(10), 1291–1303. https://doi.org/10.1016/j.bcp.2011.06.045

Chen, Y., Dong, W., Tan, L., Held, M. A., & Kieliszewski, M. J. (2015). Arabinosylation Plays a Crucial Role in Extensin Cross-linking *In Vitro*. *Biochemistry Insights*, *8s2*, BCI.S31353. https://doi.org/10.4137/BCI.S31353

Cheng, P. W., & Radhakrishnan, P. (2011). Mucin O-glycan branching enzymes: Structure, function, and gene regulation. *Advances in Experimental Medicine and Biology*, *705*, 465–492. https://doi.org/10.1007/978-1-4419-7877-6_25

Conant, D., Hsiau, T., Rossi, N., Oki, J., Maures, T., Waite, K., Yang, J., Joshi, S., Kelso, R., Holden, K., Enzmann, B. L., & Stoner, R. (2022). Inference of CRISPR Edits from Sanger Trace Data. *The CRISPR Journal*, *5*(1), 123–130. https://doi.org/10.1089/crispr.2021.0113

Concordet, J.-P., & Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research*, *46*(W1), W242–W245. https://doi.org/10.1093/nar/gky354

Cunha, N. B., Araújo, A. C. G., Leite, A., Murad, A. M., Vianna, G. R., & Rech, E. L. (2010). Short communication Correct targeting of proinsulin in protein storage vacuoles of transgenic soybean seeds. *Genetics and Molecular Research*, *9*(2), 1163–1170. https://doi.org/10.4238/vol9-2gmr849

Dalziel, M., Crispin, M., Scanlan, C. N., Zitzmann, N., & Dwek, R. A. (2014). Emerging Principles for the Therapeutic Exploitation of Glycosylation. *Science*, *343*(6166). https://doi.org/10.1126/science.1235681

Daniel, E. J. P., las Rivas, M., Lira-Navarrete, E., García-García, A., Hurtado-Guerrero, R., Clausen, H., & Gerken, T. A. (2020). Ser and Thr acceptor preferences of the GalNAc-Ts vary among isoenzymes to modulate mucin-type O-glycosylation. *Glycobiology*, *30*(11), 910–922. https://doi.org/10.1093/glycob/cwaa036

Dardelle, F., Lehner, A., Ramdani, Y., Bardor, M., Lerouge, P., Driouich, A., & Mollet, J.-C. (2010). Biochemical and Immunocytological Characterizations of Arabidopsis Pollen Tube Cell Wall . *Plant Physiology*, *153*(4), 1563–1576. https://doi.org/10.1104/pp.110.158881

Darling, R. J., Kuchibhotla, U., Glaesner, W., Micanovic, R., Witcher, D. R., & Beals, J. M. (2002). Glycosylation of Erythropoietin Affects Receptor Binding Kinetics: Role of Electrostatic Interactions. *Biochemistry*, *41*(49), 14524–14531. https://doi.org/10.1021/bi0265022

Daskalova, S. M., Radder, J. E., Cichacz, Z. A., Olsen, S. H., Tsaprailis, G., Mason, H., & Lopez, L. C. (2010). Engineering of N. benthamiana L. plants for production of N-acetylgalactosamine-glycosylated proteins - towards development of a plant-based platform for production of protein therapeutics with mucin type O-glycosylation. *BMC Biotechnology*, *10*(1), 62. https://doi.org/10.1186/1472-6750-10-62

Dauplais, M., Lecoq, A., Song, J., Cotton, J., Jamin, N., Gilquin, B., Roumestand, C., Vita, C., de Medeiros, CleaneL. C., Rowan, E. G., Harvey, A. L., & Ménez, A. (1997). On the Convergent Evolution of Animal Toxins. *Journal of Biological Chemistry*, *272*(7), 4302–4309. https://doi.org/10.1074/jbc.272.7.4302

Davey, R. T., Dodd, L., Proschan, M. A., Neaton, J., Neuhaus Nordwall, J., Koopmeiners, J. S., Beigel, J., Tierney, J., Lane, C., Fauci, A. S., Massaquoi, M. B. F., Sahr, F., & Malvy, D. (2016). A Randomized, Controlled Trial of ZMapp for Ebola Virus Infection. *New England Journal of Medicine*, *375*(15), 1448–1456. https://doi.org/10.1056/NEJMoa1604330

de Haan, N., Falck, D., & Wuhrer, M. (2020). Monitoring of immunoglobulin N- and O-glycosylation in health and disease. *Glycobiology*, *30*(4), 226–240. https://doi.org/10.1093/glycob/cwz048

de las Rivas, M., Coelho, H., Diniz, A., Lira-Navarrete, E., Compañón, I., Jiménez-Barbero, J., Schjoldager, K. T., Bennett, E. P., Vakhrushev, S. Y., Clausen, H., Corzana, F., Marcelo, F., & Hurtado-Guerrero, R. (2018). Structural Analysis of a GalNAc-T2 Mutant Reveals an Induced-Fit Catalytic Mechanism for GalNAc-Ts. *Chemistry – A European Journal*, *24*(33), 8382–8392. https://doi.org/10.1002/chem.201800701

de las Rivas, M., Daniel, E. J. P., Coelho, H., Lira-Navarrete, E., Raich, L., Compañón, I., Diniz, A., Lagartera, L., Jiménez-Barbero, J., Clausen, H., Rovira, C., Marcelo, F., Corzana, F., Gerken, T. A., & Hurtado-Guerrero, R. (2018). Structural and Mechanistic Insights into the Catalytic-Domain-Mediated Short-Range Glycosylation Preferences of GalNAc-T4. *ACS Central Science*, *4*(9), 1274–1290. https://doi.org/10.1021/acscentsci.8b00488

de las Rivas, M., Lira-Navarrete, E., Daniel, E. J. P., Compañón, I., Coelho, H., Diniz, A., Jiménez-Barbero, J., Peregrina, J. M., Clausen, H., Corzana, F., Marcelo, F., Jiménez-Osés, G., Gerken, T. A., & Hurtado-Guerrero, R. (2017). The interdomain flexible linker of the polypeptide GalNAc transferases dictates their long-range glycosylation preferences. *Nature Communications*, *8*(1), 1959. https://doi.org/10.1038/s41467-017-02006-0

de las Rivas, M., Lira-Navarrete, E., Gerken, T. A., & Hurtado-Guerrero, R. (2019). Polypeptide GalNAc-Ts: from redundancy to specificity. *Current Opinion in Structural Biology*, *56*, 87–96. https://doi.org/10.1016/j.sbi.2018.12.007

De Muynck, B., Navarre, C., Nizet, Y., Stadlmann, J., & Boutry, M. (2009). Different subcellular localization and glycosylation for a functional antibody expressed in Nicotiana tabacum plants and suspension cells. *Transgenic Research*, *18*(3), 467–482. https://doi.org/10.1007/s11248-008-9240-1

de Sousa-Pereira, P., & Woof, J. M. (2019). IgA: Structure, Function, and Developability. *Antibodies*, *8*(4), 57. https://doi.org/10.3390/antib8040057

Decker, E. L., Parsons, J., & Reski, R. (2014). Glyco-engineering for biopharmaceutical production in moss bioreactors. *Frontiers in Plant Science*, *5*(JUL). https://doi.org/10.3389/fpls.2014.00346

Diamos, A. G., & Mason, H. S. (2018). Chimeric 3' flanking regions strongly enhance gene expression in plants. *Plant Biotechnology Journal*, *16*(12), 1971–1982. https://doi.org/10.1111/pbi.12931

Dicker, M., & Strasser, R. (2015). Using glyco-engineering to produce therapeutic proteins. In *Expert Opinion on Biological Therapy* (Vol. 15, Issue 10, pp. 1501–1516). Taylor and Francis Ltd. https://doi.org/10.1517/14712598.2015.1069271

Dicker, M., Tschofen, M., Maresch, D., König, J., Juarez, P., Orzaez, D., Altmann, F., Steinkellner, H., & Strasser, R. (2016). Transient Glyco-Engineering to Produce Recombinant IgA1 with Defined N- and O-Glycans in Plants. *Frontiers in Plant Science*, *7*(JAN2016). https://doi.org/10.3389/fpls.2016.00018

Dilokpimol, A., Poulsen, C. P., Vereb, G., Kaneko, S., Schulz, A., & Geshi, N. (2014). Galactosyltransferases from Arabidopsis thaliana in the biosynthesis of type II arabinogalactan: molecular interaction enhances enzyme activity. *BMC Plant Biology*, *14*(1), 90. https://doi.org/10.1186/1471-2229-14-90

Dolan, M. C., Wu, D., Cramer, C. L., & Xu, J. (2014). Hydroxyproline-O-glycosylated peptide tags enhance recombinant protein yields in tobacco transient expression. *Process Biochemistry*, *49*(3), 490–495. https://doi.org/10.1016/j.procbio.2013.12.010

Dubé, S., Fisher, J. W., & Powell, J. S. (1988). Glycosylation at specific sites of erythropoietin is essential for biosynthesis, secretion, and biological function. *The Journal of Biological Chemistry*, *263*(33), 17516–17521. http://www.ncbi.nlm.nih.gov/pubmed/3182860

Duby, G., Degand, H., & Boutry, M. (2001). Structure requirement and identification of a cryptic cleavage site in the mitochondrial processing of a plant F 1 -ATPase β-subunit presequence. *FEBS Letters*, *505*(3), 409–413. https://doi.org/10.1016/S0014-5793(01)02856-3

Duruflé, H., Hervé, V., Balliau, T., Zivy, M., Dunand, C., & Jamet, E. (2017). Proline Hydroxylation in Cell Wall Proteins: Is It Yet Possible to Define Rules? *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01802

Edwards, K. D., Fernandez-Pozo, N., Drake-Stowe, K., Humphry, M., Evans, A. D., Bombarely, A., Allen, F., Hurst, R., White, B., Kernodle, S. P., Bromley, J. R., Sanchez-Tamburrino, J. P., Lewis, R. S., & Mueller, L. A. (2017). A reference genome for Nicotiana tabacum enables map-based cloning of homeologous loci implicated in nitrogen utilization efficiency. *BMC Genomics*, *18*(1). https://doi.org/10.1186/s12864-017-3791-6

Egelund, J., Obel, N., Ulvskov, P., Geshi, N., Pauly, M., Bacic, A., & Petersen, B. L. (2007). Molecular characterization of two Arabidopsis thaliana glycosyltransferase mutants, rra1 and rra2, which have a reduced residual arabinose content in a polymer tightly associated with the cellulosic wall residue. *Plant Molecular Biology*, *64*(4), 439–451. https://doi.org/10.1007/s11103-007-9162-y

Eidenberger, L., Kogelmann, B., & Steinkellner, H. (2023). Plant-based biopharmaceutical engineering. *Nature Reviews Bioengineering*, *1*(6), 426–439. https://doi.org/10.1038/s44222-023-00044-6

Ellies, L. G., Tsuboi, S., Petryniak, B., Lowe, J. B., Fukuda, M., & Marth, J. D. (1998a). Core 2 Oligosaccharide Biosynthesis Distinguishes between Selectin Ligands Essential for Leukocyte Homing and Inflammation. *Immunity*, *9*(6), 881–890. https://doi.org/10.1016/S1074-7613(00)80653-6

Ellies, L. G., Tsuboi, S., Petryniak, B., Lowe, J. B., Fukuda, M., & Marth, J. D. (1998b). Core 2 Oligosaccharide Biosynthesis Distinguishes between Selectin Ligands Essential for Leukocyte Homing and Inflammation. In *Immunity* (Vol. 9).

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.-M., Werner, S., Jones, J. D. G., Patron, N. J., & Marillonnet, S. (2014). A Golden Gate Modular Cloning Toolbox for Plants. *ACS Synthetic Biology*, *3*(11), 839–843. https://doi.org/10.1021/sb4001504

Estévez, J. M., Kieliszewski, M. J., Khitrov, N., & Somerville, C. (2006). Characterization of Synthetic Hydroxyproline-Rich Proteoglycans with Arabinogalactan Protein and Extensin Motifs in Arabidopsis. *Plant Physiology*, *142*(2), 458–470. https://doi.org/10.1104/pp.106.084244

Farinas, C. S., Leite, A., & Miranda, E. A. (2007). Recombinant human proinsulin from transgenic corn endosperm: solvent screening and extraction studies. *Brazilian Journal of Chemical Engineering*, *24*(3), 315–323. https://doi.org/10.1590/S0104-66322007000300002

Festari, M. F., Trajtenberg, F., Berois, N., Pantano, S., Revoredo, L., Kong, Y., Solari-Saquieres, P., Narimatsu, Y., Freire, T., Bay, S., Robello, C., Bénard, J., Gerken, T. A., Clausen, H., & Osinaga, E. (2017). Revisiting the human polypeptide GalNAc-T1 and T13 paralogs. *Glycobiology*, *27*(2), 140–153. https://doi.org/10.1093/glycob/cww111

Fragkostefanakis, S., Sedeek, K. E. M., Raad, M., Zaki, M. S., & Kalaitzis, P. (2014). Virus induced gene silencing of three putative prolyl 4-hydroxylases enhances plant growth in tomato (Solanum lycopersicum). *Plant Molecular Biology*, *85*(4–5), 459–471. https://doi.org/10.1007/s11103-014-0197-6

Fritz, T. A., Hurley, J. H., Trinh, L.-B., Shiloach, J., & Tabak, L. A. (2004). The beginnings of mucin biosynthesis: The crystal structure of UDP-GalNAc:polypeptide α- *N* -acetylgalactosaminyltransferase-T1. *Proceedings of the National Academy of Sciences*, *101*(43), 15307–15312. https://doi.org/10.1073/pnas.0405657101

Fritz, T. A., Raman, J., & Tabak, L. A. (2006). Dynamic Association between the Catalytic and Lectin Domains of Human UDP-GalNAc:Polypeptide α-N-Acetylgalactosaminyltransferase-2. *Journal of Biological Chemistry*, *281*(13), 8613–8619. https://doi.org/10.1074/jbc.M513590200

Fu, L. H., Miao, Y., Lo, S. W., Seto, T. C., Sun, S. S. M., Xu, Z.-F., Clemens, S., Clarke, L. A., Kermode, A. R., & Jiang, L. (2009). Production and characterization of soluble human lysosomal enzyme α-iduronidase with high activity from culture media of transgenic tobacco BY-2 cells. *Plant Science*, *177*(6), 668–675. https://doi.org/10.1016/j.plantsci.2009.08.016

Gao, M., Kieliszewski, M. J., Lamport, D. T. A., & Showalter, A. M. (1999). Isolation, characterization and immunolocalization of a novel, modular tomato arabinogalactan-protein corresponding to the LeAGP-1 gene. *Plant Journal*, *18*(1), 43–55. https://doi.org/10.1046/j.1365-313X.1999.00428.x

Gattinger, P., Izadi, S., Grünwald-Gruber, C., Kallolimath, S., & Castilho, A. (2021). The Instability of Dimeric Fc-Fusions Expressed in Plants Can Be Solved by

Monomeric Fc Technology. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.671728

Gerken, T. A., Jamison, O., Perrine, C. L., Collette, J. C., Moinova, H., Ravi, L., Markowitz, S. D., Shen, W., Patel, H., & Tabak, L. A. (2011). Emerging Paradigms for the Initiation of Mucin-type Protein O-Glycosylation by the Polypeptide GalNAc Transferase Family of Glycosyltransferases. *Journal of Biological Chemistry*, *286*(16), 14493–14507. https://doi.org/10.1074/jbc.M111.218701

Gerken, T. A., Revoredo, L., Thome, J. J. C., Tabak, L. A., Vester-Christensen, M. B., Clausen, H., Gahlay, G. K., Jarvis, D. L., Johnson, R. W., Moniz, H. A., & Moremen, K. (2013). The Lectin Domain of the Polypeptide GalNAc Transferase Family of Glycosyltransferases (ppGalNAc Ts) Acts as a Switch Directing Glycopeptide Substrate Glycosylation in an N- or C-terminal Direction, Further Controlling Mucin Type O-Glycosylation. *Journal of Biological Chemistry*, *288*(27), 19900–19914. https://doi.org/10.1074/jbc.M113.477877

Gerken, T. A., Zhang, J., Levine, J., & Elhammer, Å. (2002). Mucin Core O-Glycosylation Is Modulated by Neighboring Residue Glycosylation Status. *Journal of Biological Chemistry*, *277*(51), 49850–49862. https://doi.org/10.1074/jbc.M205851200

Geshi, N., Johansen, J. N., Dilokpimol, A., Rolland, A., Belcram, K., Verger, S., Kotake, T., Tsumuraya, Y., Kaneko, S., Tryfona, T., Dupree, P., Scheller, H. V., Höfte, H., & Mouille, G. (2013). A galactosyltransferase acting on arabinogalactan protein glycans is essential for embryo development in <scp>A</scp> rabidopsis. *The Plant Journal*, *76*(1), 128–137. https://doi.org/10.1111/tpj.12281

Ghag, S. B., Adki, V. S., Ganapathi, T. R., & Bapat, V. A. (2021). Plant Platforms for Efficient Heterologous Protein Production. *Biotechnology and Bioprocess Engineering*, *26*(4), 546–567. https://doi.org/10.1007/s12257-020-0374-1

Gill, D. J., Chia, J., Senewiratne, J., & Bard, F. (2010). Regulation of O-glycosylation through Golgi-to-ER relocation of initiation enzymes. *Journal of Cell Biology*, *189*(5), 843–858. https://doi.org/10.1083/jcb.201003055

Gill, D. J., Clausen, H., & Bard, F. (2011). Location, location, location: new insights into O-GalNAc protein glycosylation. *Trends in Cell Biology*, *21*(3), 149–158. https://doi.org/10.1016/j.tcb.2010.11.004

Gille, S., Hänsel, U., Ziemann, M., & Pauly, M. (2009). Identification of plant cell wall mutants by means of a forward chemical genetic approach using hydrolases. *Proceedings of the National Academy of Sciences*, *106*(34), 14699–14704. https://doi.org/10.1073/pnas.0905434106

Gille, S., Sharma, V., Baidoo, E. E. K., Keasling, J. D., Scheller, H. V., & Pauly, M. (2013). Arabinosylation of a Yariv-Precipitable Cell Wall Polymer Impacts Plant Growth as Exemplified by the Arabidopsis Glycosyltransferase Mutant ray1. *Molecular Plant*, *6*(4), 1369–1372. https://doi.org/10.1093/mp/sst029

Goderis, I. J. W. M., De Bolle, M. F. C., François, I. E. J. A., Wouters, P. F. J., Broekaert, W. F., & Cammue, B. P. A. (2002). A set of modular plant transformation vectors allowing flexible insertion of up to six expression units. In *Plant Molecular Biology* (Vol. 50).

Gomord, V., Fitchette, A., 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 Biotechnology Journal*, *8*(5), 564–587. https://doi.org/10.1111/j.1467-7652.2009.00497.x

González-Ramírez, A. M., Grosso, A. S., Yang, Z., Compañón, I., Coelho, H., Narimatsu, Y., Clausen, H., Marcelo, F., Corzana, F., & Hurtado-Guerrero, R. (2022). Structural basis for the synthesis of the core 1 structure by C1GalT1. *Nature Communications*, *13*(1), 2398. https://doi.org/10.1038/s41467-022-29833-0

Göritzer, K., Goet, I., Duric, S., Maresch, D., Altmann, F., Obinger, C., & Strasser, R. (2020). Efficient N-Glycosylation of the Heavy Chain Tailpiece Promotes the Formation of Plant-Produced Dimeric IgA. *Frontiers in Chemistry*, *8*. https://doi.org/10.3389/fchem.2020.00346

Göritzer, K., Maresch, D., Altmann, F., Obinger, C., & Strasser, R. (2017). Exploring Site-Specific N-Glycosylation of HEK293 and Plant-Produced Human IgA Isotypes. *Journal of Proteome Research*, *16*(7), 2560–2570. https://doi.org/10.1021/acs.jproteome.7b00121

Gorres, K. L., & Raines, R. T. (2010). Prolyl 4-hydroxylase. *Critical Reviews in Biochemistry and Molecular Biology*, *45*(2), 106–124. https://doi.org/10.3109/10409231003627991

Granovsk, M., Bielfeldt, T., Peters, S., Paulsen, H., Meldal, M., Brockhausen, J., & Brockhausen, I. (1994). UDPgalactose:glycoprotein– N-acetylgalactosamine 3-β-galactosyltransferase activity synthesizing O-glycan core 1 is controlled by the amino acid sequence and glycosylation of glycopeptide substrates. *European Journal of Biochemistry*, *221*(3), 1039–1046. https://doi.org/10.1111/j.1432-1033.1994.tb18822.x

Hagel, J. M., & Facchini, P. J. (2018). Expanding the roles for 2-oxoglutarate-dependent oxygenases in plant metabolism. *Natural Product Reports*, *35*(8), 721–734. https://doi.org/10.1039/C7NP00060J

Hager, K. J., Pérez Marc, G., Gobeil, P., Diaz, R. S., Heizer, G., Llapur, C., Makarkov, A. I., Vasconcellos, E., Pillet, S., Riera, F., Saxena, P., Geller Wolff, P., Bhutada, K., Wallace, G., Aazami, H., Jones, C. E., Polack, F. P., Ferrara, L., Atkins, J., … Ward, B. J. (2022). Efficacy and Safety of a Recombinant Plant-Based Adjuvanted Covid-19 Vaccine. *New England Journal of Medicine*, *386*(22), 2084–2096. https://doi.org/10.1056/NEJMoa2201300

Hahn, F., Korolev, A., Sanjurjo Loures, L., & Nekrasov, V. (2020). A modular cloning toolkit for genome editing in plants. *BMC Plant Biology*, *20*(1), 179. https://doi.org/10.1186/s12870-020-02388-2

Halim, A., Rüetschi, U., Larson, G., & Nilsson, J. (2013). LC–MS/MS Characterization of O-Glycosylation Sites and Glycan Structures of Human Cerebrospinal Fluid Glycoproteins. *Journal of Proteome Research*, *12*(2), 573–584. https://doi.org/10.1021/pr300963h

Hanania, U., Ariel, T., Tekoah, Y., Fux, L., Sheva, M., Gubbay, Y., Weiss, M., Oz, D., Azulay, Y., Turbovski, A., Forster, Y., & 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 Biotechnology Journal*, *15*(9), 1120–1129. https://doi.org/10.1111/pbi.12702

Hang, H. C., & Bertozzi, C. R. (2005). The chemistry and biology of mucin-type O-linked glycosylation. *Bioorganic & Medicinal Chemistry*, *13*(17), 5021–5034. https://doi.org/10.1016/j.bmc.2005.04.085

Hassan, H., Reis, C. A., Bennett, E. P., Mirgorodskaya, E., Roepstorff, P., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., & Clausen, H. (2000). The Lectin Domain of UDP-N-acetyl-d-galactosamine:PolypeptideN-acetylgalactosaminyltransferase-T4 Directs Its Glycopeptide Specificities.

Journal of Biological Chemistry, *275*(49), 38197–38205. https://doi.org/10.1074/jbc.M005783200

Held, M. A., Tan, L., Kamyab, A., Hare, M., Shpak, E., & Kieliszewski, M. J. (2004). Di-isodityrosine Is the Intermolecular Cross-link of Isodityrosine-rich Extensin Analogs Cross-linked in Vitro. *Journal of Biological Chemistry*, *279*(53), 55474–55482. https://doi.org/10.1074/jbc.M408396200

Helm, M., Schmid, M., Hierl, G., Terneus, K., Tan, L., Lottspeich, F., Kieliszewski, M. J., & Gietl, C. (2008). KDEL-tailed cysteine endopeptidases involved in programmed cell death, intercalation of new cells, and dismantling of extensin scaffolds. *American Journal of Botany*, *95*(9), 1049–1062. https://doi.org/10.3732/ajb.2007404

Herr, A. B., Ballister, E. R., & Bjorkman, P. J. (2003). Insights into IgA-mediated immune responses from the crystal structures of human FcalphaRI and its complex with IgA1-Fc. *Nature*, *423*(6940), 614–620. https://doi.org/10.1038/nature01685

Hiatt, A., Caffferkey, R., & Bowdish, K. (1989). Production of antibodies in transgenic plants. *Nature*, *342*(6245), 76–78. https://doi.org/10.1038/342076a0

Hieta, R., & Myllyharju, J. (2002). Cloning and Characterization of a Low Molecular Weight Prolyl 4-Hydroxylase from Arabidopsis thaliana. *Journal of Biological Chemistry*, *277*(26), 23965–23971. https://doi.org/10.1074/jbc.M201865200

Hijazi, M., Durand, J., Pichereaux, C., Pont, F., Jamet, E., & Albenne, C. (2012). Characterization of the Arabinogalactan Protein 31 (AGP31) of Arabidopsis thaliana. *Journal of Biological Chemistry*, *287*(12), 9623–9632. https://doi.org/10.1074/jbc.M111.247874

Hijazi, M., Roujol, D., Nguyen-Kim, H., del Rocio Cisneros Castillo, L., Saland, E., Jamet, E., & Albenne, C. (2014). Arabinogalactan protein 31 (AGP31), a putative network-forming protein in Arabidopsis thaliana cell walls? *Annals of Botany*, *114*(6), 1087–1097. https://doi.org/10.1093/aob/mcu038

Hijazi, M., Velasquez, S. M., Jamet, E., Estevez, J. M., & Albenne, C. (2014). An update on post-translational modifications of hydroxyproline-rich glycoproteins: toward a model highlighting their contribution to plant cell wall architecture. *Frontiers in Plant Science*, *5*(AUG). https://doi.org/10.3389/fpls.2014.00395

Hori, H. (1978). The demonstration of galactosamine in extracellular hydroxyproline-rich macromolecule purified from the culture media of tobacco cells. *Plant and Cell Physiology*, *19*(3), 501–505. https://doi.org/10.1093/oxfordjournals.pcp.a075619

Hu, S., He, W., & Wu, G. (2022). Hydroxyproline in animal metabolism, nutrition, and cell signaling. In *Amino Acids* (Vol. 54, Issue 4, pp. 513–528). Springer. https://doi.org/10.1007/s00726-021-03056-x

Huang, L. F., Liu, Y. K., Lu, C. A., Hsieh, S. L., & Yu, S. M. (2005). Production of human serum albumin by sugar starvation induced promoter and rice cell culture. *Transgenic Research*, *14*(5), 569–581. https://doi.org/10.1007/s11248-004-6481-5

Ilan, Y., Gingis-Velitski, S., Ben Ya'aco, A., Shabbat, Y., Zolotarov, L., Almon, E., & Shaaltiel, Y. (2017). A plant cell-expressed recombinant anti-TNF fusion protein is biologically active in the gut and alleviates immune-mediated hepatitis and colitis. *Immunobiology*, *222*(3), 544–551. https://doi.org/10.1016/j.imbio.2016.11.001

Iwasaki, H., Zhang, Y., Tachibana, K., Gotoh, M., Kikuchi, N., Kwon, Y.-D., Togayachi, A., Kudo, T., Kubota, T., & Narimatsu, H. (2003). Initiation of O-Glycan Synthesis in IgA1 Hinge Region Is Determined by a Single Enzyme, UDP-N-Acetyl-α-d-galactosamine:PolypeptideN-Acetylgalactosaminyltransferase 2. *Journal of Biological Chemistry*, *278*(8), 5613–5621. https://doi.org/10.1074/jbc.M211097200

James, E. A., Wang, C., Wang, Z., Reeves, R., Shin, J. H., Magnuson, N. S., & Lee, J. M. (2000). Production and Characterization of Biologically Active Human GM-CSF Secreted by Genetically Modified Plant Cells. *Protein Expression and Purification*, *19*(1), 131–138. https://doi.org/10.1006/prep.2000.1232

Jansing, J., Sack, M., Augustine, S. M., Fischer, R., & 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 Biotechnology Journal*, *17*(2), 350–361. https://doi.org/10.1111/pbi.12981

Ji, S., Samara, N. L., Revoredo, L., Zhang, L., Tran, D. T., Muirhead, K., Tabak, L. A., & Ten Hagen, K. G. (2018). A molecular switch orchestrates enzyme

specificity and secretory granule morphology. *Nature Communications*, *9*(1), 3508. https://doi.org/10.1038/s41467-018-05978-9

Jiang, M.-C., Hu, C.-C., Lin, N.-S., & Hsu, Y.-H. (2019). Production of Human IFNγ Protein in Nicotiana benthamiana Plant through an Enhanced Expression System Based on Bamboo mosaic Virus. *Viruses*, *11*(6), 509. https://doi.org/10.3390/v11060509

Jo, S.-H., Kwon, S.-Y., Park, D.-S., Yang, K.-S., Kim, J.-W., Lee, K.-T., Kwak, S.-S., & Lee, H.-S. (2006). High-yield production of functional human lactoferrin in transgenic cell cultures of siberian ginseng (Acanthopanax senticosus). *Biotechnology and Bioprocess Engineering*, *11*(5), 442–448. https://doi.org/10.1007/BF02932312

Johnson, K. L., Cassin, A. M., Lonsdale, A., Bacic, A., Doblin, M. S., & Schultz, C. J. (2017). Pipeline to Identify Hydroxyproline-Rich Glycoproteins. *Plant Physiology*, *174*(2), 886–903. https://doi.org/10.1104/pp.17.00294

Ju, T., Brewer, K., D'Souza, A., Cummings, R. D., & Canfield, W. M. (2002). Cloning and Expression of Human Core 1 β1,3-Galactosyltransferase. *Journal of Biological Chemistry*, *277*(1), 178–186. https://doi.org/10.1074/jbc.M109060200

Ju, T., & Cummings, R. D. (2002). A unique molecular chaperone Cosmc required for activity of the mammalian core 1 β3-galactosyltransferase. *Proceedings of the National Academy of Sciences*, *99*(26), 16613–16618. https://doi.org/10.1073/pnas.262438199

Julien, S., Picco, G., Sewell, R., Vercoutter-Edouart, A.-S., Tarp, M., Miles, D., Clausen, H., Taylor-Papadimitriou, J., & Burchell, J. M. (2009). Sialyl-Tn vaccine induces antibody-mediated tumour protection in a relevant murine model. *British Journal of Cancer*, *100*(11), 1746–1754. https://doi.org/10.1038/sj.bjc.6605083

Julien, S., Videira, P. A., & Delannoy, P. (2012). Sialyl-Tn in Cancer: (How) Did We Miss the Target? *Biomolecules*, *2*(4), 435–466. https://doi.org/10.3390/biom2040435

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein

structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kallolimath, S., Hackl, T., Gahn, R., Grünwald-Gruber, C., Zich, W., Kogelmann, B., Lux, A., Nimmerjahn, F., & Steinkellner, H. (2020). Expression Profiling and Glycan Engineering of IgG Subclass 1–4 in Nicotiana benthamiana. *Frontiers in Bioengineering and Biotechnology*, *8*. https://doi.org/10.3389/fbioe.2020.00825

Karki, U., Fang, H., Guo, W., Unnold-Cofre, C., & Xu, J. (2021). Cellular engineering of plant cells for improved therapeutic protein production. *Plant Cell Reports*, *40*(7), 1087–1099. https://doi.org/10.1007/s00299-021-02693-6

Karki, U., Wright, T., & Xu, J. (2022). High yield secretion of human erythropoietin from tobacco cells for ex vivo differentiation of hematopoietic stem cells towards red blood cells. *Journal of Biotechnology*, *355*, 10–20. https://doi.org/10.1016/j.jbiotec.2022.06.010

Karnoup, A. S., Turkelson, V., & Anderson, W. H. K. (2005). O-Linked glycosylation in maize-expressed human IgA1. *Glycobiology*, *15*(10), 965–981. https://doi.org/10.1093/glycob/cwi077

Kato, K., Jeanneau, C., Tarp, M. A., Benet-Pagès, A., Lorenz-Depiereux, B., Bennett, E. P., Mandel, U., Strom, T. M., & Clausen, H. (2006). Polypeptide GalNAc-transferase T3 and Familial Tumoral Calcinosis. *Journal of Biological Chemistry*, *281*(27), 18370–18377. https://doi.org/10.1074/jbc.M602469200

Kato, K., Takeuchi, H., Kanoh, A., Mandel, U., Hassan, H., Clausen, H., & Irimura, T. (2001). N-acetylgalactosamine incorporation into a peptide containing consecutive threonine residues by UDP-N-acetyl-D-galactosaminide:polypeptide N-acetylgalactosaminyltransferases. *Glycobiology*, *11*(10), 821–829. https://doi.org/10.1093/glycob/11.10.821

Kehler, P., Neumann, T., Jaekel, A., Weis, M., Gellert, J., Danielczyk, A., Hartung, E., Lischke, T., Gurka, S., Kast, N., & Weiske, M. (2022). 1346 Antibodies against GlycoTargets as novel platform approach to address unmet needs in cancer therapy. *Regular and Young Investigator Award Abstracts*, A1397–A1397. https://doi.org/10.1136/jitc-2022-SITC2022.1346

Keskiaho, K., Hieta, R., Sormunen, R., & Myllyharju, J. (2007). Chlamydomonas reinhardtii Has Multiple Prolyl 4-Hydroxylases, One of Which Is Essential for

Proper Cell Wall Assembly. *The Plant Cell*, *19*(1), 256–269. https://doi.org/10.1105/tpc.106.042739

Kim, H., Yamaguchi, Y., Masuda, K., Matsunaga, C., Yamamoto, K., Irimura, T., Takahashi, N., Kato, K., & Arata, Y. (1994). O-glycosylation in hinge region of mouse immunoglobulin G2b. *The Journal of Biological Chemistry*, *269*(16), 12345–12350. http://www.ncbi.nlm.nih.gov/pubmed/7512967

Kim, J., Park, H., Park, B. T., Hwang, H. S., Kim, J. Il, Kim, D. K., & Kim, H. H. (2016). O-glycans and O-glycosylation sites of recombinant human GM-CSF derived from suspension-cultured rice cells, and their structural role. *Biochemical and Biophysical Research Communications*, *479*(2), 266–271. https://doi.org/10.1016/j.bbrc.2016.09.057

Kim, T.-G., Lee, H.-J., Jang, Y.-S., Shin, Y.-J., Kwon, T.-H., & Yang, M.-S. (2008). Co-expression of proteinase inhibitor enhances recombinant human granulocyte–macrophage colony stimulating factor production in transgenic rice cell suspension culture. *Protein Expression and Purification*, *61*(2), 117–121. https://doi.org/10.1016/j.pep.2008.06.005

King, S. L., Joshi, H. J., Schjoldager, K. T., Halim, A., Madsen, T. D., Dziegiel, M. H., Woetmann, A., Vakhrushev, S. Y., & Wandall, H. H. (2017). Characterizing the O-glycosylation landscape of human plasma, platelets, and endothelial cells. *Blood Advances*, *1*(7), 429–442. https://doi.org/10.1182/bloodadvances.2016002121

Kittur, F. S., Hung, C.-Y., Li, P. A., Sane, D. C., & Xie, J. (2023). Asialo-rhuEPO as a Potential Neuroprotectant for Ischemic Stroke Treatment. *Pharmaceuticals*, *16*(4), 610. https://doi.org/10.3390/ph16040610

Knoch, E., Dilokpimol, A., Tryfona, T., Poulsen, C. P., Xiong, G., Harholt, J., Petersen, B. L., Ulvskov, P., Hadi, M. Z., Kotake, T., Tsumuraya, Y., Pauly, M., Dupree, P., & Geshi, N. (2013). A β–glucuronosyltransferase from Arabidopsis thaliana involved in biosynthesis of type II arabinogalactan has a role in cell elongation during seedling growth. *The Plant Journal*, *76*(6), 1016–1029. https://doi.org/10.1111/tpj.12353

Kong, Y., Joshi, H. J., Schjoldager, K. T., Madsen, T. D., Gerken, T. A., Vester-Christensen, M. B., Wandall, H. H., Bennett, E. P., Levery, S. B., Vakhrushev, S. Y., & Clausen, H. (2015). Probing polypeptide GalNAc-transferase isoform

substrate specificities by in vitro analysis. *Glycobiology*, *25*(1), 55–65. https://doi.org/10.1093/glycob/cwu089

Konkina, A., Klepadlo, M., Lakehal, A., Zein, Z. El, Krokida, A., Botros, M., Iakovidis, M., Chernobavskiy, P., Elfatih Zerroumda, M., Tsanakas, G., Petrakis, N., Dourou, A.-M., & Kalaitzis, P. (2021). An Arabidopsis Prolyl 4 Hydroxylase Is Involved in the Low Oxygen Response. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.637352

Konstantinidi, A., Nason, R., Čaval, T., Sun, L., Sørensen, D. M., Furukawa, S., Ye, Z., Vincentelli, R., Narimatsu, Y., Vakhrushev, S. Y., & Clausen, H. (2022). Exploring the glycosylation of mucins by use of O-glycodomain reporters recombinantly expressed in glycoengineered HEK293 cells. *Journal of Biological Chemistry*, *298*(4), 101784. https://doi.org/10.1016/j.jbc.2022.101784

Koski, M. K., Hieta, R., Böllner, C., Kivirikko, K. I., Myllyharju, J., & Wierenga, R. K. (2007). The Active Site of an Algal Prolyl 4-Hydroxylase Has a Large Structural Plasticity. *Journal of Biological Chemistry*, *282*(51), 37112–37123. https://doi.org/10.1074/jbc.M706554200

Koski, M. K., Hieta, R., Hirsilä, M., Rönkä, A., Myllyharju, J., & Wierenga, R. K. (2009). The Crystal Structure of an Algal Prolyl 4-Hydroxylase Complexed with a Proline-rich Peptide Reveals a Novel Buried Tripeptide Binding Motif. *Journal of Biological Chemistry*, *284*(37), 25290–25301. https://doi.org/10.1074/jbc.M109.014050

Kubota, T., Shiba, T., Sugioka, S., Furukawa, S., Sawaki, H., Kato, R., Wakatsuki, S., & Narimatsu, H. (2006). Structural Basis of Carbohydrate Transfer Activity by Human UDP-GalNAc: Polypeptide α-N-Acetylgalactosaminyltransferase (pp-GalNAc-T10). *Journal of Molecular Biology*, *359*(3), 708–727. https://doi.org/10.1016/j.jmb.2006.03.061

Kudelka, M. R., Ju, T., Heimburg-Molinaro, J., & Cummings, R. D. (2015). Simple Sugars to Complex Disease—Mucin-Type O-Glycans in Cancer. In *Advances in Cancer Research* (Vol. 126, pp. 53–135). Academic Press Inc. https://doi.org/10.1016/bs.acr.2014.11.002

Kutyrieva-Nowak, N., Leszczuk, A., Ezzat, L., Kaloudas, D., Zając, A., Szymańska-Chargot, M., Skrzypek, T., Krokida, A., Mekkaoui, K., Lampropoulou, E., Kalaitzis, P., & Zdunek, A. (2024). The modified activity of prolyl 4 hydroxylases reveals the effect of arabinogalactan proteins on changes in the

cell wall during the tomato ripening process. *Frontiers in Plant Science*, *15*. https://doi.org/10.3389/fpls.2024.1365490

Kwon, K., Sherman, A., Chang, W., Kamesh, A., Biswas, M., Herzog, R. W., & Daniell, H. (2018). Expression and assembly of largest foreign protein in chloroplasts: oral delivery of human FVIII made in lettuce chloroplasts robustly suppresses inhibitor formation in haemophilia A mice. *Plant Biotechnology Journal*, *16*(6), 1148–1160. https://doi.org/10.1111/pbi.12859

Kwon, S.-Y., Jo, S.-H., Lee, O.-S., Choi, S.-M., Kwak, S.-S., & Lee, H.-S. (2003). Transgenic Ginseng Cell Lines That Produce High levels of a Human Lactoferrin. *Planta Medica*, *69*(11), 1005–1008. https://doi.org/10.1055/s-2003-45146

Kwon, T. H., Seo, J. E., Kim, J., Lee, J. H., Jang, Y. S., & Yang, M. S. (2003). Expression and secretion of the heterodimeric protein interleukin-12 in plant cell suspension culture. *Biotechnology and Bioengineering*, *81*(7), 870–875. https://doi.org/10.1002/bit.10528

Laguia-Becher, M., Zaldúa, Z., Xu, W., Marconi, P. L., Velander, W., & Alvarez, M. A. (2019). Co-expressing Turnip Crinkle Virus-coat protein with the serine protease α-thrombin precursor (pFIIa) in Nicotiana benthamiana Domin. *In Vitro Cellular & Developmental Biology - Plant*, *55*(1), 88–98. https://doi.org/10.1007/s11627-018-09956-0

Lee, J., Lee, S.-K., Park, J.-S., & Lee, K.-R. (2023). Plant-made pharmaceuticals: exploring studies for the production of recombinant protein in plants and assessing challenges ahead. *Plant Biotechnology Reports*, *17*(1), 53–65. https://doi.org/10.1007/s11816-023-00821-0

Lee, J.-H., Kim, N.-S., Kwon, T.-H., Jang, Y.-S., & Yang, M.-S. (2002). Increased production of human granulocyte-macrophage colony stimulating factor (hGM-CSF) by the addition of stabilizing polymer in plant suspension cultures. *Journal of Biotechnology*, *96*(3), 205–211. https://doi.org/10.1016/S0168-1656(02)00044-5

Lee, K. H., Kwon, D. H., Song, J. T., & Seo, H. S. (2020). Production mechanisms, structural features and post-translational modifications of plant peptides. *Journal of Plant Biology*, *63*(4), 259–265. https://doi.org/10.1007/s12374-020-09255-5

Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., & Gascuel, O. (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Research*, *47*(W1), W260–W265. https://doi.org/10.1093/nar/gkz303

Leonard, R., Petersen, B. O., Himly, M., Kaar, W., Wopfner, N., Kolarich, D., van Ree, R., Ebner, C., Duus, J. Ø., Ferreira, F., & Altmann, F. (2005). Two Novel Types of O-Glycans on the Mugwort Pollen Allergen Art v 1 and Their Role in Antibody Binding. *Journal of Biological Chemistry*, *280*(9), 7932–7940. https://doi.org/10.1074/jbc.M410407200

Léonard, R., Wopfner, N., Pabst, M., Stadlmann, J., Petersen, B. O., Duus, J. Ø., Himly, M., Radauer, C., Gadermaier, G., Razzazi-Fazeli, E., Ferreira, F., & Altmann, F. (2010). A New Allergen from Ragweed (Ambrosia artemisiifolia) with Homology to Art v 1 from Mugwort. *Journal of Biological Chemistry*, *285*(35), 27192–27200. https://doi.org/10.1074/jbc.M110.127118

Leszczuk, A., Kalaitzis, P., Kulik, J., & Zdunek, A. (2023). Review: structure and modifications of arabinogalactan proteins (AGPs). In *BMC Plant Biology* (Vol. 23, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12870-023-04066-5

Li, X., Wang, J., Li, W., Xu, Y., Shao, D., Xie, Y., Xie, W., Kubota, T., Narimatsu, H., & Zhang, Y. (2012). Characterization of ppGalNAc-T18, a member of the vertebrate-specific Y subfamily of UDP-N-acetyl-α-d-galactosamine:polypeptide N-acetylgalactosaminyltransferases †. *Glycobiology*, *22*(5), 602–615. https://doi.org/10.1093/glycob/cwr179

Liang, G., Zhang, H., Lou, D., & Yu, D. (2016). Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. *Scientific Reports*, *6*(1), 21451. https://doi.org/10.1038/srep21451

Lindberg, L., Liu, J., Gaunitz, S., Nilsson, A., Johansson, T., Karlsson, N. G., & Holgersson, J. (2013). Mucin-type fusion proteins with blood group A or B determinants on defined O-glycan core chains produced in glycoengineered Chinese hamster ovary cells and their use as immunoaffinity matrices. *Glycobiology*, *23*(6), 720–735. https://doi.org/10.1093/glycob/cwt011

Lira-Navarrete, E., de las Rivas, M., Compañón, I., Pallarés, M. C., Kong, Y., Iglesias-Fernández, J., Bernardes, G. J. L., Peregrina, J. M., Rovira, C., Bernadó, P., Bruscolini, P., Clausen, H., Lostao, A., Corzana, F., & Hurtado-Guerrero, R.

(2015). Dynamic interplay between catalytic and lectin domains of GalNAc-transferases modulates protein O-glycosylation. *Nature Communications*, *6*(1), 6937. https://doi.org/10.1038/ncomms7937

Lira-Navarrete, E., Iglesias-Fernández, J., Zandberg, W. F., Compañón, I., Kong, Y., Corzana, F., Pinto, B. M., Clausen, H., Peregrina, J. M., Vocadlo, D. J., Rovira, C., & Hurtado-Guerrero, R. (2014). Substrate-Guided Front-Face Reaction Revealed by Combined Structural Snapshots and Metadynamics for the Polypeptide *N* -Acetylgalactosaminyltransferase 2. *Angewandte Chemie International Edition*, *53*(31), 8206–8210. https://doi.org/10.1002/anie.201402781

Liu, J., Jin, C., Cherian, R. M., Karlsson, N. G., & Holgersson, J. (2015). O -glycan repertoires on a mucin-type reporter protein expressed in CHO cell pools transiently transfected with O -glycan core enzyme cDNAs. *Journal of Biotechnology*, *199*, 77–89. https://doi.org/10.1016/j.jbiotec.2015.02.017

Liu, X., McKenna, S., Welch, L. R., & Showalter, A. M. (2020). *Bioinformatic Identification of Plant Hydroxyproline-Rich Glycoproteins* (pp. 463–481). https://doi.org/10.1007/978-1-0716-0621-6_26

Lopez-Hernandez, F., Tryfona, T., Rizza, A., Yu, X. L., Harris, M. O. B., Webb, A. A. R., Kotake, T., & Dupree, P. (2020). Calcium Binding by Arabinogalactan Polysaccharides Is Important for Normal Plant Development. *The Plant Cell*, *32*(10), 3346–3369. https://doi.org/10.1105/tpc.20.00027

Loya, A., Pnueli, L., Yosefzon, Y., Wexler, Y., Ziv-Ukelson, M., & Arava, Y. (2008). The 3'-UTR mediates the cellular localization of an mRNA encoding a short plasma membrane protein. *RNA*, *14*(7), 1352–1365. https://doi.org/10.1261/rna.867208

Lu, H. S., Clogston, C. L., Wypych, J., Parker, V. P., Lee, T. D., Swiderek, K., Baltera, R. F., Patel, A. C., Chang, D. C., Brankow, D. W., Liu, X.-D., Ogden, S. G., Karkare, S. B., Hu, S. S., Zsebo, K. M., & Langley, K. E. (1992). Post-translational processing of membrane-associated recombinant human stem cell factor expressed in Chinese hamster ovary cells. *Archives of Biochemistry and Biophysics*, *298*(1), 150–158. https://doi.org/10.1016/0003-9861(92)90106-7

Ludovic, M., Eric, N.-O., Maxime, G., Abdoul-Salam, K., Follet-Gueye, M.-L., Azeddine, D., Maïté, V.-G., & Sophie, A.-A. (2015). O-glycosylation in plant and mammal cells: the use of chemical inhibitors to understand the biosynthesis

and function of O-glycosylated proteins. *Plant Science Today*, *2*(2), 43–51. https://doi.org/10.14719/pst.2015.2.2.67

Ma, J. K., Lehner, T., Stabila, P., Fux, C. I., & Hiatt, A. (1994). Assembly of monoclonal antibodies with IgG1 and IgA heavy chain domains in transgenic tobacco plants. *European Journal of Immunology*, *24*(1), 131–138. https://doi.org/10.1002/eji.1830240120

MacAlister, C. A., Ortiz-Ramírez, C., Becker, J. D., Feijó, J. A., & Lippman, Z. B. (2016). Hydroxyproline O-arabinosyltransferase mutants oppositely alter tip growth in Arabidopsis thaliana and Physcomitrella patens. *The Plant Journal*, *85*(2), 193–208. https://doi.org/10.1111/tpj.13079

Madsen, T. D., Hansen, L. H., Hintze, J., Ye, Z., Jebari, S., Andersen, D. B., Joshi, H. J., Ju, T., Goetze, J. P., Martin, C., Rosenkilde, M. M., Holst, J. J., Kuhre, R. E., Goth, C. K., Vakhrushev, S. Y., & Schjoldager, K. T. (2020). An atlas of O-linked glycosylation on peptide hormones reveals diverse biological roles. *Nature Communications*, *11*(1), 4033. https://doi.org/10.1038/s41467-020-17473-1

Magalhães, A., Duarte, H. O., & Reis, C. A. (2021). The role of O-glycosylation in human disease. *Molecular Aspects of Medicine*, *79*, 100964. https://doi.org/10.1016/j.mam.2021.100964

Magnuson, N. S., Linzmaier, P. M., Reeves, R., An, G., HayGlass, K., & Lee, J. M. (1998). Secretion of Biologically Active Human Interleukin-2 and Interleukin-4 from Genetically Modified Tobacco Cells in Suspension Culture. *Protein Expression and Purification*, *13*(1), 45–52. https://doi.org/10.1006/prep.1998.0872

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

Maharjan, P. M., & Choe, S. (2021). Plant-Based COVID-19 Vaccines: Current Status, Design, and Development Strategies of Candidate Vaccines. *Vaccines*, *9*(9), 992. https://doi.org/10.3390/vaccines9090992

Marillonnet, S., & Grützner, R. (2020). Synthetic DNA Assembly Using Golden Gate Cloning and the Hierarchical Modular Cloning Pipeline. *Current Protocols in Molecular Biology*, *130*(1). https://doi.org/10.1002/cpmb.115

Martinez, S., & Hausinger, R. P. (2015). Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases. *Journal of Biological Chemistry*, *290*(34), 20702–20711. https://doi.org/10.1074/jbc.R115.648691

Marzol, E., Borassi, C., Bringas, M., Sede, A., Rodríguez Garcia, D. R., Capece, L., & Estevez, J. M. (2018). Filling the Gaps to Solve the Extensin Puzzle. *Molecular Plant*, *11*(5), 645–658. https://doi.org/10.1016/j.molp.2018.03.003

Mathieu-Rivet, E., Mati-Baouche, N., Walet-Balieu, M.-L., Lerouge, P., & Bardor, M. (2020). N- and O-Glycosylation Pathways in the Microalgae Polyphyletic Group. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.609993

Mattu, T. S., Pleass, R. J., Willis, A. C., Kilian, M., Wormald, M. R., Lellouch, A. C., Rudd, P. M., Woof, J. M., & Dwek, R. A. (1998). *The Glycosylation and Structure of Human Serum IgA1, Fab, and Fc Regions and the Role of N-Glycosylation on Fc Receptor Interactions**. http://www.jbc.org

McDonald, K. A., Hong, L. M., Trombly, D. M., Xie, Q., & Jackman, A. P. (2008). Production of Human α-1-Antitrypsin from Transgenic Rice Cell Culture in a Membrane Bioreactor. *Biotechnology Progress*, *21*(3), 728–734. https://doi.org/10.1021/bp0496676

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. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.00403

Merkouropoulos, G., & Shirsat, A. H. (2003). The unusual Arabidopsis extensin gene atExt1 is expressed throughout plant development and is induced by a variety of biotic and abiotic stresses. *Planta*, *217*(3), 356–366. https://doi.org/10.1007/s00425-003-1002-y

Mestecky, J., Tomana, M., Moldoveanu, Z., Julian, B. A., Suzuki, H., Matousovic, K., Renfrow, M. B., Novak, L., Wyatt, R. J., & Novak, J. (2008). Role of aberrant glycosylation of IgA1 molecules in the pathogenesis of iga nephropathy. In *Kidney and Blood Pressure Research* (Vol. 31, Issue 1, pp. 29–37). https://doi.org/10.1159/000112922

Min, S. R., Woo, J. W., Jeong, W. J., Han, S. K., Lee, Y. B., & Liu, J. R. (2006). Production of human lactoferrin in transgenic cell suspension cultures of sweet

potato. *Biologia Plantarum*, *50*(1), 131–134. https://doi.org/10.1007/s10535-005-0087-5

Minchiotti, L., Campagnoli, M., Rossi, A., Cosulich, M. E., Monti, M., Pucci, P., Kragh-Hansen, U., Granel, B., Disdier, P., Weiller, P., & Galliano, M. (2001). A nucleotide insertion and frameshift cause albumin Kénitra, an extended and O-glycosylated mutant of human serum albumin with two additional disulfide bridges. *European Journal of Biochemistry*, *268*(2), 344–352. https://doi.org/10.1046/j.1432-1033.2001.01899.x

Ming, A., Zhao, J., Liu, Y., Wang, Y., Wang, X., Li, J., & Zhang, L. (2024). O-glycosylation in viruses: A sweet tango. *MLife*, *3*(1), 57–73. https://doi.org/10.1002/mlf2.12105

Misaki, R., Fujiyama, K., & Seki, T. (2006). Expression of human CMP-N-acetylneuraminic acid synthetase and CMP-sialic acid transporter in tobacco suspension-cultured cell. *Biochemical and Biophysical Research Communications*, *339*(4), 1184–1189. https://doi.org/10.1016/j.bbrc.2005.11.130

Mócsai, R., Göritzer, K., Stenitzer, D., Maresch, D., Strasser, R., & Altmann, F. (2021). Prolyl Hydroxylase Paralogs in Nicotiana benthamiana Show High Similarity With Regard to Substrate Specificity. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.636597

Møller, S. R., Yi, X., Velásquez, S. M., Gille, S., Hansen, P. L. M., Poulsen, C. P., Olsen, C. E., Rejzek, M., Parsons, H., Yang, Z., Wandall, H. H., Clausen, H., Field, R. A., Pauly, M., Estevez, J. M., Harholt, J., Ulvskov, P., & Petersen, B. L. (2017). Identification and evolution of a plant cell wall specific glycoprotein glycosyl transferase, ExAD. *Scientific Reports*, *7*(1), 45341. https://doi.org/10.1038/srep45341

Moloney, M. M., Boothe, J., Keon, R., Nykiforuk, C., & Van Rooijen, G. (2009). Methods for the production of insulin in plants (Patent US 7,547,821 B2). In *Int. Cl. References Cited U.S. PATENT DOCUMENTS* (US 7,547,821 B2). http://www.aafp.org/

Moriguchi, R., Matsuoka, C., Suyama, A., & Matsuoka, K. (2011). Reduction of Plant-Specific Arabinogalactan-Type O-Glycosylation by Treating Tobacco Plants with Ferrous Chelator 2,2'-Dipyridyl. *Bioscience, Biotechnology, and Biochemistry*, *75*(5), 994–996. https://doi.org/10.1271/bbb.100884

Mulagapati, S., Koppolu, V., & Raju, T. S. (2017). Decoding of O-Linked Glycosylation by Mass Spectrometry. *Biochemistry*, *56*(9), 1218–1226. https://doi.org/10.1021/acs.biochem.6b01244

Myllykoski, M., Sutinen, A., Koski, M. K., Kallio, J. P., Raasakka, A., Myllyharju, J., Wierenga, R. K., & Koivunen, P. (2021). Structure of transmembrane prolyl 4-hydroxylase reveals unique organization of EF and dioxygenase domains. *The Journal of Biological Chemistry*, *296*, 100197. https://doi.org/10.1074/jbc.RA120.016542

Nabeshima, Y., Abe, C., Kawauchi, T., Hiroi, T., Uto, Y., & Nabeshima, Y. (2020). Simple method for large-scale production of macrophage activating factor GcMAF. *Scientific Reports*, *10*(1), 19122. https://doi.org/10.1038/s41598-020-75571-y

Nagata, T., Nemoto, Y., & Hasezawa, S. (1992). Tobacco BY-2 Cell Line as the "HeLa" Cell in the Cell Biology of Higher Plants. In *International Review of Cytology* (Vol. 132, pp. 1–30). Elsevier. https://doi.org/10.1016/S0074-7696(08)62452-3

Narimatsu, Y., Joshi, H. J., Nason, R., Van Coillie, J., Karlsson, R., Sun, L., Ye, Z., Chen, Y.-H., Schjoldager, K. T., Steentoft, C., Furukawa, S., Bensing, B. A., Sullam, P. M., Thompson, A. J., Paulson, J. C., Büll, C., Adema, G. J., Mandel, U., Hansen, L., … Clausen, H. (2019). An Atlas of Human Glycosylation Pathways Enables Display of the Human Glycome by Gene Engineered Cells. *Molecular Cell*, *75*(2), 394-407.e5. https://doi.org/10.1016/j.molcel.2019.05.017

Narimatsu, Y., Joshi, H. J., Schjoldager, K. T., Hintze, J., Halim, A., Steentoft, C., Nason, R., Mandel, U., Bennett, E. P., Clausen, H., & Vakhrushev, S. Y. (2019). Exploring Regulation of Protein O-Glycosylation in Isogenic Human HEK293 Cells by Differential O-Glycoproteomics. *Molecular & Cellular Proteomics*, *18*(7), 1396–1409. https://doi.org/10.1074/mcp.RA118.001121

Narimatsu, Y., Kubota, T., Furukawa, S., Morii, H., Narimatsu, H., & Yamasaki, K. (2010). Effect of Glycosylation on *Cis* / *Trans* Isomerization of Prolines in IgA1-Hinge Peptide. *Journal of the American Chemical Society*, *132*(16), 5548–5549. https://doi.org/10.1021/ja9106429

Nason, R., Büll, C., Konstantinidi, A., Sun, L., Ye, Z., Halim, A., Du, W., Sørensen, D. M., Durbesson, F., Furukawa, S., Mandel, U., Joshi, H. J., Dworkin, L. A., Hansen, L., David, L., Iverson, T. M., Bensing, B. A., Sullam, P. M., Varki, A.,

… Narimatsu, Y. (2021). Display of the human mucinome with defined O-glycans by gene engineered cells. *Nature Communications*, *12*(1), 4070. https://doi.org/10.1038/s41467-021-24366-4

Navarre, C., & Chaumont, F. (2022). *Production of Recombinant Glycoproteins in Nicotiana tabacum BY-2 Suspension Cells* (pp. 81–88). https://doi.org/10.1007/978-1-0716-2241-4_5

Navarre, C., Ghislain, M., Leterme, S., Ferroud, C., Dufour, J. P., & Goffeau, A. (1992). Purification and complete sequence of a small proteolipid associated with the plasma membrane H(+)-ATPase of Saccharomyces cerevisiae. *Journal of Biological Chemistry*, *267*(9), 6425–6428. https://doi.org/10.1016/S0021-9258(18)42713-5

Navarre, C., Orval, R., Peeters, M., Bailly, N., & Chaumont, F. (2023). Issue when expressing a recombinant protein under the control of p35S in Nicotiana tabacum BY-2 cells. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1266775

Nguema-Ona, E., Vicré-Gibouin, M., Gotté, M., Plancot, B., Lerouge, P., Bardor, M., & Driouich, A. (2014). Cell wall O-glycoproteins and N-glycoproteins: aspects of biosynthesis and function. *Frontiers in Plant Science*, *5*(OCT). https://doi.org/10.3389/fpls.2014.00499

Nguyen, J. T., Evans, D. P., Galvan, M., Pace, K. E., Leitenberg, D., Bui, T. N., & Baum, L. G. (2001). CD45 Modulates Galectin-1-Induced T Cell Death: Regulation by Expression of Core 2 *O* -Glycans. *The Journal of Immunology*, *167*(10), 5697–5707. https://doi.org/10.4049/jimmunol.167.10.5697

Niemann, M. C. E., Bartrina, I., Ashikov, A., Weber, H., Novák, O., Spíchal, L., Strnad, M., Strasser, R., Bakker, H., Schmülling, T., & Werner, T. (2015). Arabidopsis ROCK1 transports UDP-GlcNAc/UDP-GalNAc and regulates ER protein quality control and cytokinin activity. *Proceedings of the National Academy of Sciences*, *112*(1), 291–296. https://doi.org/10.1073/pnas.1419050112

Novak, J., Julian, B. A., Mestecky, J., & Renfrow, M. B. (2012). Glycosylation of IgA1 and pathogenesis of IgA nephropathy. *Seminars in Immunopathology*, *34*(3), 365–382. https://doi.org/10.1007/s00281-012-0306-z

Nuñez, A., Fishman, M. L., Fortis, L. L., Cooke, P. H., & Hotchkiss, A. T. (2009). Identification of Extensin Protein Associated with Sugar Beet Pectin. *Journal of*

Agricultural and Food Chemistry, *57*(22), 10951–10958. https://doi.org/10.1021/jf902162t

Nurzijah, I., Elbohy, O. A., Kanyuka, K., Daly, J. M., & Dunham, S. (2022). Development of Plant-Based Vaccines for Prevention of Avian Influenza and Newcastle Disease in Poultry. *Vaccines*, *10*(3), 478. https://doi.org/10.3390/vaccines10030478

Ogawa-Ohnishi, M., & Matsubayashi, Y. (2015). Identification of three potent hydroxyproline O-galactosyltransferases in Arabidopsis. *The Plant Journal*, *81*(5), 736–746. https://doi.org/10.1111/tpj.12764

Ogawa-Ohnishi, M., Matsushita, W., & Matsubayashi, Y. (2013). Identification of three hydroxyproline O-arabinosyltransferases in Arabidopsis thaliana. *Nature Chemical Biology*, *9*(11), 726–730. https://doi.org/10.1038/nchembio.1351

Oka, T., Saito, F., Shimma, Y., Yoko-o, T., Nomura, Y., Matsuoka, K., & Jigami, Y. (2009). Characterization of Endoplasmic Reticulum-Localized UDP-Galactose: Hydroxyproline O-Galactosyltransferase Using Synthetic Peptide Substrates in Arabidopsis. *Plant Physiology*, *152*(1), 332–340. https://doi.org/10.1104/pp.109.146266

Paccalet, T., Bardor, M., Rihouey, C., Delmas, F., Chevalier, C., D'Aoust, M., Faye, L., Vézina, L., Gomord, V., & Lerouge, P. (2007). Engineering of a sialic acid synthesis pathway in transgenic plants by expression of bacterial Neu5Ac-synthesizing enzymes. *Plant Biotechnology Journal*, *5*(1), 16–25. https://doi.org/10.1111/j.1467-7652.2006.00211.x

Park, S. H., Ji, K.-Y., Kim, H. M., Ma, S. H., Park, S. Y., Do, J. H., Oh, D.-B., Kang, H. S., Shim, J. S., & Joung, Y. H. (2021). Optimization of the human colorectal carcinoma antigen GA733-2 production in tobacco plants. *Plant Biotechnology Reports*, *15*(1), 55–67. https://doi.org/10.1007/s11816-020-00657-y

Park, S. H., Ji, K.-Y., Park, S. Y., Kim, H. M., Ma, S. H., Do, J. H., Kang, H., Kang, H. S., Oh, D.-B., Shim, J. S., & Joung, Y. H. (2022). Immunotherapeutic effects of recombinant colorectal cancer antigen produced in tomato fruits. *Scientific Reports*, *12*(1), 9723. https://doi.org/10.1038/s41598-022-13839-1

Parsons, J., Altmann, F., Graf, M., Stadlmann, J., Reski, R., & Decker, E. L. (2013). A gene responsible for prolyl-hydroxylation of moss-produced recombinant

human erythropoietin. *Scientific Reports*, *3*(1), 3019. https://doi.org/10.1038/srep03019

Pearce, G., Bhattacharya, R., Chen, Y.-C., Barona, G., Yamaguchi, Y., & Ryan, C. A. (2009). Isolation and Characterization of Hydroxyproline-Rich Glycopeptide Signals in Black Nightshade Leaves. *Plant Physiology*, *150*(3), 1422–1433. https://doi.org/10.1104/pp.109.138669

Peng, B., Kong, G., Yang, C., & Ming, Y. (2020). Erythropoietin and its derivatives: from tissue protection to immune regulation. *Cell Death & Disease*, *11*(2), 79. https://doi.org/10.1038/s41419-020-2276-8

Perrakis, A., Bita, C. E., Arhondakis, S., Krokida, A., Mekkaoui, K., Denic, D., Blazakis, K. N., Kaloudas, D., & Kalaitzis, P. (2019). Suppression of a prolyl 4 hydroxylase results in delayed abscission of overripe tomato fruits. *Frontiers in Plant Science*, *10*, 1–11. https://doi.org/10.3389/fpls.2019.00348

Perrine, C. L., Ganguli, A., Wu, P., Bertozzi, C. R., Fritz, T. A., Raman, J., Tabak, L. A., & Gerken, T. A. (2009). Glycopeptide-preferring Polypeptide GalNAc Transferase 10 (ppGalNAc T10), Involved in Mucin-type O-Glycosylation, Has a Unique GalNAc-O-Ser/Thr-binding Site in Its Catalytic Domain Not Found in ppGalNAc T1 or T2. *Journal of Biological Chemistry*, *284*(30), 20387–20397. https://doi.org/10.1074/jbc.M109.017236

Petersen, B. L., MacAlister, C. A., & Ulvskov, P. (2021). Plant Protein O-Arabinosylation. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.645219

Pietrosemoli, N., García-Martín, J. A., Solano, R., & Pazos, F. (2013). Genome-Wide Analysis of Protein Disorder in Arabidopsis thaliana: Implications for Plant Environmental Adaptation. *PLoS ONE*, *8*(2), e55524. https://doi.org/10.1371/journal.pone.0055524

Pinho, S. S., & Reis, C. A. (2015). Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer*, *15*(9), 540–555. https://doi.org/10.1038/nrc3982

Pinkhasov, J., Alvarez, M. L., Rigano, M. M., Piensook, K., Larios, D., Pabst, M., Grass, J., Mukherjee, P., Gendler, S. J., Walmsley, A. M., & Mason, H. S. (2011). Recombinant plant-expressed tumour-associated MUC1 peptide is immunogenic and capable of breaking tolerance in MUC1.Tg mice. *Plant*

Biotechnology Journal, *9*(9), 991–1001. https://doi.org/10.1111/j.1467-7652.2011.00614.x

Pires, A. S., Rosa, S., Castanheira, S., Fevereiro, P., & Abranches, R. (2012). Expression of a recombinant human erythropoietin in suspension cell cultures of Arabidopsis, tobacco and Medicago. *Plant Cell, Tissue and Organ Culture (PCTOC)*, *110*(1), 171–181. https://doi.org/10.1007/s11240-012-0141-x

Pires, A. S., Santos, R. B., Nogueira, A. C., & Abranches, R. (2014). Production of human lipocalin-type prostaglandin D synthase in the model plant Medicago truncatula. *In Vitro Cellular & Developmental Biology - Plant*, *50*(2), 276–281. https://doi.org/10.1007/s11627-013-9584-y

Plomp, R., Dekkers, G., Rombouts, Y., Visser, R., Koeleman, C. A. M., Kammeijer, G. S. M., Jansen, B. C., Rispens, T., Hensbergen, P. J., Vidarsson, G., & Wuhrer, M. (2015). Hinge-Region O-Glycosylation of Human Immunoglobulin G3 (IgG3). *Molecular & Cellular Proteomics*, *14*(5), 1373–1384. https://doi.org/10.1074/mcp.M114.047381

Poiroux, G., Barre, A., Van Damme, E., Benoist, H., & Rougé, P. (2017). Plant Lectins Targeting O-Glycans at the Cell Surface as Tools for Cancer Diagnosis, Prognosis and Therapy. *International Journal of Molecular Sciences*, *18*(6), 1232. https://doi.org/10.3390/ijms18061232

Poulsen, C. P., Dilokpimol, A., Mouille, G., Burow, M., & Geshi, N. (2014). Arabinogalactan Glycosyltransferases Target to a Unique Subcellular Compartment That May Function in Unconventional Secretion in Plants. *Traffic*, *15*(11), 1219–1234. https://doi.org/10.1111/tra.12203

Pratt, M. R., Hang, H. C., Ten Hagen, K. G., Rarick, J., Gerken, T. A., Tabak, L. A., & Bertozzi, C. R. (2004). Deconvoluting the Functions of Polypeptide N-α-Acetylgalactosaminyltransferase Family Members by Glycopeptide Substrate Profiling. *Chemistry & Biology*, *11*(7), 1009–1016. https://doi.org/10.1016/j.chembiol.2004.05.009

Proshlyakov, D. A., Henshaw, T. F., Monterosso, G. R., Ryle, M. J., & Hausinger, R. P. (2004). Direct Detection of Oxygen Intermediates in the Non-Heme Fe Enzyme Taurine/α-Ketoglutarate Dioxygenase. *Journal of the American Chemical Society*, *126*(4), 1022–1023. https://doi.org/10.1021/ja039113j

Qin, X., Yamauchi, R., Aizawa, K., Inakuma, T., & Kato, K. (2000). Isolation and Characterization of Arabinogalactan-protein from the Fruit of Lycium

chinense Mill. *Journal of Applied Glycoscience*, *47*(2), 155–161. https://doi.org/10.5458/jag.47.155

Qu, J., Ma, C., Xu, X.-Q., Xiao, M., Zhang, J., Li, D., Liu, D., Konkle, B. A., Miao, C. H., Li, L., & Xiao, W. (2020). Comparative glycosylation mapping of plasma-derived and recombinant human factor VIII. *PLOS ONE*, *15*(5), e0233576. https://doi.org/10.1371/journal.pone.0233576

Qu, Y., Egelund, J., Gilson, P. R., Houghton, F., Gleeson, P. A., Schultz, C. J., & Bacic, A. (2008). Identification of a novel group of putative Arabidopsis thaliana β-(1,3)-galactosyltransferases. *Plant Molecular Biology*, *68*(1–2), 43–59. https://doi.org/10.1007/s11103-008-9351-3

Radhakrishnan, P., Dabelsteen, S., Madsen, F. B., Francavilla, C., Kopp, K. L., Steentoft, C., Vakhrushev, S. Y., Olsen, J. V., Hansen, L., Bennett, E. P., Woetmann, A., Yin, G., Chen, L., Song, H., Bak, M., Hlady, R. A., Peters, S. L., Opavsky, R., Thode, C., … Wandall, H. H. (2014). Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proceedings of the National Academy of Sciences*, *111*(39), E4066–E4075. https://doi.org/10.1073/pnas.1406619111

Raman, J., Fritz, T. A., Gerken, T. A., Jamison, O., Live, D., Liu, M., & Tabak, L. A. (2008). The Catalytic and Lectin Domains of UDP-GalNAc:Polypeptide α-N-Acetylgalactosaminyltransferase Function in Concert to Direct Glycosylation Site Selection. *Journal of Biological Chemistry*, *283*(34), 22942–22951. https://doi.org/10.1074/jbc.M803387200

Ramírez-Alanis, I. A., Renaud, J. B., García-Lara, S., Menassa, R., & Cardineau, G. A. (2018). Transient co-expression with three O-glycosylation enzymes allows production of GalNAc-O-glycosylated Granulocyte-Colony Stimulating Factor in N. benthamiana. *Plant Methods*, *14*(1), 98. https://doi.org/10.1186/s13007-018-0363-y

Ramos-Martinez, E. M., Fimognari, L., & Sakuragi, Y. (2017). High-yield secretion of recombinant proteins from the microalga Chlamydomonas reinhardtii. *Plant Biotechnology Journal*, *15*(9), 1214–1224. https://doi.org/10.1111/pbi.12710

Rebelo, B. A., Folgado, A., Ferreira, A. C., & Abranches, R. (2022). Production of the SARS-CoV-2 Spike protein and its Receptor Binding Domain in plant cell suspension cultures. *Frontiers in Plant Science*, *13*. https://doi.org/10.3389/fpls.2022.995429

Reinholdt, J., Tomana, M., Mortensen, S. B., & Kilian, M. (1990). Molecular aspects of immunoglobulin A1 degradation by oral streptococci. *Infection and Immunity*, *58*(5), 1186–1194. https://doi.org/10.1128/iai.58.5.1186-1194.1990

Revoredo, L., Wang, S., Bennett, E. P., Clausen, H., Moremen, K. W., Jarvis, D. L., Ten Hagen, K. G., Tabak, L. A., & Gerken, T. A. (2016). Mucin-type O-glycosylation is controlled by short- and long-range glycopeptide substrate recognition that varies among members of the polypeptide GalNAc transferase family. *Glycobiology*, *26*(4), 360–376. https://doi.org/10.1093/glycob/cwv108

Röttger, S., White, J., Wandall, H. H., Olivo, J.-C., Stark, A., Bennett, E. P., Whitehouse, C., Berger, E. G., Clausen, H., & Nilsson, T. (1998). Localization of three human polypeptide GalNAc-transferases in HeLa cells suggests initiation of O-linked glycosylation throughout the Golgi apparatus. *Journal of Cell Science*, *111*(1), 45–60. https://doi.org/10.1242/jcs.111.1.45

Rouwendal, G. J., van der Lee, M. M., Meyer, S., Reiding, K. R., Schouten, J., de Roo, G., Egging, D. F., Leusen, J. H., Boross, P., Wuhrer, M., Verheijden, G. F., Dokter, W. H., Timmers, M., & Ubink, R. (2016). A comparison of anti-HER2 IgA and IgG1 in vivo efficacy is facilitated by high N-glycan sialylation of the IgA. *MAbs*, *8*(1), 74–86. https://doi.org/10.1080/19420862.2015.1102812

Royle, L., Roos, A., Harvey, D. J., Wormald, M. R., Van Gijlswijk-Janssen, D., Redwan, E.-R. M., Wilson, I. A., Daha, M. R., Dwek, R. A., & Rudd, P. M. (2003). Secretory IgA N- and O-Glycans Provide a Link between the Innate and Adaptive Immune Systems. *Journal of Biological Chemistry*, *278*(22), 20140–20153. https://doi.org/10.1074/jbc.M301436200

Ruhlman, T., Ahangari, R., Devine, A., Samsam, M., & Daniell, H. (2007). Expression of cholera toxin B–proinsulin fusion protein in lettuce and tobacco chloroplasts – oral administration protects against development of insulitis in non-obese diabetic mice. *Plant Biotechnology Journal*, *5*(4), 495–510. https://doi.org/10.1111/j.1467-7652.2007.00259.x

Ryan, C. A., & Pearce, G. (2003). Systemins: A functionally defined family of peptide signals that regulate defensive genes in Solanaceae species. *Proceedings of the National Academy of Sciences*, *100*(suppl_2), 14577–14580. https://doi.org/10.1073/pnas.1934788100

Rybicki, E. P. (2014). Plant-based vaccines against viruses. *Virology Journal*, *11*(1), 205. https://doi.org/10.1186/s12985-014-0205-0

Saito, F., Suyama, A., Oka, T., Yoko-o, T., Matsuoka, K., Jigami, Y., & Shimma, Y. (2014). Identification of Novel Peptidyl Serine α-Galactosyltransferase Gene Family in Plants. *Journal of Biological Chemistry*, *289*(30), 20405–20420. https://doi.org/10.1074/jbc.M114.553933

Salvá, I., & Jamet, É. (2001). Expression of the Tobacco Ext1.4 Extensin Gene Upon Mechanical Constraint and Localization of Regulatory Regions. *Plant Biology*, *3*(1), 32–41. https://doi.org/10.1055/s-2001-11746

Sanchez, J.-F., Lescar, J., Chazalet, V., Audfray, A., Gagnon, J., Alvarez, R., Breton, C., Imberty, A., & Mitchell, E. P. (2006). Biochemical and Structural Analysis of Helix pomatia Agglutinin. *Journal of Biological Chemistry*, *281*(29), 20171–20180. https://doi.org/10.1074/jbc.M603452200

Sanda, M., Morrison, L., & Goldman, R. (2021). N-and O-Glycosylation of the SARS-CoV-2 Spike Protein. *Analytical Chemistry*, *93*(4), 2003–2009. https://doi.org/10.1021/acs.analchem.0c03173

Santos, R. B., Abranches, R., Fischer, R., Sack, M., & Holland, T. (2016). Putting the Spotlight Back on Plant Suspension Cultures. *Frontiers in Plant Science*, *7*. https://doi.org/10.3389/fpls.2016.00297

Schjoldager, K. T., Joshi, H. J., Kong, Y., Goth, C. K., King, S. L., Wandall, H. H., Bennett, E. P., Vakhrushev, S. Y., & Clausen, H. (2015). Deconstruction of O-glycosylation—GalNAc-T isoforms direct distinct subsets of the O-glycoproteome. *EMBO Reports*, *16*(12), 1713–1722. https://doi.org/10.15252/embr.201540796

Schjoldager, K. T., Narimatsu, Y., Joshi, H. J., & Clausen, H. (2020). Global view of human protein glycosylation pathways and functions. *Nature Reviews Molecular Cell Biology*, *21*(12), 729–749. https://doi.org/10.1038/s41580-020-00294-x

Schjoldager, K. T., Vester-Christensen, M. B., Bennett, E. P., Levery, S. B., Schwientek, T., Yin, W., Blixt, O., & Clausen, H. (2010). O-Glycosylation Modulates Proprotein Convertase Activation of Angiopoietin-like Protein 3. *Journal of Biological Chemistry*, *285*(47), 36293–36303. https://doi.org/10.1074/jbc.M110.156950

Schmidt, A., Datta, K., & Marcus, A. (1991). Peptidyl Proline Hydroxylation and the Growth of a Soybean Cell Culture. *Plant Physiology*, *96*(2), 656–659. https://doi.org/10.1104/pp.96.2.656

Schwientek, T., Yeh, J.-C., Levery, S. B., Keck, B., Merkx, G., van Kessel, A. G., Fukuda, M., & Clausen, H. (2000). Control of O-Glycan Branch Formation. *Journal of Biological Chemistry*, *275*(15), 11106–11113. https://doi.org/10.1074/jbc.275.15.11106

Seifert, G. J. (2020). On the Potential Function of Type II Arabinogalactan O-Glycosylation in Regulating the Fate of Plant Secretory Proteins. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.563735

Shajahan, A., Supekar, N. T., Gleinich, A. S., & Azadi, P. (2020). Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology*, *30*(12), 981–988. https://doi.org/10.1093/glycob/cwaa042

Sheehan, J. K., Kirkham, S., Howard, M., Woodman, P., Kutay, S., Brazeau, C., Buckley, J., & Thornton, D. J. (2004). Identification of Molecular Intermediates in the Assembly Pathway of the MUC5AC Mucin. *Journal of Biological Chemistry*, *279*(15), 15698–15705. https://doi.org/10.1074/jbc.M313241200

Sherman, A., Su, J., Lin, S., Wang, X., Herzog, R. W., & Daniell, H. (2014). Suppression of inhibitor formation against FVIII in a murine model of hemophilia A by oral delivery of antigens bioencapsulated in plant cells. *Blood*, *124*(10), 1659–1668. https://doi.org/10.1182/blood-2013-10-528737

Shimizu, M., Igasaki, T., Yamada, M., Yuasa, K., Hasegawa, J., Kato, T., Tsukagoshi, H., Nakamura, K., Fukuda, H., & Matsuoka, K. (2005). Experimental determination of proline hydroxylation and hydroxyproline arabinogalactosylation motifs in secretory proteins. *The Plant Journal*, *42*(6), 877–889. https://doi.org/10.1111/j.1365-313X.2005.02419.x

Shin, Y., Chong, Y., Yang, M., & Kwon, T. (2011). Production of recombinant human granulocyte macrophage-colony stimulating factor in rice cell suspension culture with a human-like *N* -glycan structure. *Plant Biotechnology Journal*, *9*(9), 1109–1119. https://doi.org/10.1111/j.1467-7652.2011.00636.x

Shin, Y.-J., König-Beihammer, J., Vavra, U., Schwestka, J., Kienzl, N. F., Klausberger, M., Laurent, E., Grünwald-Gruber, C., Vierlinger, K., Hofner, M., Margolin, E., Weinhäusel, A., Stöger, E., Mach, L., & Strasser, R. (2021). N-Glycosylation of the SARS-CoV-2 Receptor Binding Domain Is Important for Functional Expression in Plants. *Frontiers in Plant Science*, *12*, 689104. https://doi.org/10.3389/fpls.2021.689104

Showalter, A. M. (1993). Structure and function of plant cell wall proteins. *The Plant Cell*, *5*(1), 9–23. https://doi.org/10.1105/tpc.5.1.9

Showalter, A. M., & Basu, D. (2016). Extensin and Arabinogalactan-Protein Biosynthesis: Glycosyltransferases, Research Challenges, and Biosensors. *Frontiers in Plant Science*, *7*(JUNE2016). https://doi.org/10.3389/fpls.2016.00814

Showalter, A. M., Keppler, B., Lichtenberg, J., Gu, D., & Welch, L. R. (2010). A Bioinformatics Approach to the Identification, Classification, and Analysis of Hydroxyproline-Rich Glycoproteins . *Plant Physiology*, *153*(2), 485–513. https://doi.org/10.1104/pp.110.156554

Shpak, E., Barbar, E., Leykam, J. F., & Kieliszewski, M. J. (2001). Contiguous Hydroxyproline Residues Direct Hydroxyproline Arabinosylation in Nicotiana tabacum. *Journal of Biological Chemistry*, *276*(14), 11272–11278. https://doi.org/10.1074/jbc.M011323200

Shpak, E., Leykam, J. F., & Kieliszewski, M. J. (1999). Synthetic genes for glycoprotein design and the elucidation of hydroxyproline- *O* -glycosylation codes. *Proceedings of the National Academy of Sciences*, *96*(26), 14736–14741. https://doi.org/10.1073/pnas.96.26.14736

Silva, J., Ferraz, R., Dupree, P., Showalter, A. M., & Coimbra, S. (2020). Three Decades of Advances in Arabinogalactan-Protein Biosynthesis. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.610377

Smargiasso, N., Nader, J., Rioux, S., Mazzucchelli, G., Boutry, M., De Pauw, E., Chaumont, F., & Navarre, C. (2019). Exploring the N-Glycosylation Profile of Glycoprotein B from Human Cytomegalovirus Expressed in CHO and Nicotiana tabacum BY-2 Cells. *International Journal of Molecular Sciences*, *20*(15), 3741. https://doi.org/10.3390/ijms20153741

Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Paul Bennett, E., Mandel, U., Brunak, S., Wandall, H. H., Levery, S. B., & Clausen, H. (2013). Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *The EMBO Journal*, *32*(10), 1478–1488. https://doi.org/10.1038/emboj.2013.79

Steentoft, C., Vakhrushev, S. Y., Vester-Christensen, M. B., Schjoldager, K. T., Kong, Y., Bennett, E. P., Mandel, U., Wandall, H., Levery, S. B., & Clausen, H. (2011).

Mining the O-glycoproteome using zinc-finger nuclease–glycoengineered SimpleCell lines. *Nature Methods*, *8*(11), 977–982. https://doi.org/10.1038/nmeth.1731

Stephen C. Fry. (1993). Chemical analysis of components of the primary cell wall. In N. Harris & K. J. Oparka (Eds.), *Plant Cell Biology - A Practical Approach* (pp. 199–219).

Sticher, L., Hofsteenge, J., Milani, A., Neuhaus, J.-M., & Meins, F. (1992). Vacuolar Chitinases of Tobacco: a New Class of Hydroxyproline-Containing Proteins. *Science*, *257*(5070), 655–657. https://doi.org/10.1126/science.1496378

Strahl-Bolsinger, S., Gentzsch, M., & Tanner, W. (1999). Protein O-mannosylation. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1426*(2), 297–307. https://doi.org/10.1016/S0304-4165(98)00131-7

Strasser, R. (2013). Engineering of human-type O-glycosylation in Nicotiana benthamiana plants. *Bioengineered*, *4*(4), 191–196. https://doi.org/10.4161/bioe.22857

Strasser, R. (2023). Plant glycoengineering for designing next-generation vaccines and therapeutic proteins. *Biotechnology Advances*, *67*, 108197. https://doi.org/10.1016/j.biotechadv.2023.108197

Strasser, R., Seifert, G., Doblin, M. S., Johnson, K. L., Ruprecht, C., Pfrengle, F., Bacic, A., & Estevez, J. M. (2021). Cracking the "Sugar Code": A Snapshot of N- and O-Glycosylation Pathways and Functions in Plants Cells. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.640919

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

Su, H., van Eerde, A., Rimstad, E., Bock, R., Branza-Nichita, N., Yakovlev, I. A., & Clarke, J. L. (2023). Plant-made vaccines against viral diseases in humans and farm animals. *Frontiers in Plant Science*, *14*. https://doi.org/10.3389/fpls.2023.1170815

Sun, Q.-Y., Ding, L.-W., Lomonossoff, G. P., Sun, Y.-B., Luo, M., Li, C.-Q., Jiang, L., & Xu, Z.-F. (2011). Improved expression and purification of recombinant human serum albumin from transgenic tobacco suspension culture. *Journal of Biotechnology*, *155*(2), 164–172. https://doi.org/10.1016/j.jbiotec.2011.06.033

Sun, W., Kieliszewski, M. J., & Showalter, A. M. (2004). Overexpression of tomato LeAGP-1 arabinogalactan-protein promotes lateral branching and hampers reproductive development. *Plant Journal*, *40*(6), 870–881. https://doi.org/10.1111/j.1365-313X.2004.02274.x

Suresh, S., Rajvanshi, P. K., & Noguchi, C. T. (2020). The Many Facets of Erythropoietin Physiologic and Metabolic Response. *Frontiers in Physiology*, *10*. https://doi.org/10.3389/fphys.2019.01534

Suzuki, T., Narciso, J. O., Zeng, W., van de Meene, A., Yasutomi, M., Takemura, S., Lampugnani, E. R., Doblin, M. S., Bacic, A., & Ishiguro, S. (2017). KNS4/UPEX1: A Type II Arabinogalactan β-(1,3)-Galactosyltransferase Required for Pollen Exine Development. *Plant Physiology*, *173*(1), 183–205. https://doi.org/10.1104/pp.16.01385

Taleb, V., Liao, Q., Narimatsu, Y., García-García, A., Compañón, I., Borges, R. J., González-Ramírez, A. M., Corzana, F., Clausen, H., Rovira, C., & Hurtado-Guerrero, R. (2022). Structural and mechanistic insights into the cleavage of clustered O-glycan patches-containing glycoproteins by mucinases of the human gut. *Nature Communications*, *13*(1), 4324. https://doi.org/10.1038/s41467-022-32021-9

Tan, L., Eberhard, S., Pattathil, S., Warder, C., Glushka, J., Yuan, C., Hao, Z., Zhu, X., Avci, U., Miller, J. S., Baldwin, D., Pham, C., Orlando, R., Darvill, A., Hahn, M. G., Kieliszewski, M. J., & Mohnen, D. (2013). An *Arabidopsis* Cell Wall Proteoglycan Consists of Pectin and Arabinoxylan Covalently Linked to an Arabinogalactan Protein. *The Plant Cell*, *25*(1), 270–287. https://doi.org/10.1105/tpc.112.107334

Tan, L., Leykam, J. F., & Kieliszewski, M. J. (2003). Glycosylation Motifs That Direct Arabinogalactan Addition to Arabinogalactan-Proteins. *Plant Physiology*, *132*(3), 1362–1369. https://doi.org/10.1104/pp.103.021766

Tan, L., Qiu, F., Lamport, D. T. A., & Kieliszewski, M. J. (2004). Structure of a Hydroxyproline (Hyp)-Arabinogalactan Polysaccharide from Repetitive Ala-

Hyp Expressed in Transgenic Nicotiana tabacum. *Journal of Biological Chemistry*, *279*(13), 13156–13165. https://doi.org/10.1074/jbc.M311864200

Tan, L., Tees, D., Qian, J., Kareem, S., & Kieliszewski, M. J. (2018). Intermolecular interactions between glycomodules of plant cell wall arabinogalactan-proteins and extensins. *The Cell Surface*, *1*, 25–33. https://doi.org/10.1016/j.tcsw.2018.03.001

Tan, L., Varnai, P., Lamport, D. T. A., Yuan, C., Xu, J., Qiu, F., & Kieliszewski, M. J. (2010). Plant O-Hydroxyproline Arabinogalactans Are Composed of Repeating Trigalactosyl Subunits with Short Bifurcated Side Chains. *Journal of Biological Chemistry*, *285*(32), 24575–24583. https://doi.org/10.1074/jbc.M109.100149

Taschwer, M., Hackl, M., Hernández Bort, J. A., Leitner, C., Kumar, N., Puc, U., Grass, J., Papst, M., Kunert, R., Altmann, F., & Borth, N. (2012). Growth, productivity and protein glycosylation in a CHO EpoFc producer cell line adapted to glutamine-free growth. *Journal of Biotechnology*, *157*(2), 295–303. https://doi.org/10.1016/j.jbiotec.2011.11.014

Taujale, R., Venkat, A., Huang, L.-C., Zhou, Z., Yeung, W., Rasheed, K. M., Li, S., Edison, A. S., Moremen, K. W., & Kannan, N. (2020). Deep evolutionary analysis reveals the design principles of fold A glycosyltransferases. *ELife*, *9*. https://doi.org/10.7554/eLife.54532

Taylor, C. M., Karunaratne, C. V., & Xie, N. (2012). Glycosides of hydroxyproline: Some recent, unusual discoveries. *Glycobiology*, *22*(6), 757–767. https://doi.org/10.1093/glycob/cwr188

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 Biotechnology Journal*, *13*(8), 1199–1208. https://doi.org/10.1111/pbi.12428

Temple, H., Mortimer, J. C., Tryfona, T., Yu, X., Lopez-Hernandez, F., Sorieul, M., Anders, N., & Dupree, P. (2019). Two members of the DUF579 family are responsible for arabinogalactan methylation in Arabidopsis. *Plant Direct*, *3*(2). https://doi.org/10.1002/pld3.117

Tenno, M., Kézdy, F. J., Elhammer, Å. P., & Kurosaka, A. (2002). Function of the lectin domain of polypeptide N-acetylgalactosaminyltransferase 1. *Biochemical*

and Biophysical Research Communications, *298*(5), 755–759. https://doi.org/10.1016/S0006-291X(02)02549-4

Thompson, N., & Wakarchuk, W. (2022). O-glycosylation and its role in therapeutic proteins. *Bioscience Reports*, *42*(10). https://doi.org/10.1042/BSR20220094

Tiainen, P., Myllyharju, J., & Koivunen, P. (2005). Characterization of a Second Arabidopsis thaliana Prolyl 4-Hydroxylase with Distinct Substrate Specificity. *Journal of Biological Chemistry*, *280*(2), 1142–1148. https://doi.org/10.1074/jbc.M411109200

Tian, E., & Ten Hagen, K. G. (2009). Recent insights into the biological roles of mucin-type O-glycosylation. *Glycoconjugate Journal*, *26*(3), 325–334. https://doi.org/10.1007/s10719-008-9162-4

Tran, D. T., & Ten Hagen, K. G. (2013). Mucin-type O-glycosylation during development. *The Journal of Biological Chemistry*, *288*(10), 6921–6929. https://doi.org/10.1074/jbc.R112.418558

Tryfona, T., Liang, H.-C., Kotake, T., Kaneko, S., Marsh, J., Ichinose, H., Lovegrove, A., Tsumuraya, Y., Shewry, P. R., Stephens, E., & Dupree, P. (2010). Carbohydrate structural analysis of wheat flour arabinogalactan protein. *Carbohydrate Research*, *345*(18), 2648–2656. https://doi.org/10.1016/j.carres.2010.09.018

Tryfona, T., Liang, H.-C., Kotake, T., Tsumuraya, Y., Stephens, E., & Dupree, P. (2012). Structural Characterization of Arabidopsis Leaf Arabinogalactan Polysaccharides . *Plant Physiology*, *160*(2), 653–666. https://doi.org/10.1104/pp.112.202309

Tryfona, T., Theys, T. E., Wagner, T., Stott, K., Keegstra, K., & Dupree, P. (2014). Characterisation of FUT4 and FUT6 α-(1→2)-Fucosyltransferases Reveals that Absence of Root Arabinogalactan Fucosylation Increases Arabidopsis Root Growth Salt Sensitivity. *PLoS ONE*, *9*(3), e93291. https://doi.org/10.1371/journal.pone.0093291

Uetz, P., Göritzer, K., Vergara, E., Melnik, S., Grünwald-Gruber, C., Figl, R., Deghmane, A.-E., Groppelli, E., Reljic, R., Ma, J. K.-C., Stöger, E., & Strasser, R. (2024). Implications of O-glycan modifications in the hinge region of a plant-produced SARS-CoV-2-IgA antibody on functionality. *Frontiers in Bioengineering and Biotechnology*, *12*. https://doi.org/10.3389/fbioe.2024.1329018

Uetz, P., Melnik, S., Grünwald-Gruber, C., Strasser, R., & Stoger, E. (2022). CRISPR/Cas9-mediated knockout of a prolyl-4-hydroxylase subfamily in *Nicotiana benthamiana* using DsRed2 for plant selection. *Biotechnology Journal*, *17*(7). https://doi.org/10.1002/biot.202100698

Üzülmez, Ö., Kalic, T., Mayr, V., Lengger, N., Tscheppe, A., Radauer, C., Hafner, C., Hemmer, W., & Breiteneder, H. (2021). The Major Peanut Allergen Ara h 2 Produced in Nicotiana benthamiana Contains Hydroxyprolines and Is a Viable Alternative to the E. Coli Product in Allergy Diagnosis. *Frontiers in Plant Science*, *12*. https://doi.org/10.3389/fpls.2021.723363

Van Der Fits, L., Deakin, E. A., Harry, J., Hoge, C., & Memelink, J. (2000). The ternary transformation system: constitutive virG on a compatible plasmid dramatically increases Agrobacterium-mediated plant transformation. In *Plant Molecular Biology* (Vol. 43).

Van Hengel, A. J., & Roberts, K. (2002). Fucosylated arabinogalactan-proteins are required for full root cell elongation in arabidopsis. *The Plant Journal*, *32*(1), 105–113. https://doi.org/10.1046/j.1365-313X.2002.01406.x

Vasilev, N., Grömping, U., Lipperts, A., Raven, N., Fischer, R., & 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 Biotechnology Journal*, *11*(7), 867–874. https://doi.org/10.1111/pbi.12079

Velasquez, S. M., Marzol, E., Borassi, C., Pol-Fachin, L., Ricardi, M. M., Mangano, S., Juarez, S. P. D., Salter, J. D. S., Dorosz, J. G., Marcus, S. E., Knox, J. P., Dinneny, J. R., Iusem, N. D., Verli, H., & Estevez, J. M. (2015). Low Sugar Is Not Always Good: Impact of Specific O-Glycan Defects on Tip Growth in Arabidopsis. *Plant Physiology*, *168*(3), 808–813. https://doi.org/10.1104/pp.114.255521

Velasquez, S. M., Ricardi, M. M., Dorosz, J. G., Fernandez, P. V., Nadra, A. D., Pol-Fachin, L., Egelund, J., Gille, S., Harholt, J., Ciancia, M., Verli, H., Pauly, M., Bacic, A., Olsen, C. E., Ulvskov, P., Petersen, B. L., Somerville, C., Iusem, N. D., & Estevez, J. M. (2011). O-Glycosylated Cell Wall Proteins Are Essential in Root Hair Growth. *Science*, *332*(6036), 1401–1403. https://doi.org/10.1126/science.1206657

Velasquez, S. M., Ricardi, M. M., Poulsen, C. P., Oikawa, A., Dilokpimol, A., Halim, A., Mangano, S., Denita Juarez, S. P., Marzol, E., Salgado Salter, J. D., Dorosz, J. G., Borassi, C., Möller, S. R., Buono, R., Ohsawa, Y., Matsuoka, K., Otegui, M. S., Scheller, H. V., Geshi, N., … Estevez, J. M. (2015). Complex Regulation of Prolyl-4-Hydroxylases Impacts Root Hair Expansion. *Molecular Plant*, *8*(5), 734–746. https://doi.org/10.1016/j.molp.2014.11.017

Vigani, G., Morandini, P., & Murgia, I. (2013). Searching iron sensors in plants by exploring the link among 2'-OG-dependent dioxygenases, the iron deficiency response and metabolic adjustments occurring under iron deficiency. *Frontiers in Plant Science*, *4*(MAY). https://doi.org/10.3389/fpls.2013.00169

Vlad, F., Spano, T., Vlad, D., Bou Daher, F., Ouelhadj, A., & Kalaitzis, P. (2007). Arabidopsis prolyl 4-hydroxylases are differentially expressed in response to hypoxia, anoxia and mechanical wounding. *Physiologia Plantarum*, *130*(4), 471–483. https://doi.org/10.1111/j.1399-3054.2007.00915.x

Vlad, F., Tiainen, P., Owen, C., Spano, T., Daher, F. B., Oualid, F., Senol, N. O., Vlad, D., Myllyharju, J., & Kalaitzis, P. (2010). Characterization of two carnation petal prolyl 4 hydroxylases. *Physiologia Plantarum*, *140*(2), 199–207. https://doi.org/10.1111/j.1399-3054.2010.01390.x

Wallace, E. L., Goker-Alpan, O., Wilcox, W. R., Holida, M., Bernat, J., Longo, N., Linhart, A., Hughes, D. A., Hopkin, R. J., Tøndel, C., Langeveld, M., Giraldo, P., Pisani, A., Germain, D. P., Mehta, A., Deegan, P. B., Molnar, M. J., Ortiz, D., Jovanovic, A., … Warnock, D. G. (2024). Head-to-head trial of pegunigalsidase alfa versus agalsidase beta in patients with Fabry disease and deteriorating renal function: results from the 2-year randomised phase III BALANCE study. *Journal of Medical Genetics*, *61*(6), 520–530. https://doi.org/10.1136/jmg-2023-109445

Walsh, G. (2018). Biopharmaceutical benchmarks 2018. *Nature Biotechnology*, *36*(12), 1136–1145. https://doi.org/10.1038/nbt.4305

Walsh, G., & Walsh, E. (2022). Biopharmaceutical benchmarks 2022. *Nature Biotechnology*, *40*(12), 1722–1760. https://doi.org/10.1038/s41587-022-01582-x

Wandall, H. H., Irazoqui, F., Tarp, M. A., Bennett, E. P., Mandel, U., Takeuchi, H., Kato, K., Irimura, T., Suryanarayanan, G., Hollingsworth, M. A., & Clausen, H. (2007). The lectin domains of polypeptide GalNAc-transferases exhibit carbohydrate-binding specificity for GalNAc: lectin binding to GalNAc-

glycopeptide substrates is required for high density GalNAc-O-glycosylation. *Glycobiology*, *17*(4), 374–387. https://doi.org/10.1093/glycob/cwl082

Wang, B., Xiao, S., Zhao, X., Zhao, L., Zhang, Y., Cheng, J., & Zhang, J. (2023). Recent Advances in the Hydroxylation of Amino Acids and Its Derivatives. In *Fermentation* (Vol. 9, Issue 3). MDPI. https://doi.org/10.3390/fermentation9030285

Wang, X., Karki, U., Abeygunaratne, H., UnnoldCofre, C., & Xu, J. (2021). Plant cell-secreted stem cell factor stimulates expansion and differentiation of hematopoietic stem cells. *Process Biochemistry*, *100*, 39–48. https://doi.org/10.1016/j.procbio.2020.09.029

Ward, B. J., Landry, N., Trépanier, S., Mercier, G., Dargis, M., Couture, M., D'Aoust, M.-A., & Vézina, L.-P. (2014). Human antibody response to N-glycans present on plant-made influenza virus-like particle (VLP) vaccines. *Vaccine*, *32*(46), 6098–6106. https://doi.org/10.1016/j.vaccine.2014.08.079

Ward, B. J., Makarkov, A., Séguin, A., Pillet, S., Trépanier, S., Dhaliwall, J., Libman, M. D., Vesikari, T., & Landry, N. (2020). Efficacy, immunogenicity, and safety of a plant-derived, quadrivalent, virus-like particle influenza vaccine in adults (18-64 years) and older adults (≥65 years): two multicentre, randomised phase 3 trials. *Lancet (London, England)*, *396*(10261), 1491–1503. https://doi.org/10.1016/S0140-6736(20)32014-6

Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., & Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. *Science (New York, N.Y.)*, *369*(6501), 330–333. https://doi.org/10.1126/science.abb9983

Weise, A., Altmann, F., Rodriguez-Franco, M., Sjoberg, E. R., Bäumer, W., Launhardt, H., Kietzmann, M., & Gorr, G. (2007). High-level expression of secreted complex glycosylated recombinant human erythropoietin in the *Physcomitrella ⊿-fuc-t ⊿-xyl-t* mutant. *Plant Biotechnology Journal*, *5*(3), 389–401. https://doi.org/10.1111/j.1467-7652.2007.00248.x

Wold, J. K., & Hillestad, A. (1976). The demonstration of galactosamine in a higher plant: Cannabis sativa. *Phytochemistry*, *15*(2), 325–326. https://doi.org/10.1016/S0031-9422(00)89015-9

Wu, Y., Williams, M., Bernard, S., Driouich, A., Showalter, A. M., & Faik, A. (2010). Functional Identification of Two Nonredundant Arabidopsis

α(1,2)Fucosyltransferases Specific to Arabinogalactan Proteins. *Journal of Biological Chemistry*, *285*(18), 13638–13645. https://doi.org/10.1074/jbc.M110.102715

Xu, J., Okada, S., Tan, L., Goodrum, K. J., Kopchick, J. J., & Kieliszewski, M. J. (2010). Human growth hormone expressed in tobacco cells as an arabinogalactan-protein fusion glycoprotein has a prolonged serum life. *Transgenic Research*, *19*(5), 849–867. https://doi.org/10.1007/s11248-010-9367-8

Xu, J., Tan, L., Goodrum, K. J., & Kieliszewski, M. J. (2007). High-yields and extended serum half-life of human interferon α2b expressed in tobacco cells as arabinogalactan-protein fusions. *Biotechnology and Bioengineering*, *97*(5), 997–1008. https://doi.org/10.1002/bit.21407

Xu, J., Tan, L., Lamport, D. T. A., Showalter, A. M., & Kieliszewski, M. J. (2008). The O-Hyp glycosylation code in tobacco and Arabidopsis and a proposed role of Hyp-glycans in secretion. *Phytochemistry*, *69*(8), 1631–1640. https://doi.org/10.1016/j.phytochem.2008.02.006

Xue, J., Zhu, L.-P., & Wei, Q. (2013). IgG-Fc N-glycosylation at Asn297 and IgA O-glycosylation in the hinge region in health and disease. *Glycoconjugate Journal*, *30*(8), 735–745. https://doi.org/10.1007/s10719-013-9481-y

Yang, Q., Hughes, T. A., Kelkar, A., Yu, X., Cheng, K., Park, S., Huang, W.-C., Lovell, J. F., & Neelamegham, S. (2020). Inhibition of SARS-CoV-2 viral entry upon blocking N- and O-glycan elaboration. *ELife*, *9*. https://doi.org/10.7554/eLife.61552

Yang, S., Onigman, P., Wu, W. W., Sjogren, J., Nyhlen, H., Shen, R.-F., & Cipollo, J. (2018). Deciphering Protein O-Glycosylation: Solid-Phase Chemoenzymatic Cleavage and Enrichment. *Analytical Chemistry*, *90*(13), 8261–8269. https://doi.org/10.1021/acs.analchem.8b01834

Yang, S., Wu, W. W., Shen, R., Sjogren, J., Parsons, L., & Cipollo, J. F. (2020). Optimization of O-GIG for O-Glycopeptide Characterization with Sialic Acid Linkage Determination. *Analytical Chemistry*, *92*(16), 10946–10951. https://doi.org/10.1021/acs.analchem.0c01346

Yang, W., Ao, M., Hu, Y., Li, Q. K., & Zhang, H. (2018). Mapping the O-glycoproteome using site-specific extraction of O-linked glycopeptides

(EXoO). *Molecular Systems Biology*, *14*(11). https://doi.org/10.15252/msb.20188486

Yang, Z., Bennett, E. P., Jørgensen, B., Drew, D. P., Arigi, E., Mandel, U., Ulvskov, P., Levery, S. B., Clausen, H., & Petersen, B. L. (2012). Toward Stable Genetic Engineering of Human O-Glycosylation in Plants. *Plant Physiology*, *160*(1), 450–463. https://doi.org/10.1104/pp.112.198200

Yang, Z., Drew, D. P., Jørgensen, B., Mandel, U., Bach, S. S., Ulvskov, P., Levery, S. B., Bennett, E. P., Clausen, H., & Petersen, B. L. (2012). Engineering Mammalian Mucin-type O-Glycosylation in Plants. *Journal of Biological Chemistry*, *287*(15), 11911–11923. https://doi.org/10.1074/jbc.M111.312918

Yang, Z., Halim, A., Narimatsu, Y., Jitendra Joshi, H., Steentoft, C., Schjoldager, K. T., Alder Schulz, M., Sealover, N. R., Kayser, K. J., Paul Bennett, E., Levery, S. B., Vakhrushev, S. Y., & Clausen, H. (2014). The GalNAc-type O-Glycoproteome of CHO Cells Characterized by the SimpleCell Strategy. *Molecular & Cellular Proteomics*, *13*(12), 3224–3235. https://doi.org/10.1074/mcp.M114.041541

Yeh, J.-C., Ong, E., & Fukuda, M. (1999). Molecular Cloning and Expression of a Novel β-1,6-N-Acetylglucosaminyltransferase That Forms Core 2, Core 4, and I Branches. *Journal of Biological Chemistry*, *274*(5), 3215–3221. https://doi.org/10.1074/jbc.274.5.3215

Yu, C., Liang, L., & Yin, Y. (2019). Structural basis of carbohydrate transfer activity of UDP-GalNAc: Polypeptide N-acetylgalactosaminyltransferase 7. *Biochemical and Biophysical Research Communications*, *510*(2), 266–271. https://doi.org/10.1016/j.bbrc.2019.01.084

Yuasa, K., Toyooka, K., Fukuda, H., & Matsuoka, K. (2005). Membrane-anchored prolyl hydroxylase with an export signal from the endoplasmic reticulum. *The Plant Journal*, *41*(1), 81–94. https://doi.org/10.1111/j.1365-313X.2004.02279.x

Zahmanova, G., Aljabali, A. A., Takova, K., Toneva, V., Tambuwala, M. M., Andonov, A. P., Lukov, G. L., & Minkov, I. (2023). The Plant Viruses and Molecular Farming: How Beneficial They Might Be for Human and Animal Health? *International Journal of Molecular Sciences*, *24*(2), 1533. https://doi.org/10.3390/ijms24021533

Zhang, N., Dolan, M., Wu, D., Phillips, G. C., & Xu, J. (2016). Dramatic secretion of recombinant protein expressed in tobacco cells with a designer glycopeptide

tag is highly impacted by medium composition. *Plant Cell Reports*, *35*(12), 2513–2522. https://doi.org/10.1007/s00299-016-2051-6

Zhang, N., Gonzalez, M., Savary, B., & Xu, J. (2016). High-yield secretion of recombinant proteins expressed in tobacco cell culture with a designer glycopeptide tag: Process development. *Biotechnology Journal*, *11*(4), 497–506. https://doi.org/10.1002/biot.201500377

Zhang, N., Wright, T., Caraway, P., & Xu, J. (2019). Enhanced secretion of human α1-antitrypsin expressed with a novel glycosylation module in tobacco BY-2 cell culture. *Bioengineered*, *10*(1), 87–97. https://doi.org/10.1080/21655979.2019.1604037

Zhang, N., Wright, T., Wang, X., Karki, U., Savary, B. J., & Xu, J. (2019). Engineering 'designer' glycomodules for boosting recombinant protein secretion in tobacco hairy root culture and studying hydroxyproline- *O* -glycosylation process in plants. *Plant Biotechnology Journal*, *17*(6), 1130–1141. https://doi.org/10.1111/pbi.13043

Zhang, P., Wang, T., Bardor, M., & Song, Z. (2013). Deciphering O-glycomics for the development and production of biopharmaceuticals. *Pharmaceutical Bioprocessing*, *1*(1), 89–104. https://doi.org/10.4155/pbp.13.7

Zhang, Y., Held, M. A., Kaur, D., & Showalter, A. M. (2021). CRISPR-Cas9 multiplex genome editing of the hydroxyproline-O-galactosyltransferase gene family alters arabinogalactan-protein glycosylation and function in Arabidopsis. *BMC Plant Biology*, *21*(1), 16. https://doi.org/10.1186/s12870-020-02791-9

Zhang, Y., Held, M. A., & Showalter, A. M. (2020). Elucidating the roles of three β-glucuronosyltransferases (GLCATs) acting on arabinogalactan-proteins using a CRISPR-Cas9 multiplexing approach in Arabidopsis. *BMC Plant Biology*, *20*(1), 221. https://doi.org/10.1186/s12870-020-02420-5

Zhang, Y., & Showalter, A. M. (2020). CRISPR/Cas9 Genome Editing Technology: A Valuable Tool for Understanding Plant Cell Wall Biosynthesis and Function. *Frontiers in Plant Science*, *11*. https://doi.org/10.3389/fpls.2020.589517

Zhang, Y., Zhao, W., Mao, Y., Chen, Y., Zheng, S., Cao, W., Zhu, J., Hu, L., Gong, M., Cheng, J., & Yang, H. (2021). O-Glycosylation Landscapes of SARS-CoV-2 Spike Proteins. *Frontiers in Chemistry*, *9*. https://doi.org/10.3389/fchem.2021.689521

Zhao, Z. D., Tan, L., Showalter, A. M., Lamport, D. T. A., & Kieliszewski, M. J. (2002). Tomato LeAGP-1 arabinogalactan-protein purified from transgenic tobacco corroborates the Hyp contiguity hypothesis. *Plant Journal*, *31*(4), 431–444. https://doi.org/10.1046/j.1365-313X.2002.01365.x

Zlocowski, N., Grupe, V., Garay, Y. C., Nores, G. A., Lardone, R. D., & Irazoqui, F. J. (2019). Purified human anti-Tn and anti-T antibodies specifically recognize carcinoma tissues. *Scientific Reports*, *9*(1), 8097. https://doi.org/10.1038/s41598-019-44601-9