Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

School of Engineering and Sciences

Engineering mammalian-specific post-translational modifications in plant-derived proteins: phosphorylation and Mucin-type *O*-glycosylation as a challenge.

A dissertation presented by

Israel Alfonso Ramírez Alanis

Submitted to the
School of Engineering and Sciences
in partial fulfillment of the requirements for the degree of

Doctor

In

Biotechnology

Monterrey Nuevo León, December 14th, 2018

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

School of Engineering and Sciences

The committee members, hereby, certify that they have read the dissertation presented by Israel Alfonso Ramírez Alanis and that it is fully adequate in scope and quality as a partial requirement for the degree of Doctor in Biotechnology

Dr. Silverio García-Lara Tecnológico de Monterrey School of Engineering and Sciences Advisor	Dr. Guy Albert Cardineau Tecnológico de Monterrey Arizona State University Advisor
Dr. Sergio Serna Saldívar Tecnológico de Monterrey School of Engineering and Sciences Committee Member	Dr. Sean Patrick Scott Tecnológico de Monterrey School of Medicine and Health Sciences Committee Member
Dr. Sasha M. Daskalova Arizona State University The Biodesign Institute External Committee Member	

Dr. Rubén Morales Menéndez
Dean of Graduate Studies
School of Engineering and Sciences

Tecnológico de Monterrey

04 DIC 2018

Dirección Nacional de Posgrado
Dr. Rubén Morales Menéndez

Monterrey, Nuevo León, December 14th, 2018

Declaration of Authorship

I, **Israel Alfonso Ramírez Alanis**, declare that this dissertation titled, Engineering mammalian-specific post-translational modifications in plant-derived proteins: phosphorylation and Mucin-type *O*-glycosylation as a challenge, and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Israel Alfonso Ramírez Alanis

Monterrey Nuevo León, December 14th, 2018

Dedication

I thank my family, my closest friends, my advisors and all of those who helped me through this period of time without even knowing me.

Thank you all for trusting me and believing in me.

Acknowledgements

I would also like to thank Tecnológico de Monterrey for the support on tuition and CONACyT with the support for living.

Engineering mammalian-specific post-translational modifications in plant-derived proteins: phosphorylation and Mucin-type *O*-glycosylation as a challenge.

by

Israel Alfonso Ramírez Alanis

Abstract

Expression of economically relevant plant-derived recombinant proteins in alternative expression platforms, especially plant expression platforms, has gained significant interest in recent years, due to the possibility to reduce production costs, or because of product quality of production. Among the different qualities that plants can offer for the production of recombinant proteins, capability to perform post-translational modifications like protein glycosylation and phosphorylation are some of the crucial ones since it has an impact on pharmaceuticals functionality and/or stability or protein activity, respectively. In this dissertation, the pharmaceutical glycoprotein human Granulocyte-Colony Stimulating Factor is transiently expressed in *N. benthamiana*, as several protein versions targeted to different compartments (apoplast, cytoplasm and as protein bodies), offering an alternative for the consideration of production of this protein. Furthermore, the glycoprotein was subjected to the native GalNAc-*O*-glycosylation, by co-expressing the pharmaceutical, together with the enzymes responsible for such glycosylation. In the case of phosphoproteins, the bovine β- and κ-caseins and their specific kinase the bovine Fam20C were also expressed for the first time in *N. benthamiana* plants, to assess the feasibility of controlling their phosphorylation pattern, which could be considered for the generation of soybean transgenic lines, enriched with such nutraceutical and nutrimental phosphoproteins.

List of Figures

List of Tables

Contents

Chapter 1

• Introduction

The sequencing of the human genome has allowed the identification and characterization of new genes and their respective proteins relevant for the study of pathologies and development of new medical treatments. Since then, it has been better understood how a specific phenotype can be the result of the expression of a specific protein, its interaction with other proteins, and the regulation in the expression of these molecules. Moreover, achievements of several groups, have made possible the understanding of the dynamics of the central dogma of life, in which a single gene not only can be translated into a single protein, but this gene can give rise to a variety of mRNA´s, which can give rise to a protein that can later on be modified and thus converted into different molecules with possibly, very different functions. The versatility of the expression of a single gene and the modifications that the coding nucleotide and amino acid sequences undergo during the production and final targeting of the protein determines the tertiary and quaternary structures and has a huge impact on the final structure and functionality of a protein, and thus the phenotype.

Among the different modifications that a protein can undergo after translation, phosphorylation is the most studied one, since it is involved in the reversible activation or deactivation of a protein, by altering protein folding and ultimately mediate or inhibit interaction with other proteins and protein signaling (Reimand, Wagih, & Bader, 2013). On the other hand, the attachment of sugar residues to an amino acid backbone, called glycosylation, is probably the most complex post-translational modification (PTM). Overall, glycosylation of proteins is an event that is usually driven at the moment when the amino acid chain is being displayed out of the ribosome, and/or along the translocation of the protein to its final destination. Nevertheless, the pattern of glycosylation that a protein undergoes depends on multiple factors, such as location of amino acid consensus and the protein itself, cell type, developmental stage, and organism (Spiro, 2002a). The most crucial event in the attachment of a carbohydrate onto a protein is the biosynthesis of the sugar-amino acid bond, which will determine the future aggregation of more sugar units, and further on can influence the protein´s biological activity (Spiro, 2002a).

As mentioned before, phosphorylation or glycosylation can alter the physicochemical structure and/or functionality of a protein. In general, it can be stated that phosphorylation dictates protein activation or deactivation and interaction with other proteins, while protein glycosylation can influence the stability of the tertiary structure and thus quaternary structure of a protein. Glycosylation can therefore modulate protein aggregation, resistance against degradation by proteases or denaturing agents like heat, modulate ligand-receptor recognition, protein clearance, and catalytic domain (Van den Steen, Rudd, Dwek, & Opdenakker, 1998).

All this information is relevant for the development of the so called Plant-Made Pharmaceuticals (PMP´s), recombinant proteins produced in transgenic plants for health purposes, with a more native structure and thus greater presumed efficacy when used by patients or consumers. This field of plant biotechnology is commonly referred to as plant molecular farming (Coku, 2007; Fischer & Emans, 2000; Twyman, Schillberg, & Fischer, 2012; Twyman, Stoger, Schillberg, Christou, & Fischer, 2003).

Plant molecular farming has emerged as an alternative expression system for the production of foreign proteins of economic relevance using plant platforms (Daniell, Singh, Mason, & Streatfield, 2009; Faye, Boulaflous, Benchabane, Gomord, & Michaud, 2005; Ko & Koprowski, 2005; Sharma & Sharma, 2009). Currently, there is a variety of products being produced in plants using genetic engineering approaches. Bioactive peptides, vaccine antigens, antibodies, diagnostic proteins, nutritional supplements, enzymes and biodegradable plastics, can be found among the different plant-made products (Sharma & Sharma, 2009).

There are several advantages to using plants as expression systems, such as low cost of production and maintenance, fast scalability, biological safety, proper protein folding and assembly (Daniell et al., 2009; Faye et al., 2005; Sharma & Sharma, 2009). However, there are also some limitations concerning the expectance of bioactivity of the produced proteins, due to post-translational modifications (Daskalova et al., 2010). Although both plants and mammalian cells are known to perform PTM´s there are differences in patterns between sorts of cells, such as phosphorylation and glycosylation. This might represent a major drawback that can impact the biochemical properties of the plant-derived recombinant protein (Bardor et al., 2003; Daskalova et al., 2010; Faye et al., 2005; Gomord et al., 2010a; Jin et al., 2008; Yang, Drew, et al., 2012).

As examples of such scenario mentioned above, we can find the case of casein proteins, like the bovine β-casein (βC), a phosphoprotein with nutrimental and nutraceutical relevance. This

protein has been produced in potato and soybean (Chong et al., 1997; Maughan, Philip, Cho, Widholm, & Vodkin, 1999; Philip, Darnowski, Maughan, & Vodkin, 2001) however the protein expressed in these plants was not phosphorylated. In this case, the lack of phosphorylation affects the binding property to Ca^{2+}, thus preventing the production of casein micelles and later products like cheese, decreasing the Ca^{2+} uptake from the diet. Furthermore, the phosphorylated free casein can cause allergenic responses in susceptible consumers (Bernard, Meisel, Creminon, & Wal, 2000; Cases et al., 2011; Johansson, 2002; Tezcucano Molina, Alli, Konishi, & Kermasha, 2007).

The κ-casein (κC), another type of casein that interacts directly with βC stabilizing the micelles (Kethireddipalli & Hill, 2015), represents another example of phosphoprotein, which just like the βC is a source of bioactive peptides with antioxidant, antithrombotic, antimicrobial, antihypertensive, and immunomodulating activity (Silva & Malcata, 2005). Phosphate groups in this protein might not only impact its nutrimental and nutraceutical properties, but they might also compromise the stability of the micelle, thus affecting the performance and sensorial quality of the products containing β- and κC´s.

In the case of glycosylation, an example of glycoprotein is the pharmaceutical protein human Granulocyte-Colony Stimulating Factor (hG-CSF). This Mucin-type *O*-glycosylated protein is a commonly used pharmaceutical to treat patients suffering neutropenia (Kaushansky, Lin, & Adamson, 1988). The product is commercially available using the traditionally well-established bacterial (*Escherichia coli*) and mammalian cell culture (Chinese Hamster Ovary or CHO cells). However, due to production costs and value of this pharmaceutical protein, alternative platforms have been explored for its production. Among these platforms, several plant systems have been used. Nevertheless, glycosylation approaches have not been explored, and it is known now that plants, as with other expression systems, such as bacterial, yeast or insect cells, do not possess the glycosylation machinery, required to imitate the human glycosylation pattern.

1.1 Hypothesis

The development of plant expression systems equipped with the heterologous genes involved in the synthesis of mammalian-specific phosphorylation and Mucin-type *O*-glycosylation, would allow the production of plant-derived recombinant proteins with mammalian-specific post-translational modifications (**Figure 1-1**).

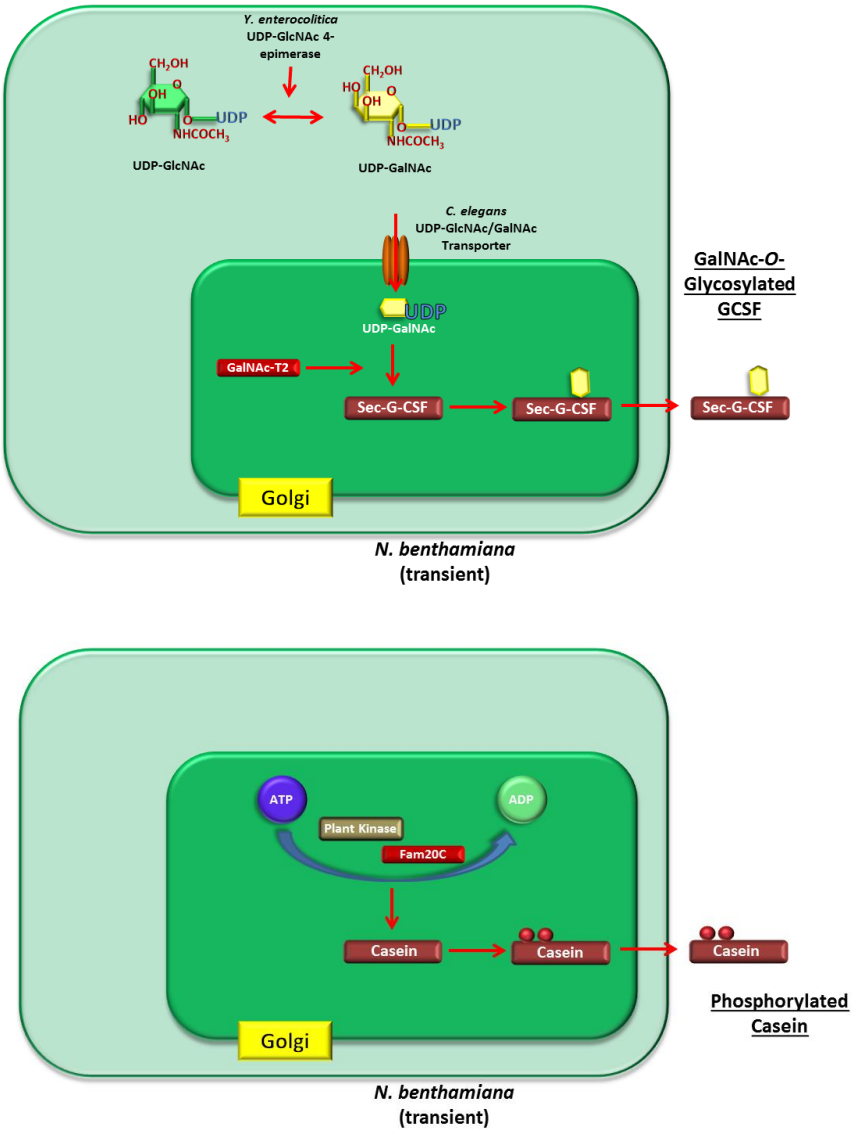


Figure 1-1 Graphical representation of the hypothesis, presented in this work

Upper diagram: Co-expression of the human Granulocyte-Colony Stimulating Factor together with a Mucin-type *O*-glycosylation machinery, would allow the production of plant-derived mammalian-specific glycosylated hG-CSF. Lower diagram: Expression of Casein proteins targeted to the secretory pathway and co-secretion of plant protein kinases or co-expression of a putative casein-specific kinase (Fam20C) would allow the production of plant-derived phosphorylated casein.

1.2 Objective

The main objective of the present dissertation was to engineer mammalian-specific PTM´s, of the nutrimental/nutraceutical β- and κ-C´s and the pharmaceutical hG-CSF, using *Nicotiana*

benthamiana as a plant expression platform. The post-translationally modified plant-derived recombinant proteins would possess the mammalian-like phosphorylation pattern and Mucin-type *O*-glycosylation.

1.2.1 Specific Objectives

1. Design and generate binary expression vectors, with the corresponding genes to be expressed in *N. benthamiana.*
2. Determine cellular localization of the expressed proteins.
3. Evaluate accumulation of the model proteins.
4. Characterization of PTM´s.

1.3 Justification

Immunogenicity or reduced product efficacy of plant-derived pharmaceuticals or nutraceuticals as a result of undesired PTM´s may result in a potential risk for the usage or consumption by the patients or consumers. The absence of PTM´s might also hinder the bioactivity of plant-derived proteins. Therefore, the lack of control of such protein PTM´s in plant expression systems hampers the broad usage of plants as bioreactors for the production of economically relevant proteins for pharmaceutical or nutrimental purposes.

So far, no efforts have been made to modulate the production of phosphorylated casein in plant expression systems, which would probably require the co-expression and colocalization of a casein-specific kinase. Similarly, no attempts have been explored for the production of the economically relevant glycoprotein hG-CSF, which would allow the human-specific Mucin-type *O*-glycosylation of such pharmaceutical, using a plant expression platform. Both cases represent an area of opportunity where modulation of specific PTM would demonstrate the feasibility of using plant expression platforms for the production of complex proteins, whose PTM´s denote an impact on the activity and stability of the plant-derived recombinant protein.

Chapter 2

- **THEORETICAL FRAMEWORK POST-TRANSLATIONAL MODIFICATIONS**

Protein PTM´s refer to those covalent and generally enzymatically modifications of proteins during or after their biosynthesis. These modifications are independent from the gene sequence, but are able to impact protein functions and activity by causing changes in structure and dynamic interactions with other proteins (Seo & Lee, 2004).

2.1 Protein glycosylation

Protein glycosylation is a type of PTM that proteins undergo after or during protein synthesis, in which a carbohydrate residue is attached to amino acid residues (Corfield, 2016; Spiro, 2002a). As a kind of PTM, glycosylation is independent from the gene sequence, but may possess an impact on protein functions and activity (Seo & Lee, 2004). Such impact is driven by the structural changes or dynamic interactions that the modified protein may suffer due to the attachment of a carbohydrate group to its backbone (Spiro, 2002a).

Glycosylation is probably the most complex PTM. It is orchestrated by several elements: accessibility of the protein amino acid consensus sequence, subcellular localization, availability of the carbohydrate substrate, enzymes responsible for such modification. All these elements will determine the glycosylation event. The latter is also highly dependent on the cell type, developmental stage, and organism where the protein is being produced (Spiro, 2002a).

It is well understood that glycosylation imposes an impact on the physico-chemical properties of a protein, therefore affecting thermal denaturation, proteolytic degradation, solubility, and biological functions, like ligand-receptor interaction, immunogenicity and clearance rate. All of these impacts on protein properties, explains why glycosylation is one of the most commonly studied PTM´s (Seo & Lee, 2004).

2.1.1 Types of glycosylation

Probably the most crucial event during protein glycosylation is the biosynthesis of the sugar-amino acid bond. This event determines the further elongation or maturation of the glycosylation

site, by remodeling the carbohydrate structure, being the final structure responsible for the protein biological identity and activity (Spiro, 2002a). According to the sugar-amino acid bond, glycosylation can be classified into several groups, however, among all the different kinds of glycosylation, *N*- and *O*-glycosylation are the most abundant ones, occurring on secreted or membrane-bound proteins, and therefore will be described here (**Figure 2-1 and Table 2-1**) (Tran & Ten Hagen, 2013)**.**

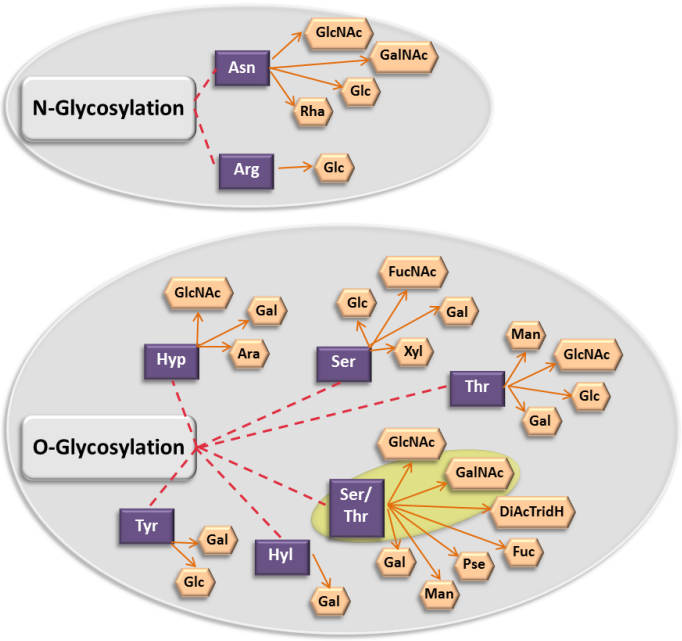


Figure 2-1 Types of sugar-peptide bonds

Representation of the most distinctive types of sugar-peptide linkages. Abbreviations: GlcNAc: *N*-Acetylglucosamine; Glc: Glucose; GalNAc: *N*-Acetylgalactosamine; Rha: Rhamnose; Gal: Galactose; Man: Mannose; Fuc: Fucose; Pse: Pseudaminic acid; DiAcTridH: 2,4,-diacetamido, 2,4,6-tridehoxyhexose; FucNAc: *N*-Acetylfucosamine; Xyl: Xylose; Ara: Arabinose. Mucin-type *O*-glycosylation is highlighted. Modified from Spiro et al 2002 (Spiro, 2002a).

Table 2-1 Kinds of glycosylation: Nature and distribution of sugar-peptide linkages

Modified from Spiro et al 2002 (Spiro, 2002a)**.**

Type of bond	Linkage			Phylogenetic distribution			
	Amino acid	Sugar	Configuration	Eukaryotes	Archaea[a]	Bacteria[a]	Examples
N-glycosyl	Asn	GlcNAc	β	+	+	+	Ovalbumin, fetuin, insulin receptor
	Asn	Glc	β	+	+	–	Laminin, *H. halobium* S-layer
	Asn	GalNAc	*	–	+	–	*H. halobium* S-layer
	Asn	Rha	*	–	–	+	*S. sanguis* cell wall
	Arg	Glc	β	+	–	–	Sweet corn amylogenin
O-glycosyl	Ser/Thr	GalNAc	α	+	–	–	Mucins, fetuin, glycophorin
	Ser/Thr	GalNAc	β	–	+	–	*A. thermoaerophilus* S-layer
	Ser/Thr	GlcNAc	β	+	–	–	Nuclear and cytoplasmic proteins
	Ser/Thr	Gal	α	+	–	+	Earthworm collagen, *B. cellulosoleum*
	Ser/Thr	Man	α	+	–	–	Yeast mannoproteins
	Ser/Thr	Man	*	+	–	+	α-dystroglycan, *F. meningosepticum*
	Ser/Thr	Fuc	α	+	–	–	Coagulation and fibrinolytic factors
	Ser/Thr	Pse[b]	α	–	–	+	*C. jejuni* flagellins
	Ser/Thr	DiActrideoxyhexose[c]	*	–	–	+	*N. meningitidis* pili
	Ser	Glc	β	+	–	–	Coagulation factors
	Ser	FucNAc	β	–	–	+	*P. aeruginosa* pili
	Ser	Xyl	β	+	–	–	Proteoglycans
	Ser	Gal	α	+	–	–	Cell walls of plants
	Thr	Man	α	–	–	+	*M. tuberculosis* secreted glycoproteins
	Thr	Man	*	+	–	–	Clamworm collagen
	Thr	GlcNAc	α	+	–	–	Dictyostelium, *T. cruzi*
	Thr	GlcNAc	*	+	–	–	Rho proteins (GTPases)
	Thr	Glc	*	+	–	–	Rho proteins (GTPases)
	Thr	Gal	*	+	+	–	*H. halobium* S-layer, vent worm collagen
	Hyl[d]	Gal	β	+	–	–	Collagen, C1q complement, core specific lectin
	Hyp[d]	Ara	α	+	–	–	Plant cell walls
	Hyp	Ara	β	+	–	–	Potato lectin
	Hyp	Gal	β	+	–	–	Wheat endosperm
	Hyp	GlcNAc	*	+	–	–	*Dictyostelium* cytoplasmic proteins
	Tyr	Glc	α	+	–	–	Muscle and liver glycogenin
	Tyr	Glc	β	–	–	+	*C. thermohydrosulfuricum* S-layer
	Tyr	Gal	β	–	–	+	*T. thermohydrosulfuricus* S-layer

* Configuration has not been established yet.

[a] Also known as archaebacteria and eubacteria, respectively.

[b] Pse refers to pseudaminic acid (5,7-diacetamido-3,5,7,9-tetradeoxy-L-glycero-L-manno-nonulosinic acid).

[c] DiActrideoxyhexose refers to 2,4-diacetamido-2,4,6-trideoxyhexose.

[d] The abbreviations Hyl and Hyp refer to hydroxylysine and hydroxyproline, respectively.

2.1.1.1 N-Glycosylation

The attachment of carbohydrates to the Asn amino group is defined as *N*-glycosylation. This type of glycosylation is the most studied and best characterized kind of glycosylation, and is also one of the most widely distributed with the GlcNAc-β-Asn linkage and its complex carbohydrate branches attached to proteins, the most representative sugar-amino acid bond of this glycosylation (Spiro, 2002a; Varki, 1993)**.**

In eukaryotes, initiation of the GlcNAc-β-Asn bond occurs in the Endoplasmic Reticulum (ER), where a preassembled dolichol-linked triglucosylate polymannose oligosaccharide is attached to the target protein in the lumen of the ER by the oligosaccharyltransferase (OST) complex (Burda & Aebi, 1999). In this kind of glycosylation, a consensus of amino acid has been identified to be Ans-X-Ser/Thr (**Table 2-3**) (Corfield, 2016). The enzyme responsible for the transfer of the oligosaccharide in the ER, has been found in several organisms, where *N*-glycosylation occurs (**Table 2-4**) (Spiro, 2002a)**.** Subsequent processing of the transferred *N*-glycan takes place in the lumen of the ER, where glucosidases remove terminal and penultimate glucose residues. Further maturation steps allow remodelling the *N*-glycan, which starts in the lumen of the ER along the secretory pathway in the Golgi apparatus, by a series of glycosidases and glycosyltransferases. During the first maturation steps, the precursor *N*-glycan is deglucosylated by the α-glucosidases I and II, then reglucosylated by an UDP-glucose: glycoprotein glucosyltransferase (UGGT). Further translocation to the Golgi apparatus allows removal of mannose residues. *N*-acetylglucosaminyltransferase I (GnT I) then transfers a first *N*-acetylglucosaminyl (GlcNAc) residue and initiates the synthesis of a large variety of structurally different compley-type *N*-glycans. Typically, further maturation steps involve the removal of mannosyl residues and transfer of a second terminal GlcNAc residue. The resulted *N*-glycan is then typically decorated with α(1,6)-fucose and terminal sialic acids (Dicker, Schoberer, Vavra, & Strasser, 2015; Mathieu-Rivet et al., 2014; Vukušić, Šikić, & Balen, 2016).

Table 2-2 Synthesis of glycopeptide bonds. (Spiro, 2002b)

Linkage[a]	Enzyme[b]	Glycosyl donor	Location	Source[c]	Cloned
GlcNAc-β-Asn	Oligo[d]-tr	Dol-PP-Oligo	ER	Liver, pancreas oviduct, yeast	Yes (multiple subunits)
GalNAc-α-Ser/Thr	GalNAc-tr	UDP-GalNAc	Golgi	Colostrum, submaxillary gland	Yes (multiple enzymes)
GlcNAc-α-Thr	GlcNAc-tr	UDP-GlcNAc	Golgi	Trypanosomes, *Dictyostelium*	No
GlcNAc-β-Ser/Thr	GlcNAc-tr	UDP-GlcNAc	Cytosol, nucleus	Liver, blood	Yes
Man-α-Ser/Thr	Man-tr	Dol-P-Man	ER	Yeast	Yes (multiple enzymes)
Fuc-α-Ser/Thr	Fuc-tr	GDP-Fuc	Golgi	CHO cells, liver	Yes
Xyl-β-Ser	Xyl-tr	UDP-Xyl	ER, Golgi	Cartilage, choriocarcinoma	Yes
Glc-Thr	Clostridial cytotoxin	UDP-Glc	Cytosol	*C. difficile* and *sordelli*	Yes
GlcNAc-Thr	Clostridial cytotoxin	UDP-GlcNAc	Cytosol	*C. novyi*	Yes
Gal-β-Hyl	Gal-tr	UDP-Gal	Golgi	Kidney, cartilage	No
GlcNAc-Hyp	GlcNAc-tr	UDP-GlcNAc	Cytosol	*Dictyostelium*	Yes
Glc-α-Tyr	Glycogenin[e]	UDP-Glc	Cytosol	Liver, muscle	Yes

[a] See Table I for nature of the bonds.

[b] *tr* refers to saccharide:polypeptide transferase.

[c] Tissues or cells from which enzyme has been examined after varying degrees of purification.

[d] Abbreviations are Oligo, $Glc_3Man_9GlcNAc_2$; GPI, glycosylphosphatidylinositol.

[e] Autoglucosylation.

2.1.1.2 O-glycosylation

In the case of ***O*-glycosylation**, the sugars are attached to amino acids that contain a hydroxyl group. This kind of glycosylation not only occurs in a great variety of proteins, but also in a broader anomeric configuration, in comparison to *N*-glycosylation (**Figure 2-1, Table 2-1**). *O*-glycosidic bonds have been found in all amino acids with a hydroxyl group, such as Ser, Thr, Tyr, Hyp (hydroxyproline) and Hyl (hydroxylysine), without a specifically defined amino acid consensus (**Figure 2-1, Table 2-3**) (Corfield, 2016; Spiro, 2002a; Van den Steen et al., 1998).

In this case, the **GalNAc-α-Ser/Thr** is perhaps the most popular *O*-carbohydrate bond and the most widely identified *O*-glycosylation in mammalian proteins such as fetuin, human gonadotropins, glycophorin, and antifreeze glycoproteins (**Table 2-1**). The **GalNAc-α-Ser/Thr** *O*-glycosylation is usually called **Mucin-type *O*-glycosylation** (**Figure 2-1, Figure 2-2**), since it has been related to the glycosylation pattern present in mucins, where it occurs as clusters (**Table 2-1**) (E. Tian & Hagen, 2009; Tran & Ten Hagen, 2013).

2.1.1.2.1 Mucin-type *O*-glycosylation

Mucins are eukaryotic proteins of high molecular weight, they are produced in abundance by epithelial and goblet cells, where the MUC genes codify for the different polypeptides of this family (Dell, Galadari, Sastre, & Hitchen, 2010). The main characteristic of these proteins is the presence of tandem repeats of Ser-Thr-Pro rich sequences, which are highly *O*-glycosylated. Mucins are responsible for forming gel-like structures, and are key components of most gel-like secretions in eukaryotes, having a function as lubrication or receptor for microorganisms (Dell et al., 2010).

The GalNAc-α-Ser/Thr linkage has been extensively studied in eukaryotic cells, due to the relevance of these proteins where it has been found. More than nine GalNAc-transferases, responsible for this carbohydrate bond have been identified. Because of the way in which the mucin-type *O*-glycosylation is synthesized, it is suggested that these glycosyltransferases act in a hierarchical manner in order to create such linkages (E. Tian & Hagen, 2009)**.** Although some GalNAc-transferases have been cloned, there is no amino acid consensus so far established. In general, it appears that the GalNAc-α-Ser/Thr (Mucin-type) bond is found in clusters of Ser/Thr residues with a β-turn near Pro and at a distance from charged amino acids (**Table 2-3**). Immunoelectron microscopic studies have indicated that GalNAc attachment to Ser/Thr occurs in the cis-Golgi. However, due to the multiple isoforms, it is feasible that some GalNAc-transferases act in the ER or pre-Golgi compartments. It is also possible that not all GalNAc-transferases are expressed in every cell type or species, or even at the same developmental stage (Spiro, 2002b; Van den Steen et al., 1998).

Table 2-3 Consensus sequences. Modified from Spiro et al 2002 (Spiro, 2002a).

Glycopeptide bond	Consensus sequence or peptide domain
GlcNAc-β-Asn	*Asn*-X-Ser/Thr (X = any amino acid except Pro)
Glc-β-Asn	*Asn*-X-Ser/Thr
GalNAc-α-Ser/Thr	Repeat domains rich in Ser, Thr, Pro, Gly, Ala in no special sequence
GlcNAc-α-Thr	Thr rich domain near Pro residues
GlcNAc-β-Ser/Thr	Ser/Thr rich domains near Pro, Val, Ala, Gly
Man-α-Ser/Thr	Ser/Thr rich domains
Fuc-α-Ser/Thr	EGF modules (Cys-X-X-Gly-Gly-*Thr*/*Ser*-Cys)
Glc-β-Ser	EGF modules (Cys-X-*Ser*-X-Pro-Cys)
Xyl-β-Ser	*Ser*-Gly (Ala) (in the vicinity of one or more acidic residues)
Glc/GlcNAc-Thr	Rho: *Thr*-37[c]; Ras, Rac and Cdc42: *Thr*-35[c]
Gal-Thr	Gly-X-*Thr* (X = Ala, Arg, Pro, Hyp, Ser) (vent worm)
Gal-β-Hyl	Collagen repeats (X-*Hyl*-Gly)
Ara-α-Hyp	Repetitive Hyp rich domains (e.g., Lys-Pro-*Hyp*-Hyp-Val)
GlcNAc-Hyp	Skp1: *Hyp*-143
Glc-α-Tyr	Glycogenin: Tyr-194

As mentioned in the case of *N*-glycans, the GalNAc-α-Ser/Thr, also known as **Tn antigen**, is often modified and further elongated (**Figure 2-2; Figure 2-3**). The most abundant modification of the Tn antigen is known as the core 1 or **T antigen** (Gal-β1-3-GalNAc-α-Ser/Thr) (**Figure 2-2**). The glycosyltransferase in mammals, responsible for the core 1 structures is known as core 1 β1-3 galactosyltransferase (core 1 β-3Gal-T), and has been found as widely expressed and in all developmental stages, indicating the biological relevance of this *O*-glycosylation. The Core 1 β-3Gal-T is evolutionary conserved in *Caenorhabditis elegans* and *Drosophila* (Corfield, 2016; E. Tian & Hagen, 2009).

Another kind of modification of the Tn antigen is known as **core 3** (GlcNAc-β1-3-GalNAc-α-Ser/Thr) (**Figure 2-2**). This structure is catalyzed by the β1-3 *N*-acetylglucosaminyltransferase (β-3GlcNAc-T6). The expression of β-3GlcNac-T6 is primarily restricted to the stomach and small intestine of mammals and its orthologues have not been found in *C. elegans* nor *Drosophila* (Gomord et al., 2010b; E. Tian & Hagen, 2009).

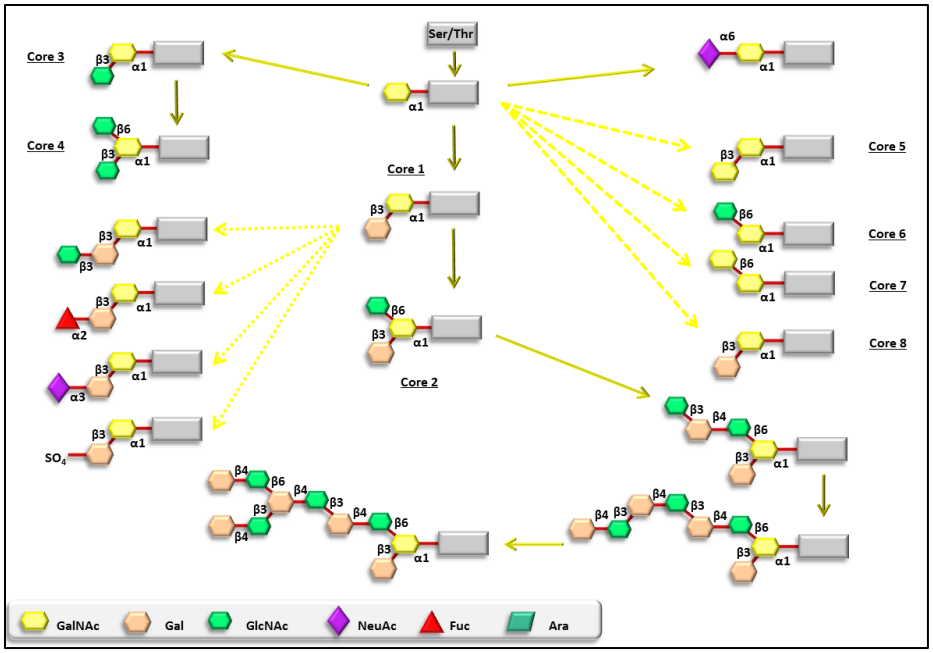


Figure 2-2 Mucin-type *O*-Glycosylation varieties

Modified from Gomord et al 2010 (Gomord et al., 2010b).

The **core 2 and 4** Mucin-like *O*-glycan branching are modifications that occur to the core 1 and 3, respectively (**Figure 2-2**). This event is controlled by the β1-6 *N*-acetylglucosaminyltransferases (β-6GlcNAc-T's). There are three core 2 β-6GlcNAc-T's that have been identified in mammals, two are responsible for the core 2 synthesis and the other one is able to catalyze both cores (core 2 and 4) (Sangadala, Sivakami, & Mendicino, 1991). The expression patterns differ from one isoform to the other, suggesting diverse biological impacts. The isoform 1 is expressed ubiquitously; while the isoform 2 (responsible for the synthesis of core 2 and 4) is expressed in the kidney, pancreas, stomach, intestines and colon; the isoform 3 is expressed in the thymus and T cells. So far there is almost no knowledge about the respective orthologues in *C. elegans* and *Drosophila* (E. Tian & Hagen, 2009)**.** The above Mucin-type *O*-glycosydic branches can be further extended by the addition of Gal and GlcNAc, and further modified by fucosylation and sialylation (**Figure 2-2**) (E. Tian & Hagen, 2009)**.**

2.1.2 Glycosylation in different recombinant protein expression platforms

As mentioned above, glycosylation may play a crucial role in the stability and activity of a protein. The impact of a glycosylation site on a protein might be critical for the production of a recombinant product, especially in the case of recombinant proteins in foreign organism. As mentioned before, other non-human species might glycosylate their proteins in a different way, as human cells do. Overall, the most evolutionary distant organisms, like bacteria, yeasts, fungi, insects and plants, possess glycosylation machineries quite different to those found in human. Some human glycosylation pathways may not be present in some of the previously mentioned organisms. Essentially, lack of specific glycosydases, glycosyltransferases, sugar donors, sugar transporters, or even cellular compartments that favor glycosylation, such as ER and Golgi might be absent. Therefore a specific glycosylation found in human cells, was not associated with proteins produced by these organisms. However, these organisms are commonly used for the production of recombinant human proteins, resulting in the absence of either glycosylation or modification with glycoconjugates with a conformation and structure never found in human cells, which could be immunogenic for humans and/or rapidly eliminated from the organism (Brooks, 2004).

2.1.2.1 Prokaryotes

Bacterial glycosylation machinery differs tremendously from human, simply because they do not possess ER and Golgi apparatus. Instead, they glycosylate some proteins in the outside of the cell membrane and the glycans are typically very different from the human ones (**Table 2-1**). Until recently, it was believed that bacteria would not glycosylate at all, but it is now known that they can perform such PTM, especially on proteins of the cell wall. On the other hand, archaebacteria and eubacteria are able to synthesize *O*- and *N*-glycosylation, but little is known of their detailed structure and biosynthesis (Brooks, 2004).

2.1.2.2 Yeasts

Yeasts have arisen as potential platforms for the production of recombinant glycoproteins for human use, due to the fact that they can glycosylate and secrete proteins, which make protein harvesting easier. However, *O*-linked glycosylation is very different from human cells (**Table 2-1**). Overall, yeast *O*-linked glycans are more similar to eukaryotic *N*-linked glycans. Typically, *O*-glycosylation starts with the attachment of Man to Ser/Thr, and is further elongated with

sequential Man residues. Since this pattern of glycosylation is not observed in humans, it is undesired in the production of recombinant proteins (Brooks, 2004).

2.1.2.3 Filamentous fungi

Filamentous fungi represent an interesting expression platform for the production of recombinant proteins, since they can glycosylate and secrete proteins in reasonable amounts. However, not much is known about the glycosylation pattern, except for *Aspergillus niger*, where some studies have been performed already. Typically, filamentous fungi can *N*- and *O*-glycosylate, although the side chains are different from the highly mannosylated structures found in yeasts. However, *O*-glycan structures seem to consist of small *O*-linked Man residues, which appear to attach single Gal, Glc, phosphate, and sulfate groups, producing some *O*-glycan structures different from those found in humans. Sialylation has not been observed so far, but since genetic manipulation is relatively easy, attempts are in progress to suppress some glycosylation steps (hyper mannosylation) and introduce human glycosylation patterns (Brooks, 2004)**.**

2.1.2.4 Insect cells

Insect cells as expression systems for recombinant proteins have emerged as highly versatile systems, well established, technically straightforward, and can produce modest quantities of cloned proteins quickly. However, one significant limitation of insect cell systems is the glycosylation patterns, which are different from human cells. Little is known about their ability for *O*-glycosylation, but it is believed that they typically synthesize Gal-β1-3-GalNAc-Ser/Thr. It is also believed that they sialylate their *O*-glycans, but in a different manner than that found in humans. Diverse attempts have also started to introduce human glycosyltransferases into insect cells, in order to mimic human glycosylation machinery (Brooks, 2004)**.**

2.1.2.5 Mammalian cells

Due to the higher similarity between the glycans present in mammalian proteins, as compared to those present in other more distant organisms, like bacteria, yeasts, insects or plants there has been a special interest in producing human glycoproteins in mammalian cells. The first glycoproteins expressed in these cells were interleukin-2, interferon β, and interferon γ.

Commonly, CHO cells and baby hamster kidney cells (BHK-21) are most frequently used for the expression of human glycoproteins, since these expression platforms possess the machinery to glycosylate proteins with typical mammalian *N*- and *O*-glycans. However, difference between the glycosylation pattern found in recombinant proteins produced in these systems and the human ones, can be found. An example of the latter is the case of the production of human erythropoietin by CHO or BHK-21 cells, where only 20-25% of the produced recombinant protein was properly glycosylated for its usage in humans, making the rest of the protein population undesirable and thus requiring separation and purification. This case, exemplifies the potential difference of glycosylation patterns in related mammalian cells to human cells (Brooks, 2004).

2.1.2.6 Higher plants

Plants possess a number of attractive advantages, when it comes to production in large-scale of industrial and pharmaceutical glycoproteins. Transgenic plants like tobacco, soybeans or maize, or even plant cells can act as “bioreactors”, in a process known as “molecular farming”, and present several advantages over bacterial, yeast or mammal cell platforms. Among the main features, one can find a low cost for their culture, and safer systems, since pathogens are not shared with humans (e.g., virus, prions). The ability of plant cells to process the proper folding or several complex proteins, and to target recombinant proteins to different compartments, like seeds or storage organelles, represent an interesting advantage over most of the other expression systems (Sharma & Sharma, 2009).

Despite all the advantages of using plants as expression platforms for human proteins, they still present some barriers concerning the proper glycosylation pattern, finding usually different glycan structures, which often produce allergenic reactions in humans (**Figure 2-3**). Among the most crucial differences, plants do not seem to sialylate glycans. Incorporation of glycans not seen in human also occurs, like xylose, rhamnose, and arabinose (**Figure 2-3**) (Gomord et al., 2010b; Gomord & Faye, 2004).

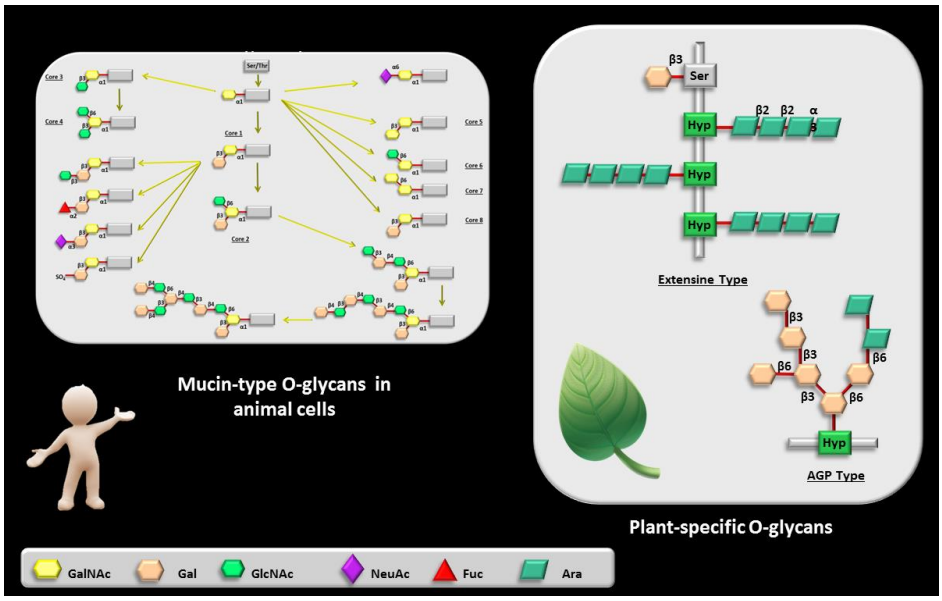


Figure 2-3 Mammalian-specific and plant-specific *O*-glycans

Modified from Gomord et al 2010 (Gomord et al., 2010b).

Plant *N*-glycosylation in comparison to the human type, differs mainly in the attachment of α(1,3)-fucose and β(1,2)-xylose to the core *N*-glycan, while human and mammalian types present α(1,6)-fucose and no xylose is attached. Further maturation of plant *N*-glycans results in a biantennary structure, occasionally with terminal Lewis A epitopes, which are rarely observed in human proteins. Human maturation of *N*-glycans typically result in a multi-antennary structure with two or more terminal branches, further extended with galactose and sialic acid (Castilho et al., 2013; Dicker et al., 2016, 2015; Khan, Bayat, Rajabibazl, Sabri, & Rahimpour, 2017; Vukušić et al., 2016).

Plant *O*-glycans are usually linked to Ser/Thr, like in humans, but also to Hyp. The typical *O*-glycosylation patterns in plants are extensins (involved in plant cell wall synthesis), lectins, and soluble arabinogalactans (present in exudates, cell walls, and intracellular proteins) (**Figure 2-3**). In the case of extensins, typically short Ara chains link to Hyp, and Gal (either mono- or di-saccharide) links to Ser. Soluble arabinogalactans are highly glycosylated with rhamnose, arabinose, Gal, glucuronic acid, galactouronic acid, and methylated derivatives of these. Overall, the *O*-linkage is often between a Gal or Glc residue and a Hyp (**Figure 2-3)** (Brooks, 2004; Saint-Jore-Dupas, Faye, & Gomord, 2007).

2.1.3 Engineering glycosylation

Concerning the engineering of glycans, major efforts have been aimed to imitate ***N*-glycan** structures from human proteins, in plant-derived proteins. The strategies have been based on retaining the recombinant protein in the ER, or by blocking the plant-specific glycosylation machinery in the Golgi apparatus, thus eliminating β(1,2)-xylose and α(1,3)-fucose residues to be added to the plant-derived protein (Daskalova et al., 2010). Recently, the production of human recombinant proteins in plants without the Lewis A epitope on *N*-glycans has also been reported. Lewis A is rarely present on human glycoproteins of healthy adults, but it is widely present on plant glycoproteins (Parsons et al., 2013). Further achievements have been made in plant *N*-glycosylation, incorporating heterologous enzymes like the human β1,4 galatosyltransferase and others, necessary for the imitation of human sialylation (Parsons et al., 2013).

In the case of ***O*-glycosylation**, fewer but significant advances have been achieved concerning engineering of *O*-glycosylation in plants, whose *O*-glycans differ significantly from the typical mucin-type glycans (Gomord et al., 2010b; Parsons et al., 2013). Among the recent achievements, Daskalova and co-workers (Daskalova et al., 2010) successfully expressed the human GalNAc-T2 in *Nicotiana benthamiana* plants, where it retained its localization in the Golgi compartment, and its activity. The GalNAc-T2 specifically O-glycosylated an acceptor protein, the 113-136 amino acid fragment of the chorionic gonadotropin β-subunit, at the native Ser-121 and Ser-127 GalNAc attachment sites, *in vitro*. The transiently expressed GalNAc-T2 was also able to perform GalNAc-glycosylation at the Thr-119 site of the endogenous enzyme endochitinase, *in planta*. However, the GalNAc-glycosylation of a human mucin 1 tandem repeat derived was barely glycosylated, presumably due to the low endogenous UDP-GalNAc substrate pool and the poor translocation of this substrate to the Golgi lumen. Further genetic engineering consisted of co-expressing the *Y. enterocolitica* UDP-GlcNAc 4-epimerase, in charge of converting UDP-GlcNAc to UDP-GalNAc in the cytoplasm, and a *C. elegans* UDP-GlcNAc/GalNAc transporter, responsible for the transport of the sugar donor to the Golgi lumen. The latter strategy conferred the GalNAc-type glycosylation capability to the *N. benthamiana* plants, specifically *O*-glycosylating the human Mucin 1 peptide derivative, which was detected exclusively as a glycoform (Daskalova et al., 2010).

Recently, Yang and colleagues (Yang, Drew, et al., 2012) were also able to generate mucin-type *O*-glycosylation in *N. benthamiana*, using a similar approach. They transiently expressed a *Pseudomonas aeruginosa* UDP-GlcNAc 4-epimerase, together with the human GalNAc-T 2 and 4, and a human 3.5 tandem repeat of Mucin1. The derivative mucin peptide was GalNAc-type O-glycosylated with up to three and five GalNAc residues, when expressed with GalNAc-T2 and GalNAc-T 2 and 4, respectively. The mucin-type *O*-glycosylation was also demonstrated by mass spectrometry on other protein substrates, tandem repeat of MUC16 and interferon α2b. It also resulted astonishing that no glycosylated Hyp were detected (Yang, Drew, et al., 2012).

Yang and colleagues (2012) have also been able to generate stable transgenic lines of soil grown *Arabidopsis thaliana* and *Nicotiana tabacum* Bright Yellow 2 (BY-2) suspension culture cells, using the strategy above mentioned (Yang, Bennett, et al., 2012). Surprisingly, even though the stable transgenic lines were capable to GalNAc-glycosylate the mucin substrate (MUC1), a high degree of Pro hydroxylation and Ara-Hyp on the MUC1 derivative was observed (Yang, Bennett, et al., 2012).

As mentioned previously, elongation and branching of the T antigen might also be desired for several PMP's, in order to fully imitate the human specific mucin-type *O*-glycosylation. Recently, Castilho and colleagues (Castilho et al., 2013) were able to produce mucin-type mono- and di-sialylated core 1 *O*-linked glycans in transiently transformed *Nicotiana benthamiana* plants, as well as biantennary sialic acid residues at N-glycan on a recombinant human erythropoietin fused to an IgG heavy chain fragment (EPO-Fc). For the generation of the sialylated core 1 structures, it was required to transiently express the human polypeptide:*N*.acetylgalactosaminyltransferase 2 (ppGalNAc-T2), the *Drosophila melanogaster* core 1 β1,3-galactosyltransferase (C1 Gal-T1), the human α2,3-sialyltransferase (ST3Gal-I), the *Mus musculus* α2,6-sialyltransferase (ST6GalNAc-III/IV), and the human CosmC chaperon, together with EPO-Fc as substrate peptide and the machinery for sialylation of *N*-glycans. This study demonstrated the feasibility of plants to produce well-defined human-type *O*- and *N*-linked glycans on recombinant therapeutic proteins.

Recently, the genes responsible for the plant-specific 4-trans-hydroxyproline (Hyp), the main anchor for *O*-glycosylation in plants, have been identified and knocked-out in *P. patens,* allowing the production of human recombinant erythropoietin, free of plant-specific Hyp moieties. Hyp is synthesized post-translationally by prolyl 4-hydroxylases (P4Hs) via hydroxylation of the γ

carbon of proline, but the recognition sequences differ between plant and animal cells (Parsons et al., 2013).

2.1.4 Human Granulocyte-Colony Stimulating Factor as case of study for feasibility of production of a pharmaceutical *O*-glycosylated protein in plants

The glycoprotein hG-CSF is a good example of an economically relevant pharmaceutical, for which alternative expression platforms are continuously being explored for production (Bae, Yang, Lee, & Park, 1999; Hong, Kwon, Jang, Kim, & Yang, 2006; Hong, Kwon, Lee, Jang, & Yang, 2002; Kraševec et al., 2014; Lasnik, Porekar, & Stalc, 2001; Nair, Chidambareswaren, & Manjula, 2014; Sharifi Tabar, Akbar Habashi, & Rajabi Memari, 2013; Suk Yang, Soon Bae, & Lee, 1997; Tabar, Solouki, Tohidfar, & Sadeghizadeh, 2012; Talebkhan, Samadi, Samie, Barkhordari, & Azizi, 2016; L. Tian & Sun, 2011; Vanz et al., 2008). This glycoprotein is the main cytokine that regulates granulopoiesis or production of granulocytes (Hattori, Orita, Oheda, Tamura, & Ono, 1996; Williams, Smith, Spooncer, Dexter, & Taylor, 1990; Yuo et al., 1989). As such, this cytokine plays an important role to protect the organism against bacterial, fungal and viral infections, and is clinically used for the treatment of patients with neutropenia, typically cancer patients that undergo chemotherapy (Vanz et al., 2008).

The mature G-CSF is a 19.6 kDa glycoprotein, made up 174 amino acids, two intramolecular disulphide bonds, one free cysteine at the residue 17, and a carbohydrate chain bound to the Thr amino acid at residue 133 infections (Barreda, Hanington, & Belosevic, 2004; Chavez-Tapia et al., 2015; Hermesh, Moran, Jain, & López, 2012; Houston, Stevens, & Cour, 1999; Reeves, 2014). Monocytes, macrophages, stromal cells, endothelial cells and fibroblasts are the cells responsible for the production of the G-CSF. The expression of this cytokine is driven under the exposure to endotoxins and inflammation reactions (Broudy, Kaushansky, Harlan, & Adamson, 1987; Kaushansky et al., 1988; Nicola & Metcalf, 1985; Palmblad, Gyllenhammar, Lindgren, & Malmsten, 1984; Vellenga, Rambaldi, Ernst, Ostapovicz, & Griffin, 1988).

2.1.4.1 Recombinant Human G-CSF Expression Systems

Currently, recombinant hG-CSF is commercially produced using both *E. coli*, and mammalian cells. The *E. coli* version is non-glycosylated (Neupogen), and it possesses an initial extra amino acid (Met). The mammalian version is produced in CHO Cells, and thus it is glycosylated

(Granocyte) with a mucin-type *O*-glycosylation site, in the corresponding Thr residue. The GalNAc-*O*-glycosylated site is decorated with SA-Gal-GalNAc, presumably as it is glycosylated in human cells (Bönig et al., 2001).

2.1.4.2 Relevance of glycosylation on G-CSF

It is important to determine the possible effects that structural differences in the recombinant protein might confer on patients. It is stipulated that glycosylation represents important consequences in terms of cytokine efficacy (Bönig et al., 2001). Because of the structural differences present in the commercially available recombinant hG-CSFs (**Table 2-1**), a study was performed, in which the biological activity of this cytokine was tested in humans (127 patients with lymphoma or myeloma). In this study, it was found that in order to obtain 2 x 10^6 $CD34^+$ cells/Kg (constitutive level known as optimal for a good quality cytaphereris event), it was necessary to administer 1 vial of Granocyte per day (an average of 3.5 µg/kg/day), compared to 10 µg/kg/day of Neupogen for the same result. Therefore, at least 3 times more Neupogen were needed in the same period of time to obtain the same physiological effect (Duncan, Hewetson, Atra, Dick, & Pinkerton, 1997).

It has been demonstrated that the non-glycosylated recombinant hG-CSF is susceptible to degradation caused by the human neutrophil derived elastase (HNE) in the serum. Such degradation produces a reduction in the biological activity of the cytokine. The glycosylated cytokine also suffers degradation by the HNE, but this enzymatic destruction occurs at a slower rate. This data was corroborated by producing a non-glycosylated G-CSF in yeast, and a glycosylated G-CSF which was enzymatically treated in order to remove the carbohydrate. In this study, it was found that the glycosylic residue was critical for the cytokine protection against protease degradation (Hattori et al., 1996).

2.1.4.3 hG-CSF Expressed in Yeasts

Alternative expression platforms have also been explored for the production of this pharmaceutical. Yeast platforms offer as an advantage a secretion system similar to that of higher order eukaryotes. By using such secretion systems, the recovery/purification process can be simplified and, therefore the recovery levels can be significantly elevated (Böer, Steinborn, Kunze, & Gellissen, 2007).

Expression of hG-CSF was achieved in *Saccharomyces cerevisiae*, obtaining a concentration of over 60 mg/L after 40 hours. Although, no consisted studies about glycosylation were accomplished, it was estimated that the hG-CSF did not present any O-glycosylation (Suk Yang et al., 1997). However, the secreted hG-CSF was usually found as large multimers in the medium, due to strong hydrophobic interactions. Such multimers diminish the specific union activity to the hG-CSF receptor, thus causing undesired problems in the process of recovery/purification (Bae et al., 1999).

Other yeast systems have been used for the production of hG-CSF, such as *Pichia pastoris*, in which an expression level of 10 mg/L has been obtained. However, this production platform presented a similar problem as with *S. cerevisiae* (Lasnik et al., 2001), and glycosylation analysis was not addressed.

2.1.4.4 hG-CSF expressed in Fungi

The filamentous fungus *Aspergillus niger* has also been explored for the production of hG-CSF. In this case, 5-10 mg/L of secreted protein were obtained. But similar to the yeast examples mentioned above, hG-CSF was not glycosylated, and was present mainly as multimers (Kraševec et al., 2014).

2.1.4.5 hG-CSF expressed in Plants

In recent experiments, the production of the hG-CSF has been achieved in tobacco suspension cultures. In the last system, a production of 105 µg/L of hG-CSF has been obtained, after 9 days from the suspension culture initiation. However, after the ninth day, the production of the hG-CSG was diminished. The latter suggests that the contact of proteases in the medium with the cytokine, during a long period, could have caused such a production effect (Hong et al., 2002).

In studies achieved later on by Yang et al., 2006, it was possible to obtain a higher production of the hG-CSF. After 13 days of culturing rice cells in suspension, a concentration of 2.5 mg/L (0.7% of the total secreted protein) was achieved, using the α-amylase 3D promoter, which is inducible in the absence of sucrose. In this way, the expression and secretion to the medium of the hG-CSF was achieved due to the signal sequence of the leader peptide Ramy3D (Shin, Hong, Kwon, Jang, & Yang, 2003). Furthermore, the biological activity of the secreted cytokine

was similar to that observed by the hG-CSF expressed in *E. coli* (Hong et al., 2006), however, no studies to determine the *O*-glycosylation were performed.

G-CSF has also been produced in stably transgenic tobacco plants (*Nicotiana tabacum*) (Tabar et al., 2012), ER-targeted in BY-2 cells (Nair et al., 2014), in lettuce chloroplasts (Sharifi Tabar et al., 2013), and transiently in *N. benthamiana* (Zvereva et al., 2009). However, none of these reports addresses glycosylation characterization.

2.2 Protein phosphorylation

Protein phosphorylation is the most studied PTM involved in the reversible activation or deactivation of a protein, and it usually has an impact on the bioactivity of the protein as well as in cellular signaling processes (Seo & Lee, 2004; Sun, Eun, Yoon, Ji, & Park, 2006). It can occur anywhere along the protein structure at amino acid residues S, Y, T, H and D (Seo & Lee, 2004). Typically, phosphorylation is determined in eukaryotic cells, by the activation of protein kinases and phosphatases on their target protein (**Figure 2-4**). Once the covalent conjugation of phosphate groups to a protein backbone is achieved, it alters the protein function by inducing conformational changes and/or by affecting protein-protein/protein-substrate interactions (Wang, Peterson, & Loring, 2014).

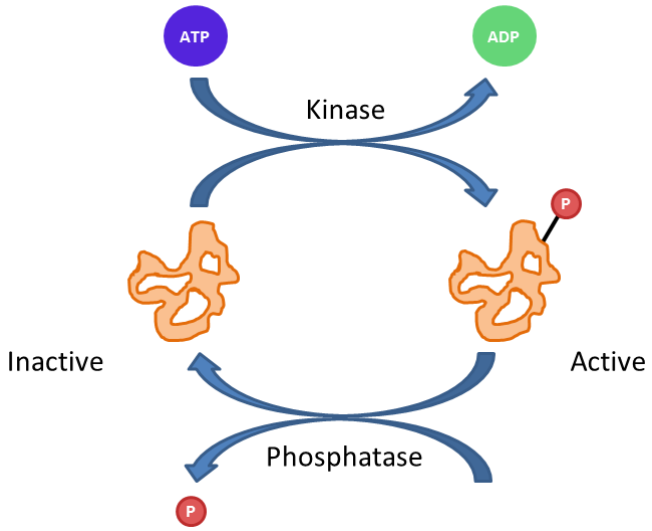


Figure 2-4 Protein phosphorylation

Graphical representation of protein phosphorylation, where a protein kinase attaches a phosphate group from ATP to amino acid side chains, activating or inactivating a target protein. The phosphorylated protein can be again dephosphorylated by phosphatases. ATP: Adenosine Triphosphate; ADP: Adenosine diphosphate; P: phosphate group.

2.2.1 Caseins as case study of phosphoproteins

One of the most popular phosphoproteins is probably casein. Caseins belong to a group of proteins that evolved from the so called secreted calcium (phosphate)-binding phosphoproteins (SCPP) and usually four different types of caseins are found in milk, namely α_{S1}-, α_{S2}-, β- and κ-casein (Holt, 2013). These phosphoproteins are the main proteinaceous component of milk, accounting for almost 80% of the protein fraction in bovine´s milk. Casein proteins are widely known as a source of essential amino acids required for neonates, prevention of pathological calcification of the mammary gland by forming casein micelles, and various peptides contained in the amino acid sequence of caseins have been increasingly studied for their bioactivity roles as antioxidant, antithrombotic, antimicrobial, antihypertensive and immune modulator (Smyth, Clegg, & Holt, 2004).

Caseins have also extensively been used for cheese production and dairy products, due to the clotting properties of these proteins in presence of acids and proteolytic enzymes. The importance that they have in this industry has probably made them the most widely studied food proteins (Ginger & Grigor, 1999). In milk, the different types of caseins interact with each other

(α_{S1}-, α_{S2}-, β- and κ-casein: 4:1:3.5:1.5, respectively), forming colloidal casein-calcium phosphate particles, traditionally called micelles (Kethireddipalli & Hill, 2015). The primary structure of a casein micelle is formed by the hydrophobic interactions of several α_{S1}-, α_{S2}- and βC in the core together with amorphous calcium phosphate, and the hydrophobic N-terminus of κC (para-kappa-casein) at the surface, while the hydrophilic C-terminus of κC (caseinomacropeptide or CMP) extends at the surface of the micelle as a glycosylated hairy layer (**Figure 2-5**). It is due to this polar hydrophilic layer formed by the C-terminal regions of κC that the casein micelles remain stable in milk, instead of coming closer and clotting (De Kruif, 1999; Kethireddipalli & Hill, 2015).

However, when chymosin (enzyme present in rennet) is added to the milk, a clotting reaction is induced within minutes, causing destabilization of the solubilized casein micelles and then forming a gel. This transition is achieved by the proteolytic action of chymosin, breaking the peptide bond in κ-casein releasing the CMP, and thus disrupting the strong intermicelle repulsive forces. After releasing the CMP, the para-casein micelles come closer one another and in the presence of ionic calcium, they begin to aggregate via hydrophobic interactions aimed by calcium bridging, forming a gel structure that traps serum and fat globules. After coagulation is induced, several treatments can be employed in order to obtain a large variety of cheese products (Kethireddipalli & Hill, 2015).

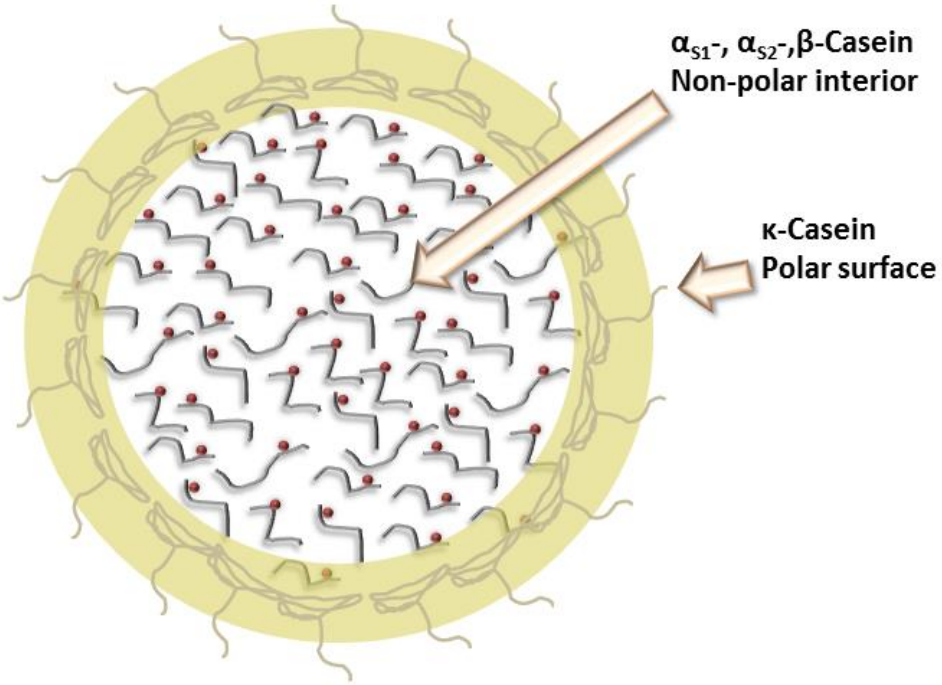


Figure 2-5 Casein micelle structure

Graphical representation of a casein micelle formed by phosphorylated (red dots) α_{S1}-, α_{S2}-, β-casein in the core and κ-casein at the surface.

As stated above, among the different members of the family of caseins, βC plays a pivotal role in forming the casein micelle, while κC has an impact on stabilizing the micelle, and therefore the structure of these two bovine proteins will be explained with more detail.

Bovine βC is a 24 kDa phosphoprotein with 5 phosphorylated Serine residues. This phosphoprotein is rich in proline residues and does not possess disulfide bridges, favoring a disorganized secondary structure (Portnaya et al., 2008). Analysis of the amino acid charge frequency indicates that βC is a linear amphiphilic protein. In fact it is the most “soap-like” casein among the different members of this family, being the N-terminus (residues 1 to 25) of the molecule the only portion with a net charge due to the phosphoserine residues, which bind calcium, while the rest of the molecule is hydrophobic (Farrell, Qi, Wickham, & Unruh, 2001). Such amphiphilic structure makes it also possible to form βC (only) micelles, which can form a gel structure by reducing the pH to 5.5, lowering the temperature and adding sodium polyphosphate (Panouillé, Durand, & Nicolai, 2005; Portnaya et al., 2008).

Bovine κC is a 19 kDa phosphoprotein. This casein differs from the rest of the caseins in the sense that this is not a calcium-sensitive casein as the others, being the only casein that remains soluble in the presence of calcium ions and it also has a much smaller phosphate component than the other members of the family. The phosphorylated sites possess also the peculiarity of being confined in the C-terminal region of the protein, present as single sites rather than clusters, as found in the other calcium-sensitive caseins, and the phosphorylated residues are located among glycosylation sites, being this casein the only member of the family with carbohydrate moieties (Ginger & Grigor, 1999).

KC has in total three phosphoserine residues and up to six *O*-glycosylated threonine sites, all located at the C-terminus. *O*-glycosylation sites belong to the so called Mucin-type *O*-glycosylation, which will be further described later on. In this particular case, the *O*-glycosylated threonine sites consist of *N*-acetyl galactosamine, galactose and *N*-acetyl neuraminic acid (Ginger & Grigor, 1999). Phosphorylation of caseins is now believed to occur in the Golgi cisternae while being secreted by specific casein kinases, and start sequestering calcium during the secretion pathway. In the case of κC, glycosylation starts in the Golgi apparatus, and it is completed before secretion (Farrell et al., 2001; Ginger & Grigor, 1999).

2.2.1.1 Production of recombinant caseins

Due to the industrial interest of cheese production, major efforts to produce recombinant casein proteins have been employed to address structural studies of these proteins or for research purposes. Typically, expression of βC has been achieved with *E. coli,* yeast and CHO cells platforms (Fahimirad, Abtahi, Razavi, Alizadeh, & Ghorbanpour, 2017; Hansson et al., 1993; Jiang, Li, Yang, & Geng, 2004; Jimenez-Flores, Richardson, & Bisson, 1990; Thurmond et al., 1997). Food proteins, as opposed to pharmaceutical proteins, have not been extensively produced in heterologous systems because of their abundance, variety of sources, and availability in nature (Choi & Jime, 2001). However, in terms of genetically engineering manipulation of caseins for human consumption, some efforts have been made for the production of transgenic farm animals with altered casein expression levels, in order to modulate milk quality production and therefore dairy products (Brophy et al., 2003; Jabed, Wagner, McCracken, Wells, & Laible, 2012).

In the case of κC, it has also been produced recombinantly in *E. coli* for structural characterization studies and nutritional improvement by directed mutagenesis, rebalancing the

amino acid content from a nutritional perspective (Kang, 1988; Oh & Richardson, 1991). In this case, particular attention has been paid on producing recombinant bioactive peptides contained within the κC amino acid sequence, which has the potential to be used in practical foods or medicinal nutritional products. The highly nutritious human CMP is a known promoter of beneficial bacteria (*Bifidobacteria*), antithrombotic, and adhesion inhibitor of oral *Actinomycetes* to human cell membranes as well as cholera toxin to its receptors. This κC derived peptide has been produced in *E. coli* (Liu, Liao, & Chen, 2008) and yeast (Kim, Oh, Kang, Lee, & Park, 2005). Similarly, lactaptin analogs, a proteolytic fragment derived from the human κC, capable to reduce cell viability and induce apoptosis of the mammary adenocarcinoma cell line MCF-7, have been expressed in *E. coli* (Semenov et al., 2010).

2.2.1.2 Production of recombinant caseins in the context of PTM

It has to be noticed that the βC expressed in *E. coli* was not phosphorylated, therefore being unable to sequester calcium which is relevant for clotting of cheeses (Hansson et al., 1993). But when the βC was co-expressed with the human casein kinase II, *E. coli* was able to produce phosphorylated βC (Thurmond et al., 1997).Thus, it is feasibility to produce a phosphorylated βC when a proper kinase is functionally available. When produced in yeast, the βC expressed by *P. pastoris* was observed to be properly phosphorylated but also *N*-glycosylated with several mannose residues at a single Asn residue (Asn-68) (Choi & Jime, 2001). A similar pattern was observed when the βC was expressed in *S. cerevisiae* (Jimenez-Flores et al., 1990), where the casein was also present as a heterogeneous phosphorylated recombinant protein, but it seemed to be *O*-glycosylated, instead of *N*-glycosylated. On the other hand, when βC was expressed in CHO cells, the casein was found phosphorylated (Jiang et al., 2004). The latter examples offer an overview of the capabilities of different expression systems to produce properly functional (phosphorylated) recombinant βC.

A similar scenario could be observed with κC, which is natively phosphorylated and *O*-glycosylated. The whole protein and the CMP were expressed in *E. coli*, however no phosphorylation studies were addressed (Kang, 1988; Oh & Richardson, 1991) and only 1 study proposed that the recombinant CMP was not glycosylation, as would be expected (Liu et al., 2008). On the other hand, only the CMP has been expressed in *S. cerevisiae* and *P. pastoris* (Kim, Oh, et al., 2005). In this case, the CMP produced in *S. cerevisiae* was *O*-glycosylated but

to a lesser extent in comparison to its native counterpart and no phosphorylation analysis were performed (Kim, Park, et al., 2005).

2.2.1.3 Phosphorylation of casein using plant expression systems

Even though plants are known to possess a vast variety of kinases and phosphatases and therefore capable of performing such PTM, it has been reported that the expression of phosphoproteins such as βC in potato and soybean, lacks the presence of phosphate groups attached to the recombinant protein (Chong et al., 1997; Maughan et al., 1999; Philip et al., 2001). The latter might suggest the role of specific kinases or absence of these in certain compartments of plant cells. Recently, Tagliabracci and colleagues (Tagliabracci et al., 2012) identified a protein kinase called FAM20C (Family with sequence similarity 20, member C), which is co-secreted together with the human βC by the mammalian cells, and is responsible for the specific phosphorylation of several proteins at the S-X-E/S domains, including βC and other proteins important for biomineralization. The presence of FAM20C in human gland cells, might explain the dependency of βC on the FAM20C to get specifically phosphorylated.

Plants do possess mechanisms to phosphorylate and dephosphorylate proteins, having an impact on protein functionality and signaling, as in animal cells. However, although no FAM20C homologues have been reported in plants, the protein kinases I and II (CKI and CKII), among others, are present in plants and are known to be able to phosphorylate *in vivo* a variety of proteins, as well as the βC *in vitro*. However, CKI and CKII are found in the nucleus, cytoplasm, membranes and mitochondria, subcellular locations where they act (protein conformational activation/deactivation) (Lôbo Machado, Da Silva, Da Silva, Leite, & Ottoboni, 2002). Lack of co-localization of an appropriate kinase and target protein, as well as appropriate compartment conditions might dictate this PTM, and therefore properties of recombinant proteins produced in foreign organisms, as it was the case of the βC produced in soybean by Philip and colleagues (Philip et al., 2001).

Neither κC nor CMP have been successfully expressed in plant systems so far. Although as it was the case with βC, it would be expected that proper control of PTM with this protein might also occur, for both, phosphorylation and *O*-glycosylation (which will be further explained later on) and thus should be considered for heterologous expression.

2.2.1.4 Modulation of protein mammalian-like phosphorylation in plants

In the case of phosphorylation, no efforts have been made so far to allow site-specific phosphorylation on recombinant proteins produced in plants. The expression of βC in potato and soybean seeds was achieved (Chong et al., 1997; Maughan et al., 1999; Philip et al., 2001), but no efforts were made to obtain this protein in its phosphorylated original form, the active form. This case represents a clear example, in which the absence of specific kinases, or proper localization of such enzyme, leads to the production of foreign proteins without their respective mammalian-specific PTM´s, and thus a potential risk for a deficient activity. The human protein kinase FAM20C recently identified by Tagliabracci and colleagues (Tagliabracci et al., 2012), which is localized to the Golgi apparatus and secreted along with other proteins, might represent an alternative to specifically phosphorylate recombinants proteins expressed in plants, like β- and κC´s (Tagliabracci et al., 2012). Although no plant FAM20C-like proteins have been reported so far, FAM20C homologues might be present in other animals, presenting a similar function, and thus its heterologous co-expression might offer the possibility to produce recombinant proteins with proper phosphorylation. The bovine Dentin Matrix Protein 4 (DMP4), also recently renamed as FAM20C (XP_614520.6), is highly homologous (86.6%) to the human FAM20C. Although not proved so far by any study, this putative casein-specific kinase might be responsible for the specific phosphorylation of bovine caseins and thus its co-expression with casein proteins in plant system might solve the issue in respect to the lack of phosphorylation.

2.3 Nicotiana benthamiana as an expression platform for engineering post-translational modifications of recombinant proteins

All the previous plant expression platforms explored for the production of caseins or their derivatives and for hG-CSF, have used a stable transformation approach. This traditionally adopted strategy for expression of recombinant proteins in plants, although widely accepted, presents major drawbacks such as the laborious time required for the generation of stable transgenic lines and the effect of transgene copy number and positional effects. These problems negatively impact on protein yields, and therefore the screening for the most efficient transgenic lines making the whole process more laborious (Ahmad, Pereira, Conley, Richman, & Menassa, 2010).

Transient expression systems, like agro-infiltration of *N. benthamiana* leaves, represent an alternative for the production of recombinant proteins in plants, lightening the issues mentioned

above with stable transgenic lines. These systems allow the rapid production of recombinant proteins with high yields, reaching gram-sized quantities in a time frame of days or a couple of weeks (Desai, Shrivastava, & Padh, 2010; Twyman et al., 2012). Moreover, strategies for the production of highly complex products such as multiple co-expression of proteins or pathways, can be explored in order to optimize the generation of the desired product (Ahmad et al., 2010).

Transient expression via agro-infiltration of *N. benthamiana* leaves is achieved by directly introducing the *Agrobacterium* strain harboring the expression vector, into the plant tissue using vacuum infiltration or direct injection. Once the bacteria is introduced in the tissue, single stranded T-DNA is transferred to the nucleus of several cells, allowing a highly heterogeneous population of transformed cells, and therefore avoiding positional effects, while allowing high-level expression (Ahmad et al., 2010; Desai et al., 2010).

2.4 Conclusions

Several mammalian proteins for human interest, as pharmaceutical products or food supplements are commonly found with PTM´s, like glycosylation and phosphorylation. Plants have emerged as alternative expression systems for the production of such proteins. Traditionally, expression of recombinant proteins, do not address such issues, not only in plant systems but also in other platforms. Lack of post-translational machineries in the expression system might hinder the proper expression of the recombinant protein, having impact on protein stability, activity or specificity, and thus they must be considered.

Chapter 3

- **Experimental Procedures**

3.1 General strategy

Figure 3-1 depicts the general strategy used in the present work. The first part consisted on building the different expression vectors used in this work. The different expression cassettes were transiently expressed in *N. benthamiana* plants via *Agrobacterium* infiltration. Protein expression was detected by confocal microscopy, and corroborated by Western Blot. The model proteins were further analyzed in order to characterize the post-translational modification.

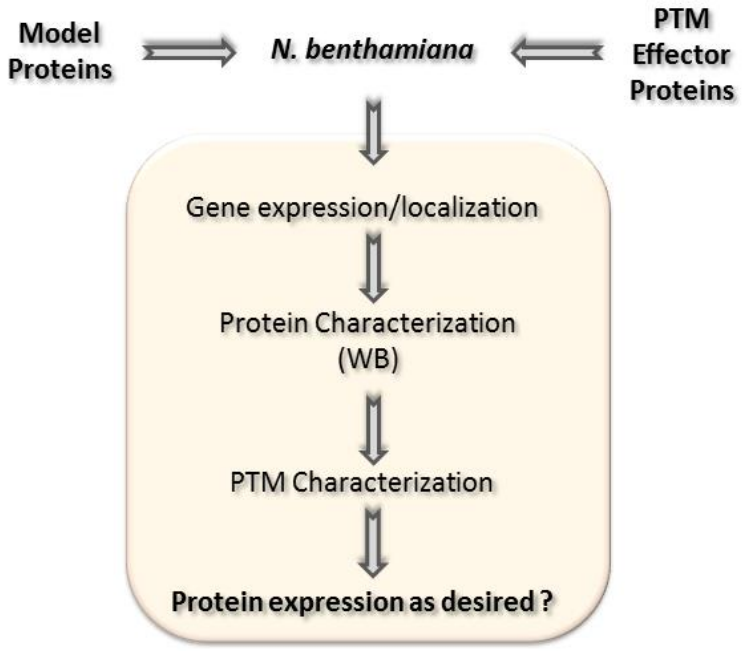


Figure 3-1 General strategy followed in this work

3.2 DNA manipulation

The different coding sequences used in this work were PCR amplified to introduce the flanking BbsI A (CATG) and B (AGCT) sites, and removal of the stop codon to allow in frame fusion with the appropriate tag. The resulting amplified products were cloned in pUC57 vector via EcoRV, and the mutations sites were verified by DNA sequencing and restriction digestions. The resulting pUC57 "CDS" (Coding Sequence) BbsI A/B without Stop codon (Cloning Stage Level 0; pUC57 "*CDS*" BbsI A/B w/o Stop) was then used to clone the coding sequence into the pENTR4 BbsI AB, via GoldenGate, using the restriction enzyme BbsI and DNA ligase in a single cutting-ligation reaction. The resulting pENTR4 "*CDS*" without Stop codon (Cloning Stage Level 1; pENTR4 "*CDS*" w/o Stop) was then used to pass the corresponding coding sequence to the binary vector via Gateway LR reaction, allowing the in-frame cloning of the coding sequence with the desired C-terminus tag, eYFP (pGWB 641) or c-Myc (pGWB 617) (Cloning Stage Level 2) (Nakamura et al., 2010) (**Figure 3-2**). The binary vectors were used to transform the agrobacterium strain AGL-1, via electroporation, and positive clones were stored at -80 C and used for further transient expression assays in *N. benthamiana* plants.

Besides of the binary vectors containing the coding sequences corresponding to the model proteins, empty pGWB 642 was used as control for expression of only the eYFP tag, for confocal purposes, which would allow the detection of eYFP in the cytoplasm, referred to as CyteYFP (Cytoplasmic eYFP) (**Figure 5-4**). *Agrobacterium tumefaciens* EHA105 carrying a binary vector containing the post-transcriptional gene silencing suppressor (PTGSS) p19 gene from the *Cymbidium ringspot tombusvirus* (Silhavy et al., 2002) was used to co-express genes of interest, allowing higher accumulation levels (**Figure 4-5; Figure 5-3**). *Agrobacterium tumefaciens* EHA105 strains carrying binary vectors containing SecGFP (Secretory GFP), GFP:KDEL (Secretory GFP fused to the ER retention signal KDEL) (Conley, Joensuu, Menassa, & Brandle, 2009), and STM:mRFP (Syalyltransferase fused to mRFP) (Latijnhouwers et al., 2005) were also used for confocal purposes (**Figure 4-10**).

Finally, *Agrobacterium tumefaciens* AGL-1 carrying the binary vectors containing the genes corresponding to the glycosylation machinery, *Yersinia enterocolitica* UDPGlcNAc/UDPGalANc epimerase (gne), *Caenorhabditis elegans* UDPGlcNAc/UDPGalANc Transporter and GalNAc-Transferase 2 (Daskalova et al., 2010), were used to perform glycosylation studies. Schematic representation of the constructs containing these genes is depicted in **Figure 4-12** and further represented in **Figure 1-1**.

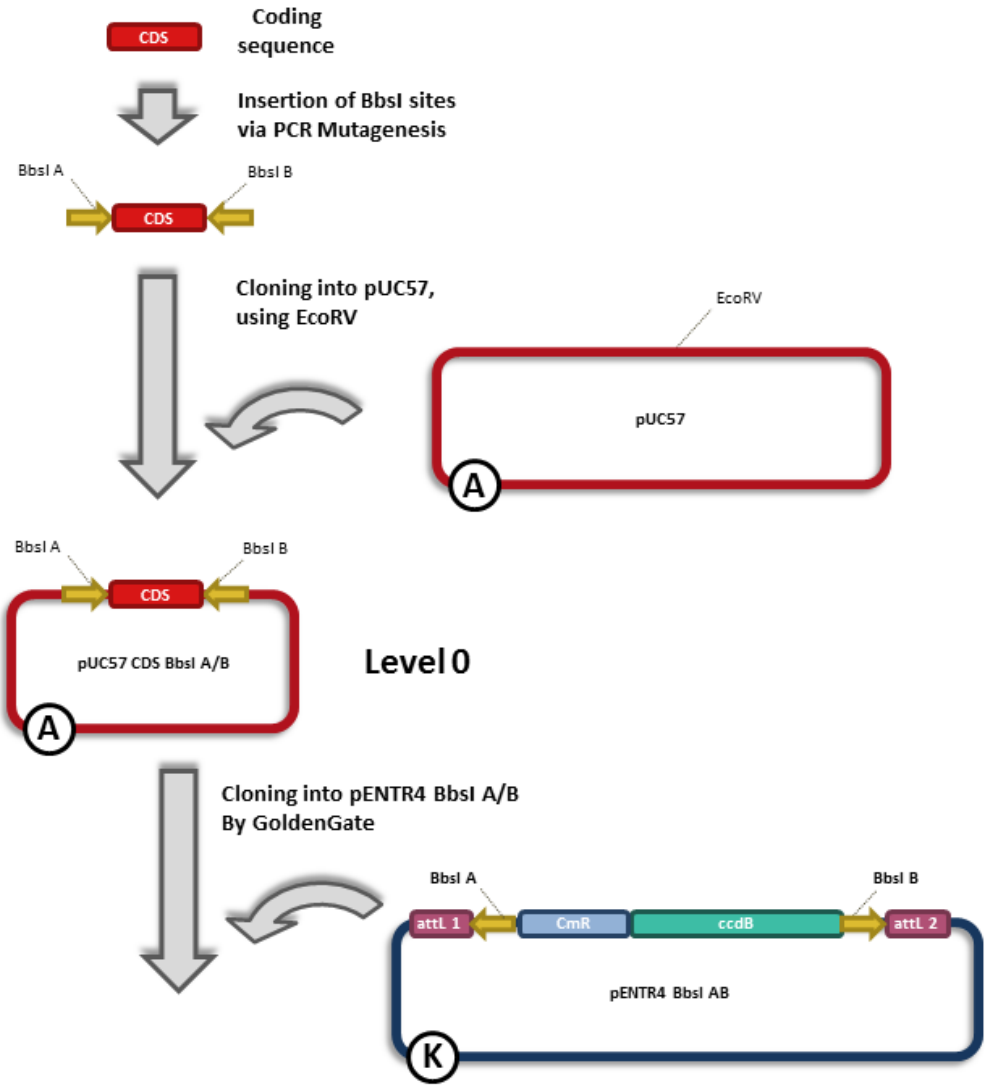


Figure 3-2 Cloning strategy

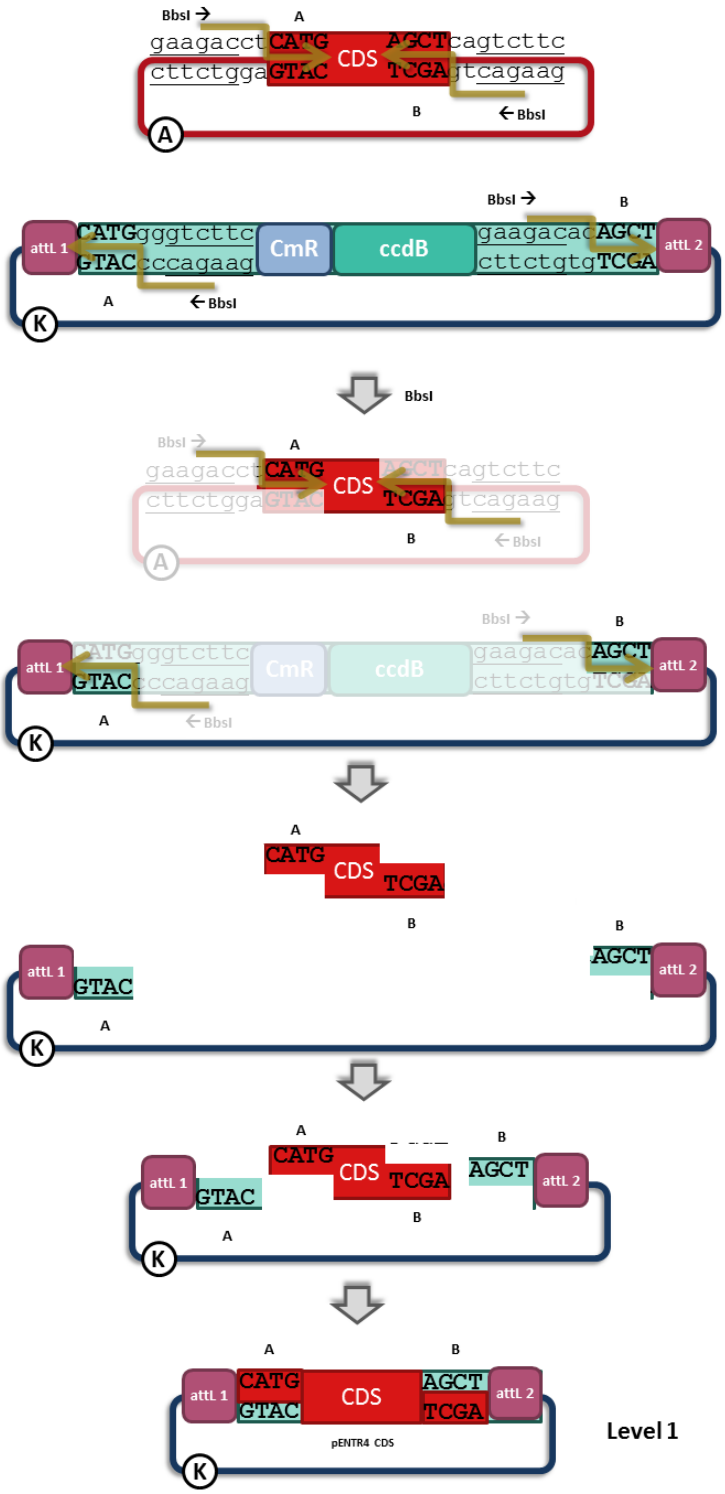


Figure 3-2 Cloning strategy (Continuation)

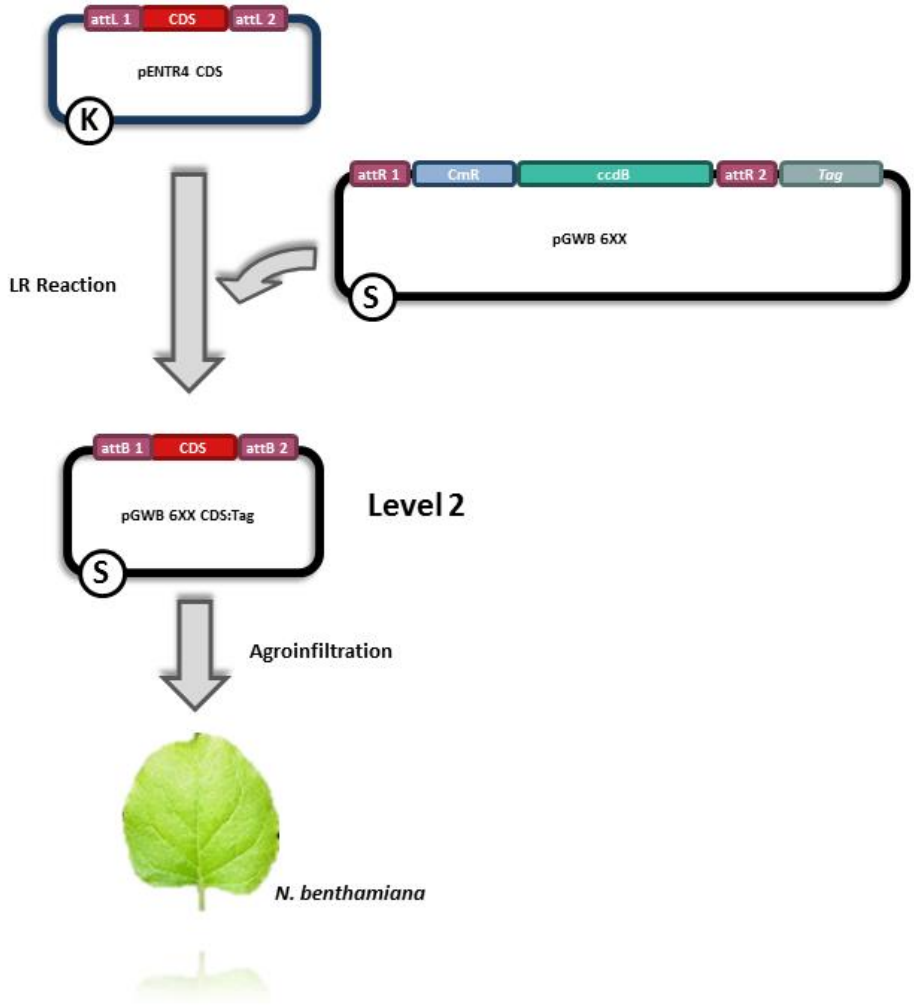


Figure 3-2 Cloning strategy (Continuation)

All coding sequences (CDS) corresponding to the different proteins used in this study were cloned using the strategy represented in the diagram. Level 0: The coding sequence was amplified via mutagenesis PCR, adding the BbsI restriction sites with the corresponding specific overhangs A (CATG) and B (AGCT), and the amplified fragment was introduced in a pUC57 vector via EcoRV. Level 1: corroborated coding sequences were cut using BbsI, and cloned into pENTR4 BbsI A/B, via GoldenGate (cut-ligation reaction). Level 3: The resulting pENTR4 “CDS” w/o Stop was then subjected to an LR reaction to allow recombination of the coding sequence into the binary vector. A (red vector): ampicillin resistance; K (blue vector): kanamycin resistance; S (black vector): spectinomycin resistance. CDS: Coding Sequence. Gate Way (GW) cassette (*ccdB* suicidal gene and Chloramphenicol (CmR) marker) pGWB 6xx: binary vector pGWB 600 series (Nakamura et al., 2010).

3.3 *N. benthamiana plants growth and maintenance*

N. benthamiana plants were grown for 7 weeks and used for transient expression. Plants were grown in a growth chamber at 22 C with a 16 h photoperiod at a light intensity of 110 µmol/m^2/s. Plants were watered with a water soluble fertilizer (N:P:K = 20:8:20) at 0.25 g/L (Plant products, Brampton, ON, Canada).

3.4 Agroinfiltration of N. benthamiana plants

Agrobacterium tumefaciens strains were cultured to an optical density of 600 nm (OD600) of 0.5-0.8. The cells were then collected by centrifugation at 1000 g for 30 min. The pellets were then resuspended in Agro-infiltration solution (3.2 g/L Gamborg´s B5 medium and vitamins, 20 g/L sucrose, 10 mM MES pH 5.6, 200 µM 4´-Hydroxy-3´-5´-dimethoxyacetophenone) to a final OD600 of 1.0, followed by incubation at room temperature for 1 h, with gentle agitation. The suspension was then used for need-less infiltration of the abaxial leaf epidermis through the stomata of *N. benthamiana* plants (Kapila, De Rycke, Van Montagu, & Angenon, 1997).

3.5 Tissue sampling

N. benthamiana leaf tissue samples were collected at 2-8 dpi. Four leaf discs were collected from 3-5 biological replicates in 2 mL tubes with 3 small zircon/glass/plastic beads (11079125z, BioSpec Products Inc). For confocal microscopy purposes, leaf discs were collected right before monitoring. For protein extraction purposes, the leaf tissue was weighed and immediately frozen in liquid Nitrogen and stored at -80 C until used.

3.6 Confocal microscopy

Protein subcellular visualization was determined by imaging the abaxial epidermal cells of leaf samples, with an Olympus LSM FV1200. Different lasers allowed for imaging of the different fluorescent tag fusion proteins. For GFP imaging, the tag was excited with a 488 Argon laser and detected at 500-545 nm. For eYFP imaging, the tag was excited at 515 and detected at 530-545 nm. For mRFP imaging, the tag was excited at 559 nm with a He/Ne laser and detected at 570-545 nm. The Imaris software (version 7.6.1, Bitplane Scientific Software,

Bitplane, Zurich, Switzerland) was used to generate 3D images from z-stack confocal images. Line-sequential scanning mode was used for co-localization imaging.

3.7 Protein extraction and quantification

Four leaf discs (approximate fresh weight of 30 mg) from at least three biological replicates per sample were collected with a 7 mm diameter cork borer and put into a 2 ml tube with three 2.3 mm zirconia/silica beads (Bio Spec Products Inc, Cat. No. 11079125z) and frozen in liquid N_2. Collected leaf discs were pulverized in a Mixer Mill (Retsch, Haan, Germany) for 1 min at 30 Hz, in previously frozen homogenizer blocks. Pulverized tissue was spun down for 1 min at 1000 g and 300 µL Protein Extraction Buffer (PBST0.1%, 2% PVPP, 1mM EDTA pH 8.0, 1mM PMSF, 1µg/mL Leupeptin, 100 mM Sodium L-ascorbate) or Reducing Extraction Buffer (50 mM Tris, pH 8.0, 1% SDS, 20 mM DTT) were added to the sample. Samples were vigorously vortexed 3 times for 5 sec and centrifuged at 20000 g for 15 min at 4°C. The resulting cleared supernatant was transferred to a new tube and TSP (Total Soluble Protein) was quantified using the Bio-Rad Bradford (Bradford, 1976) protein assay reagent (Bio-Rad, Cat. No. 5000006).

3.8 Serological procedures

Serological assays were performed by resolving the protein sample in a sodium dodecylsulphate-polyacrylamide gel electrophoresis (SDS-PAGE), and transferred to a PVDF or Nitrocellulose membrane. Recombinant proteins were detected with a 1:5000 dilution of the primary mouse anti-c-Myc monoclonal antibody (GenScript, Cat. No. A00864) or 1:5000 dilution of the primary rabbit anti-GCSF polyclonal antibody (GeneTex, Cat. No. GTX31157), and 1:3000 dilution of the goat anti-mouse IgG HRP-conjugated secondary antibody (Bio-Rad, Cat. No. 170-6516) or 1:3000 dilution of the goat anti-rabbit IgG-conjugated secondary antibody (Bio-Rad, Cat. No. 1706515). In the case of lectin blots, a 1:1000 VVA-HRP conjugate (EY Laboratories, Cat. No. H-4601-1) was used. Blotted membranes were visualized with the enhanced chemiluminescence (ECL) detection kit (GE Healthcare, Cat. No. RPN2232), following the manufacturer recommendations, and imaged with the DNR Bio-Imaging System MicroChemi (RANCOM A/S, Birkerød, Denmark). Band analysis was performed using the TotalLab TL 100 software (Nonlinear Dynamics, Durham, NC).

3.9 Protein salt precipitation

One gr or 10 gr leaf was homogenized with pestle and mortar applying sufficient Protein Extraction Buffer, until acquiring a well homogenized extract. The recovered extract was pellet by centrifugation at 12000 rpm, 12 min, and cleared pellet was passed to an appropriate tube. Several ammonium sulfate concentrations were tested, ranging from 0.6 to 1.6 M. Salts were added slowly to the protein extract under slow agitation. Protein sample was then centrifuged 12000 rpm, 12 min, and the recovered pellet was resuspended in $PBST^{0.1\%}$. The resolubilized sample was centrifuged again under the same conditions, and the cleared supernatant was recovered. Resolubilized pellet and soluble fractions were used for serological assays or gel staining. Protein gel staining was performed by using the GelCode Blue Staining Reagent (Cat Nr 24590, Bio-Rad, Hercules, CA, USA).

3.10 Protein purification

Protein extract was transferred to a spin column, and anti-c-Myc beads were added to the column, following the manufacturer instructions (MBL International, Cat. No. 3305). The sample was incubated at 4°C for 1 h, with gentle agitation end-over-end. The column was briefly centrifuged for 10 sec, and the flow through was recovered. Three washes were performed by adding 200 µL of the washing buffer provided with the kit, and briefly centrifuging. Finally, 20 µL of elution peptide was added to the column, the sample was incubated for 5 min at 4°C, and the purified protein was recovered by centrifuging the column for 10 sec. Elution was repeated twice.

3.11 Analysis of O-glycosylation by mass spectrometry

SDS-PAGE excised bands were Trypsin/Chymotrypsin (6.25 mg/L each) digested (Finnie, 2007) at 30°C overnight. The peptide digests were analyzed using an Easy-nLC 1000 nano-flow system with a 100 µm x 2 cm Acclaim C18 PepMap™ trap column and a 75 µm x 15 cm Acclaim C18 PepMap™ analytical (Thermo Scientific, MA, USA) coupled to a Q-Exactive™ Quadrupole Orbitrap mass spectrometer (Thermo Scientific, MA, USA). The flow rate was 300 nL min^{-1} and 10 µL of the protein digest was injected. 97 % mobile phase A (LC/MS Optima

water, 0.1 % formic acid) was decreased to 90 % over 3 min. Peptides were eluted with a linear gradient from 10-35 % mobile phase B (LC/MS Optima acetonitrile 0.1 % formic acid) over 21 min followed by 35-90% over 3 min and maintained for 8 min. The nanospray voltage was set at 2.1 kV, capillary temperature 275°C, and S-lens RF level 55. The Q-Exactive was operated in top 10 data-dependent acquisition mode with a full scan mass range of 400 to 2000 *m/z* at 70,000 resolution, automatic gain control (AGC) of 1e6 and maximum injection time (IT) of 250 ms. The MS/MS scans were acquired at 17,500 resolution, AGC of 2e5, maximum IT of 50 ms, intensity threshold of 8e4, normalized collision energy of 27 and isolation window of 1.2 *m/z.* Unassigned, singly and >4 charged peptides were not selected for MS/MS and a 20 s dynamic exclusion was used. The Thermo .raw files were converted to mascot generic format using Proteowizard v2 (Kessner, Chambers, Burke, Agus, & Mallick, 2008) and the MS/MS scans were searched against the target/reverse human G-CSF amino acid sequence and the *N. benthamiana* proteome (Sol Genomics Network, accessed Jan 10th, 2015) using X! tandem (Craig & Beavis, 2004) search algorithm operated from the SearchGUI v.2.35 (Vaudel, Barsnes, Berven, Sickmann, & Martens, 2011) interface and processed in PeptideShaker v1.3.6 (Vaudel et al., 2015). A 3 ppm precursor ion mass error and a 0.02 Da product ion error were used along with carbamidomethylation as a constant modification and oxidation of methionine, Hex(1)NAc(1) of Thr, HexNAc of Thr as variable modifications. A 1 % FDR rate was used at the protein, peptide and peptide spectrum match level.

3.12 Statistical analysis

The statistical analyses were performed with the Minitab 18 software (Minitab Inc., PA, USA). A one-way analysis of variance (one-way ANOVA) was performed followed by Tukey test to find significance differences between means (statistical difference was defined as $p \leq 0.05$).

Chapter 4

- **Transient Expression of Human Granulocyte-Colony Stimulating Factor Variants in *N. benthamiana* Plants, as Non-Glycosylated and *O*-glycosylated Plant-Derived Proteins.**

4.1 Abstract

Expression of economically relevant proteins in alternative expression platforms, especially plant expression platforms, has gained significant interest in recent years, due to the possibility of reducing production costs and enhancing product quality. Among the different qualities that plants can offer for the production of recombinant proteins, protein glycosylation is one of the crucial ones since it has an impact on pharmaceuticals functionality and/or stability. In this work, the pharmaceutical glycoprotein hG-CSF is transiently expressed in *N. benthamiana*, as several versions, being targeted to the apoplast, cytoplasm and as protein bodies, offering an alternative for the consideration of production of this protein. Furthermore, the glycoprotein was subjected to the native GalNAc-*O*-glycosylation, by co-expressing the pharmaceutical, together with the enzymes responsible for such glycosylation.

4.2 Introduction

Plants have emerged as alternative expression systems for the production of pharmaceutical proteins (Daniell et al., 2009; Faye et al., 2005; Ko & Koprowski, 2005; Sharma & Sharma, 2009). There are several advantages concerning the usage of plants as expression systems, such as low cost of production and maintenance, fast scalability, biological safety, proper protein folding and assembly (Daniell et al., 2009; Faye et al., 2005; Sharma & Sharma, 2009). Nevertheless, there are also some limitations concerning the expectance of bioactivity of the produced proteins, due to PTM´s (Daskalova et al., 2010). Although plants are known to perform PTM´s, just like mammalian cells do, differences in glycosylation might represent a major drawback that can impact the biochemical properties of the plant-derived recombinant

protein (Bardor et al., 2003; Daskalova et al., 2010; Faye et al., 2005; Gomord et al., 2010a; Jin et al., 2008; Yang, Drew, et al., 2012).

Concerning the engineering of glycosylation in plants, major efforts have been focused to imitate *N*-glycan structures from human proteins, in plant-derived proteins (Daskalova et al., 2010; Parsons et al., 2013). Recently, fewer but significant advances have been achieved concerning engineering of *O*-glycosylation in plants, whose *O*-glycans differ significantly from the typical Mucin-type glycans, and thus stimulates the production of antibodies in mammals (Parsons et al., 2013). Among the recent achievements, Daskalova and co-workers (Daskalova et al., 2010) successfully expressed an *O*-glycosylated human Mucin 1 peptide derivative, which was detected exclusively as a glycoform (Daskalova et al., 2010). This was achieved by transiently co-expressing in *N. benthamiana,* the *Yersinia enterocolitica* UDP-GlcNAc 4-epimerase, in charge of converting UDP-GlcNAc to UDP-GalNAc in the cytoplasm, a *Caenorhabditis elegans* UDP-GlcNAc/GalNAc transporter, responsible for the transport of the sugar donor to the Golgi lumen, and the human GalNAc-Transferase 2, responsible for the specific attachment of the GalNAc sugar to the substrate.

In 2012, Yang and colleagues (Yang, Drew, et al., 2012) were also able to generate Mucin-type *O*-glycosylation in *N. benthamiana*, using a similar approach. They transiently expressed a *Pseudomonas aeruginosa* UDP-GlcNAc 4-epimerase, together with the human GalNAc-T 2 and 4, and a human 3.5 tandem repeat of Mucin1. The derivative mucin peptide was GalNAc-type *O*-glycosylated with up to three and five GalNAc residues, when expressed with GalNAc-T2 and GalNAc-T 2 and 4, respectively. The mucin-type *O*-glycosylation was also demonstrated by mass spectrometry on other protein substrates, tandem repeat of MUC16 and interferon α2b.

The hG-CSF is a cytokine that stimulates the production, proliferation, differentiation and activation of the neutrophil stem cells, raising the levels of neutrophils in blood stream, and thus protecting the organism against microbial infections (Broudy et al., 1987; Kaushansky et al., 1988; Nicola & Metcalf, 1985; Vellenga et al., 1988). This cytokine is an important glycoprotein that is used to treat patients with neutropenia, in order to reduce opportunistic infections, especially in patients under chemotherapy or radiotherapy, or other pathologies causing neutropenia. Due to this clinical relevance, this protein is one the most widely sold pharmaceutical, and it is the most common pharmaceutical therapeutic used in cancer treatment, thus making the production of this pharmaceutical for clinical use very important

(Duncan et al., 1997). hG-CSF is a Mucin-type *O*-glycosylated, with a GalNAc glycan at the residue T-133. It has been successfully expressed in CHO cells, *E. coli*, yeast and plants (Bae et al., 1999; Lasnik et al., 2001; Mire-Sluis, Das, & Thorpe, 1995; Sharifi Tabar et al., 2013; Suk Yang et al., 1997; Tabar et al., 2012). Nevertheless, only the mammalian platform version is reported to be glycosylated. Several studies have shown the impact of the glycosylated version vs non-glycosylated (C. R D Carter, Keeble, & Thorpe, 2004; Clive R D Carter, Whitmore, & Thorpe, 2004; Höglund, 1998; Mattii et al., 2005), pointing out the relevance of production and glycosylation of this pharmaceutical.

The aim of this project was to transiently express the hG-CSF, as a model pharmaceutical protein in *N. benthamiana* plants, via *Agrobacterium* infiltration. The expression of the recombinant protein was explored in 3 different versions in order to compare expression levels: cytoplasm targeted (Cyt-G-CSF), secreted (Sec-G-CSF), and fused with the Zera protein (Zera-G-CSF) (Conley, Joensuu, Richman, & Menassa, 2011; Joseph et al., 2012; Torrent et al., 2009) (**Figure 4-1**). The secretion gene version was later co-expressed with the genes required for the synthesis of the *N*-acetyl-galactosamine (GalNAc)-*O*-glycosylation: human GalNAc-Transferase (GNT2), *Yersinia enterocolitica* UDP-GalNAc 4-epimerase, and *Caenorhabditis elegans* UDP-GlcNAc/GalNAc transporter (**Figure 1-1**).

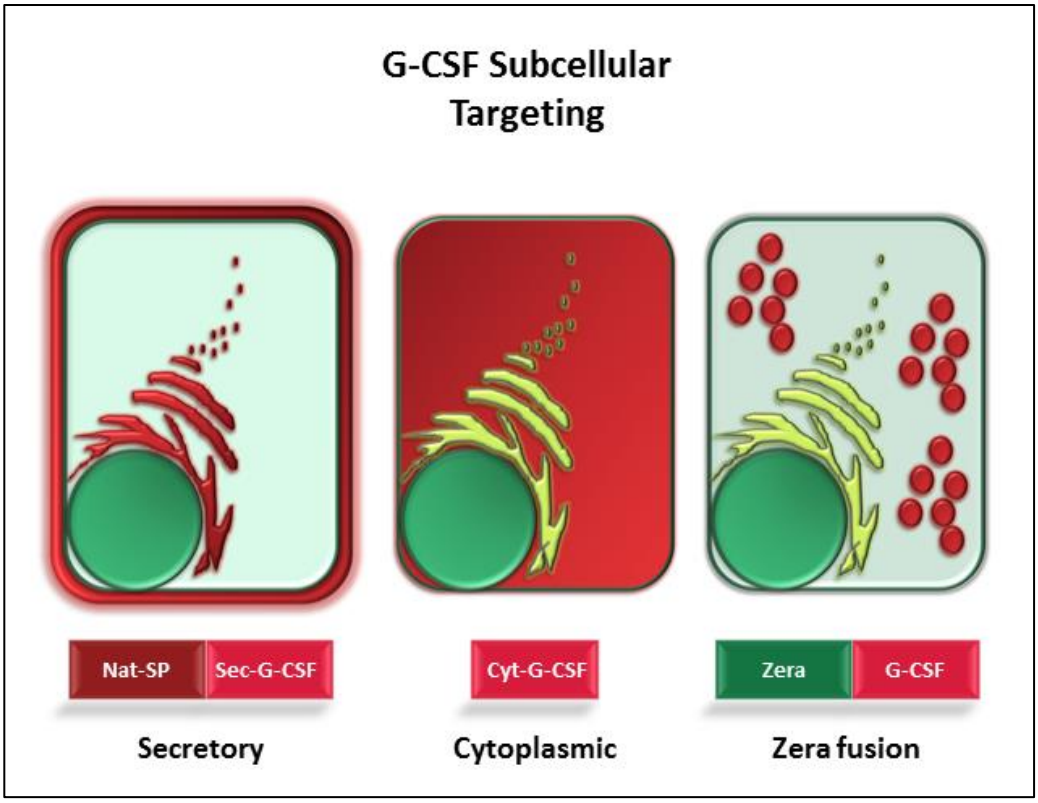


Figure 4-1 Subcellular targeting of G-CSF variants

Schematic representation of the subcellular final destination for each G-CSF variant. Sec-G-CSF: Secretory G-CSF using the native secretory signal peptide (Nat-SP) was targeted to the secretory pathway (represented in red): ER-Golgi-Apoplast. Cyt-G-CSF: cytoplasmic G-CSF was targeted to the cytoplasm (represented in red). Zera-G-CSF: Zera fusion G-CSF was targeted into ER-derived protein bodies (represented in red).

4.3 Results and discussion

4.3.1 Generation of constructs with G-CSF variants

The coding sequence corresponding to the hG-CSF variant 2 (NCBI: NM_172219) was chemically synthesized (GenScript, Piscataway, NJ, USA) in three different fragments in order to build the variants used in this study. The three fragments consisted on the nucleotide sequences corresponding to the secretory signal peptide (pUC57 G-CSF Nat-SP), the mature G-CSF fragment with a mutated glycosylation site that codifies for an Ala and an extra 5´-end

Met (pUC57 Cyt-G-CSF A), and a 5´-end enterokinase coding sequence fused to the mature G-CSF fragment with the native glycosylation site (pUC57 Ek-G-CSF). The Sec-G-CSF coding sequence was built by digesting the pUC57 G-CSF Nat-SP with NcoI and Bsu361 and using the digested fragment to replace the 5´-end enterokinase-G-CSF fragment of the pUC57 Ek-G-CSF, which was released with the same restriction enzymes, producing the pUC57 Sec-G-CSF (**Figure 4-2**). The Cyt-G-CSF coding sequence was built by releasing the 3´-end of the pUC57 Cyt-G-CSF A with SapI cutting upstream the non-glycosylation site and downstream the stop codon. The same restriction enzyme was used to digest the pUC57 Ek-G-CSF, releasing the same fragment but containing the native glycosylation site (Thr), and the released fragment was cloned into the SapI opened pUC57 Cyt-G-CSF A, generating the pUC57 Cyt-G-CSF with the glycosylation site (**Figure 4-3**). The Zera-G-CSF coding sequence was built by digesting the pUC57 Ek-G-CSF with NcoI and SacI, releasing the whole fragment, and then it was cloned into a pUC18 ZeraF1V vector available in the lab (Alvarez, Topal, Martin, & Cardineau, 2010), by releasing the F1V fragment with the same enzymes and thus generating the pUC18 Zera-G-CSF (**Figure 4-4**). The different G-CSF coding sequences generated were then PCR amplified to introduce the flanking BbsI A (5´-end Sec-G-CSF: Mt114F – GGGGGGGAAGACATCATGGCTGGACCTGCCACTC), (5´-end Cyt-G-CSF: Mt23F - GGGGGGGAAGACATCATGACACCCTTAGGACC), (5´-end Zera-G-CSF: Mt20F - GGGGGGGAAGACGTCATGAGGGTGTTGCTCGTTGC) and B (3´-end of all G-CSF variants: M113R - CCCCCCGAAGACAGAGCTCGGGTTGAGCAAGGTGAC) sites, and removal of the stop codon to allow in frame fusion with the appropriate tag. The BbsI A site contains the recognition sequence for the restriction enzyme, followed by the specific overhang CATG, and the BbsI B site contains the specific overhang AGCT, allowing future GoldenGate cloning and disruption of the stop codon for further in-frame cloning with the respective tag. The resulting amplified product was cloned in pUC57 vector via EcoRV, and the mutated sites were verified by DNA sequencing and restriction digestions. The resulting pUC57 G-CSF variants without Stop codon (pUC57 Sec-G-CSF BbsI AB w/o Stop, pUC57 Cyt-G-CSF BbsI AB w/o Stop, pUC57 Zera-G-CSF BbsI AB w/o Stop) were then used to clone in the final expression binary vector as previously described (**Figure 3-2**). The binary vectors were used to transform the agrobacterium strain AGL-1 (**Figure 4-5**).

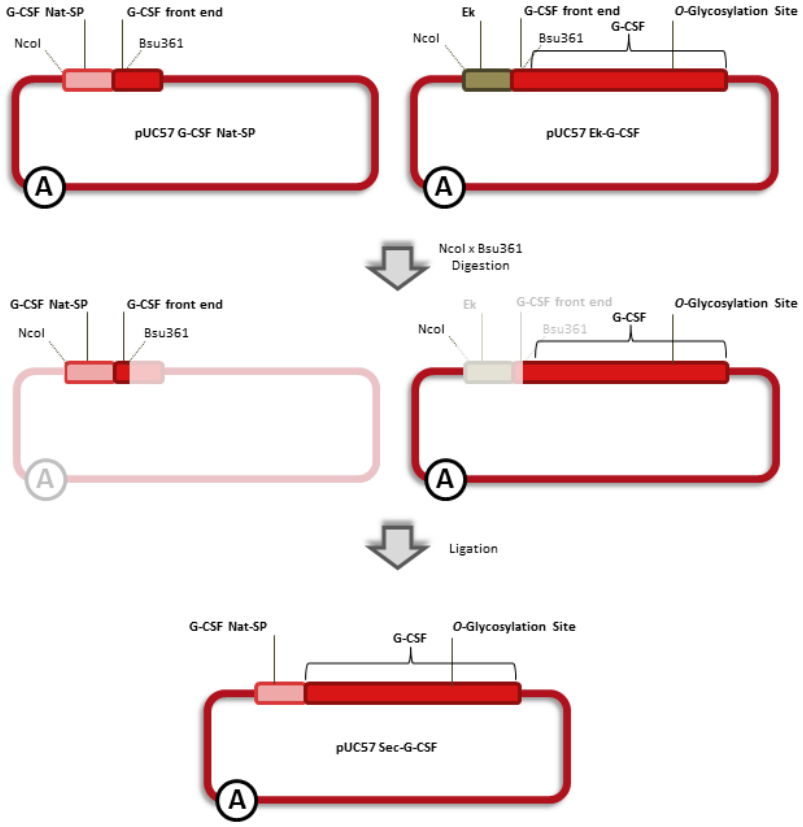


Figure 4-2 Generation of Sec-G-CSF coding sequence

pUC57 G-CSF Nat-SP and pUC57 Ek-G-CSF were digested with NcoI and Bsu361, releasing the G-CSF Nat-SP (signal peptide sequence) and G-CSF front end from pUC57 G-CSF Nat-SP, which was used to substitute the Ek and G-CSF front end from pUC57 Ek-GCSF, generating the pUC57 Sec-GCSF, which contains the native signal peptide sequence (G-CSF Nat-SP) and the mature G-CSF coding sequence. NcoI and Bsu361: restriction enzyme sites. G-CSF Nat-SP: Native signal peptide sequence from G-CSF. Ek: enterokinase cleavage site. G-CSF: mature coding sequence for G-CSF. *O*-glycosylation site: codon location (Thr) for glycosylation. A: ampicillin selectable marker.

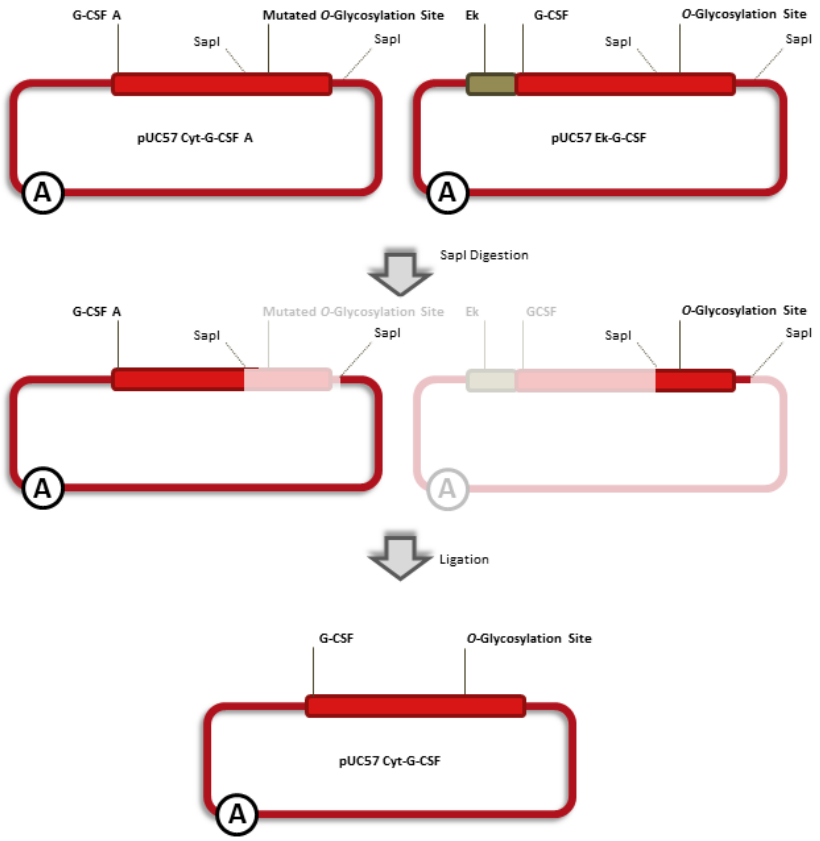


Figure 4-3 Generation of Cyt-G-CSF coding sequence

pUC57 Cyt-G-CSF A and pUC57 Ek-G-CSF were digested with SapI, releasing the G-CSF 3´-end coding sequence which contains the *O*-glycosylation site and part of the pUC57 backbone downstream the G-CSF coding sequence from pUC57 Ek-G-CSF; which was used to substitute the same fragment from pUC57 Cyt-G-CSF A, which does not have a glycosylation site, generating the pUC57 Cyt-G-CSF, which contains the mature G-CSF coding sequence with an extra Met at the 5´-end for cytoplasmic localization. SapI: restriction enzyme sites. G-CSF A: G-CSF mature coding sequence with an extra Met at 5´-end, and a mutation at the glycosylation site for an Ala instead of Thr. Ek: enterokinase cleavage site. G-CSF: mature coding sequence for G-CSF. Mutated *O*-glycosylation site: codon localization for an Ala, which substitute the original Thr sire for glycosylation. *O*-glycosylation site: codon location (Thr) for glycosylation. A: ampicillin selectable marker.

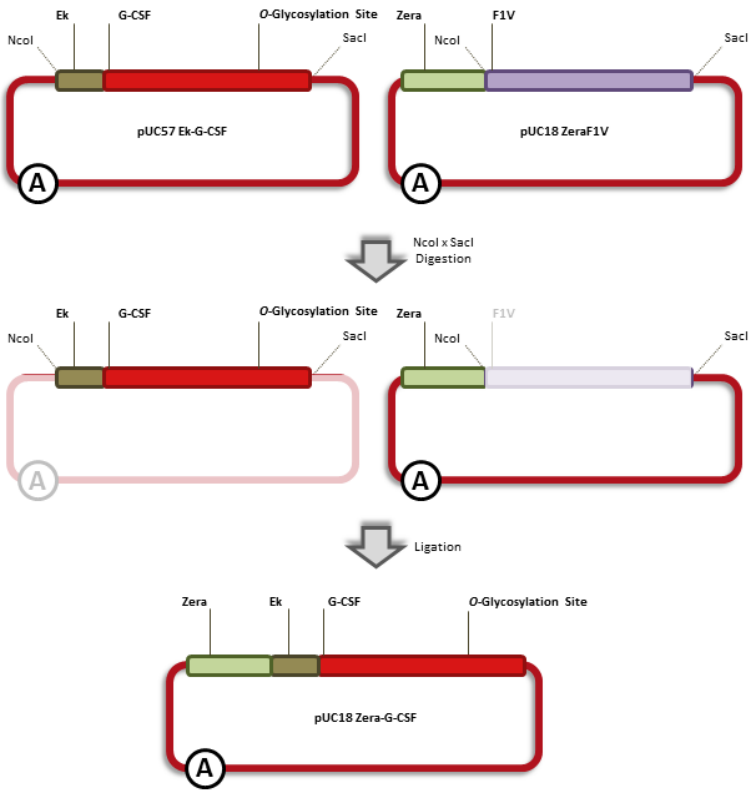


Figure 4-4 Generation of Zera-G-CSF coding sequence

pUC57 Ek-G-CSF and pUC18 ZeraF1V were digested with NcoI and SacI, releasing the F1V coding sequence from pUC18 ZeraF1V, which was then substituted by the Ek-G-CSF coding sequence from pUC57 Ek-G-CSF, to form the pUC18 Zera-G-CSF. NcoI and SacI: restriction enzyme sites. Ek: enterokinase cleavage site. G-CSF: mature coding sequence for G-CSF. *O*-glycosylation site: codon localization (Thr) for glycosylation. Zera: coding sequence for Zera® peptide. F1V: coding sequence for the fusion antigen F1 and V (Alvarez et al., 2010). A: ampicillin selectable marker.

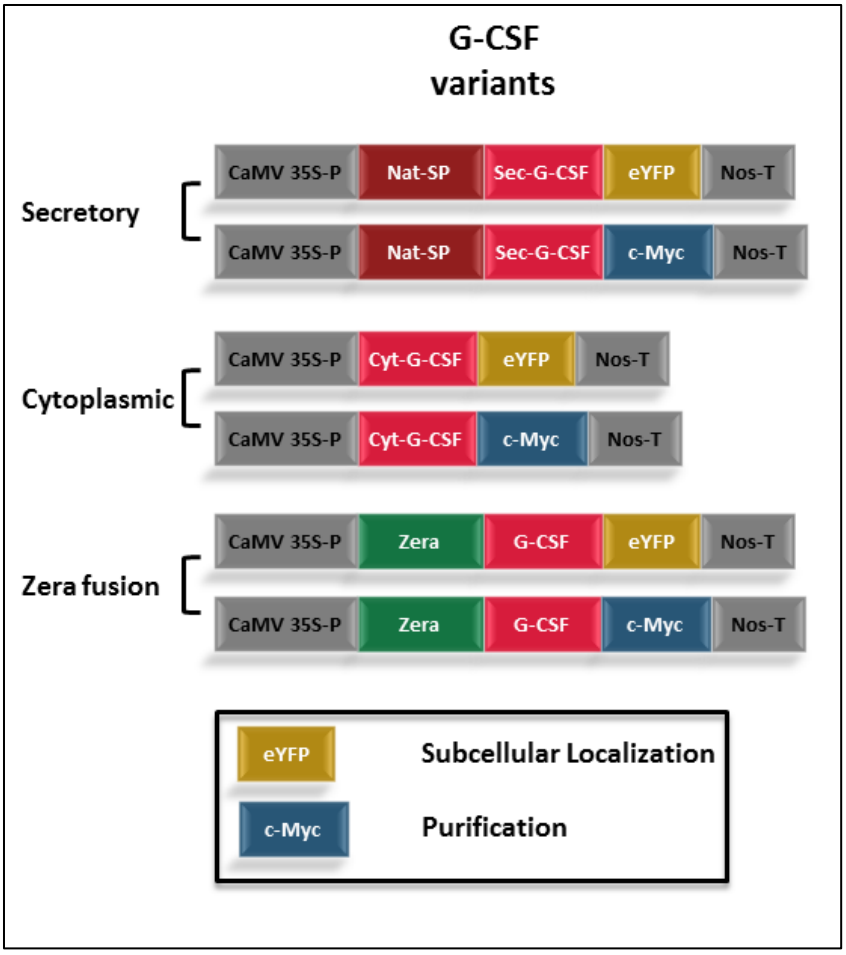


Figure 4-5 G-CSF variants constructs

Schematic representation of the different G-CSF expression cassettes used in this work. All expression unites were under the control of the CaMV 35S promoter (CaMV 35S-P) and Nos Terminator (Nos-T). Secretory G-CSF (Sec-G-CSF) contains the native secretory signal peptide (Nat-SP). Cytoplasmic G-CSF (Cyt-G-CSF) corresponds to the mature protein sequence, and an N-terminus Met. Zera fusion G-CSF (Zera-G-CSF) corresponds to the N-terminally fused Zera peptide to the sequence corresponding to the mature G-CSF. Coding sequences were cloned in frame with eYFP (binary vector pGWB 641), for cellular localization purposes; or c-Myc (binary vector pGWB 617) for purification purposes.

4.3.2 Transient expression of G-CSF variants in *N. benthamiana* plants.

Expression of the different G-CSF variants was first tested by using the G-CSF:eYFP versions, in order to determine the functionality of the expression vector, expression and accumulation of the protein and subcellular localization. This would allow determining the functionality of the vectors being used, cloning strategy, gene design, as well as the functionality of the signal peptide (Sec-G-CSF) and Zera fusion peptide (Zera-G-CSF).

Expression of the different G-CSF:eYFP variants was monitored from 2 to 8 days post-infiltration (dpi), with 3-6 dpi the optimal time frame for signal detection. G-CSF:eYFP variants were successfully expressed and detected by confocal microscopy (**Figure 4-6**). All G-CSF:eYFP variants showed the expected subcellular localization. Sec-G-CSF:eYFP was observed in the apoplast and ER-Golgi network (**Figure 4-6, A-C**), suggesting secretion of the protein and translocation of the protein along the secretory pathway. Sec-G-CSF:eYFP was expressed with and without p19, which would help to decrease accumulation of the protein in the ER, and ease visualization of signal in the apoplast (**Figure 4-6, C**). Cyt-G-CSF:eYFP showed a typical cytoplasmic localization (**Figure 4-6, D**), which is also observed when the tag alone is expressed (CyteYFP: **Figure 4-6, G; Figure 4-10**). Both, Cyt-G-CSF and CyteYFP also showed a clear apoplast (**Figure 4-6, E-F, H-I**), indication of no protein secretion, in comparison to Sec-G-CSF:eYFP, which shows localization in the apoplast (**Figure 4-6, C**). Zera-G-CSF:eYFP showed the expected ER-derived protein body formation due to the fusion of G-CSF to the Zera® protein, and strong signal, suggestion high accumulation levels (**Figure 4-6, J-L**).

Identity of the transiently expressed G-CSF:eYFP variants was further corroborated by western blot (**Figure 4-6, M-N**). Sec-G-CSF:eYFP was detected as a monomeric form with an apparent molecular weight of 49 kDa, which corresponds to the mature G-CSF:eYFP protein (after signal peptide cleavage). The same molecular weight was observed with Cyt-G-CSF:eYFP, confirming signal peptide processing of the Sec-G-CSF:eYFP (**Figure 4-6, M; Figure 4-7**). Furthermore, multimers were observed with Cyt-G-CSF:eYFP (**Figure 4-6, M; Figure 4-7**). In the case of Zera-G-CSF:eYFP a weak band was detected when protein extraction was performed with PBST (PBS+Tween), but the Zera® fusion variant was properly extracted when the plant tissue sample was extracted in denaturing and reducing conditions, using an SDS-DTT containing extraction buffer. Multimers were also observed with Zera-G-CSF:eYFP (**Figure 4-6, M; Figure 4-7**). Cleavage of eYFP was observed with all the different variants, and the cleaved eYFP band was corroborated using the CyteYFP expressing samples (**Figure 4-6, M**). It has to be noted that both Sec-G-CSF and Cyt-G-CSF were successfully extracted with PBST extraction buffer

(**Figure 4-6, N; Figure 4-7; Figure 4-9**), but Zera-G-CSF could only be properly extracted with a reducing extraction buffer (**Figure 4-6, M; Figure 4-7**).

G-CSF:eYFP variants accumulation was then assessed by densitometry analysis (**Figure 4-6, N; Figure 4-10**), with samples collected at 4, 6 and 8 dpi for each variant. We found that Sec-G-CSF reached an accumulation level of 17 mg/Kg F.W. (Fresh Weight), followed by Cyt-G-CSF:eYFP with 40 mg/Kg F.W., while Zera-G-CSF:eYFP reached 117 mg/Kg F.W., at 6 dpi, under reducing and denaturing conditions (**Figure 4-6, N; Figure 4-8**). We also observed that accumulation was stable at 4 and 6 dpi for Sec-G-CSF and Cyt-G-CSF, beginning to drop at 8 dpi, but not significantly (**Figure 4-8, A-B, D**). Zera-G-CSF was the only variant that showed a significant increase in accumulation levels on the time-course, in comparison to the other two variants, reaching up to 135 mg/Kg F.W. at 8 dpi (**Figure 4-8, C-D**).

Nonetheless, the PBST extraction buffer, which is not suitable for the extraction of Zera-G-CSF:eYFP, showed a significant increase in the extraction of Cyt-G-CSF:eYFP vs. the SDS-DTT containing extraction buffer, obtaining up to 123 mg/kg F.W. vs. 40 mg/kg F.W., respectively at 4 dpi. No significant difference was observed in the case of Sec-G-CSF:eYFP (**Figure 4-9**).

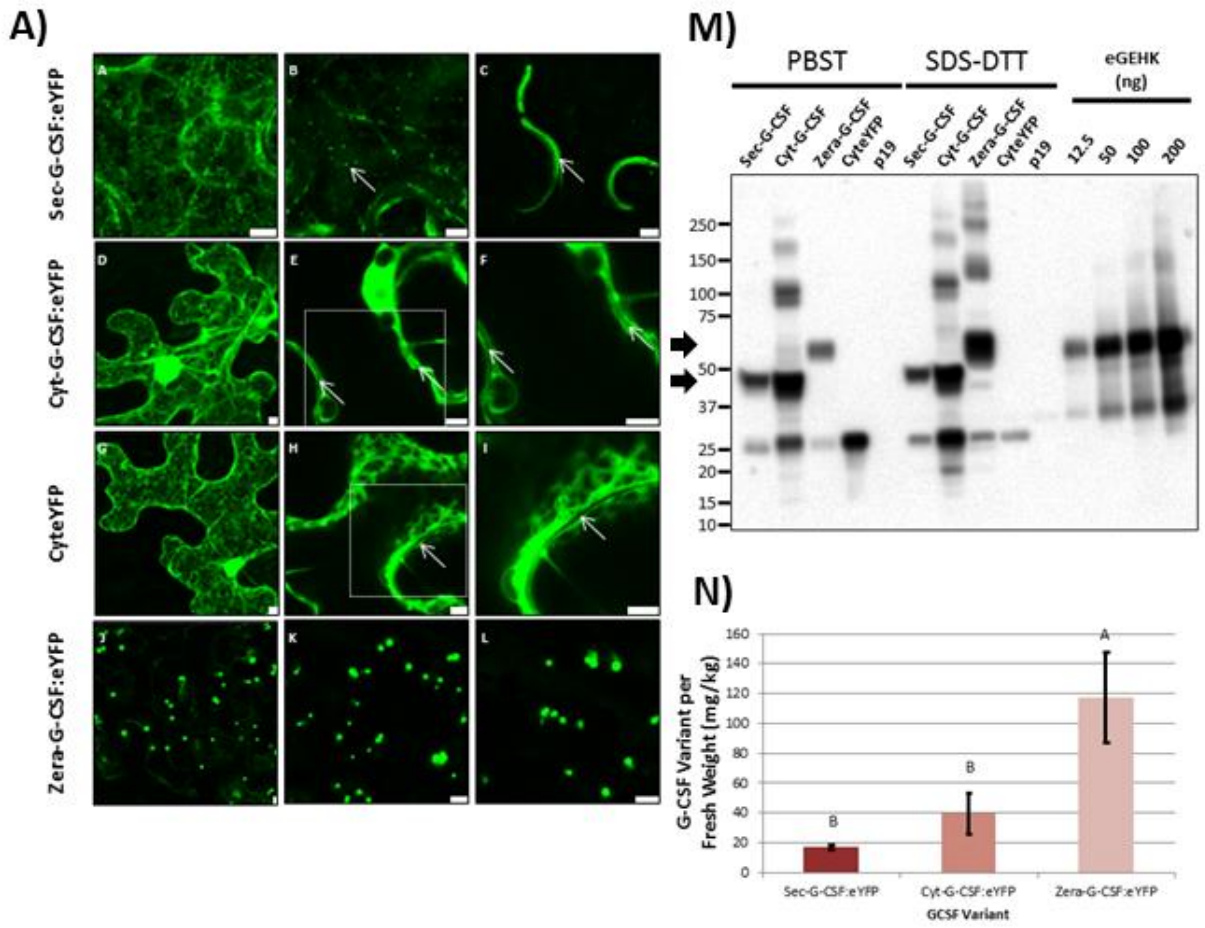


Figure 4-6 Transient expression of G-CSF:eYFP variants in *N. benthamiana* plants

Expression of Sec-G-CSF:eYFP (**A-C**), Cyt-G-CSF:eYFP (**D-F**), CyteYFP (**G-I**), Zera-G-CSF:eYFP (**J-L**). White arrows indicate apoplast localization (**C**) or absence of signal in the apoplast (**E-F, H-I**). White arrow indicating Golgi localization (**B**). Inset white boxes (**E-F**) depict microscopic zoom in shown in **F** and **I**, respectively. Samples collected at 4 dpi. Size bar: 5 µm. **M)** Western Blot detection of Sec-G-CSF:eYFP (Secretory variant, 49 kDa), Cyt-G-CSF:eYFP (Cytoplasmic and mature protein variant, 49 kDa), Zera-G-CSF:eYFP (Zera fused variant, 61 kDa), eYFP (Cytoplasmic eYFP, 29 kDa). All samples were treated with PBS+Tween (PBST) extraction buffer and reducing extraction buffer (SDS-DTT). **N)** Band quantification of SDS-DTT treated samples. Twenty (20) µg TSP of PBS samples were loaded on the gel. Nine (9) µL (equalized volume to the smallest PBS sample amount) of SDS-DTT samples were loaded on the gel. Black arrows denote monomeric G-CSF:eYFP variants. eGEHK: protein standard. Proteins were detected with GFP antibody. Samples collected at 6 dpi. Columns denoted with a different letter are significantly different ($p \leq 0.05$) using one-way ANOVA and followed by Tukey test. Error bars are standard deviations of the means.

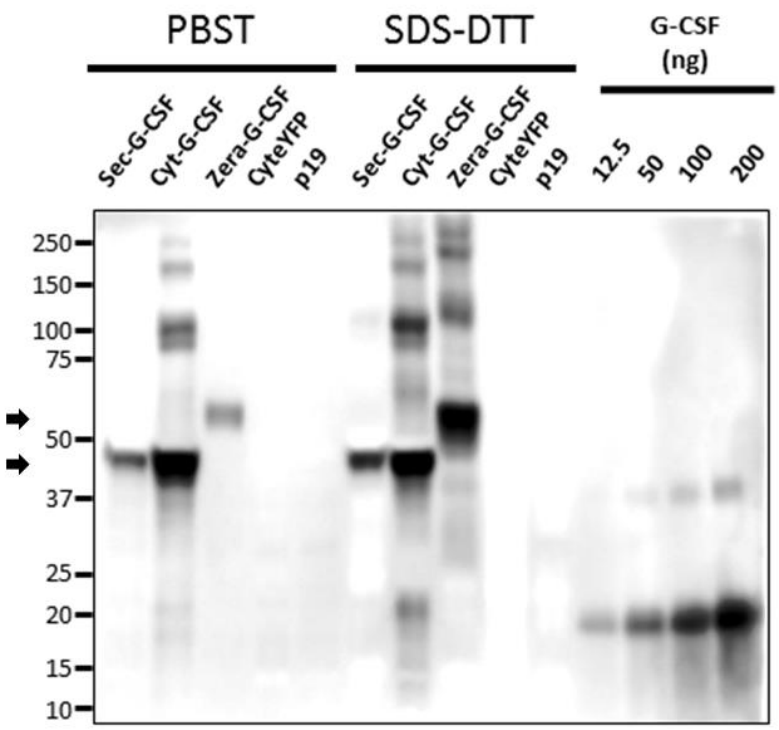


Figure 4-7 Western Blot detection of G-CSF:eYFP variants and multimers with a G-CSF-specific antibody

Sec-G-CSF:eYFP (Secretory variant, 49 kDa), Cyt-G-CSF:eYFP (Cytoplasmic and mature protein variant, 49 kDa), Zera-G-CSF:eYFP (Zera fused variant, 61 kDa), eYFP (Cytoplasmic eYFP, 29 kDa). All samples were treated with PBS+Tween (PBST) extraction buffer and reducing extraction buffer (SDS-DTT). Twenty (20) µg TSP of PBS samples were loaded on the gel. Nine (9) µL (equalized volume to the smallest PBS sample amount) of SDS-DTT samples were loaded on the gel. Black arrows denote monomeric G-CSF:eYFP variants. G-CSF: protein standard. Proteins were detected with G-CSF antibody. Samples collected at 6 dpi.

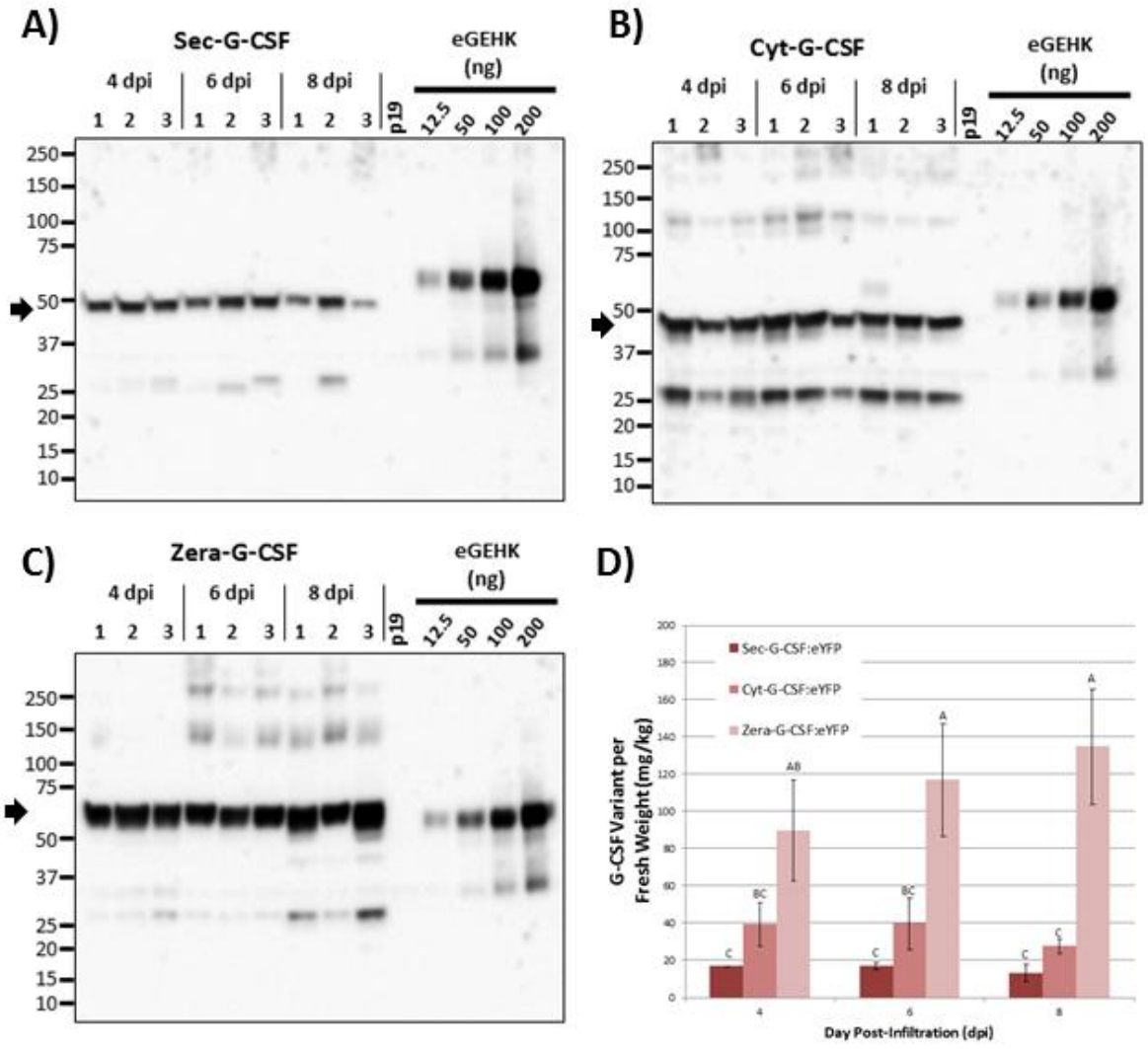
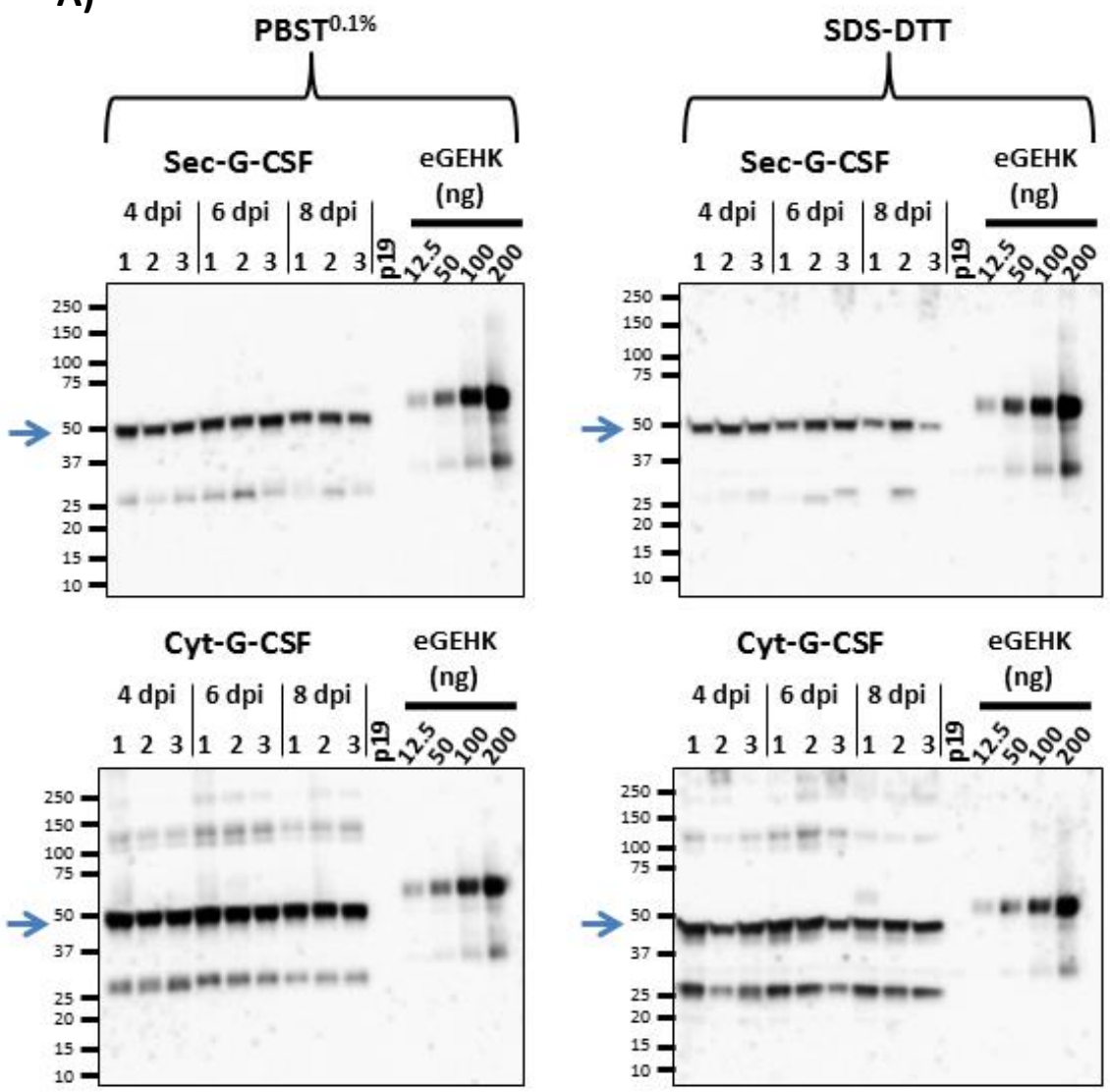


Figure 4-8 Time-course accumulation of transiently expressed G-CSF:eYFP variants in *N. benthamiana* plants

Western blot detection of Sec-G-CSF:eYFP (Secretory variant, 49 kDa) (**A**), Cyt-G-CSF:eYFP (Cytoplasmic and mature protein variant, 49 kDa) (**B**), Zera-G-CSF:eYFP (Zera fused variant, 61 kDa) (**C**). (**D**) Band quantification of Western blot detected proteins. Samples collected at 4, 6 and 8 dpi. Four leaf discs from three different plants were collected for each sample. All samples were treated with reducing extraction buffer (SDS-DTT). Equal amounts were used for all samples. Black arrows denote monomeric G-CSF:eYFP variants. eGEHK: protein standard. Proteins were detected with GFP antibody. Columns denoted with a different letter are significantly different ($p \leq 0.05$) using one-way ANOVA and followed by Tukey test. Error bars are standard deviations of the means.

A)


PBST0.1%
SDS-DTT
Sec-G-CSF
Cyt-G-CSF
eGEHK (ng)
4 dpi
6 dpi
8 dpi
1 2 3 1 2 3 1 2 3
p19
12.5 50 100 200
250
150
100
75
50
37
25
20
15
10

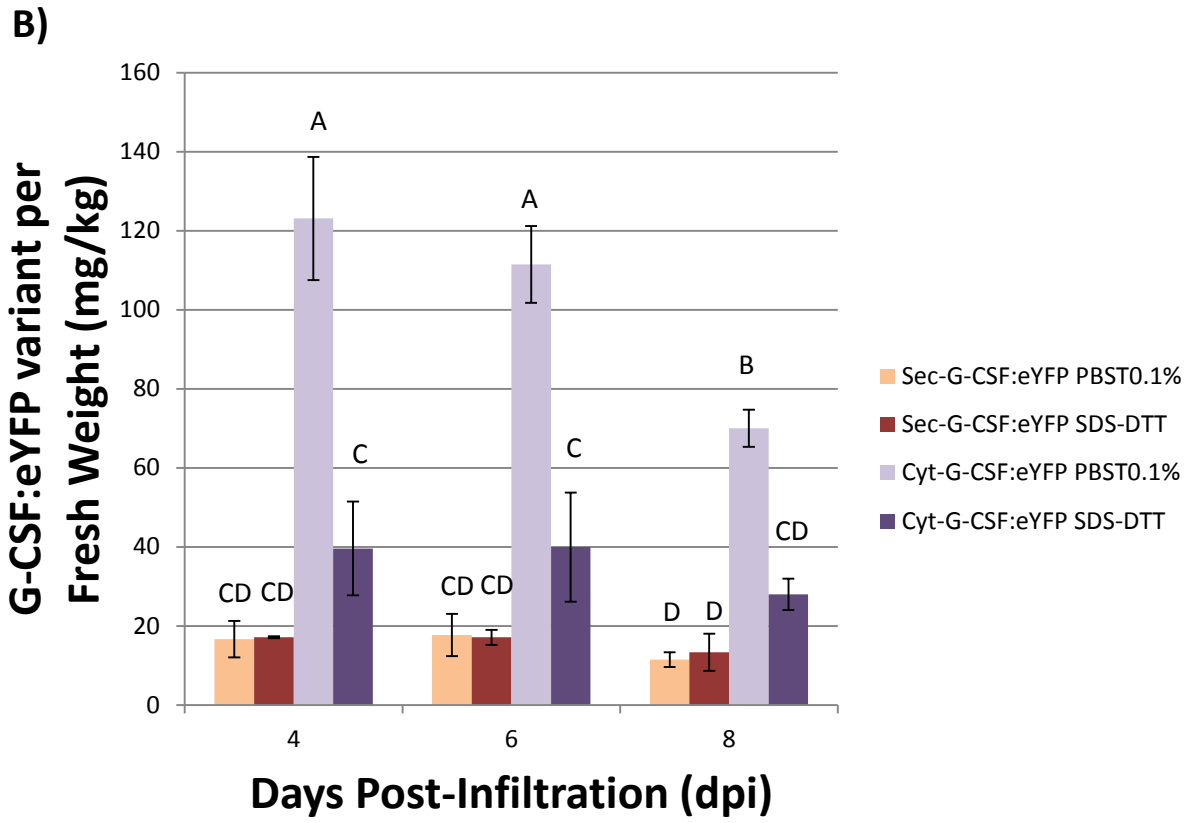


Figure 4-9 Impact of extraction buffer on G-CSF variant

A) Western Blot detection of Sec-G-CSF:eYFP and Cyt-G-CSF:eYFP extracted with PBST0.1% (gels on the left) or reducing extraction buffer (SDS-DTT) (gels on the right). **B)** Band quantification of Western blot detected proteins. Samples collected at 4, 6 and 8 dpi. Four leaf discs from different leaves were collected from each biological sample. 20 µL TSP of $PBST^{0.1\%}$ treated sample or 10 µL SDS-DTT treated sample were loaded on the gel. Blue arrows denote monomeric G-CSF:eYFP variant. eGEHK: protein standard. Proteins were detected with GFP antibody. Band quantification of Western blot detected proteins (Graph). Columns denoted with a different letter are significantly different ($p \leq 0.05$) using one-way ANOVA and followed by Tukey test. Error bars are standard deviations of the means.

4.3.3 Native secretory signal peptide of the recombinant human G-CSF retains its functionality *in planta*

Among the three G-CSF variants produced in this research, only the Sec-G-CSF was a potential candidate for glycosylation, since the *O*-glycosylation event, specifically, the addition of the GalNAc to Ser or Thr substrate sites, only takes place in the secretory pathway, and more specifically, in the Golgi. Therefore, the Sec-G-CSF variant was chosen to be subjected to *O*-glycosylation, by co-expression of the *O*-glycosylation machinery, which is absent in plants. In order to determine the potential glycosylation of Sec-G-CSF *in planta*, secretion of this variant had to be confirmed in a more proper manner.

In order to corroborate Sec-G-CSF Golgi localization and secretion to the apoplast, this variant was co-expressed with a Golgi marker (Rat sialyltransferase or STM:mRFP), and colocalization was determined by confocal microscopy. The *Agrobacterium* strains for this assay were coinfiltrated with and without p19, in order to avoid potential excessive accumulation of the proteins (**Figure 4-10**). Sec-G-CSF:eYFP was found to colocalize in the Golgi together with the Golgi marker STM:mRFP (**Figure 4-11, A-C)**. Although STM:mRFP shows a clear Golgi pattern and very weak signal in the ER, suggesting preference localization for the Golgi, Sec-G-CSF:eYFP was observed in ER and also Golgi. However, SecGFP was used as a positive control for secretion in plant, since it has a plant-specific signal peptide for secretion (PRIb tobacco secretory signal peptide), showing a similar pattern as that observed with Sec-G-CSF:eYFP, and also observed to colocalize with STM:mRFP (**Figure 4-11, E-G)**. GFP:KDEL was used as a negative control for secretion, since this GFP version is retained in the ER and does not get translocated to the Golgi nor apoplast. GFP:KDEL co-expressed with STM:mRFP showed a clear red Golgi (STM:mRFP) over a well-defined green ER network (GFP:KDEL), but not colocalized (**Figure 4-11, I-K**). The latter corroborated that Sec-G-CSF:eYFP was actually localized in the secretory pathway, being observed in the ER and Golgi.

Localization of Sec-G-CSF:eYFP was corroborated subsequently by determining localization in the apoplast by confocal microscopy (**Figure 4-11, D**). As previously described, SecGFP was used as positive control for localization in the apoplast (secretory pathway) (**Figure 4-11, H**) and GFP:KDEL as a negative control for secretion, which clearly showed GFP:KDEL did not localize in the apoplast (**Figure 4-11, L**). The latter corroborated the secretion of Sec-G-CSF:eYFP and therefore the functionality of the native mammalian secretory signal peptide of G-CSF.

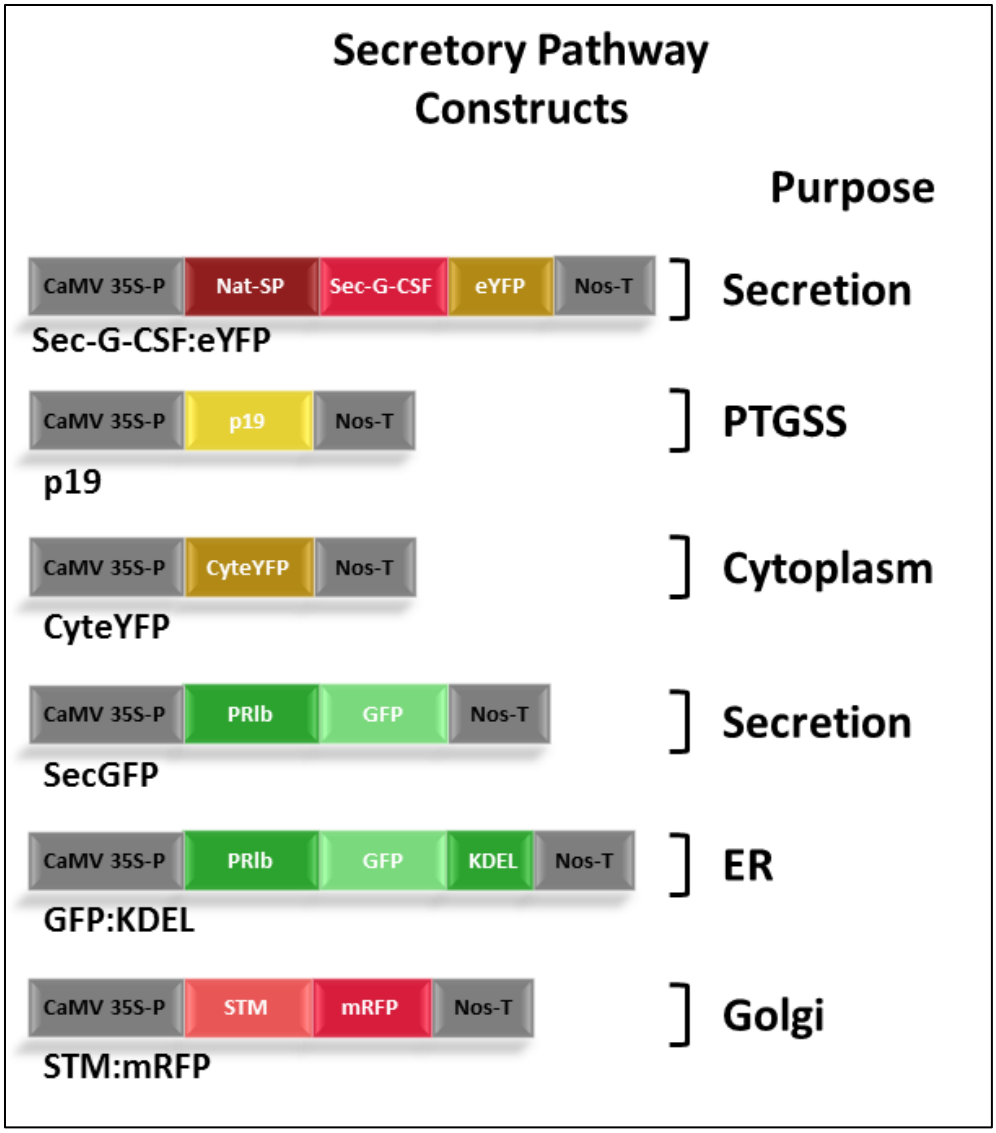


Figure 4-10 Constructs used for secretion corroboration

Schematic representation of the vectors used for confocal purposes and secretion corroboration. Sec-G-CSF:eYFP: Secretory G-CSF:eYFP. p19 was used to allow higher accumulation levels. CyteYFP: Cytoplasmic eYFP (pGWB 642, empty vector). SecGFP: Secretory GFP. GFP:KDEL: ER retained GFP. STM:mRFP: Sialyltransferase fused to mRFP used as Golgi marker. All expression units are under the control of the CaMV 35S Promoter and Nos Terminator. PRlb: Tobacco secretory signal peptide. PTGSS: Post-Transcriptional Gene Silencing Suppressor.

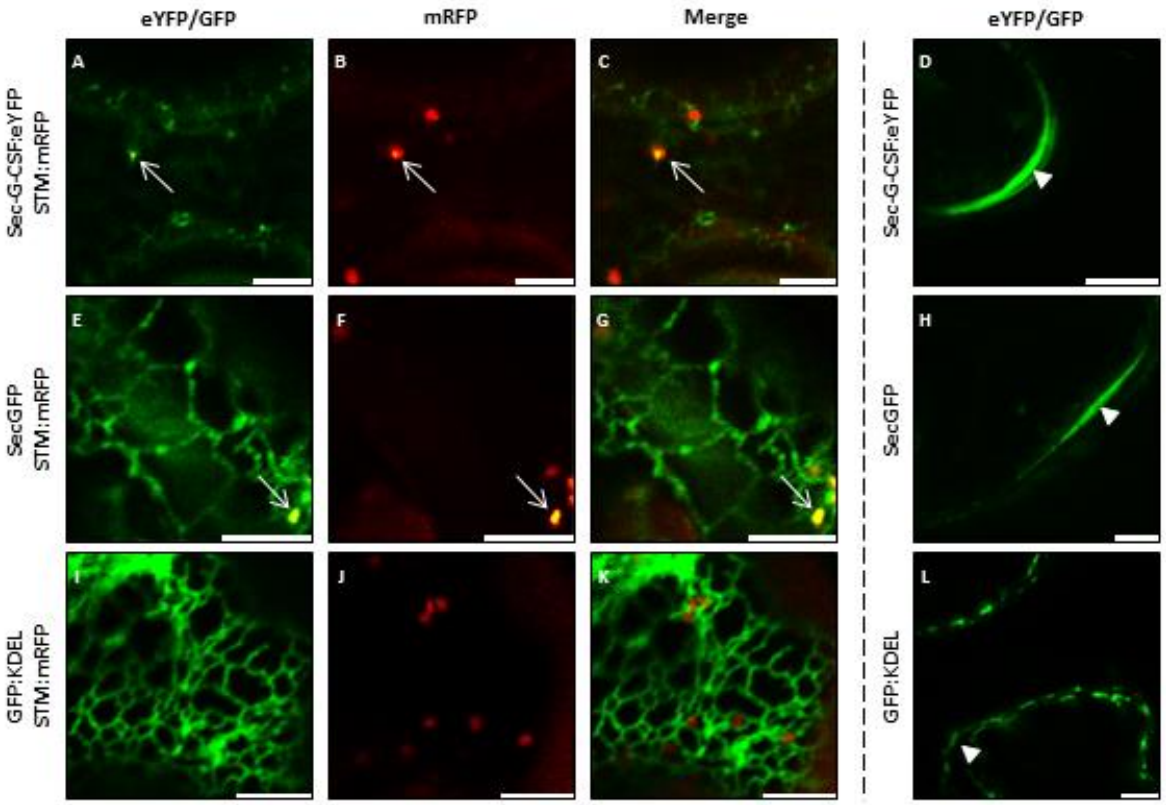


Figure 4-11 Golgi and apoplast localization of Sec-G-CSF in *N. benthamiana* plants

Sec-G-CSF:eYFP (**A, C-D**), SecGFP (**E, G-H**), GFP:KDEL (**I, K-L**). Sec-G-CSF, SecGFP and GFP:KDEL were co-infiltrated with STM:mRFP (**B-C, F-G, J-K**). White arrows depict Golgi localization (**A-B, E-F**) and colocalization (**C, G**) with STM:mRFP. White arrowheads(images on the right) depict localization in apoplast (**D, H**), or apoplast without presence of recombinant protein (**L**). *Agrobacterium* AGL-1 was carrying Sec-G-CSF. SecGFP, GFP:KDEL, STM:mRFP and p19 were delivered by EHA105. 4 dpi. Size bar: 5 µm.

4.3.4 Expression of the Sec-G-CSF variant to evaluate accumulation and co-expression with the Glyco-machinery

Once proper expression and localization of Sec-G-CSF was determined, a C-terminal c-Myc tagged version (Sec-G-CSF:cMyc) was produced for the following assays in which it would be co-expressed with the *O*-glycosylation synthetic pathway (**Figure 4-12**). Expression of Sec-G-CSF:cMyc was first tested and verified by Western Blot (**Figure 4-13, upper panel**). Accumulation of Sec-G-CSF:cMyc was monitored at 4, 6 and 8 dpi (days post infiltration), since a good fluorescence signal by confocal microscopy was usually observed at 4 dpi, but drastically weaker at 3 dpi. Sec-G-CSF:cMyc showed the strongest accumulation at 4 dpi, decreasing at 6 dpi, and almost undetectable at 8 dpi (**Figure 4-13, upper panel**). Molecular size of Sec-G-CSF:cMyc corresponded to the expected one (28 kDa), indicating proper maturation of the premature protein.

When co-expressed with the Glyco-machinery, Sec-G-CSF:cMyc showed a decrease of accumulation in comparison to that observed without the Glyco-machinery (**Figure 4-13, upper *vs.* lower panel**), no increase of accumulation was observed at 6 nor 8 dpi, and Sec-G-CSF:cMyc was barely observed at 6 dpi. Therefore 4 dpi was chosen as the time frame for sample processing for further assays to determine the glycosylation of the protein, but 6 dpi was not discarded.

Since it is assumed that co-expression of Sec-G-CSF with the Glycosylation machinery has a negative impact on protein accumulation due to exhaustive usage of the translational machinery, cell density of the *Agrobacterium* strain containing Sec-G-CSF was increased from OD_{600} 0.4 to 0.7. Increase in Sec-G-CSF ratio improved accumulation levels of Sec-G-CSF (**Figure 4-14**). Finally, Sec-G-CSF:cMyc was found to reach an accumulation level of 49 mg/kg F.W. at 4 dpi, when extracted with PBST buffer, dropping to 31 mg/kg F.W. at 6 dpi, however this slight decrease in accumulation level is not significant (**Figure 4-24**).

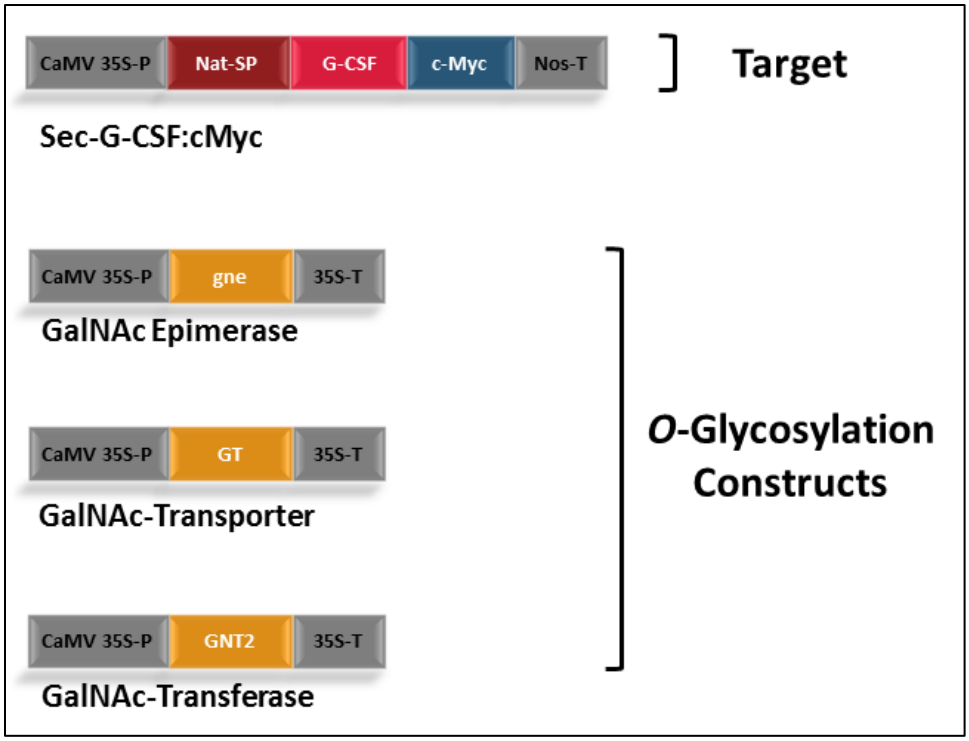


Figure 4-12 *O*-Glycosylation constructs

Schematic representation of the vectors used for the co-expression of the glycosylation machinery. Sec-G-CSF:cMyc: Secretory G-CSF:cMyc, target protein for *O*-glycosylation. gne: *Yersinia enterocolitica* UDPGlcNAc/UDPGalNAc epimerase. GT: *Caenorhabditis elegans* UDPGlcNAc/UDPGalNAc Transporter. GNT2: human GalNAc Transferase 2. All genes were under the control of the CaMV 35S Promoter and 35S Terminator or Nos Terminator.

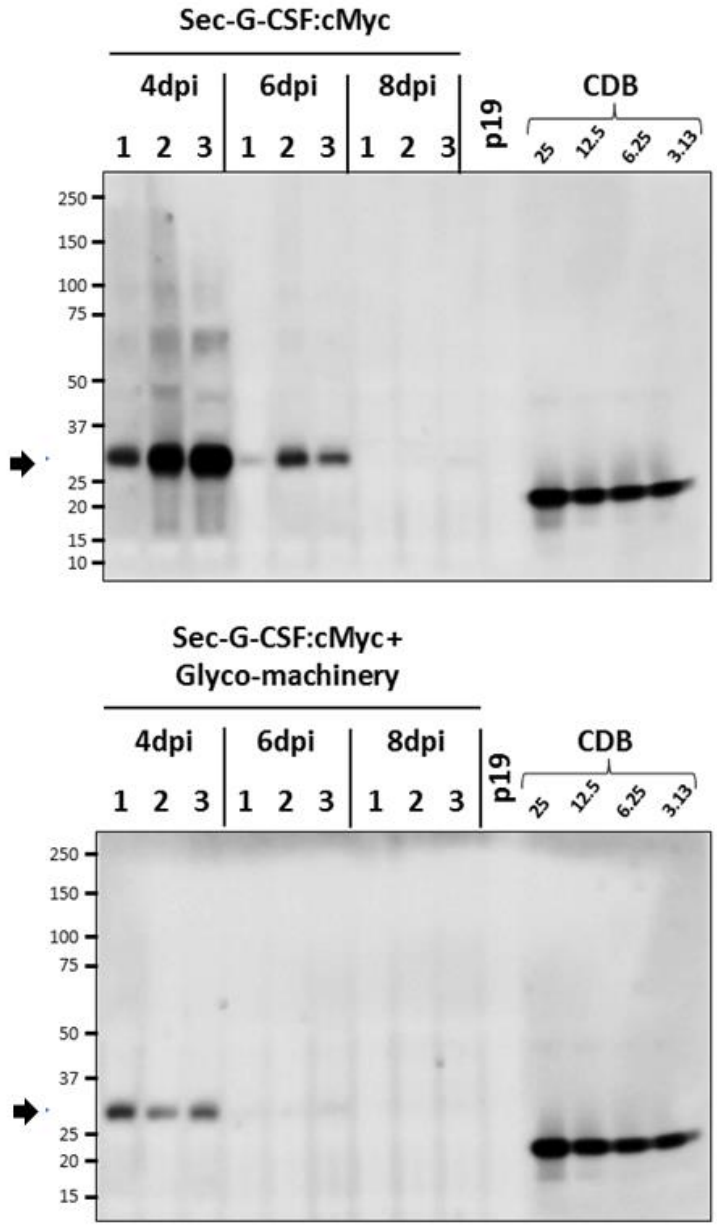


Figure 13-4 Western Blot of transiently expressed G-CSF:cMyc with Glyco-machinery

Upper section: Sec-G-CSF:cMyc co-expressed with p19 (OD_{600} ratios; Sec-G-CSF:cMyc: 0.4; p19: 0.1). **Lower section:** Sec-G-CSF:cMyc co-expressed with p19 and glyco-machinery (O-Glycosylation machinery: UDP-GlcNAc/GalNAc epimerase, UDP-GlcNAc/GalNAc transporter, and GalNAc-glycosyltransferase) (OD_{600} ratios; Sec-G-CSF:cMyc: 0.4; p19: 0.1; glyco-machinery: 0.1/strain). Samples collected at 4, 6, 8 dpi. 1, 2, 3: biological replicates. Forty (40) µg TSP loaded for each sample. CDB: standard. Black arrows: Sec-G-CSF:cMyc (28 kDa). Proteins were detected with c-Myc antibody.

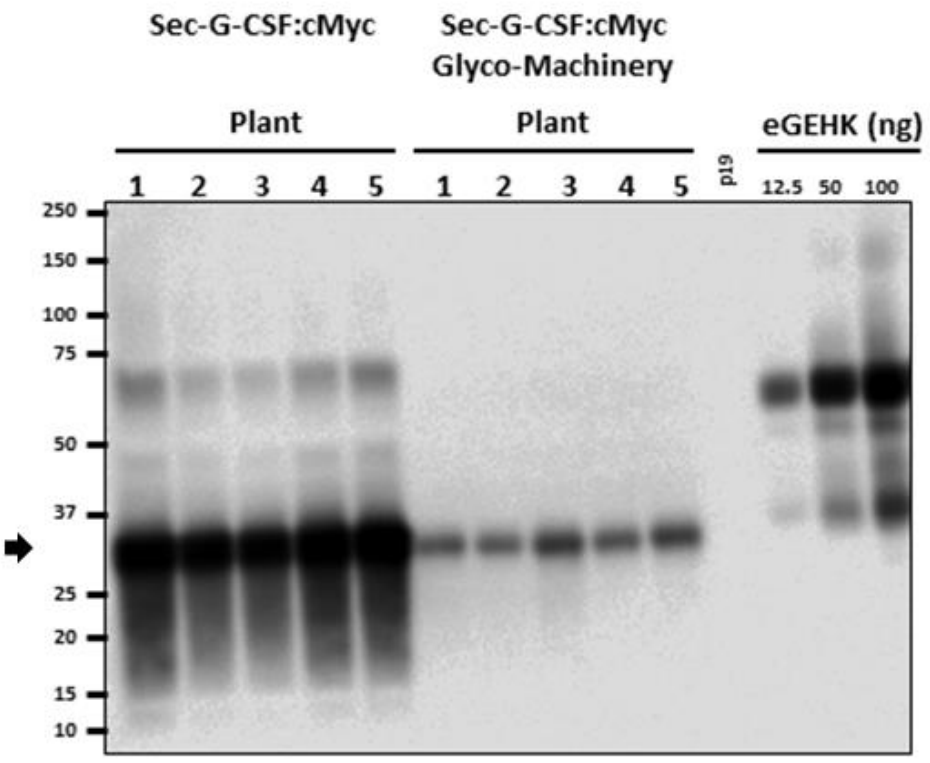


Figure 4-14 Western Blot of transiently expressed G-CSF:cMyc with Glyco-machinery

Sec-G-CSF:cMyc co-expressed with p19 (OD_{600} ratios; Sec-G-CSF:cMyc: 0.7; p19: 0.1). Sec-G-CSF:cMyc co-expressed with p19 and glyco-machinery (O-Glycosylation machinery: UDP-GlcNAc/GalNAc epimerase, UDP-GlcNAc/GalNAc transporter, and GalNAc-glycosyltransferase) (OD_{600} ratios; Sec-G-CSF:cMyc: 0.7; p19: 0.1; glyco-machinery: 0.1/strain). Samples collected at 4 dpi. 1, 2, 3, 4, 5: biological replicates. Twenty (20) µg TSP loaded for each sample. eGEHK: standard. Black arrow: Sec-G-CSF:cMyc (28 kDa). Proteins were detected with c-Myc antibody.

4.3.5 Purification of Sec-G-CSF

In order to evaluate the glycosylation of the model protein Sec-G-CSF, and specifically the GalNAc-*O*-glycosylation, Sec-G-CSF had to be purified and further characterized. For this step, the Sec-G-CSF:cMyc variant expressed was purified by a c-Myc affinity chromatography column.

One of the issues of purifying G-CSF, and the use of the purified fraction to check for glycosylation, is the small amounts obtained after purification (**Figure 4-15; Figure 4-16**). gly-G-CSF (glycosylated form) is expressed about 10-30x less than G-CSF (**Figure 4-15; Figure 4-16**).

By tracking G-CSF during the purification procedures, G-CSF can be observed in the flow through, and wash steps (**Figure 4-17**). The latter might indicate that the column might be saturated. However, multimerization has also been detected, indicating that to some extent, during the purification procedure, G-CSF might be multimerizing, not only into dimers, but also into several multimers or aggregates (**Figure 4-17; Figure 4-18**), which decreases the yields of purified monomeric protein. Gly-G-CSF was barely detected during the purification process (**Figure 4-17**), but that might be due to the GCSF:gly-G-CSF ratio (being gly-G-CSF about 10-30x less than G-CSF).

If such multimerization occurs with gly-G-CSF, but the multimers are not detected since gly-G-CSF is 10-30x less concentrated, and thus multimers are not even visualized by WB (a dimer is barely observed), the low amount of purified gly-G-CSF detected after purification, is thus due in part to the protein aggregation nature, making glycosylation characterization a challenging task.

On a GelCode Blue stained gel, the purified G-CSF and gly-G-CSF clearly show a visible dimer band (G-CSF) (**Figure 4-16**). In the case of gly-G-CSF, the dimer band is almost the same amount as the monomer band. It is necessary to mention that dimers from G-CSF and gly-G-CSF differ in their molecular weight from one another and also that monomeric G-CSF is observed as two bands, while gly-G-CSF is observed as a single but broad band, probably due to glycosylation (**Figure 4-15; Figure 4-16**).

Other platforms previously studied for the expression of human G-CSF, like *S. cerevisiae*, *P. pastoris* and *A. niger*, have reported issues related to protein aggregation, probably due to a lack of proper protein folding during secretion (Bae et al., 1999; Kraševec et al., 2014; Lasnik et al., 2001).

Concerning plant platforms, G-CSF has been expressed in tobacco and rice suspension cells (Hong et al., 2002; Shin et al., 2003), where the protein was secreted to the medium; tobacco plants (Tabar et al., 2012); ER-targeted in BY-2 cells (Nair et al., 2014); and in lettuce chloroplast (Sharifi Tabar et al., 2013). Although none of these works have verified glycosylation of G-CSF, it is known that plants do not possess the glycosylation machinery required to perform GalNAc-*O*-glycosylation (Daskalova et al., 2010), which is the specific glycosylation naturally present in this human protein. Furthermore, these works do not report protein structural properties related to folding and/or multimerization.

4.3.5.1 Salt precipitation of G-CSF

In order to improve the purification of the protein, salt precipitation was assessed. Both G-CSF and gly-G-CSF were best precipitated with 1.6 M ammonium sulfate (**Figure 4-19; Figure 4-20**). G-CSF and gly-G-CSF were found not only in the precipitated fraction but also in the solubilized fraction (**Figure 4-19; Figure 4-20**). This indicates that (gly)-G-CSF could be concentrated by using the ammonium sulfate concentration of 1.6 M, and the resolubilized fraction could be used for purification purposes.

However, purification of a salt precipitated whole-plant (approximately 10 gr) protein extract did not improve the yield of the recovered protein, in comparison to smaller scale (approximately 0.2 gr) purification without salt precipitation (**Figure 4-15 *vs.* Figure 4-16**). Although the salt precipitated sample was dialyzed to remove salts and avoid potential interference during the affinity chromatography, structural changes due to salt precipitation or salts still present in the sample could have hindered a proper affinity binding. Therefore, small-scale protein purification was used in this work.

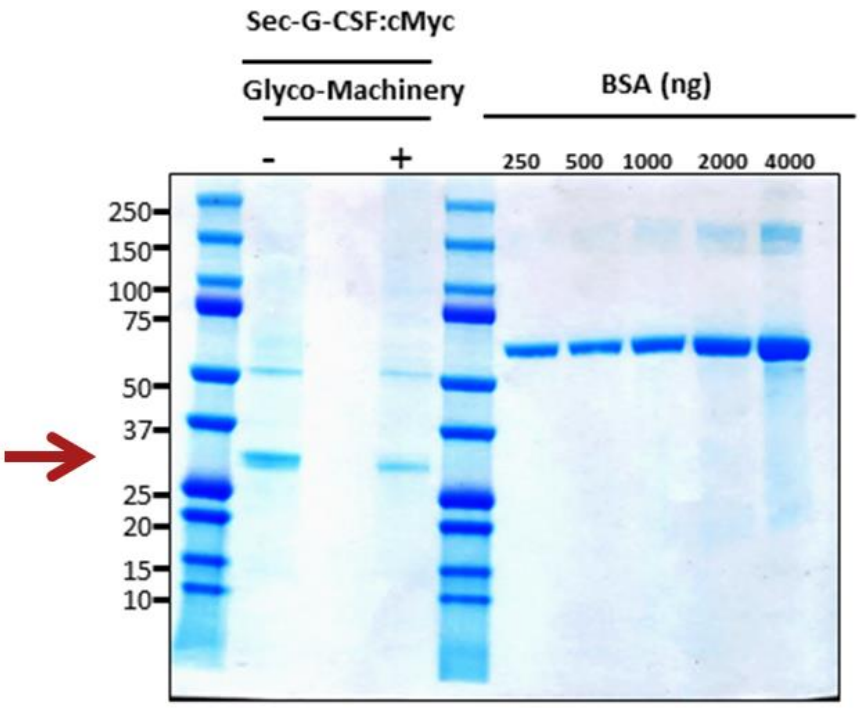


Figure 4-15 GelCode Blue stained gel of purified Sec-G-CSF (Salt Precipitated)

Ten (10) gr of whole plant infiltrated leaves were homogenized and salt precipitated as mentioned in the methodology section, and purified. Purified G-CSF and gly-G-CSF: 20 µL eluent after c-Myc purification was loaded per lane. BSA used as concentration standard. Red Arrow indicates Monomers. Sec-G-CSF:cMyc co-expressed with p19 and glyco-machinery (-/+) (O-Glycosylation machinery: UDP-GlcNAc/GalNAc epimerase, UDP-GlcNAc/GalNAc transporter, and GalNAc-glycosyltransferase) (OD_{600} ratios; Sec-G-CSF:cMyc: 0.7; p19: 0.1; GNT2: 0.2; epimerase/transporter: 0.1/strain).

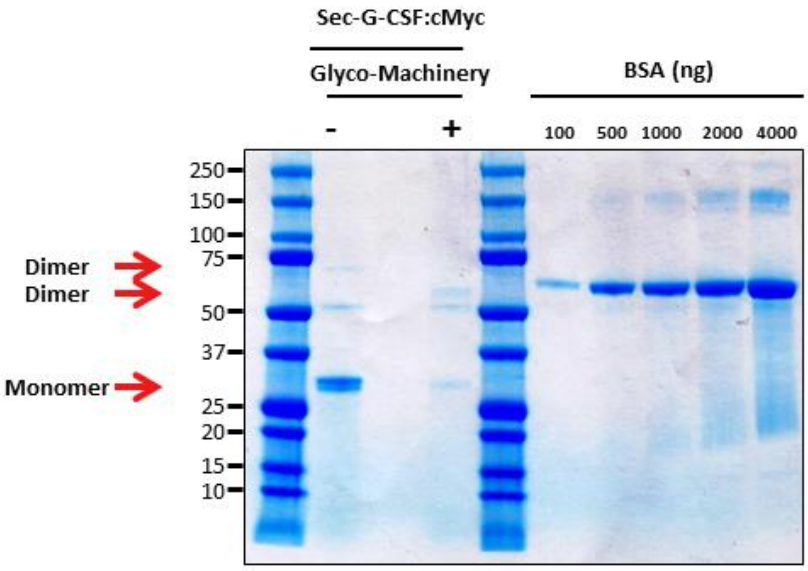


Figure 4-16 GelCode Blue stained gel of purified Sec-G-CSF

Approximately 400 mg leaf discs were homogenized, and TSP was extracted as described in the methodology section, and further purified. Purified G-CSF and gly-G-CSF: 20 µL eluent after c-Myc purification. Sec-G-CSF:cMyc co-expressed with p19 and glyco-machinery (-/+) (O-Glycosylation machinery: UDP-GlcNAc/GalNAc epimerase, UDP-GlcNAc/GalNAc transporter, and GalNAc-glycosyltransferase) (OD_{600} ratios; Sec-G-CSF:cMyc: 0.7; p19: 0.1; GNT2: 0.2; epimerase/transporter: 0.1/strain). BSA used as concentration standard. Red Arrows indicate Monomers and dimers.

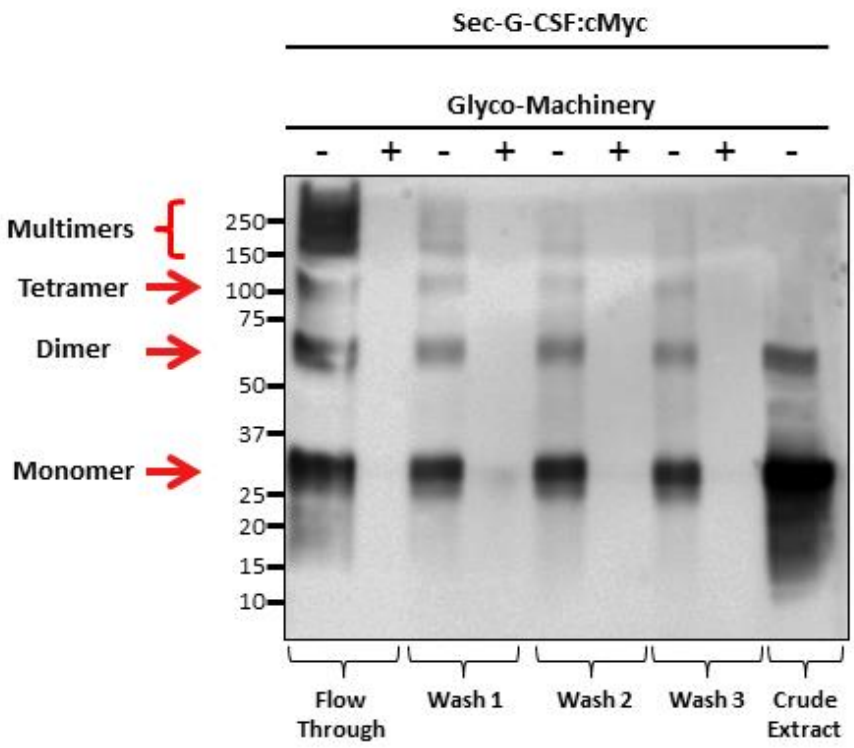


Figure 17-4 G-CSF multimerizes during cMyc purification

Western blot of Sec-G-CSF:cMyc co-expressed with p19 and glyco-machinery (O-Glycosylation machinery: UDP-GlcNAc/GalNAc epimerase, UDP-GlcNAc/GalNAc transporter, and GalNAc-glycosyltransferase) (OD_{600} ratios; Sec-G-CSF:cMyc: 0.5; p19: 0.1; GNT2: 0.2; epimerase/transporter: 0.1/strain). Flow through and wash samples: 30 µL per lane. Crude extract: 20 µg TSP. GCSF monomer and multimers are indicated with red arrows and brace.

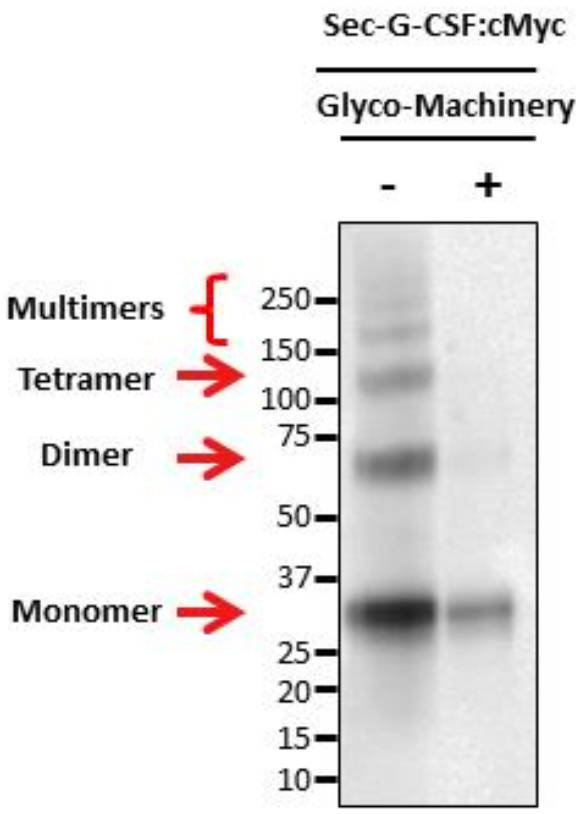


Figure 18-4 Western Blot of purified Sec-G-CSF

Sec-G-CSF:cMyc co-expressed with p19 and glyco-machinery (-/+) (O-Glycosylation machinery: UDP-GlcNAc/GalNAc epimerase, UDP-GlcNAc/GalNAc transporter, and GalNAc-glycosyltransferase) (OD_{600} ratios; Sec-G-CSF:cMyc: 0.5; p19: 0.1; GNT2: 0.2; epimerase/transporter: 0.1/strain). Sec-G-CSF (- Glyco-machinery): 0.2 µL eluent after c-Myc purification; Sec-G-CSF (+ Glyco-machinery): 1 µL eluent (5x more than negative control) after c-Myc purification. Detection using the c-Myc antibody. G-CSF monomer and multimers are indicated with red arrows and brace.

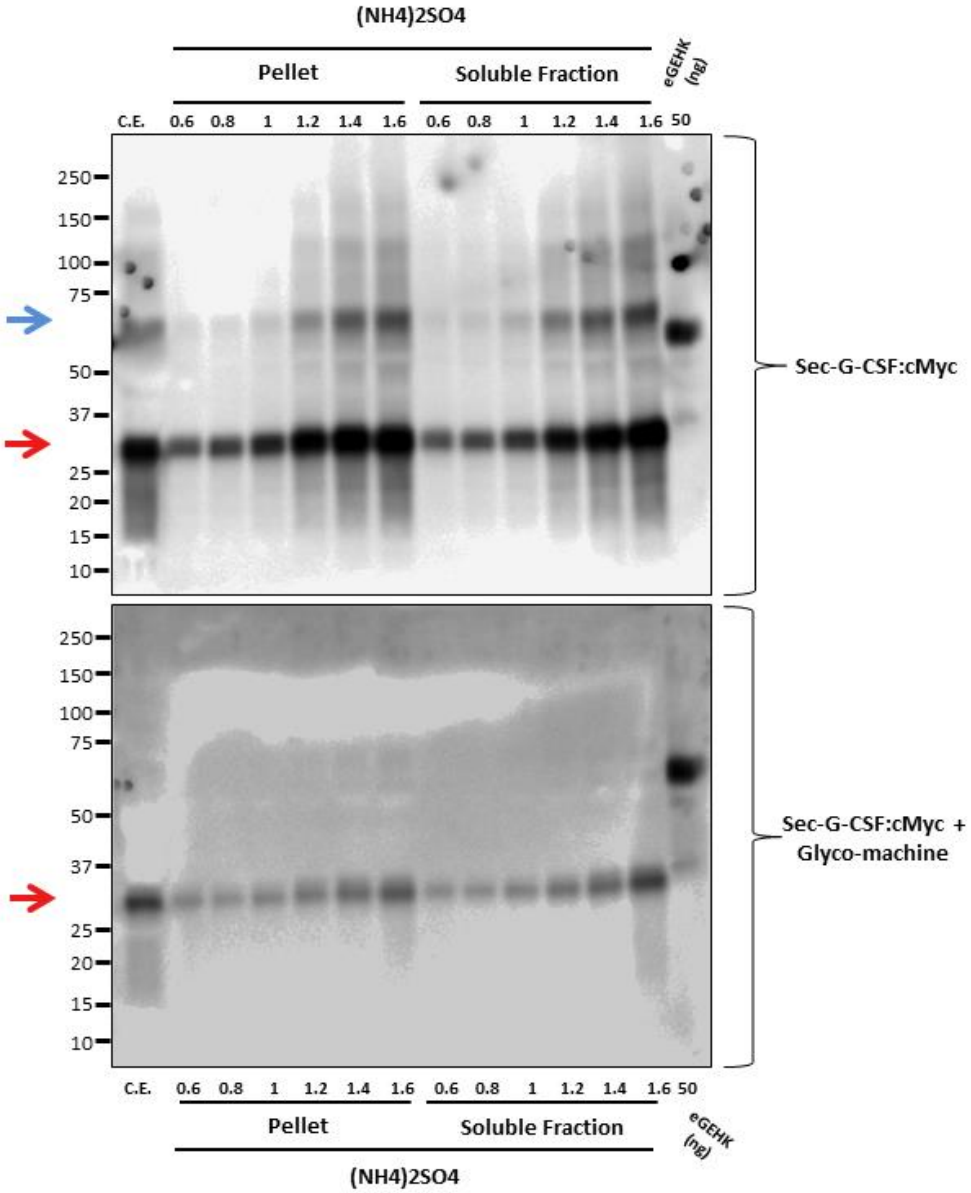


Figure 4-19 Salt precipitation of G-CSF

Western blot of Sec-G-CSF:cMyc co-expressed with p19 and glyco-machinery (*O*-Glycosylation machinery: UDP-GlcNAc/GalNAc epimerase, UDP-GlcNAc/GalNAc transporter, and GalNAc-glycosyltransferase) (OD_{600} ratios; Sec-G-CSF:cMyc: 0.7; p19: 0.1; GNT2: 0.2; epimerase/transporter: 0.1/strain). **Upper panel:** G-CSF without glycosylation machinery. **Lower panel:** G-CSF with glycosylation machinery. Crude extract from 1 gr infiltrated leaf. Equal volumes were subjected to salt precipitation in a range of 0.6 to 1.6 M ammonium sulfate. The gel was loaded with 30 µL TSP, resuspended pellet or solubilized fraction per lane. Red arrows denote monomeric Sec-G-CSF:cMyc. Blue arrow denotes Sec-G-CSF:cMyc dimer. eGEHK standard: 50 ng.

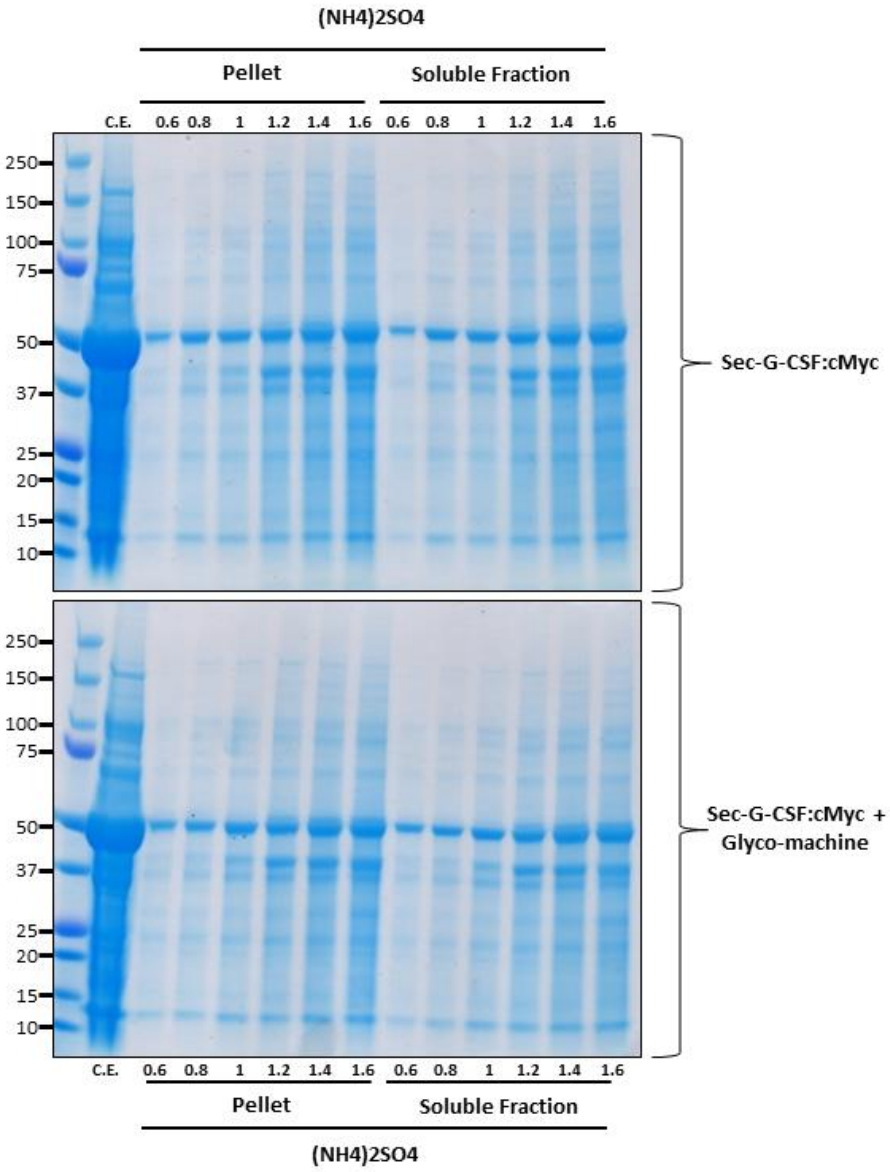


Figure 4-20 Salt precipitation of G-CSF

Coomasie stained gel of Sec-G-CSF:cMyc co-expressed with p19 and glyco-machinery (O-Glycosylation machinery: UDP-GlcNAc/GalNAc epimerase, UDP-GlcNAc/GalNAc transporter, and GalNAc-glycosyltransferase) (OD_{600} ratios; Sec-G-CSF:cMyc: 0.7; p19: 0.1; GNT2: 0.2; epimerase/transporter: 0.1/strain). **Upper panel:** G-CSF (without glycosylation machinery). **Lower panel:** gly-G-CSF. Crude extract from 1 gr infiltrated leaf. Equal volumes were subjected to salt precipitation in a range of 0.6 to 1.6 M ammonium sulfate. The gel was loaded with 30 µL TSP, resuspended pellet or solubilized fraction per lane.

4.3.6 Glycosylation characterization of Sec-G-CSF

After expression and purification of Sec-G-CSF:cMyc with and without co-expression of the *O* -glycosylation machinery, glycosylation was assessed. As the first indication of glycosylation, the co-expression of Sec-G-CSF:cMyc with the Glyco-machinery but without p19, showed a slight band shift. Although lower expression levels are obtained without p19, small band shifts at different dpi might be more evident. In this case, the glycosylation-treated Sec-G-CSF showed the highest accumulation level at 4 dpi, but a band shift was observed at 6 dpi and at 8 dpi (**Figure 4-21**). Furthermore, a smear can be observed above Sec-G-CSF at 4 dpi, and below Sec-G-CSF at 6 dpi, a potential indication of glycosylation during this time frame.

Finally, the GalNAc-*O*-glycosylation of Sec-G-CSF was corroborated by Lectin Blot. In this assay, the purified fractions of G-CSF and gly-G-CSF were detected by c-Myc antibody to confirm the presence of both proteins on the blot, and the same samples were detected using the *Vicia villosa* agglutinin lectin (VVA). This lectin has specific affinity against GalNAc-*O*-glycosylated Ser/Thr residues. In this case, a VVA conjugated to HRP (VVA-HRP) was used to specifically recognize the presence of this specific glycosylation. VVA-HRP reacted only with the Sec-G-CSF:cMyc co-expressed with the glycosylation machinery, but not with the Sec-G-CSF:cMyc expressed alone, indicating proper GalNAc-*O*-glycosylation of Sec-G-CSF:cMyc (**Figure 4-22**). To further corroborate the specific attachment of the GalNAc to the Thr-163 residue (**Figure 4-23, A**), Sec-G-CSF:cMyc co-expressed with and without the *O*-glycosylation machinery was purified and in-gel Trypsin/Chymotrypsin digested. A Trypsin/Chymotrypsin digestion system was used to generate a smaller peptide around the predicted *O*-glycosylation site that would be more amenable to MS/MS analysis. The resulting peptides were identified by nano LC-MS/MS and GalNAc attachment to the specific Thr-163 site was confirmed by the identification of an *N*-acetylhexosamine (HexNAc) attached to Thr-163, as indicated by the characteristic HexNAc product ion at *m/z* 204 (**Figure 4-23, C**).

Although accumulation levels of Sec-G-CSF:cMyc seem to drop at 8 dpi, a drastic band shift was also observed at this time, when expressed without the glycosylation machinery (**Figure 4-24, A**). Such a band shift might be an indication of plant-specific glycosylation, which are typically observed on Western Blots due to presence of multiple branchy arabinose polysaccharides or arabinogalactans polysaccharides. These may be present at technically any Pro residue, which can be modified to Hyp´s and further glycosylated (Dolan, Wu, Cramer, & Xu, 2014; Gomord et al., 2010a; Parsons et al., 2013). However, plant-specific glycosylation should be characterized in future investigations.

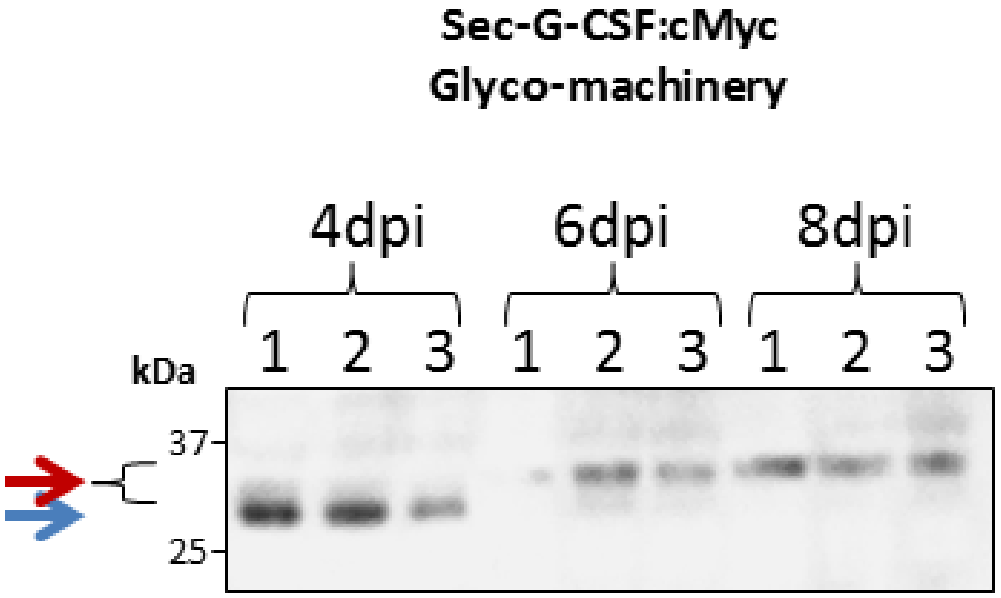


Figure 4-21 Western Blot of transiently expressed Sec-G-CSF:cMyc with Glyco-machinery, but no p19

Sec-G-CSF:cMyc co-expressed with glyco-machinery (*O*-Glycosylation machinery: UDP-GlcNAc/GalNAc epimerase, UDP-GlcNAc/GalNAc transporter, and GalNAc-glycosyltransferase) (OD_{600} ratios; Sec-G-CSF:cMyc: 0.4; glyco-machinery: 0.1/strain). Samples collected at 4, 6, 8 dpi with biological replicates of 1, 2 and 3 and 40 µg TSP loaded per sample. CDB: standard. Blue arrow: Sec-G-CSF:cMyc (28 kDa). Red arrow: potential Sec-G-CSF:cMyc glycosylation. Proteins were detected with c-Myc antibody.

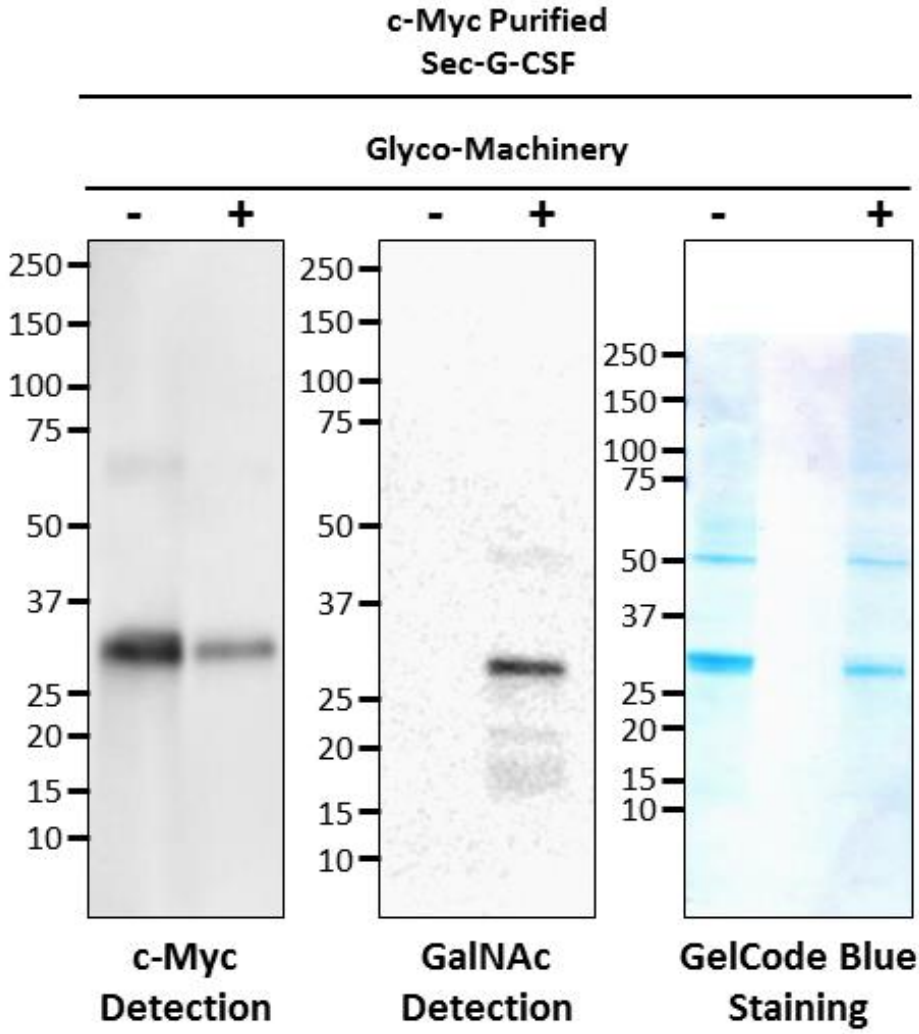


Figure 4-22 Western and Lectin Blot of *O*-glycosylated Sec-G-CSF

c-Myc Detection: 0.25 µL eluent (Sec-G-CSF without glyco-machinery: -) after c-Myc purification; and 2.5 µL eluent (10x more than negative control) (Sec-G-CSF with glyco-machinery: +) after c-Myc purification. Detection using the c-Myc antibody. **GalNAc Detection (Lectin Blot):** Equalized amounts of c-Myc purified Sec-G-CSF (without and with glyco-machinery: -, +). GalNAc detection using VVA-HRP. **GelCode Blue Staining:** GelCode Blue Staining of c-Myc purified Sec-G-CSF (without and with glyco-machinery: -, +). Sec-G-CSF:cMyc co-expressed with p19 and glyco-machinery (-/+) (*O*-Glycosylation machinery: UDP-GlcNAc/GalNAc epimerase, UDP-GlcNAc/GalNAc transporter, and GalNAc-glycosyltransferase) (OD_{600} ratios; Sec-G-CSF:cMyc: 0.5; p19: 0.1; GNT2: 0.2; epimerase/transporter: 0.1/strain).

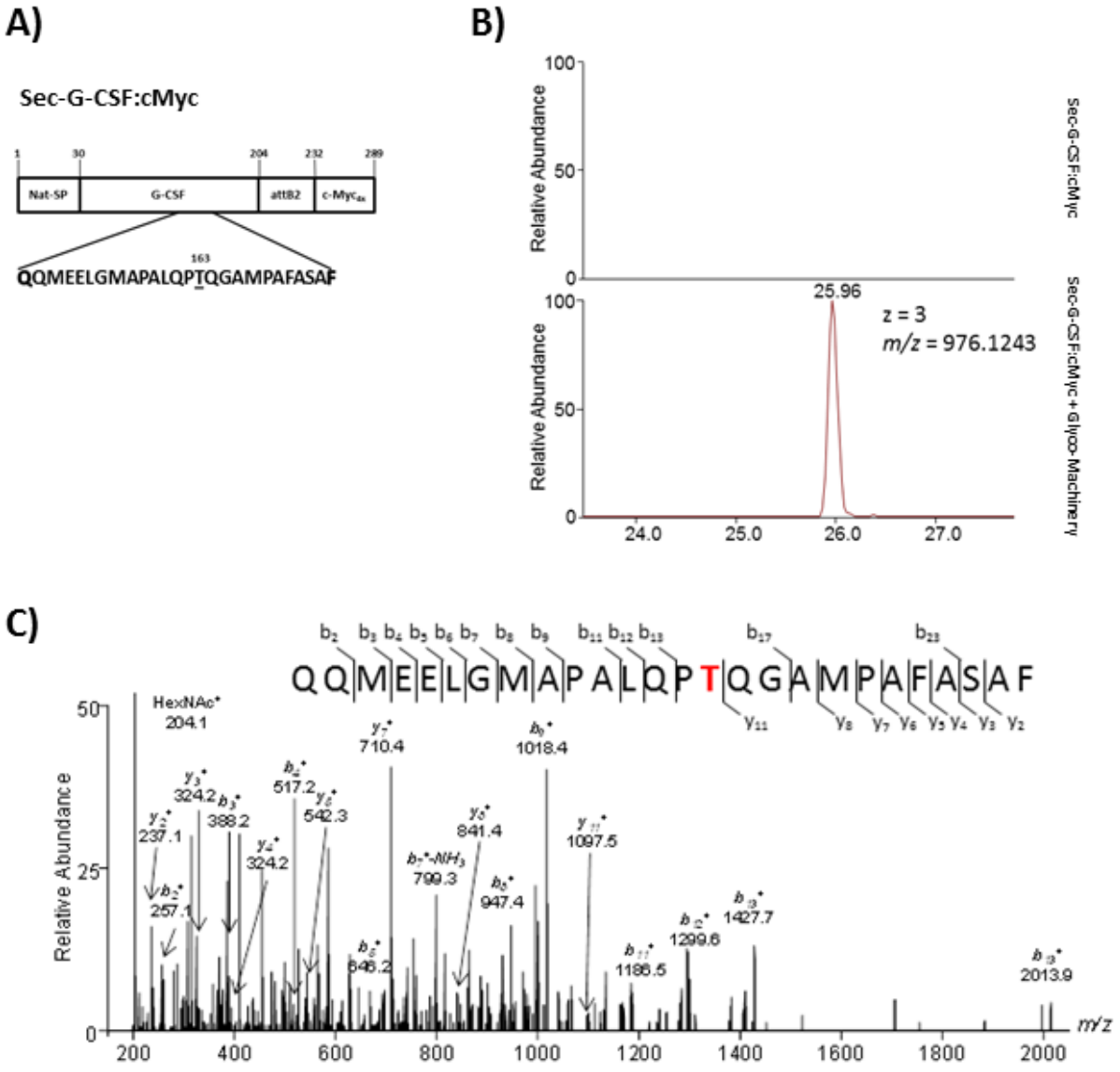


Figure 2-43 Mass Spec characterization of glycosylation site

A) Schematic illustration of released peptide after Trypsin/Chymotrypsin in gel digestion of c-Myc purified Sec-G-CSF:cMyc expressed alone or co-expressed with the *O*-glycosylation machinery. Native glycosylation site (Thr-163) is denoted by underlining. **B)** Extracted ion chromatograms of predicted QQMEELGMAPALQPTQGAMPAFASAF peptide derived from Sec-G-CSF:cMyc expressed alone (upper panel), not being detected; or co-expressed with the *O*-glycosylation machinery (lower panel), where it was detected. **C)** MS/MS of c-Myc purified and Trypsin/Chymotrypsin in gel digested Sec-G-CSF:cMyc-derived peptide (QQMEELGMAPALQPTQGAMPAFASAF) co-expressed with the *O*-glycosylation machinery, showing identified b- and y-ions. Modified Thr-163 is marked in red. Complete list of detected product ions is shown in Table 4-1.

Table 4-1 y- and b- product ions detected in GalNAc-O-glycosylated QQMEELGMAPALQPTQGAMPAFASAF derived peptide and associated mass errors

ion	*m/z*	ppm	ion	*m/z*	ppm
b2	257.12548	-1.0883	*y2*	237.12408	-0.6845
b2-NH3	240.09944	-1.5415	*y3*	324.15616	-0.7584
b3	388.16653	-1.6344	*y3-H2O*	306.14578	-0.9825
b3-NH3	371.13919	-0.7907	*y4*	395.19284	-0.3406
b4	517.21539	-7.8931	*y4-H2O*	377.18256	-0.6562
b4++	259.10741	0.08	*y5*	542.26090	2.9654
b4-H2O	499.19873	-1.8305	*y5-H2O*	474.24353	-8.8303
b4-NH3	500.181763	-0.7627	*y6-H2O*	605.28046	-5.2565
b5	646.25013	-0.07	*y7*	710.35080	-2.7156
b5-H2O	628.23822	1.2798	*y7-H2O*	676.31329	-0.9935
b5-NH3	629.22632	-2.7184	*y8*	841.39130	2.9943
b6	759.33844	-4.2399	*y8-H2O*	733.33893	-5.1282
b6-H2O	741.32812	-4.525	*y11*	1097.50800	-2.98221
b6-NH3	742.311462	-3.8624	*y13-NH3*	1409.65723	-1.43794
b7	816.354858	0.7416			
b7-NH3	799.33258	-3.4806			
b8	947.39948	-3.3751			
b8-H2O	929.38800	-2.5005			
b8-NH3	930.37634	-6.7428			
b9	1018.4365	-3.262			
b9++	509.72036	0.27			
b9-H2O	1000.42694	-4.241			
b10++	558.24667	-0.115			
b11	1186.52039	2.714			
b12	1299.60278	4.417			
b13	1427.66675	-1.048			
b13++	714.33659	0.08			
b14-H2O	1410.64880	-6.80826			
b17	2013.92859	-2.989			
b17-H2O	1995.91186	3.135			

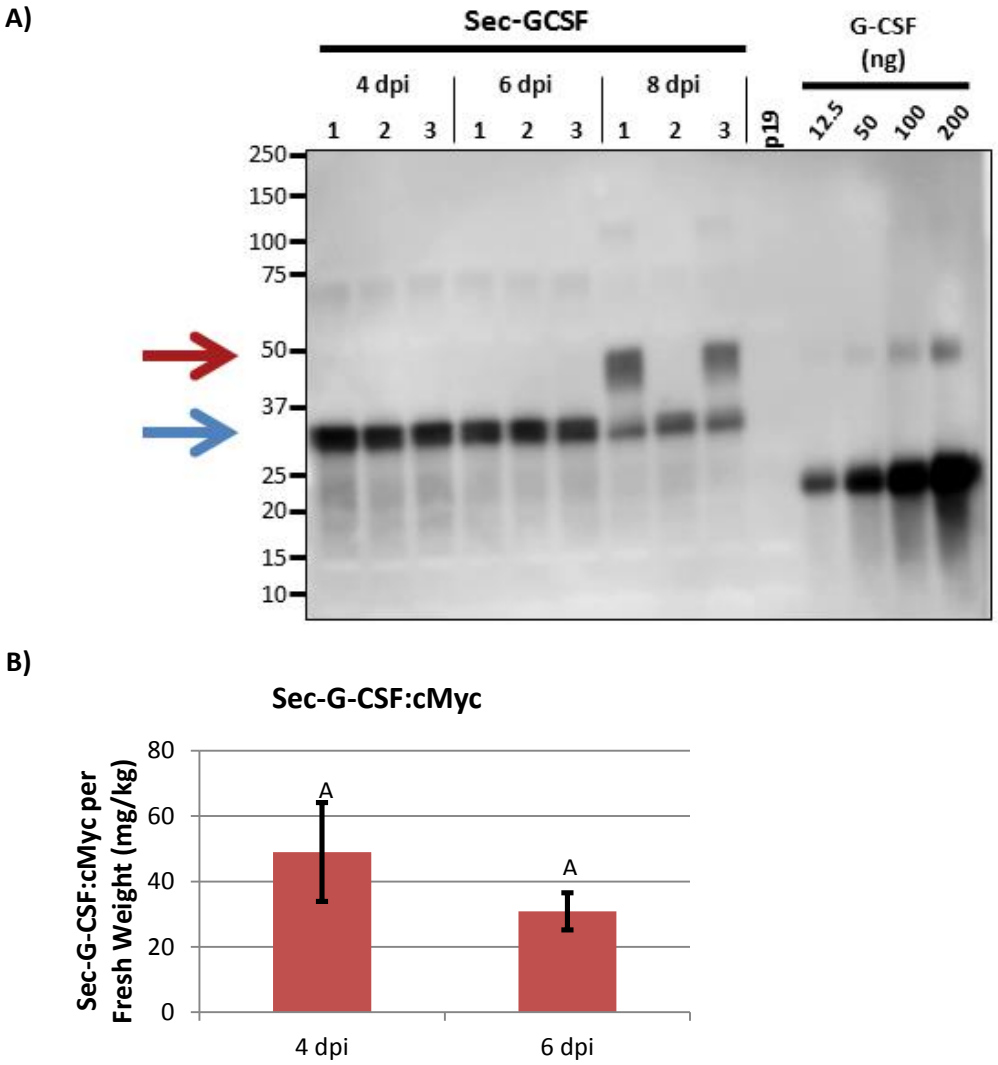


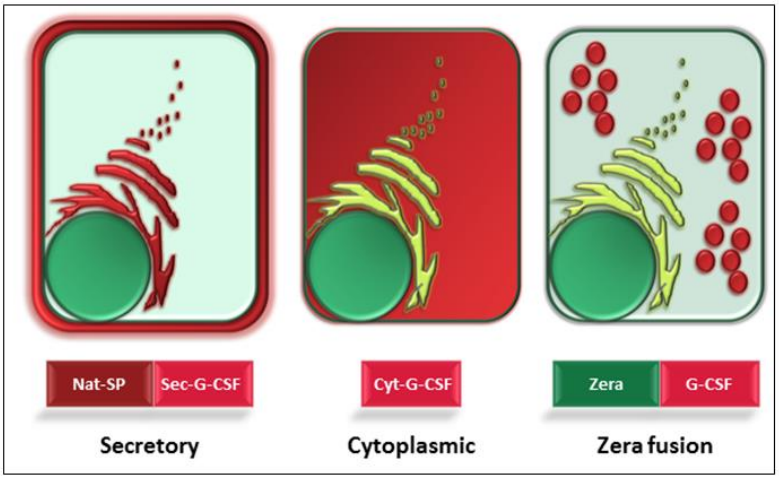
Figure 4-24 Sec-G-CSF:cMyc quantification and plant-specific glycosylation

A) Western blot detection of Sec-G-CSF:cMyc (Secretory variant, 28 kDa). Samples collected at 4, 6 and 8 dpi. Four leaf discs from three different leaves were collected for each sample. Twenty (20) µg TSP of PBS samples were loaded on the gel. Blue arrow denotes monomeric Sec-G-CSF:cMyc. Red arrow denotes potential plant-specific *O*-glycosylation. G-CSF: protein standard. Proteins were detected with G-CSF antibody. **B)** Band quantification of Western blot detected Sec-G-CSF:cMyc. Columns denoted with a different letter are significantly different ($p \leq 0.05$) using one-way ANOVA and followed by Tukey test. Error bars are standard deviations of the means.

4.4 Conclusions

We describe here the successful transient expression of a variety of plant-derived G-CSF variants in *N. benthamiana* plants. The G-CSF variants were targeted to the apoplast, cytoplasm, and ER-derived protein bodies formed by fusion of G-CSF with the Zera protein. Proper subcellular localization was demonstrated by confocal analysis. All variants showed the expected subcellular localization. In the case of the secretory variant, the native secretory signal was used, and it was demonstrated that G-CSF was effectively secreted to the apoplast. A high level of multimerization was observed in the case of the cytoplasmic G-CSF and Zera-G-CSF variants, as compared to the secretory variant, indicating a more proper folding of G-CSF when secreted. Such multimerization has an impact on purification and downstream processes. Among the three variants, the highest accumulation level when using denaturing and reducing conditions was obtained with Zera-G-CSF, followed by Cyt-G-CSF and Sec-G-CSF. However, when using a PBST-based extraction buffer, the Cyt-G-CSF variant showed the highest accumulation level, followed by Sec-G-CSF. The Zera-G-CSF could not be properly extracted under these conditions. Neither set of extraction conditions seemed to have a significant impact on the extraction of Sec-G-CSF. The secretory variant was further co-expressed together with a synthetic *O*-glycosylation machinery, which allowed the proper human-specific Mucin-type *O*-glycosylation of the plant-derived G-CSF. The *O*-glycosylation was corroborated via lectin Blot, using VVA lectin and by mass spectrometry analysis (**Figure 4-22, Figure 4-23**). Since we have demonstrated in this work the feasibility to imitate human-specific *O*-glycosylation, offering an alternative system for the production of this pharmaceutical protein, it would be advisable to explore elongation of the glycosylation moiety and determine the biological impact. It is recommended to improve the accumulation levels of the secretory variant, by using stable transgenic lines expressing the glycosylation machinery, or by transiently expressing all the constructs in a single expression unit (single binary vector). The latter would improve the efficiency of co-expression in the same individual cell.

Conclusions



Accumulation	+	+++ (++)	+++
Extraction	PBS	PBS (SDS)	SDS
Multimerization	(+)	+++	+++
Glycosylation	✓	✗	✗

Figure 4-25 Graphical representation of the conclusions obtained in this study

Chapter 5

- **Transient Expression of Bovine Beta and Kappa Caseins in *N. benthamiana* Plants.**

5.1 Abstract

Plant expression systems for the production of valuable proteins have become more competitive in recent years. Although major efforts have been focused on increasing protein accumulation, few but substantial recent studies have addressed modulation of PTM´s, such as glycosylation. Phosphorylation is another important PTM involved in protein activity; as yet, this modification has not been adequately addressed. An example of mammalian phosphoproteins, where the absence of phosphate groups negatively impacts the activity of the protein, can be found with β- and κC. Lack of phosphorylation of these proteins affects proper Ca^{2+} binding and might trigger allergic reactions on vulnerable consumers. Even though plants are equipped with kinases to allow protein phosphorylation, lack of a specific kinase in the subcellular localization where the recombinant protein is targeted might hinder such PTM. In this study, we transiently express in *N. benthamiana* plants bovine β- and κC´s as model phosphoproteins, which are bioactive peptides and relevant for the food industry, in order to assay the feasibility of plants to phosphorylate these proteins when co-expressed with a putative bovine kinase analogue of human FAM20C.

5.2 Introduction

In the last couple of decades, plants have become more attractive platforms for the expression of recombinant proteins (Vukušić et al., 2016). Although plants are known to perform PTM´s, differences in such modifications in respect to mammalian organisms have sometimes been a concern for some researches. Among the major limitations to the broad usage of plant as expression systems, increasing accumulation levels has been one of the major drawbacks and therefore this issue has received more attention, while recent works have focused on addressing PTM´s (Ahmad et al., 2010).

Not all proteins are or must be post-translationally modified; however, in some cases, PTM´s might have a major impact on protein activity. Glycosylation has been a major PTM concern and focus of engineering in plant expression systems, being this modification mainly involved in protein stability, protein-protein recognition and therefore biological activity (Ahmad et al., 2010). So far, to the best of our knowledge, phosphorylation has not been addressed in molecular pharming. However, when bovine β-casein, a model phosphoprotein with 5 phosphorylation sites, was expressed in potato and soybean, no evidence of successful phosphorylation was provided (Chong et al., 1997; Maughan et al., 1999; Philip et al., 2001). Production of a recombinant βC for human consumption as diet supplement, milk formula or additive for cheese-like products, must be properly phosphorylated. Lack of phosphorylation in this protein affects Ca^{2+} binding, which is involved in the production of micelles, important protein complexes for cheese production, Ca^{2+} uptake in the diet, and allergenic responses in susceptible consumers (Bernard et al., 2000; Cases et al., 2011; Johansson, 2002; Tezcucano Molina et al., 2007).

KC is another type of casein that interacts directly with βC stabilizing the micelles in liquid and avoiding clotting of the milk (Kethireddipalli & Hill, 2015), represents another example of phosphoprotein, which just like the βC is a source of bioactive peptides with antioxidant, antithrombotic, antimicrobial, antihypertensive, and immunomodulating activity (Silva & Malcata, 2005). Phosphate groups on this protein might not only impact its nutrimental and nutraceutical properties, but they might also influence the stability of the micelle, thus affecting the sensorial quality of the products containing β- and κC´s.

Caseins have been extensively used in cheese production and dairy products, due to the clotting properties of these proteins in presence of acids and proteolytic enzymes like rennin. The importance that they have in this industry has probably made them the most widely studied food proteins (Ginger & Grigor, 1999). Food proteins, as opposed to pharmaceutical proteins, have not been extensively produced in heterologous systems because of their abundance, variety of sources, and availability in nature (Choi & Jime, 2001). Due to the industrial interest of cheese production, research groups have considered to produce recombinant casein proteins. Research studies have successfully expressed in *E. coli,* yeast and CHO cells (Fahimirad et al., 2017; Hansson et al., 1993; Jiang et al., 2004; Jimenez-Flores et al., 1990; Thurmond et al., 1997). However, in terms of genetically engineering manipulation of caseins for human consumption, some efforts have been made for the production of transgenic farm animals with

altered casein expression levels, in order to modulate milk quality production and therefore dairy products (Brophy et al., 2003; Jabed et al., 2012).

In the case of κC, it has also been produced recombinantly in *E. coli* for structural characterization studies and nutritional improvement by directed mutagenesis, rebalancing the amino acid content from a nutritional perspective (Kang, 1988; Oh & Richardson, 1991). In this case, particular attention has been paid to producing recombinant bioactive peptides derived from κC. These peptides could be used as practical food or medicinal nutritional products for patients. One example is the highly nutritional human caseinomacropeptide (CMP). This peptide promotes the growth of beneficial bacteria (Bifidobacteria), acts as adhesion inhibitor of oral Actinomycetes to human cell membranes and cholera toxin to its receptors, and exhibits an antithrombotic effect. This κC derived peptide has been produced in *E. coli* (Liu et al., 2008) and yeast (Kim, Oh, et al., 2005). Similarly, lactaptin analogs, a proteolytic fragment derived from the human κ-casein, capable to reduce cell viability and induce apoptosis of the mammary adenocarcinoma cell line MCF-7, have been expressed in *E. coli* (Semenov et al., 2010).

It is necessary to indicate that the *E. coli* produced βC was not phosphorylated, therefore making it unable to sequester calcium which is important for clotting behavior and as nutritional property (Hansson et al., 1993). When βC was co-expressed with human casein kinase II, *E. coli* was able to produce phosphorylated βC (Thurmond et al., 1997). This result indicated the feasibility of producing a phosphorylated βC, when a proper kinase is functionally coexpressed with βC to perform such PTM´s. When produced in yeast, βC expressed by *P. pastoris* was observed to be properly phosphorylated but also *N*-glycosylated with several mannose residues at a single Asn residue (Asn-68) (Choi & Jime, 2001). A similar pattern was observed when βC was expressed in *S. cerevisiae* (Jimenez-Flores et al., 1990), where the casein was also present as a heterogeneous phosphorylated recombinant protein, but it seemed to be *O*-glycosylated, instead of *N*-glycosylated. On the other hand, when βC was expressed in CHO cells, the casein was found phosphorylated (Jiang et al., 2004). The latter examples offer an overview of the capabilities of different expression systems to produce properly functional (phosphorylated) recombinant βC.

A similar scenario could be observed with κC, which is natively phosphorylated (3 phosphoserines and 1 phosphothreonine) and *O*-glycosylated (up to 8 Mucin-type glycosylation sites). The whole protein and the CMP were expressed in *E. coli*, however no phosphorylation studies were addressed (Kang, 1988; Oh & Richardson, 1991) and only 1 study proposed that

the recombinant CMP was not glycosylated, as would be expected (Liu et al., 2008). On the other hand, only the CMP has been expressed in *S. cerevisiae* and *P. pastoris* (Kim, Oh, et al., 2005). In this case, the CMP produced in *S. cerevisiae* was *O*-glycosylated but to a lesser extent in comparison to its native counterpart and no phosphorylation analysis were performed (Kim, Park, et al., 2005).

Even though plants are known to possess a vast variety of kinases and phosphatases, thus capable of performing such PTM, it has been reported that the expression of phosphoproteins like βC in potato and soybean results in peptides that lack the presence of phosphate groups attached to the recombinant protein (Chong et al., 1997; Philip et al., 2001).

Lack of phosphorylation of βC expressed in potato and soybean (Chong et al., 1997; Maughan et al., 1999; Philip et al., 2001), might be the result of the absence of plant kinases in the cellular compartments where the recombinant protein was expressed. Recently, Tagliabracci and colleagues (Tagliabracci et al., 2012) identified a protein kinase called FAM20C, which is co-secreted together with the human βC by the mammary cells, and is responsible for the specific phosphorylation of several proteins at the S-X-E/S domains, including βC and other proteins important for biomineralization. The presence of FAM20C in human gland cells, might explain the dependency of βC on FAM20C to get specifically phosphorylated.

Plants do possess mechanisms to phosphorylate and dephosphorylate proteins, having an impact on protein functionality and signaling, as is the case in animal cells. However, although no FAM20C homologues have been reported in plants, the protein kinases I and II (CKI and CKII), among others, are present in plants and are known to be able to phosphorylate a variety of proteins *in vivo*, as well as βC *in vitro*. However, CKI and CKII are found in the nucleus, cytoplasm, membranes and mitochondria, subcellular locations where they act (protein conformational activation/deactivation) (Lôbo Machado et al., 2002). Lack of co-localization of an appropriate kinase and target protein, as well as appropriate compartment conditions might proscribe this PTM, and therefore negatively impact properties of recombinant proteins produced in foreign organisms, as was the case of the βC produced in soybean by Philip and colleagues (Philip et al., 2001).

Neither κC nor CMP have been successfully expressed in plant systems so far. Although as was the case with βC, it would be expected that proper control of PTM´s with this protein might also occur, for both phosphorylation and *O*-glycosylation and thus should be considered for heterologous expression.

In the case of phosphorylation, there are no reported efforts to allow site-specific phosphorylation on recombinant proteins produced in plants. The expression of βC in potato and soybean seeds was achieved (Chong et al., 1997; Maughan et al., 1999; Philip et al., 2001), but no efforts were made to obtain this protein in its phosphorylated and active original form, the active form. This case represents a clear example, in which the absence of specific kinases, or proper localization of such enzymes, leads to the production of foreign proteins without their respective mammalian-specific PTM´s, and thus a potential risk for a deficient activity. The human protein kinase FAM20C recently identified by Tagliabracci and colleagues (Tagliabracci et al., 2012), which is localized to the Golgi apparatus and secreted along with other proteins, might represent an alternative to specifically phosphorylate recombinants proteins expressed in plants, like β- and κC (Tagliabracci et al., 2012). Although no plant FAM20C-like proteins have been reported so far, FAM20C homologues might be present in other animals, presenting a similar function, and thus its heterologous co-expression might offer the possibility to produce recombinant proteins with proper phosphorylation. Bovine DMP4, also recently renamed as FAM20C (XP_614520.6), is highly homologous (86.6%) to the human FAM20C. Although not proved so far by any study, this putative casein-specific kinase might be responsible for the specific phosphorylation of bovine caseins and thus its co-expression with casein proteins in plant system might solve the issue in respect to the lack of phosphorylation.

The aim of this study was to transiently express the bovine β- and κC as model heterologous phosphoproteins and the putative casein-specific bovine FAM20C in *N. benthamiana* plants, via *Agrobacterium* infiltration. The model phosphoproteins and putative kinase were targeted to the secretory pathway with their own native signal peptides, in an attempt to simulate the secretion of these proteins in the mammalian cells, where caseins are phosphorylated by the FAM20C kinase (**Figure 1-1; Figure 5-1**).

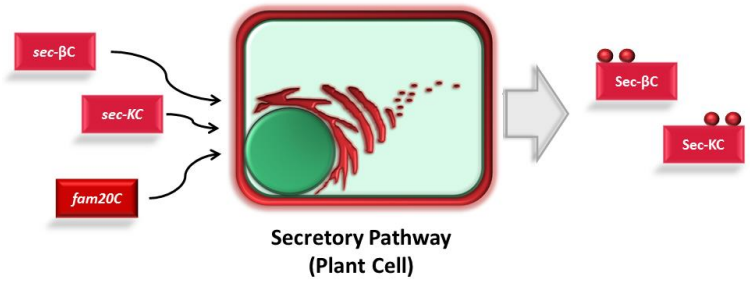


Figure 5-1 Graphical representation of the subcellular localization of the recombinant protein

Coding sequences corresponding to the bovine β- and κ-casein and Fam20C would be targeted to the secretory pathway in the plant cells, as occurs during their synthesis in mammalian cells, where casein proteins are phosphorylated by their specific kinase.

5.3 Results and discussion

5.3.1 Generation of constructs with GCSF variants

The bovine β- and κC (NCBI Acc Nr: AAA30430.1 M15132.1 and NP_776719.1, respectively) as well as the bovine Fam20C (XP_614520.5) coding sequences used in this work were chemically synthesized (GenScript, Piscataway, NJ, USA)**.** The casein-related constructs used in this work were generated by removing the stop codon of the synthesized sequences and introducing BbsI flanking sites with specific overhang sequences via mutagenesis PCR, using the forward primer Mt14F **(**GGGGGGGAAGACCTCATGAAGGTGCTGATACTTGC) and reverse primer Mt16R (GGGGGGGAAGACTGAGCTCCACGATGATAGG) for beta casein; Mt11F (CCCCGAAGACATCATGATGAAGTCATTCTTTCTC) and Mt13R (GGGGGAAGACTGAGCTCACTGCGGTGCTTG) for kappa casein; and Mt17F (GGGGGGGAAGACCTCATGAAGATGATGCTTG) and Mt19R (GGGGGGGAAGACTGAGCTCTCTGCCTGCGG) for Fam20C. The resulting amplified product was cloned in the pUC57 vector via EcoRV, and the mutated sites were verified by DNA sequencing and restriction digestions. The resulting pUC57 caseins or Fam20C without Stop codon vector (pUC57 Sec-βC BbsI AB w/o Stop, pUC57 Sec-κC BbsI AB w/o Stop, pUC57 Fam20C BbsI AB w/o Stop) was then used to clone the coding sequence into the pENTR4 BpiI AB vector (a GoldenGate compatible pENTR4 vector generated in the lab), via GoldenGate, using the restriction enzyme BbsI and DNA ligase in a single cutting-ligation reaction. The resulting pENTR4 construct without Stop codon (pENTR4 Sec-βC w/o Stop, pENTR4 Sec-κC w/o Stop, pENTR4 Fam20C w/o Stop) was then used to clone the corresponding coding

sequence into the binary vector via Gateway LR reaction, allowing the in-frame cloning of the coding sequence with the desired C-terminal tag, eYFP (pGWB 641) or c-Myc (pGWB 617) (Nakamura et al., 2010) (**Figure 3-2**).

Positive binary vectors carrying either a casein or Fam20C gene were transformed into *Agrobacterium tumefaciens* AGL-1, via electroporation, and positive clones were stored at -80 C and used for further transient expression assays in *N. benthamiana* plants (**Figure 5-3**).

Besides the binary vectors containing the caseins or Fam20C genes, empty pGWB 642 was used as control for expression of only the eYFP tag, for confocal purposes, which would allow the detection of eYFP in the cytoplasm, referred to as CyteYFP (Cytoplasmic eYFP) (**Figure 5-4**). *Agrobacterium tumefaciens* EHA105 carrying a binary vector containing the p19 gene from *Cymbidium ringspot tombusvirus* (Silhavy et al., 2002) was used to co-express genes of interest, allowing higher accumulation levels (**Figure 5-4**).

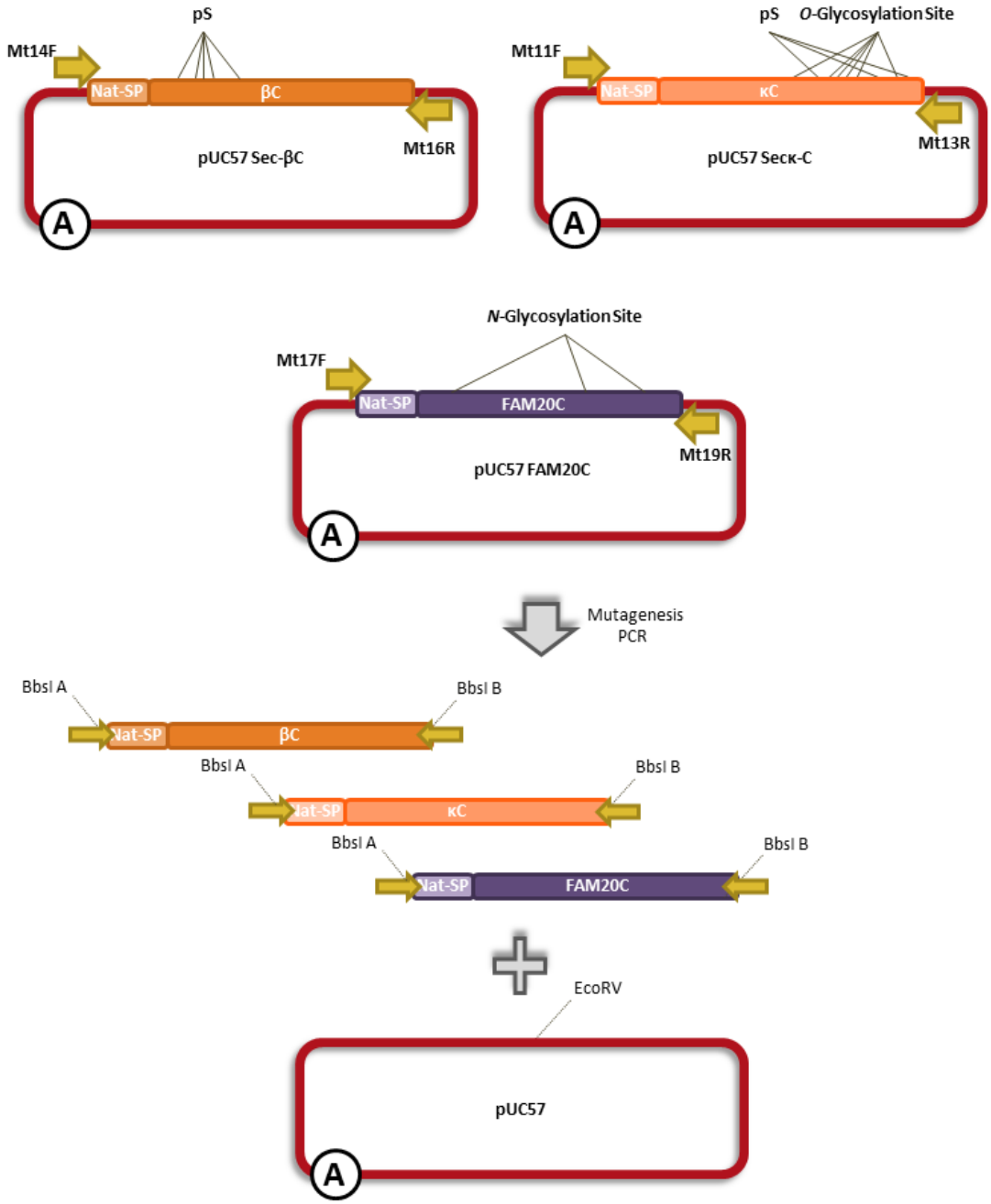


Figure 5-2 Schematic representation for the construction of caseins and kinase expression cassettes

Mutagenesis PCR of pUC57 Sec-βC, -Sec-κC and –Fam20C, adding the flanking BbsI A and B sites, and removing the stop codons and later cloning into the pUC57 vector via EcoRV.

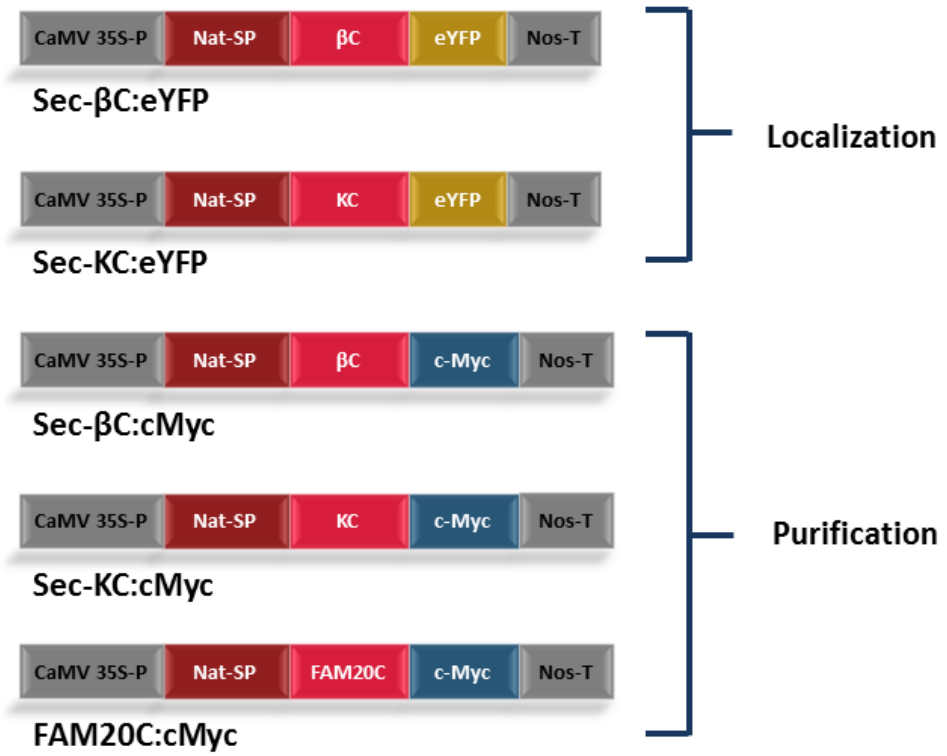


Figure 5-3 Schematic representation of the different caseins and kinase expression cassettes used in this work

All expression unites were under the control of the CaMV 35S promoter (CaMV 35S-P) and Nos Terminator (Nos-T). Native secretory signal peptide (Nat-SP) was used for each case. Coding sequences were cloned in frame with eYFP (binary vector pGWB 641), for subcellular localization purposes, or c-Myc (binary vector pGWB 617) for purification purposes.

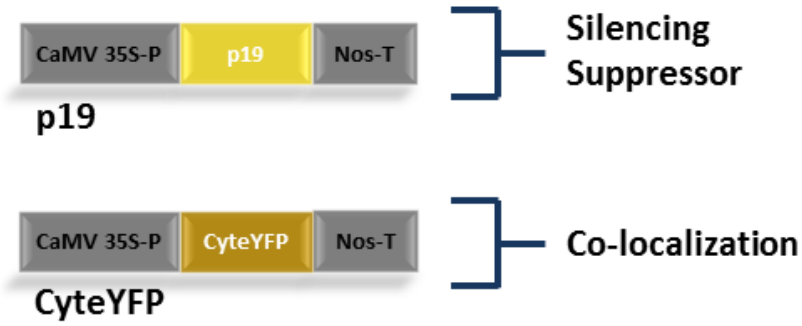


Figure 5-4 Additional constructs used for protein expression and subcellular localization

Schematic representation of the vectors used for protein expression and confocal purposes. p19 was used to allow higher accumulation levels. CyteYFP: Cytoplasmic eYFP (pGWB 642, empty vector). All expression units are under the control of the CaMV 35S Promoter and Nos Terminator.

5.3.2 Transient expression of βC in *N. benthamiana* plants and subcellular localization.

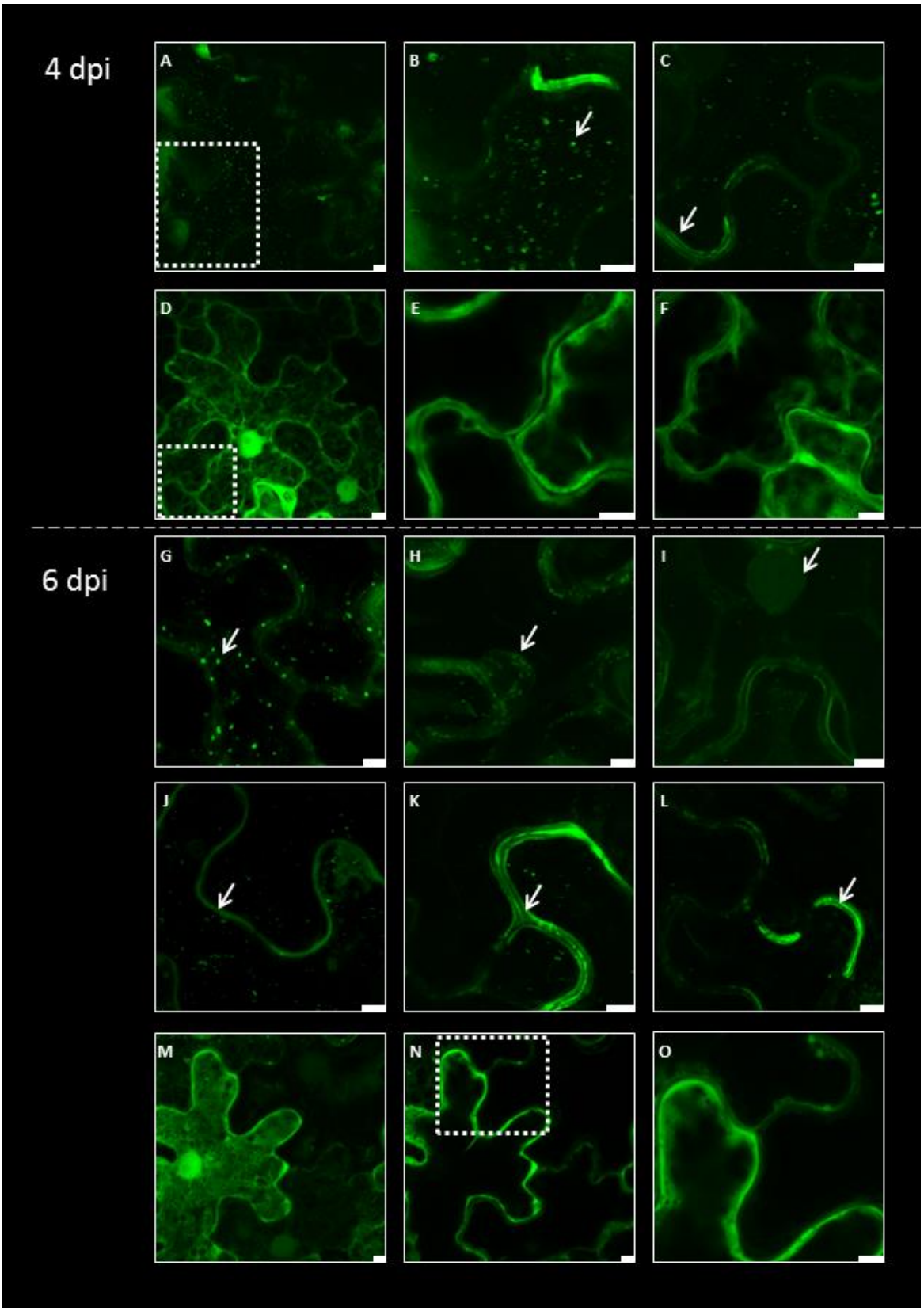
Expression of beta casein was first determined by confocal visualization. For this purpose, the Sec-βC:eYFP construct was used, and protein expression was monitored at 2 to 8 dpi. βC was found as multiple dots in the cytoplasm (**Figure 5-5, A-B)**, suggesting Golgi localization, as well as in the apoplast (**Figure 5-5, C, J-L**), and signal was observed in some nuclei (**Figure 5-5, G-I**). The latter observations suggest that βC was secreted to some extent, but it might also be cleaved during translation and thus eYFP could be observed in the nucleus. In contrast, the CyteYFP used as negative control shows a cytoplasmic localization pattern as well as nuclear localization (**Figure 5-5, D, M**), and a clear apoplast (**Figure 5-5, E-F, N-O**).

Protein expression was corroborated by Western blot, where beta casein was observed with an apparent molecular weight of 54 kDa. Three main degradation products were found, one of them was matching the molecular weight of CyteYFP, indicated cleavage of tag. The other two main degradation products were observed above the tag, indicating cleavage elsewhere in the βC sequence (**Figure 5-5, P**).

In order to determine an optimal harvesting time for downstream purposes, accumulation of βC was also monitored at 4, 6 and 8 dpi. The recombinant casein was found to reach an accumulation level of 7 mg/kg Fresh Weight. This accumulation level was maintained at 4 and 6 dpi, dropping to almost undetectable levels at 8 dpi. Sin this casein is vulnerable to degradation,

this behaviour is likely to be caused by the secretion of the protein to the apoplast, where an active proteolytic environment is found (**Figure 5-6**).

Although a weak signal in confocal microscopy was observed, and low levels were detected at 4 and 6 dpi, further accumulation was not observed at 8 dpi. BC could be observed in the apoplast, indication of proper subcellular targeting to the secretory pathway, where βC is supposed to be phosphorylated.


4 dpi
A
B
C
D
E
F
6 dpi
G
H
I
J
K
L
M
N
O

P)

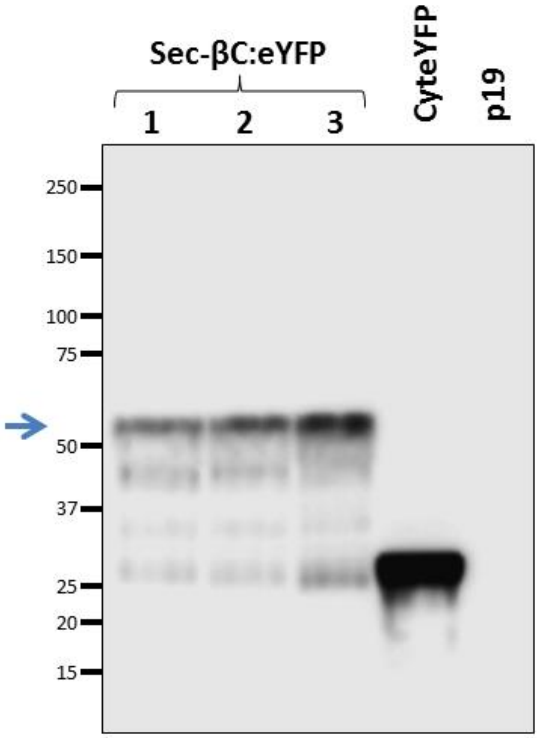


Figure 5-5 Transient expression of Sec-βC:eYFP in N. benthamiana plants

A) Expression of Sec-βC:eYFP (Secretory) (**A- C, G-L**). CyteYFP (**D-F, M-O**). Inset white box in **A**, **D** and **N** depicts microscopic zoom, shown in **B, E** and **O**, respectively. White arrows depict Sec-βC:eYFP localization in the Golgi (**B, G**), apoplast (**C, J-L**), and nucleus (**H, I**). **A-F:** tissue collected at 4 dpi. **G-O:** tissue collected at 6 dpi. Inset bar: 5 µm. **P**) Western Blot of transiently expressed Sec-βC:eYFP. Samples collected at 4 dpi. Sec-βC:eYFP (Secretory-mature, 54 kDa) is denoted by a blue arrow. CyteYFP (Cytoplasmic eYFP, 29 kDa). Samples were treated with PBS extraction buffer. Fifty (50) µg TSP of PBS samples were loaded on the gel. Proteins were detected with GFP antibody. *Agrobacterium* strain AGL-1 was used to express Sec-βC:eYFP, CyteYFP construct was carried by the *Agrobacterium* strain GV3101, and p19 by EHA105. Both Sec-βC:eYFP and CyteYFP were co-infiltrated with p19.

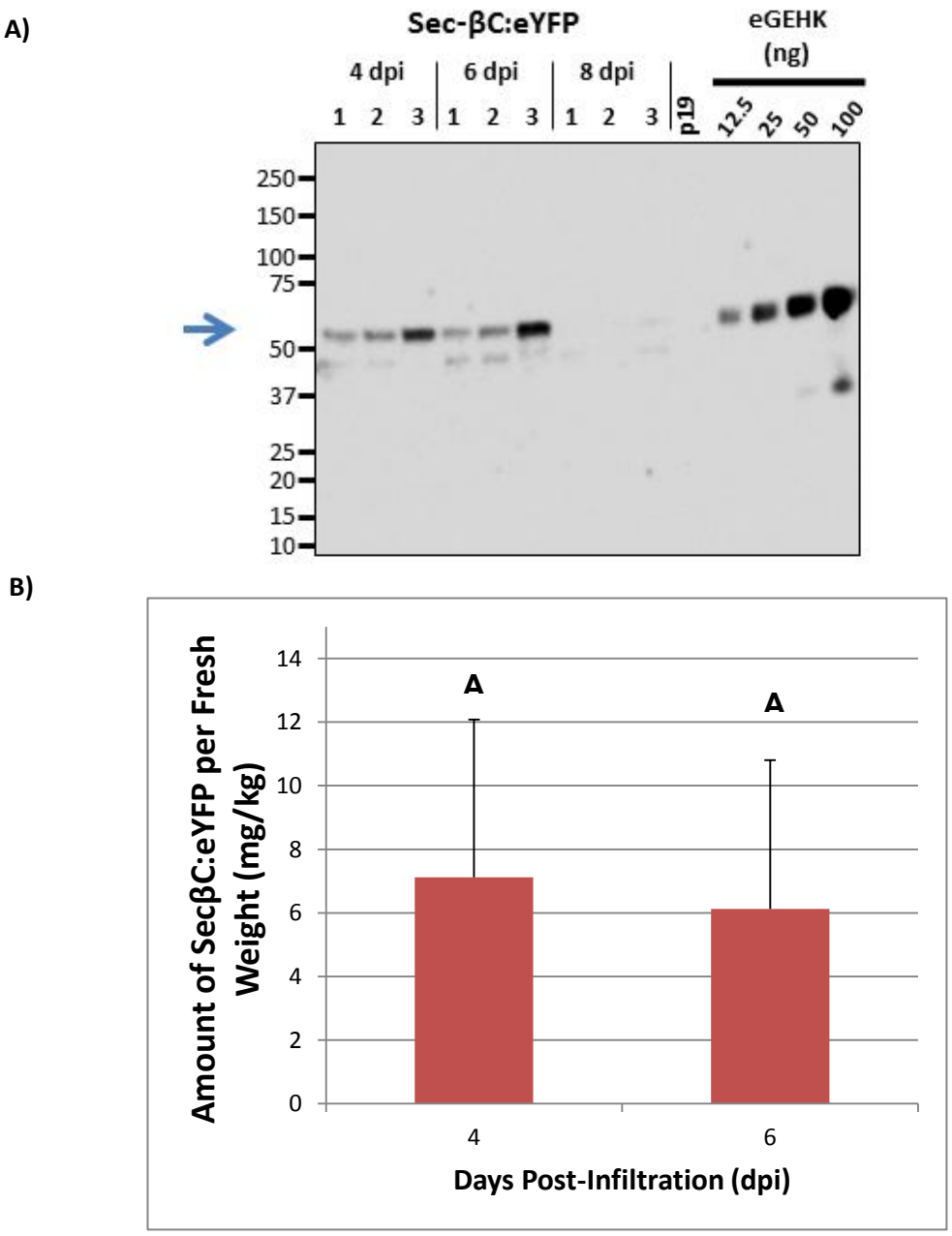


Figure 5-6 Time-course accumulation of transiently expressed Sec-βC:eYFP in *N. benthamiana* plants

A) Western blot detection of Sec-βC:eYFP (mature βC:eYFP, 54 kDa). **B)** Band quantification of Western blot detected proteins. Samples collected at 4, 6 and 8 dpi. Four leaf discs from three different plants were collected for each sample. All samples were treated with $PBST^{0.1\%}$ extraction buffer. Thirty (30) µg TSP of $PBST^{0.1\%}$ samples were loaded on the gel. Red arrow denotes monomeric Sec-βC:eYFP. eGEHK: protein standard. Proteins were detected with GFP antibody. Columns denoted with a different letter are significantly different ($p \leq 0.05$) using one-way ANOVA and followed by Tukey test. Error bars are standard deviations of the means.

5.3.3 Evaluation of phosphorylation of βC and co-expression with Fam20C.

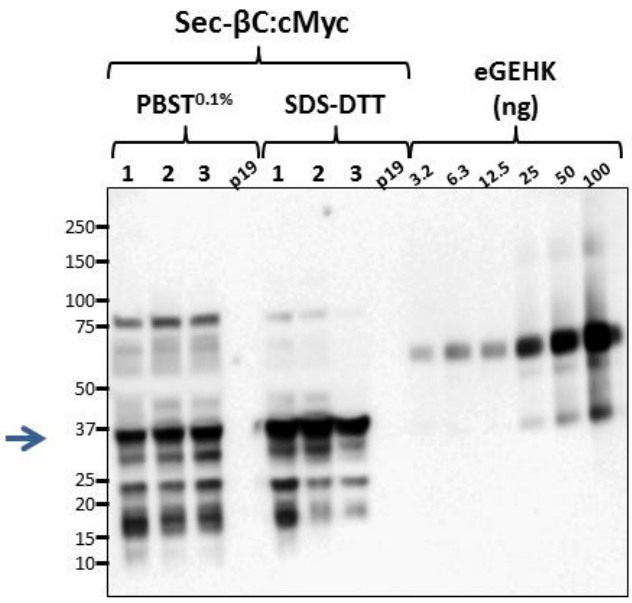
After determining that βC was directed to the secretory pathway, we then tried to evaluate the potential of this plant-derived casein to be phosphorylated while being secreted. For this purpose, we first fused the βC coding sequence with a c-Myc tag for detection and purification purposes. Expression of Sec-βC:cMyc was detected at 4 dpi with the expected molecular weight of 33 kDa, suggesting cleavage of secretory signal peptide. Degradation products were observed as it was the case with the eYFP version, and some dimerization was also observed (**Figure 5-7**). Since βC is a relatively hydrophobic protein, a reducing and denaturing extraction buffer containing SDS and DTT was also tested to observe the impact on protein extraction. Extracting the protein under denaturing and reducing conditions showed an increase of almost two fold in comparison to using a PBS-based extraction buffer, and it also showed a reduction of dimerization, suggesting that interaction of beta casein molecules was occurring after protein accumulation and extraction (**Figure 5-7**).

The bovine putative casein-specific kinase Fam20C was also expressed as a c-Myc fusion recombinant protein. In this case, both the PBS-based and SDS-DTT-containing extraction buffers were tested, and protein expression was screened at 4, 6 and 8 dpi. Fam20C:cMyc was detected by Western blot, using both buffers, PBS-based and SDS-DTT-containing buffer. As it was the case with βC, extraction of Fam20C was slightly improved by using denaturing and reduction conditions, especially at 4 dpi. Although the molecular weight of the mature Fam20C:cMyc protein is predicted to be 72.5 kDa, and the premature protein (without cleavage of signal peptide) is predicted to be 75.6 kDa, the plant-derived recombinant Fam20C:cMyc protein produced here showed an apparent molecular weight of approximately 86 kDa. These observations are not surprising since Fam20C is *N*-glycosylated, having three glycosylation sites on its mature amino acid sequence, and thus a significant increase in molecular weight can be expected. Unfortunately, at this point we cannot determine whether Fam20C:cMyc was secreted to the apoplast where it is expected to phosphorylate the secreted caseins. On a time-course screening, Fam20C:cMyc reached the highest accumulation levels at 4 dpi, and accumulation levels decreased at 6 dpi and almost undetectable at 8 dpi (**Figure 5-8**).

Considering that not only Fam20C might be responsible for specific phosphorylation of βC, but also plant kinases might be present in the secretory pathway which might phosphorylate beta casein as well, we then tried to co-express Sec-βC:cMyc with and without Fam20C. Both recombinant proteins were detected when co-expressed together, with their respective molecular weights as observed when expressed alone (**Figure 5-9, A**). Both Sec-βC:cMyc and

Sec-βC:cMyc + DMP4:cMyc were c-Myc purified for mass spectrometry analysis in order to determine phosphorylation of βC. Gel Blue Code staining of purified Sec-βC:cMyc showed a very poor monomeric band of Sec-βC:cMyc, without or with co-expression of DMP4:cMyc (**Figure 5-9, B**). The latter might be due in part due to the low accumulation levels obtained with this protein and probably due to protein degradation, since several degradation products were detected. Since beta casein is a relatively hydrophobic protein, and we have also observed dimerization of this recombinant protein, both hydrophobic interactions and protein cleavage were responsible for the generation of multiple complexes with different molecular weights. Thus, the monomeric portion for mass spectrometry analysis was reduced. This scenario is even worse when Sec-βC:cMyc was co-expressed with Fam20C, as the monomeric Sec-βC:cMyc was almost undetectable by Gel Code Blue staining or Western blot, after purification (**Figure 5-9, B-C**). Despite the poor recovery of Sec-βC:cMyc , bands corresponding to the c-Myc purified monomeric Sec-βC:cMyc expressed on its own and co-expressed with Fam20C were in-gel trypsin digested and submitted to mass spectrometry analysis. However, no casein peptide could be identified.

A)



B)

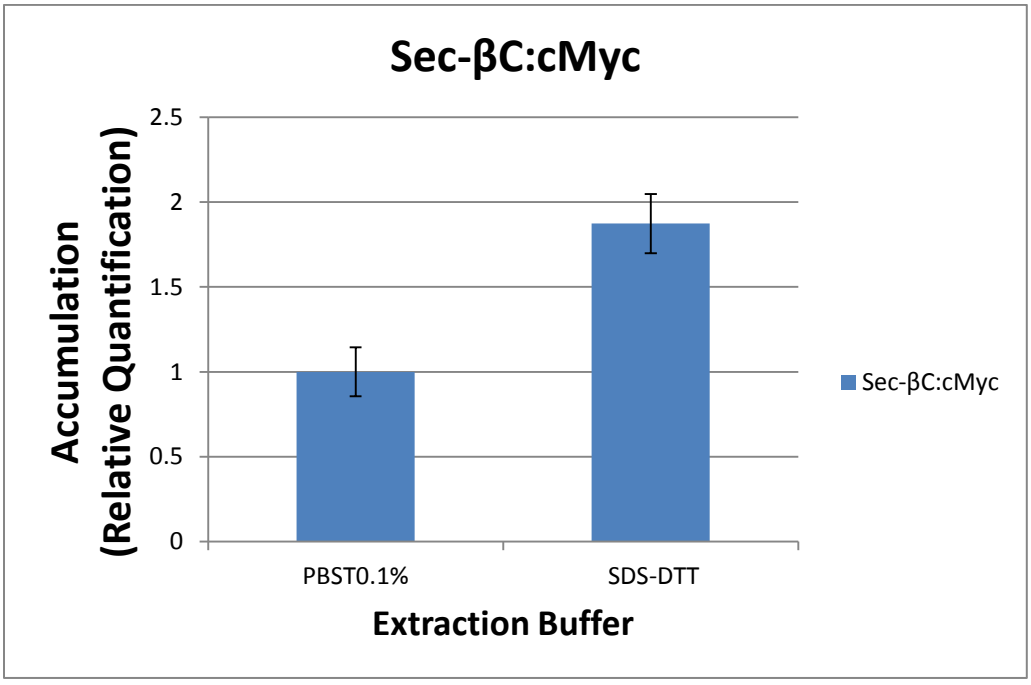


Figure 5-7 Transient expression of Sec-βC:cMyc in *N. benthamiana* plants

A) Western blot detection of Sec-βC:cMyc (premature Sec-βC:cMyc: 37.7 kDa; mature Sec-βC:cMyc, 33 kDa). **B)** Band relative quantification of Western blot detected Sec-βC:cMyc. Samples collected at 4. Four leaf discs from three different plants were collected for each sample. Samples of 40 µg TSP of PBST$^{0.1\%}$ (PBS+Tween20$^{0.1\%}$) were loaded on the gel, and equalized amount of reducing extraction buffer (SDS-DTT) samples were used. Blue arrow denotes monomeric Sec-βC:cMyc. eGEHK: protein standard. Proteins were detected with c-Myc antibody. Error bars are standard deviations of the means.

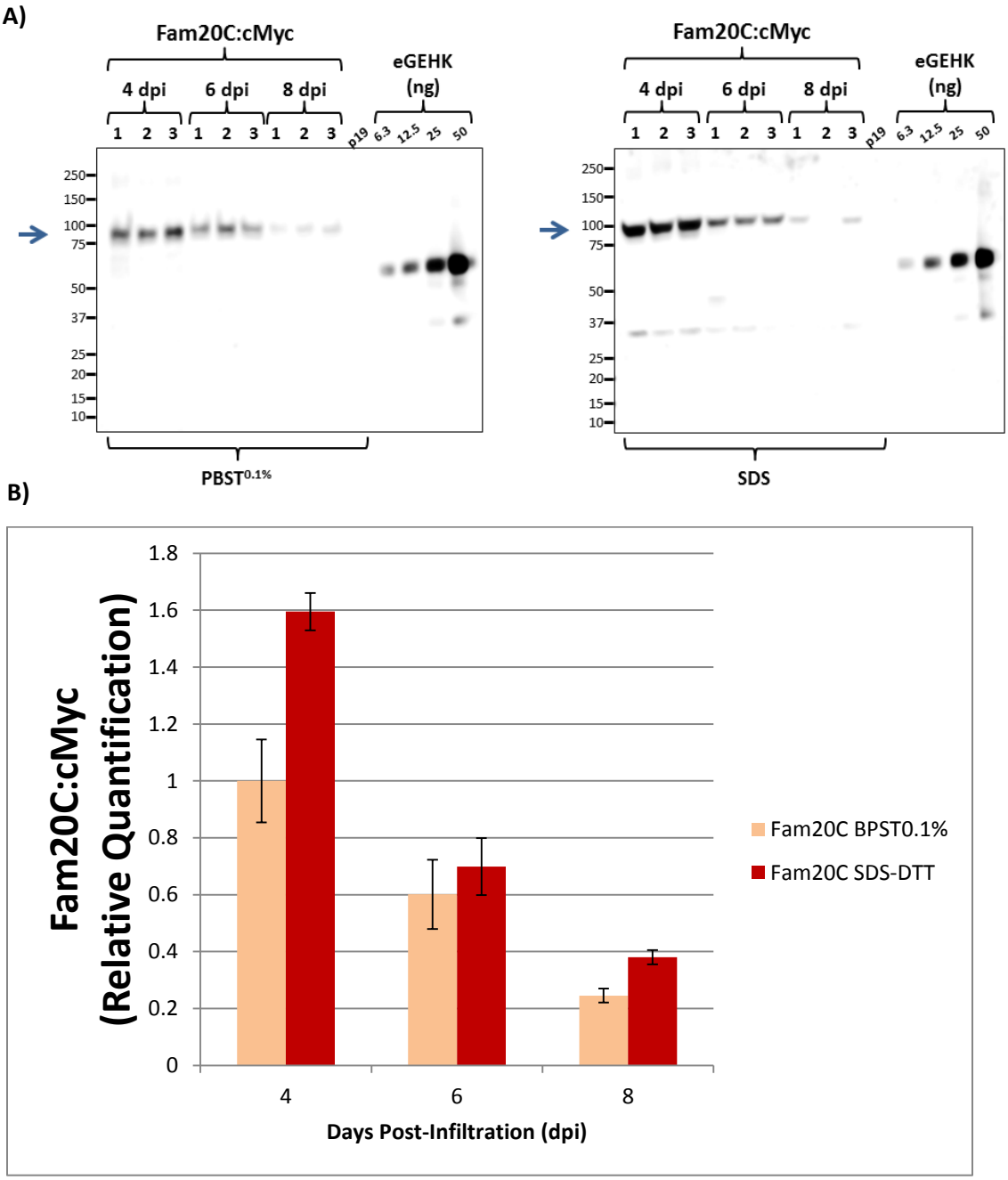


Figure 5-8 Time-course accumulation of transiently expressed Fam20C:cMyc in *N. benthamiana* plants

A) Western blot detection of Fam20C:cMyc (premature Fam20C:cMyc: 75.6 kDA; mature Fam20C:cMyc, 72.5 kDa). **B**) Band relative quantification of Western blot detected proteins. Samples collected at 4, 6 and 8 dpi. Four leaf discs from three different plants were collected for each sample. All samples were treated with PBST$^{0.1\%}$ extraction buffer. A total of 30 µL TSP (total capacity of the well, ranging from 33-54 µg of PBST$^{0.1\%}$ samples) sample, and equal volume (30 µL) of each sample but extracted with reducing extraction buffer (SDS-DTT) were loaded on the gel. Blue arrows denote Fam20C:cMyc. eGEHK: protein standard. Proteins were detected with c-Myc antibody. Error bars are standard deviations of the means.

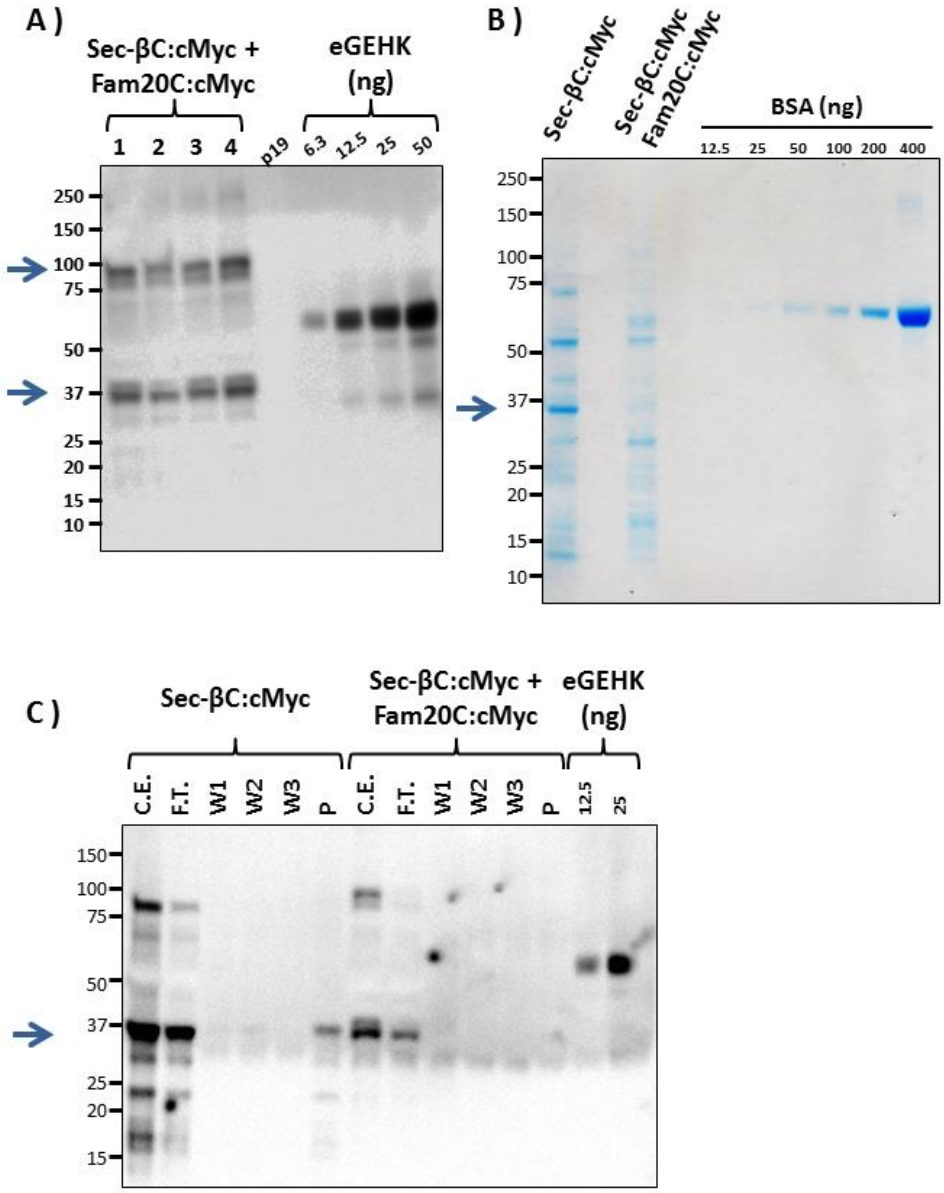


Figure 5-9 Transient expression of β-casein with Fam20C

A) Western blot detection of Sec-βC-cMyc and Fam20C:cMyc co-expressed with p19 (OD600 ratios; Sec-βC:cMyc: 0.55; Fam20C:cMyc: 0.55; p19:0.1). Samples collected at 4 dpi. 1, 2, 3, 4: biological replicates. Samples of 40 µg TSP were loaded per well. eGEHK: protein standard. Blue arrows denote Sec-βC:cMyc (33 kDa) and Fam20C:cMyc (72.5 kDa). Proteins were detected with c-Myc antibody. **B**) Approximately 0.3 gr (Sec-βC:cMyc) and 0.5 gr (Sec-βC:cMyc + Fam20C:cMyc) of whole plant infiltrated leaves were homogenized and c-Myc purified. Samples of 30 µL of eluted purified fraction were loaded on the SDS-PAGE. BSA (ng): bovine serum albumin used as standard. Blue arrow denotes Sec-βC:cMyc. **C**) Western blot detection of Sec-βC:cMyc and Fam20C:cMyc through purification steps. C.E: Crude extract; F.T.: Flow Through; W1, W2, W3: Wash 1, 2 and 3; P.: Purified fraction. eGEHK: protein standard. Blue arrow denotes Sec-βC:cMyc.

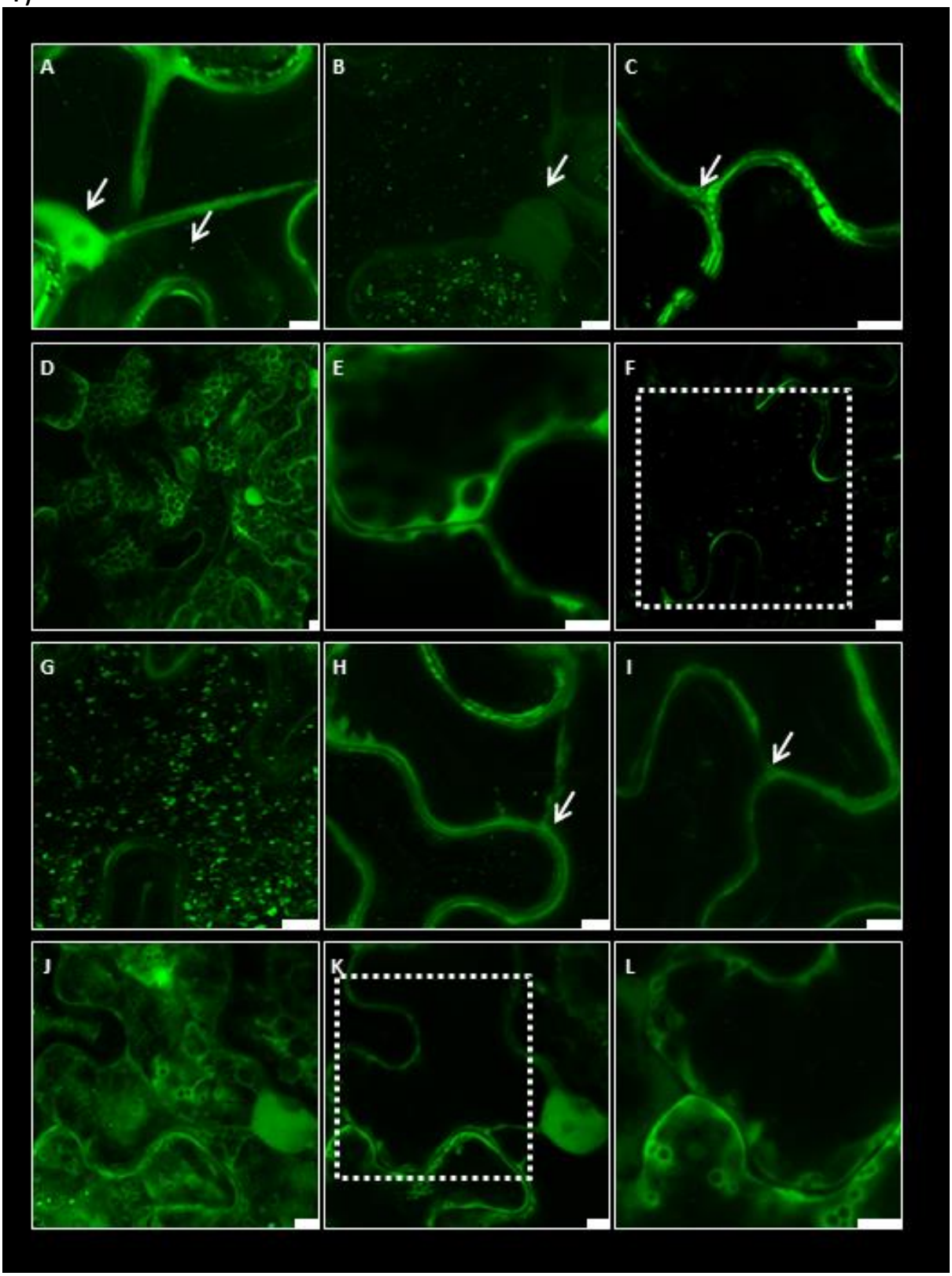
5.3.4 Transient expression of κC in *N. benthamiana* plants and subcellular localization.

Similar to βC, κC was first expressed as a fusion protein with eYFP, Sec-κC:eYFP, in order to determine subcellular localization. Sec-κC:eYFP was found to localize in the secretory pathway, being observed as multiple dots in the cytoplasm, denoting a potential Golgi localization (**Figure 5-10, A, F, G**). It was also observed in the apoplast (**Figure 5-10, C,H-I**), indication of proper secretion with the native secretory signal peptide. However, similar to βC, κC was also observed in the nuclei (**Figure 5-10, A-B**).

Nonetheless, when samples expressing Sec-κC:eYFP were subjected to western blot analysis at 4, 6 and 8 dpi, only a band corresponding to the eYFP tag could be detected, indicating cleavage of the protein and low accumulation levels, since the biological samples showed only faint bands in comparison to the control treatment (Cytoplasmic eYFP) (**Figure 5-10, II**).

In order to test effect of the tag and improve accumulation by affinity purification, Sec-κC was then expressed fused with the c-Myc tag. Expression of Sec-κC:cMyc was also assessed by co-expression with Sec-βC:cMyc in an attempt to allow the interaction of both caseins and perhaps provide structural stability and/or protection against proteases. In these both cases, and knowing the low accumulation levels of both caseins, protein extraction was performed under denaturing and reducing conditions, since it was observed to improve protein extraction of beta casein and Fam20C, and the SDS-PAGE wells were loaded to their maximum volume capacity (60 µL). Protein extraction was performed at 4, 6 and 8 dpi and all samples were assessed at the same time. A potential band corresponding to the mature κC (29 kDa) could be detected when expressed on its own and co-expressed with βC (33 kDa) (**Figure 5-11)**. Nevertheless, these assays should be repeated in the future with an appropriate negative control. Due to the number of samples and wells available, a negative control could not be included in this experiment. On the other hand, since protein degradation and low accumulation levels seem to be a common case for both caseins, although κC appears to be more vulnerable than βC to protein degradation, the expression strategy should be redesigned in order to provide protection against proteolysis and increase the accumulation levels, therefore improving the yields during purification and mass spectrometry analysis.

I)



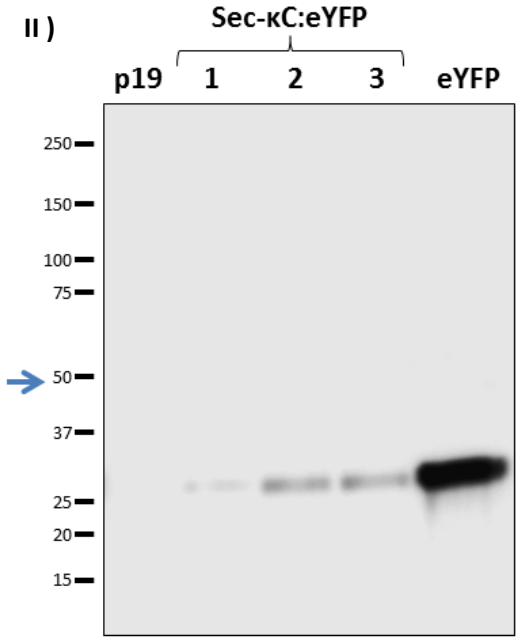


Figure 5-10 Transient expression of Sec-κC:eYFP in *N. benthamiana* plants

I) Transient expression of Sec-κC**:**eYFP (Secretory) (**A-C, F-I**). CyteYFP (**D-E, J-L**). Inset white box in **F** and **K** depicts microscopic zoom, shown in **G** and **L**, respectively. White arrows depict Sec-κC:eYFP localization in the Golgi (**A**), apoplast (**C, H-I**), and nucleus (**A-B**). **A-E**: Tissue collected at 4 dpi. **F-L**: Tissue collected at 6 dpi. Inset bar: 5 µm. **II)** Western Blot of transiently expressed Sec-κC:eYFP. Samples collected at 4 dpi. Sec-κC:eYFP (mature secretory, 49 kDa). CytYFP (Cytoplasmic eYFP, 29 kDa). Samples were treated with $PBST^{0.1\%}$ extraction buffer. Samples of 50 µg TSP of $PBST^{0.1\%}$ were loaded on the gel. Proteins were detected with GFP antibody. *Agrobacterium* strain AGL-1 was used to express Sec-κC:eYFP, CytYFP construct was carried by the *Agrobacterium* strain GV3101, and p19 by EHA105. Both Sec-κC:eYFP and CytYFP were co-infiltrated with p19.

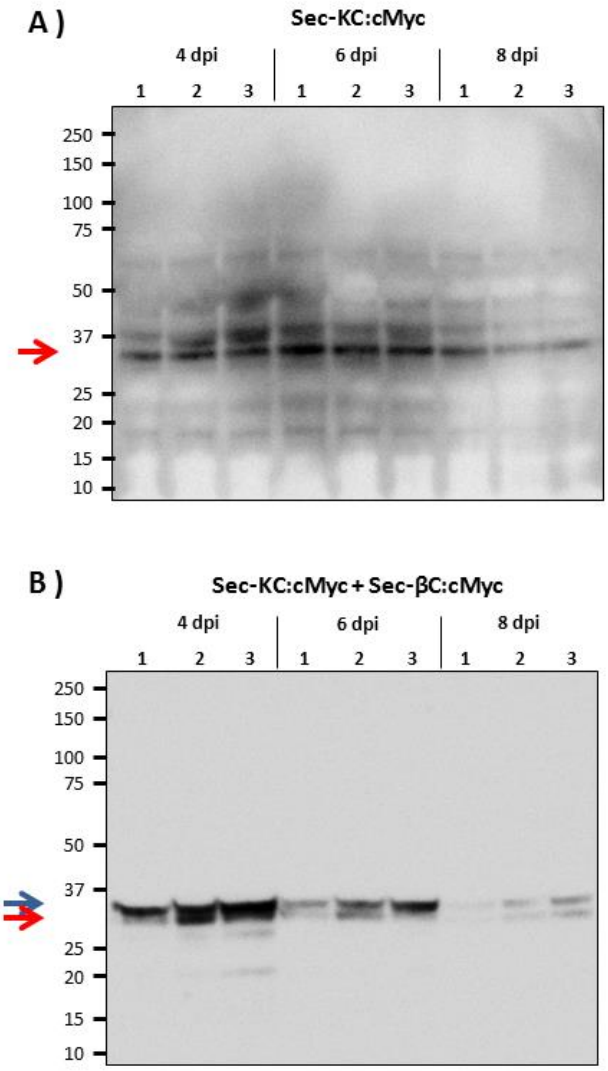


Figure 5-11 Transient expression of Sec-κC:cMyc in *N. benthamiana* plants

A) Western blot detection of Sec-κC:cMyc (premature Sec-κC:cMyc: 31 kDa; mature Sec-κC:cMyc, 29 kDa). **B)** Western blot detection of Sec-κC:cMyc co-expressed with Sec-βC:cMyc (premature Sec-βC:cMyc: 35 kDa; mature Sec-βC:cMyc, 33 kDa). Samples collected at 4, 6 and 8 dpi. Four leaf discs from three different plants were collected for each sample. Samples of 60 µL (maximum well capacity) treated with reducing extraction buffer (SDS-DTT) were loaded on the gel. Blue arrow denotes Sec-βC:cMyc; Red arrow denotes Sec-κC:cMyc. Proteins were detected with c-Myc antibody.

5.4 Conclusions

In the present study, the bovine βC was successfully expressed transiently in *N. benthamiana* plants via Agroinfiltration. The bovine βC was targeted to the secretory pathway using its own native signal peptide, and it was detected in the Golgi apparatus and apoplast. During secretion in the mammalian cells, beta casein was specifically phosphorylated by the casein kinase FAM20C. A putative bovine FAM20C was also expressed in this plant system, however further studies are required to characterize the kinase activity. Due to low accumulation levels and severe protein degradation of βC, especially when co-expressed with the bovine FAM20C, βC and its phosphorylation pattern were not confirmed by mass spectrometry. A similar situation was observed with κC, which could be observed by confocal microscopy, but the protein could not be properly identified by serology, since only the cleaved tag was detected. Further analysis allowed the identification of a putative band corresponding to κC. However, low accumulation levels due to proteolysis would not make amenable downstream procedures to allow proper characterization of both types of caseins. Therefore, it would be advisable to redesign the expression strategy for both caseins, as ER retention or other subcellular compartments as the cytoplasm and vacuoles and co-expression with the putative bovine kinase FAM20C to allow proper phosphorylation.

Chapter 6

- ## General Conclusions

Several mammalian proteins of human interest, as pharmaceutical products or food supplements are commonly produced with PTM´s, like glycosylation and phosphorylation. Plants have emerged as alternative expression systems for the production of such proteins. Traditionally, expression of recombinant proteins, usually do not usually address such issues, not only in plant systems but also in other platforms. Lack of post-translational machineries in the expression system or protein targeting might hinder the proper maturation (PTM´s) of the recombinant protein, having a potential impact on protein stability, activity or specificity, and thus must be considered when designing a strategy of protein expression in plant platforms.

Chapter 7

- ## Future Work and Perspectives

Production of PMP´s with engineered glycosylation.

Based on the data and conclusions obtained in this study, it would be advisable to:

A. Further elongate GalNAc-*O*-Glycosylated G-CSF with Galactose and sialic acid. This is a common Mucin-type *O*-glycosylation moiety present in human proteins and it is the glycosylation moiety present in the CHO cells-derived G-CSF.

B. Building a multiple expression cassette in order to increase the expression of all genes involved in the expression of G-CSF and its Mucin-type *O*-glycosylation in the same cell, increasing the ratio of glycosylated G-CSF.

C. Generate *N. benthamiana* transgenic lines expressing the *O*-glycosylation machinery, allowing the glyco-enzymes to be pre-localized in the subcellular compartments before G-CSF gets expressed, and being therefore available and active to recognize and modify G-CSF as soon as it gets expressed.

D. Alternatively, relocalization of the glycosylation machinery to the ER could help to increase accumulation levels, while potentially still performing the specific glycosylation.

Production of mammalian-specific phosphorylated caseins.

Based on the observations and results obtained herein, it would be advisable to:

A. Retarget the casein proteins to other compartments, which would protect them against protein degradation, since they seemed to be extremely vulnerable to the proteolytic environment in the apoplast. As options for this type of cases, and considered that it is required to obtained the caseins in a phosphorylated state, the ER and cytoplasm could be considered, where still contact with plant specific kinases or co-expression with the casein-specific kinase Fam20C, might allow the proper phosphorylation pattern required.

B. The bovine Fam20C, which was expressed in this study for the first time, should be further characterized to determine its identity and functionality. It could also be considered for colocalization together with the casein proteins and test its capability to phosphorylate in such compartments.

Chapter 8

- **Appendix A**

Abbreviations and acronyms

CHO – Chinese Hamster Ovary

CMP - Caseinomacropeptide

Cyt-G-CSF - Cytoplasmic G-CSF

DMP4 - Dentin Matrix Protein 4. Also called FAM20C

EK - Enterokinase

ER – Endoplasmic Reticulum

FAM20C – Family with sequence similarity 20, member C. Also called DMP4

GFP:KDEL - Secretory GFP fused to the ER retention signal KDEL

GlcNAc - N-acetylglucosaminyl

GnT I - Acetylglucosaminyltransferase I

HexNAc - N-Acetylhexosamine

hG-CSF – Human Granulocyte-Colony Stimulating Factor

HRP - Horseradish Peroxidase

Hyl - Hydroxylysine

Hyp - Hydroxyproline

Man – Mannose

N. benthamiana – Nicotiana benthamiana

Nat-SP - Native Secretory Signal Peptide

OST – Oligosaccharyltransferase

PMP – Plan-Made Pharmaceuticals

PTGSS - Post-transcriptional gene silencing suppressor

PTM – Post-translational modification

Sec-G-CSF - Secretory G-CSF

SecGFP - Secretory GFP

STM:mRFP - Rat Syalyltransferase fused to mRFP. Also called STtmd for Rat Syalyltransferase and transmembrane domain.

UGGT - Glycoprotein glucosyltransferase

VVA - Vicia villosa agglutinin lectin

Zera-G-CSF - G-CSF fused with the Zera protein

βC – Beta casein

κC – Kappa casein

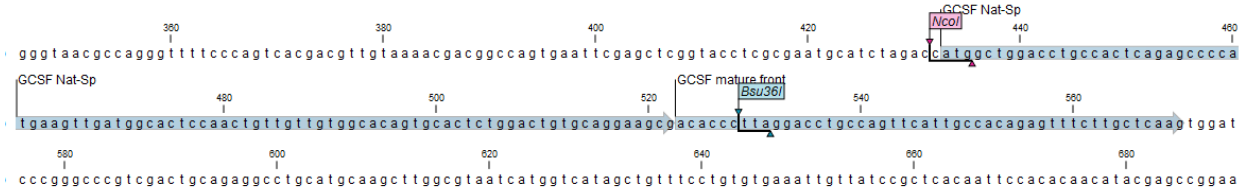
Chapter 9

- **Sequences**

Chemically synthesized sequences (only sequences of interest are depicted within pUC57):

pUC57 G-CSF Nat-SP

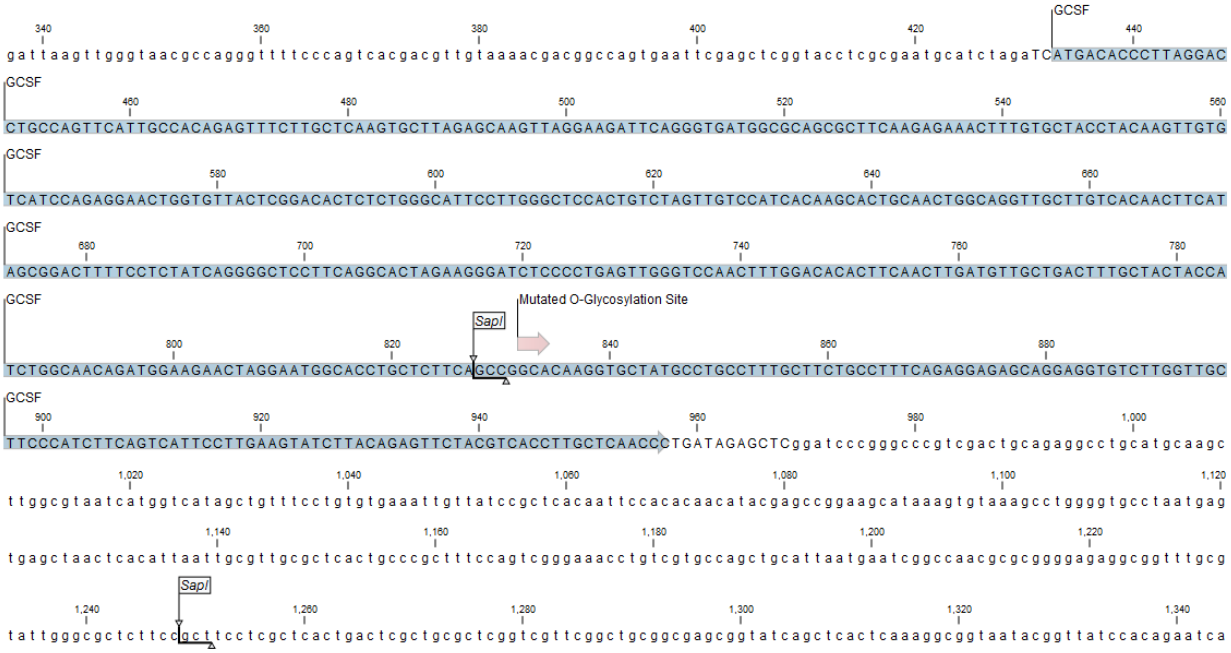
Native signal peptide DNA coding sequence for human Sec-G-CSF and corresponding amino acid sequence.



…MAGPATQSPMKLMALQLLLWHSALWTVQEA…

pUC57 Cyt-G-CSF

DNA sequence corresponding to the mature coding sequence of human G-CSF (without signal peptide) and corresponding amino acid sequence.



…MTPLGPASSLPQSFLLKCLEQVRKIQGDGAALQEKLCATYKLCHPEELVLLGHSLGIPWAPLS
SCPSQALQLAGCLSQLHSGLFLYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQMEELGM
APALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVSYRVLRHLAQP…

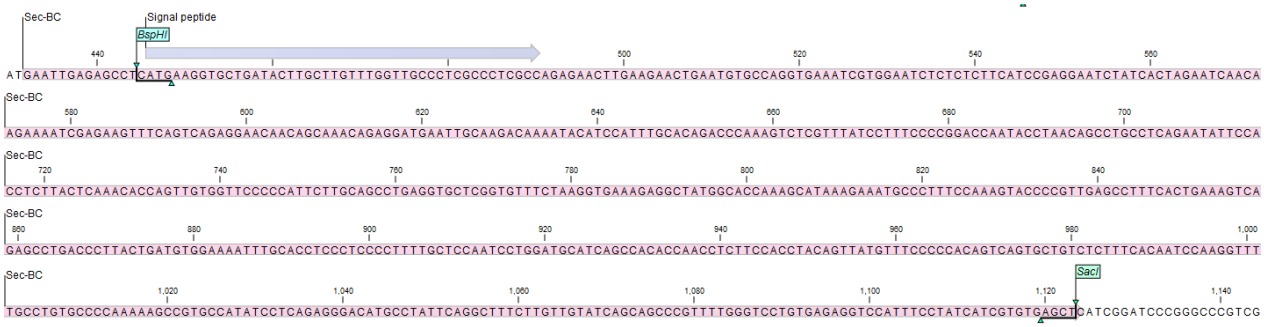
pUC57 Ek-G-CSF

DNA sequence corresponding to the mature coding sequence of human G-CSF (without signal peptide), and a 3´-end enterokinase coding sequence used as linker to fuse G-CSF with Zera® coding sequence and corresponding amino acid sequence.

…DDDDKTPLGPASSLPQSFLLKCLEQVRKIQGDGAALQEKLCATYKLCHPEELVLLGHSLGIP
WAPLSSCPSQALQLAGCLSQLHSGLFLYQGLLQALEGISPELGPTLDTLQLDVADFATTIWQQM
EELGMAPALQPTQGAMPAFASAFQRRAGGVLVASHLQSFLEVSYRVLRHLAQP…

pUC57 Sec-βC

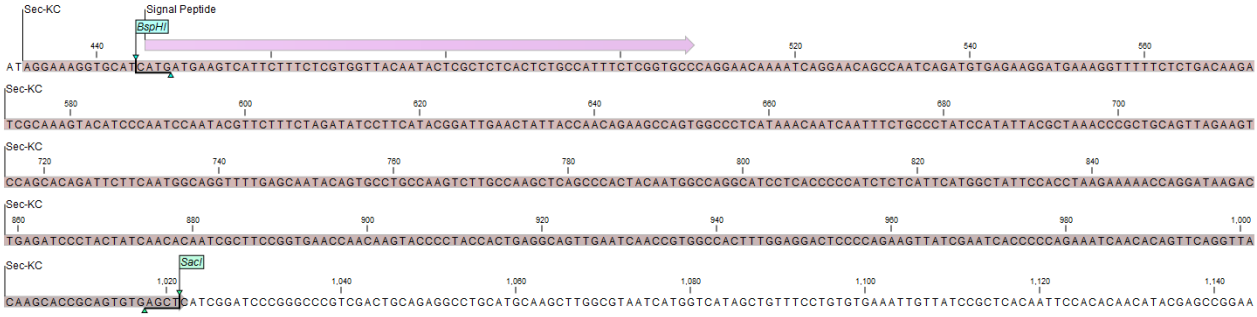
DNA coding sequence for expression of secretory bovine βC and corresponding amino acid sequence. Native signal peptide coding sequence is marked with an arrow (DNA) and underlined (amino acid sequence).



…MKVLILACLVALALARELEELNVPGEIVESLSSSEESITRINKKIEKFQSEEQQQTEDELQDKIH
PFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPFLQPEVLGVSKVKEAMAPKHKEMPFPKY
PVEPFTESQSLTLTDVENLHLPPLLLQSWMHQPHQPLPPTVMFPPQSVLSLSQSKVLPVPQKA
VPYPQRDMPIQAFLLYQQPVLGPVRGPFPIIV…

pUC57 Sec-κC

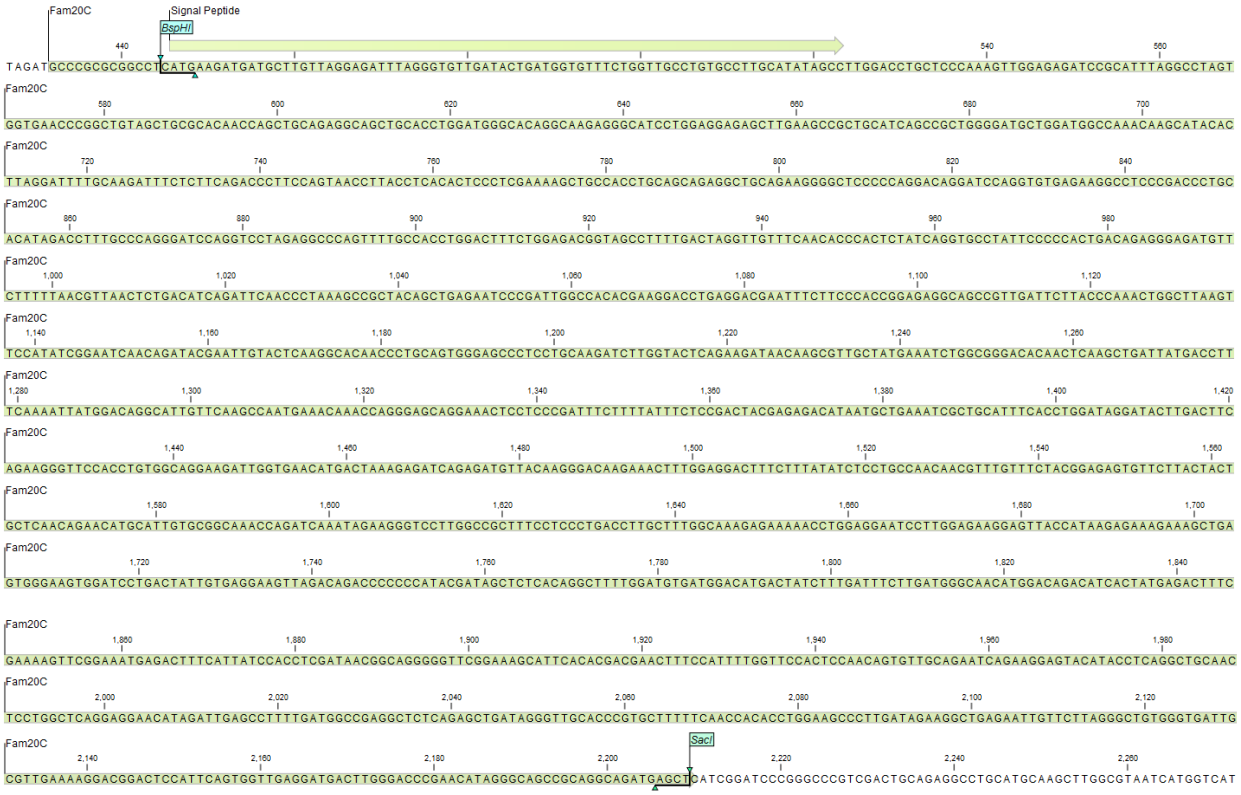
DNA coding sequence for expression of secretory bovine κC and corresponding amino acid sequence. Native signal peptide coding sequence is marked with an arrow (DNA) and underlined (amino acid sequence).



…MMKSFFLVVTILALTLPFLGAQEQNQEQPIRCEKDERFFSDKIAKYIPIQYVLSRYPSYGLNYY
QQKPVALINNQFLPYPYYAKPAAVRSPAQILQWQVLSNTVPAKSCQAQPTTMARHPHPHLSFM
AIPPKKNQDKTEIPTINTIASGEPTSTPTTEAVESTVATLEDSPEVIESPPEINTVQVTSTAV…

pUC57 Fam20C

DNA coding sequence for expression of secretory bovine Fam20C and corresponding amino acid sequence. Native signal peptide coding sequence is marked with an arrow (DNA) and underlined (amino acid sequence).



…MKMMLVRRFRVLILMVFLVACALHIALDLLPKLERSAFRPSGEPGCSCAQPAAEAAAPGWAQ
ARGHPGGELEAAASAAGDAGWPNKHTLRILQDFSSDPSSNLTSHSLEKLPPAAEAAEGAPPG
QDPGVRRPPDPAHRPLPRDPGPRGPVLPPGLSGDGSLLTRLFQHPLYQVPIPPLTEGDVLFNV
NSDIRFNPKAATAENPDWPHEGPEDEFLPTGEAAVDSYPNWLKFHIGINRYELYSRHNPAVGA
LLQDLGTQKITSVAMKSGGTQLKLIMTFQNYGQALFKPMKQTREQETPPDFFYFSDYERHNAEI
AAFHLDRILDFRRVPPVAGRLVNMTKEIRDVTRDKKLWRTFFISPANNVCFYGECSYYCSTEHA
LCGKPDQIEGSLAAFLPDLALAKRKTWRNPWRRSYHKRKKAEWEVDPDYCEEVRQTPPYDSS
HRLLDVMDMTIFDFLMGNMDRHHYETFEKFGNETFIIHLDNGRGFGKHSHDELSILVPLQQCCR
IRRSTYLRLQLLAQEEHRLSLLMAEALRADRVAPVLFQPHLEALDRRLRIVLRAVGDCVEKDGL
HSVVEDDLGPEHRAAAGR…

Chapter 10

- **Literature**

Ahmad, A., Pereira, E. O., Conley, A. J., Richman, A. S., & Menassa, R. (2010). Green biofactories: recombinant protein production in plants. *Recent Patents on Biotechnology*, *4*(May 2014), 242–259.

Alvarez, M. L., Topal, E., Martin, F., & Cardineau, G. a. (2010). Higher accumulation of F1-V fusion recombinant protein in plants after induction of protein body formation. *Plant Molecular Biology*, *72*, 75–89.

Bae, C. S., Yang, D. S., Lee, J., & Park, Y. H. (1999). Improved process for production of recombinant yeast-derived monomeric human G-CSF. *Applied Microbiology and Biotechnology*, *52*(3), 338–44.

Bardor, M., Faveeuw, C., Fitchette, A. C., Gilbert, D., Galas, L., Trottein, F., … Lerouge, P. (2003). Immunoreactivity in mammals of two typical plant glyco-epitopes, core α(1,3)-fucose and core xylose. *Glycobiology*, *13*(6), 427–434.

Barreda, D. R., Hanington, P. C., & Belosevic, M. (2004). Regulation of myeloid development and function by colony stimulating factors. *Developmental and Comparative Immunology*, *28*, 509–554.

Bernard, H., Meisel, H., Creminon, C., & Wal, J. M. (2000). Post-translational phosphorylation affects the IgE binding capacity of caseins. *FEBS Letters*, *467*, 239–244.

Böer, E., Steinborn, G., Kunze, G., & Gellissen, G. (2007). Yeast expression platforms. *Applied Microbiology and Biotechnology*, *77*(3), 513–523.

Bönig, H., Silbermann, S., Weller, S., Kirschke, R., Körholz, D., Janssen, G., … Nürnberger, W. (2001). Glycosylated vs non-glycosylated granulocyte colony-stimulating factor (G-CSF)--results of a prospective randomised monocentre study. *Bone Marrow Transplantation*, *28*(3), 259–64.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*, 248–254.

Brooks, S. A. (2004). Appropriate Glycosylation of Recombinant Proteins for Human Use Implications of Choice of Expression System. *Molecular Biotechnology, 28,* 241-255.

Brophy, B., Smolenski, G., Wheeler, T., Wells, D., L´Huillier, P., & Laible, G. (2003). Cloned transgenic cattle produce milk with higher levels of β-casein and κ-casein. *Nature*, *21*(February), 157–162.

Broudy, V. C., Kaushansky, K., Harlan, J. M., & Adamson, J. W. (1987). Interleukin 1 stimulates human endothelial cells to produce granulocyte-macrophage colony-stimulating factor and granulocyte colony-stimulating factor. *Journal of Immunology (Baltimore, Md. : 1950)*, *139*(2), 464–8.

Burda, P., & Aebi, M. (1999). The dolichol pathway of N-linked glycosylation. *Biochimica et Biophysica Acta*, *1426*(2), 239–57.

Carter, C. R. D., Keeble, J. R., & Thorpe, R. (2004). Human serum inactivates non-glycosylated but not glycosylated granulocyte colony stimulating factor by a protease dependent mechanism: Significance of carbohydrates on the glycosylated molecule. *Biologicals*, *32*, 37–47.

Carter, C. R. D., Whitmore, K. M., & Thorpe, R. (2004). The significance of carbohydrates on G-CSF: differential sensitivity of G-CSFs to human neutrophil elastase degradation. *Journal of Leukocyte Biology*, *75*(March), 515–522.

Cases, B., García-Ara, C., Boyano, M., Pérez-Gordo, M., Pedrosa, M., Vivanco, F., … Pastor-vargas, C. (2011). Phosphorylation reduces the allergenicity of cow casein in children with selective allergy to goat and sheep milk. *Journal of Investigational Allergology and Clinical Immunology*, *21*(5), 398–400.

Castilho, A., Neumann, L., Gattinger, P., Strasser, R., Vorauer-Uhl, K., Sterovsky, T., … Steinkellner, H. (2013). Generation of Biologically Active Multi-Sialylated Recombinant Human EPOFc in Plants. *PLoS ONE*, *8*(1), 1–13.

Chavez-Tapia, N. C., Mendiola-Pastrana, I., Ornelas-Arroyo, V. J., Noreña-Herrera, C., Vidaña-Perez, D., Delgado-Sanchez, G., … Barrientos-Gutierrez, T. (2015). Granulocyte-colony stimulating factor for acute-on-chronic liver failure : systematic review and meta-analysis, *14*(5), 631–641.

Choi, B., & Jime, R. (2001). Expression and Purification of Glycosylated Bovine -Casein (L70S / P71S) in Pichia pastoris. *Journal of Agricultural and Food Chemistry*, *49*(4), 1761–1766.

Chong, D. K., Roberts, W., Arakawa, T., Illes, K., Bagi, G., Slattery, C. W., & Langridge, W. H. (1997). Expression of the human milk protein beta-casein in transgenic potato plants. *Transgenic Research*, *6*, 289–296.

Coku, A. (2007). Molecular Farming of Vaccines from Transgenic Plants. *Basic Biotechnology eJournal Molecular*, *3*, 110–116.

Conley, A. J., Joensuu, J. J., Menassa, R., & Brandle, J. E. (2009). Induction of protein body formation in plant leaves by elastin-like polypeptide fusions. *BMC Biology*, *7*, 48.

Conley, A. J., Joensuu, J. J., Richman, A., & Menassa, R. (2011). Protein body-inducing fusions for high-level production and purification of recombinant proteins in plants. *Plant Biotechnology Journal*, *9*, 419–433.

Corfield, A. (2016). Eukaryotic protein glycosylation: a primer for histochemists and cell biologists. *Histochemistry and Cell Biology*.

Craig, R., & Beavis, R. C. (2004). TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics*, *20*(9), 1466–1467.

Daniell, H., Singh, N. D., Mason, H., & Streatfield, S. J. (2009). Plant-made vaccine antigens and biopharmaceuticals. *Trends in Plant Science*, *14*(12), 669–79.

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*, 62.

De Kruif, C. G. (1999). Casein micelle interactions. *International Dairy Journal*, *9*, 183–188.

Dell, A., Galadari, A., Sastre, F., & Hitchen, P. (2010). Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *International Journal of Microbiology*, *2010*.

Desai, P. N., Shrivastava, N., & Padh, H. (2010). Production of heterologous proteins in plants: Strategies for optimal expression. *Biotechnology Advances*, *28*(4), 427–435.

Dicker, M., Schoberer, J., Vavra, U., & Strasser, R. (2015). Subcellular Targeting of Proteins Involved in Modification of Plant N- and O-Glycosylation. In A. Castilho (Ed.), *Methods in Molecular Biology* (Vol. 1321, pp. 249–267). New York, NY: Humana Press, New York, NY.

Dicker, M., Tschofen, M., Maresch, D., Konig, J., Juarez, P., Orzaez, D., … Strasser, R. (2016). Transient Glyco-Engineering to Produce Recombinant IgA1 with Defined N- and O-Glycans in Plants. *Front Plant Sci*, *7*(January), 18.

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*, 490–495.

Duncan, N., Hewetson, M., Atra, A., Dick, G., & Pinkerton, R. (1997). An economic evaluation of the use of granulocyte colony-stimulating factor after bone marrow transplantation in children. *PharmacoEconomics*, *11*(2), 169–74. Retrieved from

Fahimirad, S., Abtahi, H., Razavi, S., Alizadeh, H., & Ghorbanpour, M. (2017). Production of Recombinant Antimicrobial Polymeric Protein Beta Casein and Its Antimicrobial Synergistic Effects Assessment with Thymol. *Molecules*, *22*, 822.

Farrell, H. M., Qi, P. X., Wickham, E. D., & Unruh, J. J. (2001). Secondary structural studies of bovine caseins : temperature dependence of b-casein structure as analyzed by circular dichroism and FTIR spectroscopy and correlation with micellization q, *15*(5), 341–354.

Faye, L., Boulaflous, A., Benchabane, M., Gomord, V., & Michaud, D. (2005). Protein modifications in the plant secretory pathway: Current status and practical implications in molecular pharming. *Vaccine*, *23*, 1770–1778.

Finnie, C. (2007). Combining Chymotrypsin/Trypsin Digestion to Identify Hydrophobic Proteins from Oil Bodies. *Plant Proteomics*, *1072*, 1–253.

Fischer, R., & Emans, N. (2000). Molecular farming of pharmaceutical proteins. *Transgenic Research*, *9*(4–5), 279–99; discussion 277.

Ginger, M. R., & Grigor, M. R. (1999). Comparative aspects of milk caseins. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, *124*, 133–145.

Gomord, V., & Faye, L. (2004). Posttranslational modification of therapeutic proteins in plants. *Current Opinion in Plant Biology*, *7*(2), 171–81.

Gomord, V., Fitchette, A.-C., Menu-Bouaouiche, L., Saint-Jore-Dupas, C., Plasson, C., Michaud, D., & Faye, L. (2010a). Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnology Journal*, *8*(5), 564–87.

Gomord, V., Fitchette, A. C., Menu-Bouaouiche, L., Saint-Jore-Dupas, C., Plasson, C., Michaud, D., & Faye, L. (2010b). Plant-specific glycosylation patterns in the context of therapeutic protein production. *Plant Biotechnology Journal*, *8*, 564–587.

Hansson, L., Bergström, S., Hernell, O., Lönnerdal, B., Nilsson, A. K., & Strömqvist, M. (1993). Expression of Human Milk β-Casein in Escherichia coli: Comparison of Recombinant Protein with Native Isoforms. *Protein Expression and Purification*, *4*, 373–381.

Hattori, K., Orita, T., Oheda, M., Tamura, M., & Ono, M. (1996). Comparative study of the effects of granulocyte colony-stimulating factor and granulocyte-macrophage colony-stimulating factor on generation and mobilization of neutrophils in cyclophosphamide-treated neutropenic mice. *In Vivo (Athens, Greece)*, *10*(3), 319–27.

Hermesh, T., Moran, T. M., Jain, D., & López, C. B. (2012). Granulocyte colony-stimulating factor protects mice during respiratory virus infections. *PLoS ONE*, *7*(5), 2–9.

Höglund, M. (1998). Glycosylated and non-glycosylated recombinant human granulocyte colony-stimulating factor (rhG-CSF)-what is the difference? *Medical Oncology*, *15*, 229–33.

Holt, C. (2013). Invited review : Caseins and the casein micelle : Their biological functions , structures , and behavior in foods Invited review : Caseins and the casein micelle : Their biological functions , *96*, 6127–6146.

Hong, S.-Y., Kwon, T.-H., Jang, Y.-S., Kim, S.-H., & Yang, M.-S. (2006). Production of bioactive human granulocyte-colony stimulating factor in transgenic rice cell suspension cultures. *Protein Expression and Purification*, *47*(1), 68–73.

Hong, S.-Y., Kwon, T.-H., Lee, J.-H., Jang, Y.-S., & Yang, M.-S. (2002). Production of biologically active hG-CSF by transgenic plant cell suspension culture. *Enzyme and Microbial Technology*, *30*(6), 763–767.

Houston, A. C., Stevens, L. A., & Cour, V. (1999). Pharmacokinetics of glycosylated recombinant human granulocyte colony-stimulating factor (lenograstim) in healthy male volunteers. *British Journal of Clinical Pharmacology*, *47*, 279–84.

Jabed, A., Wagner, S., McCracken, J., Wells, D. N., & Laible, G. (2012). Targeted microRNA expression in dairy cattle directs production of β-lactoglobulin-free, high-casein milk. *PNAS*, *109*(42), 16811–16816. Retrieved from

Jiang, Y. J., Li, Q. Z., Yang, H. B., & Geng, L. J. (2004). Expression and Bioactivity Analysis of Recombinant Beta-CPP Dimer. *Journal of Dairy Science*, *87*(10), 3198–3208.

Jimenez-Flores, R., Richardson, T., & Bisson, L. F. (1990). Expression of bovine β-casein in Saccharomyces cerevisiae and characterization of the protein produced in vivo. *Journal of Agricultural and Food Chemistry*, *38*(4), 1134–1141.

Jin, C., Altmann, F., Strasser, R., Mach, L., Schähs, M., Kunert, R., … Steinkellner, H. (2008). A plant-derived human monoclonal antibody induces an anti-carbohydrate immune response in rabbits. *Glycobiology*, *18*(3), 235–241.

Johansson, I. (2002). Milk and dairy products: possible effects on dental health. *Food & Nutrition Research*, (12), 119–122.

Joseph, M., Ludevid, M. D., Torrent, M., Rofidal, V., Tauzin, M., & Rossignol, M. (2012). Proteomic characterisation of endoplasmic reticulum-derived protein bodies in tobacco leaves. *BMC Plant Biology*, *12*(1), 36.

Kang, Y. C. (1988). Molecular Cloning and Expression of Bovine κ-Casein in Escherichia coli. *Journal of Dairy Science*, *71*(1), 29–40.

Kapila, J., De Rycke, R., Van Montagu, M., & Angenon, G. (1997). An Agrobacterium-mediated transient gene expression system for intact leaves. *Plant Science*, *122*(1), 101–108.

Kaushansky, K., Lin, N., & Adamson, J. W. (1988). Interleukin 1 stimulates fibroblasts to synthesize granulocyte-macrophage and granulocyte colony-stimulating factors. Mechanism for the hematopoietic response to inflammation. *Journal of Clinical Investigation*, *81*(January), 92–97.

Kessner, D., Chambers, M., Burke, R., Agus, D., & Mallick, P. (2008). ProteoWizard: Open source software for rapid proteomics tools development. *Bioinformatics*, *24*(21), 2534–2536.

Kethireddipalli, P., & Hill, A. (2015). Rennet coagulation and cheese making properties of

thermally processed milk: overview and recent developments. *Journal of Agricultural and Food Chemistry*.

Khan, A. H., Bayat, H., Rajabibazl, M., Sabri, S., & Rahimpour, A. (2017). Humanizing glycosylation pathways in eukaryotic expression systems. *World Journal of Microbiology and Biotechnology*, *33*(1), 1–12.

Kim, Y.-J., Oh, Y.-K., Kang, W., Lee, E. Y., & Park, S. (2005). Production of human caseinomacropeptide in recombinant Saccharomyces cerevisiae and Pichia pastoris. *Journal of Industrial Microbiology & Biotechnology*, *32*, 402–8.

Kim, Y.-J., Park, S., Oh, Y.-K., Kang, W., Kim, H. S., & Lee, E. Y. (2005). Purification and characterization of human caseinomacropeptide produced by a recombinant Saccharomyces cerevisiae. *Protein Expression and Purification*, *41*, 441–6.

Ko, K., & Koprowski, H. (2005). Plant biopharming of monoclonal antibodies. *Virus Research*, *111*, 93–100.

Kraševec, N., Milunović, T., Lasnik, M. a., Lukančič, I., Komel, R., & Porekar, V. G. (2014). Human granulocyte colony stimulating factor (G-CSF) produced in the filamentous fungus aspergillus niger. *Acta Chimica Slovenica*, *61*, 709–717.

Lasnik, M. A., Porekar, V. G., & Stalc, A. (2001). Human granulocyte colony stimulating factor (hG-CSF) expressed by methylotrophic yeast pichia pastoris. *Pflugers Archiv : European Journal of Physiology*, *442*(6 Suppl 1), R184-6.

Latijnhouwers, M., Hawes, C., Carvalho, C., Oparka, K., Gillingham, A. K., & Boevink, P. (2005). An Arabidopsis GRIP domain protein locates to the trans-Golgi and binds the small GTPase ARL1. *Plant Journal*, *44*, 459–470.

Liu, F., Liao, L., & Chen, J. (2008). Preparation of unglycosylated human caseinomacropeptide by engineering DAB escherichia coli. *Journal of Agricultural and Food Chemistry*, *56*, 889–893.

Lôbo Machado, E., Da Silva, A. C., Da Silva, M. J., Leite, A., & Ottoboni, L. M. M. (2002). Endogenous protein phosphorylation and casein kinase activity during seed development in Araucaria angustifolia. *Phytochemistry*, *61*, 835–842.

Mathieu-Rivet, E., Kiefer-Meyer, M.-C., Vanier, G., Ovide, C., Burel, C., Lerouge, P., & Bardor, M. (2014). Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Frontiers in Plant Science*, *5*(July).

Mattii, L., Azzarà, A., Fazzi, R., Carulli, G., Chimenti, M., Cecconi, N., … Petrini, M. (2005). Glycosylated or non-glycosylated G-CSF differently influence human granulocyte functions through RhoA. *Leukemia Research*, *29*, 1285–1292.

Maughan, P. J., Philip, R., Cho, M.-J., Widholm, J. M., & Vodkin, L. O. (1999). Biolistic transformation, expression, and inheritance of bovine β-casein in soybean (Glycine max). *In Vitro Cellular & Developmental Biology - Plant*, *35*(4), 344–349.

Mire-Sluis, A. R., Das, R. G., & Thorpe, R. (1995). The international standard for granulocyte colony stimulating factor (G-CSF). Evaluation in an international collaborative study. Participants of the Collaborative Study. *Journal of Immunological Methods*, *179*(1), 117–26.

Nair, N. R., Chidambareswaren, M., & Manjula, S. (2014). Enhanced heterologous expression of biologically active human granulocyte colony stimulating factor in transgenic tobacco BY-2 cells by localization to endoplasmic reticulum. *Molecular Biotechnology*, *56*, 849–862.

Nakamura, S., Mano, S., Tanaka, Y., Ohnishi, M., Nakamori, C., Araki, M., … Ishiguro, S. (2010). Gateway binary vectors with the bialaphos resistance gene, bar, as a selection

marker for plant transformation. *Bioscience, Biotechnology, and Biochemistry*, *74*(6), 1315–9.

Nicola, N. A., & Metcalf, D. (1985). Binding of125I-labeled granulocyte colony-stimulating factor to normal murine hemopoietic cells. *Journal of Cellular Physiology*, *124*(2), 313–321.

Oh, S., & Richardson, T. (1991). Genetic engineering of bovine. kappa.-casein to improve its nutritional quality. *Journal of Agricultural and Food Chemistry*, *39*(Figure 1), 422–427.

Palmblad, J., Gyllenhammar, H., Lindgren, J. A., & Malmsten, C. L. (1984). Effects of leukotrienes and f-Met-Leu-Phe on oxidative metabolism of neutrophils and eosinophils. *Journal of Immunology (Baltimore, Md. : 1950)*, *132*(6), 3041–5. Retrieved from

Panouillé, M., Durand, D., & Nicolai, T. (2005). Jamming and gelation of dense ??-casein micelle suspensions. *Biomacromolecules*, *6*, 3107–3111.

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*, 3019.

Philip, R., Darnowski, D. W., Maughan, P. J., & Vodkin, L. O. (2001). Processing and localization of bovine beta-casein expressed in transgenic soybean seeds under control of a soybean lectin expression cassette. *Plant Science : An International Journal of Experimental Plant Biology*, *161*(2), 323–335.

Portnaya, I., Ben-Shoshan, E., Cogan, U., Khalfin, R., Fass, D., Ramon, O., & Danino, D. (2008). Self-assembly of bovine β-casein below the isoelectric pH. *Journal of Agricultural and Food Chemistry*, *56*, 2192–2198.

Reeves, G. (2014). Overview of Use of G-CSF and GM-CSF in the Treatment of Acute Radiation Injury. *Health Physics*, *106*, 699–703.

Reimand, J., Wagih, O., & Bader, G. D. (2013). The mutational landscape of phosphorylation signaling in cancer. *Scientific Reports*, *3*, 2651.

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

Sangadala, S., Sivakami, S., & Mendicino, J. (1991). UDP-GlcNAc: Gal beta 3GalNAc-mucin: (GlcNAc----GalNAc) beta 6-N-acetylglucosaminyltransferase and UDP-GlcNAc: Gal beta 3(GlcNAc beta 6) GalNAc-mucin (GlcNAc----Gal)beta 3-N-acetylglucosaminyltransferase from swine trachea epithelium. *Molecular and Cellular Biochemistry*, *101*(2), 125–43.

Semenov, D. V., Fomin, A. S., Kuligina, E. V., Koval, O. a., Matveeva, V. a., Babkina, I. N., … Richter, V. a. (2010). Recombinant analogs of a novel milk pro-apoptotic peptide, lactaptin, and their effect on cultured human cells. *Protein Journal*, *29*, 174–180.

Seo, J., & Lee, K.-J. (2004). Post-translational modifications and their biological functions: proteomic analysis and systematic approaches. *Journal of Biochemistry and Molecular Biology*, *37*(1), 35–44.

Sharifi Tabar, M., Akbar Habashi, A., & Rajabi Memari, H. (2013). Human granulocyte colony-stimulating factor (hg-csf) expression in plastids of lactuca sativa. *Iranian Biomedical Journal*, *17*(May 2012), 158–164.

Sharma, A. K., & Sharma, M. K. (2009). Plants as bioreactors: Recent developments and emerging opportunities. *Biotechnology Advances*, *27*(6), 811–32.

Shin, Y.-J., Hong, S.-Y., Kwon, T.-H., Jang, Y.-S., & Yang, M.-S. (2003). High level of expression of recombinant human granulocyte-macrophage colony stimulating factor in

transgenic rice cell suspension culture. *Biotechnology and Bioengineering*, *82*(7), 778–83.

Silhavy, D., Molnár, A., Lucioli, A., Szittya, G., Hornyik, C., Tavazza, M., & Burgyán, J. (2002). A viral protein suppresses RNA silencing and binds silencing-generated, 21- to 25-nucleotide double-stranded RNAs. *EMBO Journal*, *21*(12), 3070–3080.

Silva, S. V., & Malcata, F. X. (2005). Caseins as source of bioactive peptides. *International Dairy Journal*, *15*(1), 1–15.

Smyth, E., Clegg, R. A., & Holt, C. (2004). A biological perspective on the structure and function of caseins and casein micelles. *International Journal of Dairy Technology*, *57*(2), 121–126.

Spiro, R. G. (2002a). Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, *12*(4), 43R–56R. Retrieved from

Spiro, R. G. (2002b). Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, *12*(4), 43R–56R.

Suk Yang, D., Soon Bae, C., & Lee, J. (1997). Production of recombinant human granulocyte-colony-stimulating factor in high cell density yeast cultures. *Biotechnology Letters*, *19*(7), 655–659.

Sun, J. K., Eun, Y. C., Yoon, J. C., Ji, H. A., & Park, O. K. (2006). Proteomics studies of post-translational modifications in plants. *Journal of Experimental Botany*, *57*(7), 1547–1551.

Tabar, M. S., Solouki, M., Tohidfar, M., & Sadeghizadeh, M. (2012). Expression of human granulocyte-colony stimulating factor (hG-CSF) gene in tobacco (Nicotiana tabacum). *Australian Journal of Crop Science*, *6*(1), 135–140. Retrieved from

Tagliabracci, V. S., Engel, J. L., Wen, J., Wiley, S. E., Worby, C. a., Kinch, L. N., … Dixon, J. E. (2012). Secreted Kinase Phosphorylates Extracellular Proteins That Regulate Biomineralization. *Science*, *336*, 1150–1153.

Talebkhan, Y., Samadi, T., Samie, A., Barkhordari, F., & Azizi, M. (2016). Expression of granulocyte colony stimulating factor (GCSF) in Hansenula polymorpha, *8*(1), 21–28.

Tezcucano Molina, A. C., Alli, I., Konishi, Y., & Kermasha, S. (2007). Effect of dephosphorylation on bovine casein. *Food Chemistry*, *101*(3), 1263–1271.

Thurmond, J. M., Hards, R. G., Seipelt, C. T., Leonard, a E., Hansson, L., Strömqvist, M., … Mukerji, P. (1997). Expression and characterization of phosphorylated recombinant human beta-casein in Escherichia coli. *Protein Expression and Purification*, *10*, 202–8.

Tian, E., & Hagen, K. G. Ten. (2009). NIH Public Access, *26*(3), 325–334.

Tian, L., & Sun, S. S. (2011). Ubiquitin fusion expression and tissue-dependent targeting of hG-CSF in transgenic tobacco. *BMC Biotechnology*, *11*, 91.

Torrent, M., Llompart, B., Lasserre-Ramassamy, S., Llop-Tous, I., Bastida, M., Marzabal, P., … Ludevid, M. D. (2009). Eukaryotic protein production in designed storage organelles. *BMC Biology*, *7*, 5.

Tran, D. T., & Ten Hagen, K. G. (2013). Mucin-type o-glycosylation during development. *Journal of Biological Chemistry*, *288*(10), 6921–6929.

Twyman, R. M., Schillberg, S., & Fischer, R. (2012). Molecular Farming in Plants: Recent Advances and Future Prospects, 145–159.

Twyman, R. M., Stoger, E., Schillberg, S., Christou, P., & Fischer, R. (2003). Molecular farming in plants: host systems and expression technology. *Trends in Biotechnology*, *21*(12), 570–8.

Van den Steen, P., Rudd, P. M., Dwek, R. a, & Opdenakker, G. (1998). Concepts and principles of O-linked glycosylation. *Critical Reviews in Biochemistry and Molecular Biology*, *33*(3), 151–208.

Vanz, A. L., Renard, G., Palma, M. S., Chies, J. M., Dalmora, S. L., Basso, L. a, & Santos, D. S. (2008). Human granulocyte colony stimulating factor (hG-CSF): cloning, overexpression, purification and characterization. *Microbial Cell Factories*, *7*, 13.

Varki, a. (1993). Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*, *3*(2), 97–130.

Vaudel, M., Barsnes, H., Berven, F. S., Sickmann, A., & Martens, L. (2011). SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. *Proteomics*, *11*, 996–999.

Vaudel, M., Burkhart, J. M., Zahedi, R. P., Oveland, E., Berven, F. S., Sickmann, A., … Barsnes, H. (2015). PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nature Biotechnology*, *33*(1), 22–24.

Vellenga, E., Rambaldi, A., Ernst, T. J., Ostapovicz, D., & Griffin, J. D. (1988). Independent regulation of M-CSF and G-CSF gene expression in human monocytes. *Blood*, *71*(6), 1529–32.

Vukušić, K., Šikić, S., & Balen, B. (2016). Recombinant therapeutic proteins produced in plants: Towards engineering of human-type O- and N-glycosylation. *Periodicum Biologorum*, *118*(2), 75–90.

Wang, Y.-C., Peterson, S. E., & Loring, J. F. (2014). Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Research*, *24*, 143–60.

Williams, G. T., Smith, C. A., Spooncer, E., Dexter, T. M., & Taylor, D. R. (1990). Haemopoietic colony stimulating factors promote cell survival by suppressing apoptosis. *Nature*, *343*(6253), 76–79.

Yang, Z., Bennett, E. P., Jørgensen, B., Drew, D. P., Arigi, E., Mandel, U., … Petersen, B. L. (2012). Toward stable genetic engineering of human O-glycosylation in plants. *Plant Physiology*, *160*(1), 450–63.

Yang, Z., Drew, D. P., Jørgensen, B., Mandel, U., Bach, S. S., Ulvskov, P., … Petersen, B. L. (2012). Engineering mammalian mucin-type O-glycosylation in plants. *The Journal of Biological Chemistry*, *287*(15), 11911–23.

Yuo, A., Kitagawa, S., Ohsaka, A., Ohta, M., Miyazono, K., Okabe, T., … Takaku, F. (1989). Recombinant human granulocyte colony-stimulating factor as an activator of human granulocytes: potentiation of responses triggered by receptor-mediated agonists and stimulation of C3bi receptor expression and adherence. *Blood*, *74*(6), 2144–9.

Zvereva, a S., Petrovskaya, L. E., Rodina, a V, Frolova, O. Y., Ivanov, P. a, Shingarova, L. N., … Atabekov, J. G. (2009). Production of biologically active human myelocytokines in plants. *Biochemistry. Biokhimiia*, *74*(11), 1187–1194.

Chapter 11

• Internships

London Research and Development Centre, Agriculture and Agri-Food Canada (AAFC).

1391 Sandford St., London, ON, N5V 4T3 Canada

International Supervisor: Dr. Rima Menassa

Period: June 1, 2016 – April 9, 2017

• Conferences

Talk Transient expression of human Granulocyte-ColonyStimulating Factor variants in N. benthamiana plants, as non-glycosylated and O-glycosylated plant-derived proteins. 48 Congreso de Investigación y Desarrollo. ITESM. Jan 2018.

Poster. Engineering Plant *O*-Glycosylation. Biology Graduate Research Forum, Western University, London, Canada. Oct 2016.

Poster. Plant-derived *O*-glycosylated pharmaceuticals. Plant Biotech 2016, Queen´s University, Kingston, Canada. Jun 2016.

Talk. Design of the recombinant human Granulocyte-Colony Stimulating Factor (rhG-CSF) gene for its expression in NT1 cells. Congreso de Investigación y Desarrollo ITESM. Jan 2010

Chapter 12

- **Paper submitted and accepted**

Ramírez‐Alanis et al. Plant Methods (2018) 14,98

Ramírez-Alanis *et al. Plant Methods* (2018) 14:98
https://doi.org/10.1186/s13007-018-0363-y

Plant Methods

RESEARCH Open Access

Transient co-expression with three *O*-glycosylation enzymes allows production of GalNAc-*O*-glycosylated Granulocyte-Colony Stimulating Factor in *N. benthamiana*

Israel A. Ramírez-Alanis[1], Justin B. Renaud[3], Silverio García-Lara[1], Rima Menassa[3,4] and Guy A. Cardineau[1,2*]

Abstract

Background: Expression of economically relevant proteins in alternative expression platforms, especially plant expression platforms, has gained significant interest in recent years. A special interest in working with plants as bioreactors for the production of pharmaceutical proteins is related to low production costs, product safety and quality. Among the different properties that plants can also offer for the production of recombinant proteins, protein glycosylation is crucial since it may have an impact on pharmaceutical functionality and/or stability.

Results: The pharmaceutical glycoprotein human Granulocyte-Colony Stimulating Factor was transiently expressed in *Nicotiana benthamiana* plants and subjected to mammalian-specific mucin-type *O*-glycosylation by co-expressing the pharmaceutical protein together with the glycosylation machinery responsible for such post-translational modification.

Conclusions: The pharmaceutical glycoprotein human Granulocyte-Colony Stimulating Factor can be expressed in *N. benthamiana* plants via agroinfiltration with its native mammalian-specific mucin-type *O*-glycosylation.

Keywords: Mucin-type *O*-glycosylation, G-CSF, Granulocyte-Colony Stimulating Factor, *Nicotiana benthamiana*, Pharmaceutical glycoprotein, Molecular farming

Background

Plants have emerged as alternative expression systems for the production of pharmaceutical proteins [1–4]. While there are several advantages to using plants as expression systems, such as low cost of production and maintenance, fast scalability, biological safety, and proper protein folding and assembly [1–3, 5], there are also some limitations concerning bioactivity or quality of the produced proteins [6]. Although plants are known to perform post-translational modifications, just like mammalian cells do, differences in glycosylation patterns between plant and animal cells might represent a major drawback that can impact the biochemical properties of the plant-derived recombinant protein [5–10].

With regard to the differences in glycosylation patterns between humans and plants, plant *N*-glycosylation differs mainly in the attachment of α(1,3)-fucose and β(1,2)-xylose to the core *N*-glycan, while *N*-glycans in human and mammalian cells present α(1,6)-fucose and no xylose is attached. Further maturation of the plant core *N*-glycan typically results in a biantennary structure, occasionally with terminal Lewis A epitopes. Terminal Lewis A epitopes are rarely observed in human proteins, but are widely distributed in plant glycoproteins. Maturation of human *N*-glycans will present a multi-antennary structure with two or more terminal branches, further extended with galactose and sialic (neuraminic) acid [4, 11]. Major efforts to humanize *N*-glycosylation patterns

*Correspondence: guy.cardineau@tec.mx; guy.cardineau@asu.edu
[1] School of Engineering and Sciences, Tecnológico de Monterrey, Campus Monterrey, Av. Eugenio Garza Sada 2501 Sur, C.P. 64849 Monterrey, NL, Mexico
Full list of author information is available at the end of the article

in plants have focused on retaining the recombinant protein in the ER to avoid further modification of the core *N*-glycan with α(1,3)-fucose, β(1,2)-xylose and Lewis A, abolishing expression of glycosyltransferases responsible for the attachment of α(1,3)-fucose and β(1,2)-xylose, and further humanization has included the expression of human β(1,4)-galactosyltransferase and heterologous enzymes required to generate sialylation, which are present in human cells but absent in plants [4, 11–14].

In the case of *O*-glycosylation, patterns differ significantly between plants and human cells, and specifically from the human mucin-type *O*-glycosylation. In plants, protein *O*-glycosylation is characterized by the presence of *O*-glycans attached to hydroxyprolines (Hyp) and Ser residues. Typically, Hyp residues in plants are decorated with arabinogalactan polysaccharides or arabinose oligosaccharides, and Ser residues with galactose [6, 11, 15, 16]. While human mucin-type *O*-glycosylation is characterized by the attachment of *N*-acetylgalactosamine (GalNAc) to Ser or Thr residues, which can be further elongated with other sugars to form specific and highly complex glycans. Although mucin-type *O*-glycosylation is widely present in human proteins, plants lack the enzymatic machinery to perform this specific glycosylation [11, 15, 16].

Recently, a few, but significant, advances in engineering *O*-glycosylation in plants to produce mucin-type glycans were reported. Among the recent achievements, Daskalova et al. [6] successfully expressed an *O*-glycosylated human Mucin 1 peptide derivative, which was detected exclusively as a glycoform in *Nicotiana benthamiana*. This was achieved by co-expressing the *Yersinia enterocolitica* UDP-GlcNAc 4-epimerase, in charge of converting UDP-GlcNAc to UDP-GalNAc in the cytoplasm, and a *Caenorhabditis elegans* UDP-GlcNAc/UDP-GalNAc transporter, responsible for the transport of the sugar donor to the Golgi lumen, together with the human GalNAc transferase 2.

In 2012, Yang et al. [10] were also able to generate mucin-type *O*-glycosylation in *N. benthamiana*, using a similar approach. They transiently expressed a *Pseudomonas aeruginosa* UDP-GlcNAc 4-epimerase, together with the human GalNAc-T 2 and 4, and a human 3.5 tandem repeat of Mucin1. The derivative mucin peptide was GalNAc-*O*-glycosylated with up to three and five GalNAc residues, when expressed with GalNAc-T2 and GalNAc-T 2 together with GalNAc-T 4, respectively. The mucin-type *O*-glycosylation was also demonstrated by mass spectrometry on a tandem repeat of MUC16 and on interferon α2b. These authors [17] were also able to establish mucin-type *O*-glycosylation in *Arabidopsis thaliana* and *Nicotiana tabacum* BY-2 cells, and encountered a high degree of proline hydroxylation and Hydroxyproline linked arabinosides, plant-specific *O*-glycosylation, on the substrate model protein.

The human Granulocyte-Colony Stimulating Factor (G-CSF) is a cytokine that stimulates the production, proliferation, differentiation and activation of neutrophil stem cells, raising the levels of neutrophils in the blood stream, and thus protecting the organism against bacterial, fungal and viral infections [18–22]. This cytokine is an important glycoprotein that is used as treatment in patients with neutropenia to reduce opportunistic infections, especially in cancer patients undergoing chemotherapy or radiotherapy [20, 23, 24]. Due to this clinical relevance, this protein is one of the most widely sold pharmaceuticals, thus making its production for clinical use very important [24–27]. G-CSF is mucin-type *O*-glycosylated at a single Thr residue. It has been successfully expressed in CHO cells, *Escherichia coli*, yeast and plants [28–32]. Nevertheless, only the mammalian platform variant is reported to be mucin-type *O*-glycosylated, while the yeast-derived variant is decorated with mannose residues. Several studies have shown the impact of glycosylation on protein stability [33–37], pointing out the relevance of production and correct glycosylation of this pharmaceutical.

In this project, our objective is to transiently express the human Granulocyte-Colony Stimulating Factor as a model pharmaceutical protein in *N. benthamiana* plants, via *Agrobacterium* infiltration, together with the genes required for the synthesis of *N*-acetylgalactosamine (GalNAc)-*O*-glycosylation, allowing the production of a plant-derived human-specific *O*-glycosylated G-CSF.

Experimental procedures

Construction of binary vectors

The coding sequence corresponding to the human G-CSF variant 2 (NCBI: NM_172219) was chemically synthesized (GenScript, Piscataway, NJ, USA) in three different fragments to build the variants used in this study. The three fragments consisted of the nucleotide sequences corresponding to the secretory signal peptide (pUC57 Nat-SP), the mature G-CSF fragment with a mutated glycosylation site that encodes an Ala, instead of the native Thr, and an extra 5′-end Met (pUC57 G-CSF A), and a 5′-end enterokinase coding sequence fused to the mature G-CSF fragment with the native glycosylation site (pUC57 EkG-CSF). The Sec-G-CSF coding sequence was built by digesting the pUC57 Nat-SP with NcoI and Bsu361 and using the digested fragment to replace the 5′-end EkG-CSF fragment of the pUC57 EkG-CSF, which was released with the same restriction enzymes, producing the pUC57 Sec-G-CSF. The Cyt-G-CSF coding sequence was built by releasing the 3′-end of the pUC57 G-CSF A with SapI cutting upstream

the non-glycosylation site and downstream the stop codon. The same restriction enzyme was used to digest the pUC57 EkG-CSF, releasing the same fragment but containing the native glycosylation site (Thr), and the released fragment was cloned into the SapI opened pUC57 G-CSF A, generating the pUC57 Cyt-G-CSF with the glycosylation site. The Zera-G-CSF coding sequence was built by digesting the pUC57 EkG-CSF with NcoI and SacI, releasing the whole fragment, and ligating it into a pUC18 ZeraF1V vector available in the lab [38], by releasing the F1V fragment with the same enzymes and thus generating the pUC18 Zera-G-CSF. The different G-CSF coding sequences generated were then PCR amplified to introduce the flanking BbsI A (5′-end Sec-G-CSF: Mt114F—GGGGGGGAAGACATCATGGCT GGACCTGCCACTC), (5′-end Cyt-G-CSF: Mt23F—GGGGGGGAAGACATCATGACACCCTTAGGACC), (5′-end Zera-G-CSF: Mt20F—GGGGGGGAAGACGTC ATGAGGGTGTTGCTCGTTGC) and B (3′-end of all G-CSF variants: M113R—CCCCCCGAAGACAGAGCT CGGGTTGAGCAAGGTGAC) sites, and removal of the stop codon to allow in frame fusion with the appropriate tag. The BbsI A site contains the recognition sequence for the restriction enzyme, followed by the specific overhang CATG, and the BbsI B site contains the specific overhang AGCT, allowing future GoldenGate cloning and disruption of the stop codon for further in-frame cloning with the respective tag. The resulting amplified product was cloned in pUC57 vector via EcoRV, and the mutated sites were verified by DNA sequencing and restriction digestions. The resulting pUC57 G-CSF variant without Stop codon (pUC57 Sec-G-CSF BbsI AB w/o Stop, pUC57 Cyt-G-CSF BbsI AB w/o Stop, pUC57 Zera-G-CSF BbsI AB w/o Stop) was then used to clone the coding sequence into the pENTR4 BbsI AB vector (a GoldenGate compatible pENTR4 vector generated in the lab), via GoldenGate, using the restriction enzyme BbsI and DNA ligase in a single cutting-ligation reaction. The resulting pENTR4 G-CSF variant without Stop codon (pENTR4 Sec-G-CSF w/o Stop, pENTR4 Cyt-G-CSF w/o Stop, pENTR4 Zera-G-CSF w/o Stop) was then used to clone the corresponding coding sequence into the binary vector via Gateway LR reaction, allowing the in-frame cloning of the G-CSF variant with the desired C-terminal tag, eYFP (pGWB 641) or c-Myc (pGWB 617) [39]. The binary vectors were used to transform *Agrobacterium tumefaciens* strain AGL-1.

Other constructs used in this work were the suppressor of post-transcriptional gene silencing p19 from *Cymbidium ringspot tombusvirus* [40] to increase accumulation levels; SecGFP (as protein secretion control), SecGFP:KDEL (for ER localization signal) previously published [41], and STtmd:mRFP (as Golgi marker for co-localization) [42]. *Agrobacterium* strain AGL-1 was used for pGWB 642, and *Agrobacterium* strain EHA105 was used for p19, GFP, and STtmd:mRFP constructs. The binary vectors used to generate *O*-glycosylation were as follows: pH7WG2 GNE, containing the *Y. enterocolitica gne* gene coding for UDPGlcNAc-4 epimerase; pH7WG2 GT, containing the *C. elegans* UDPGlcNAc/UDPGalNAc Transporter gene; and pH7WG2 GNT2, containing the human GalNAc Transferase 2 coding sequence, previously published [6]. All were electroporated into the *Agrobacterium* strain AGL-1.

Agrobacterium infiltration

Agrobacterium tumefaciens strains were cultured to an optical density at 600 nm (OD_{600}) of 0.5–0.8. The cells were then collected by centrifugation at 1000 g for 30 min. The pellets were resuspended in Agro-infiltration solution (3.2 g/L Gamborg's B5 medium and vitamins, 20 g/L sucrose, 10 mM MES pH 5.6, 200 µM 4′-Hydroxy-3′-5′-dimethoxyacetophenone) to a final OD_{600} of 1.0, followed by incubation at room temperature for 1 h, with gentle agitation. The suspension was then used for needle-less infiltration of the abaxial leaf epidermis through the stomata of *N. benthamiana* plants [43].

Confocal analysis

Protein subcellular visualization was determined by imaging the abaxial epidermal cells of leaf samples, with an Olympus LSM FV1200. Different lasers allowed for imaging of the different fluorescent tag fusion proteins. For GFP imaging, the tag was excited with a 488 Argon laser and detected at 500–545 nm. For eYFP imaging, the tag was excited at 515 and detected at 530–545 nm. For mRFP imaging, the tag was excited at 559 nm with a He/Ne laser and detected at 570–545 nm. The Imaris software (version 7.6.1, Bitplane Scientific Software, Bitplane, Zurich, Switzerland) was used to generate 3D images from z-stack confocal images. Line-sequential scanning mode was used for co-localization imaging.

Preparation of total protein extracts

Four leaf discs (approximate fresh weight of 30 mg) from at least three biological replicates per sample were collected with a 7 mm diameter cork borer, put into a 2 ml tube with three 2.3 mm zirconia/silica beads (Bio Spec Products Inc, Cat. No. 11079125z) and frozen in liquid N_2. Collected leaf discs were pulverized in a Mixer Mill (Retsch, Haan, Germany) for 1 min at 30 Hz, in previously frozen homogenizer blocks. Pulverized tissue was spun down for 1 min at 1000 g and 300 µL Protein Extraction Buffer (PBST0.1%, 2% PVPP, 1 mM EDTA pH 8.0, 1 mM PMSF, 1 µg/mL Leupeptin, 100 mM Sodium L-ascorbate) or Reducing Extraction Buffer (50 mM Tris,

pH 8.0, 1% SDS, 20 mM DTT) were added to the sample. Samples were vigorously vortexed 3 times for 5 s and centrifuged at 20,000*g* for 15 min at 4 °C. Cleared supernatant was transferred to a new tube and TSP (Total Soluble Protein) was quantified using the Bio-Rad Bradford [44] protein assay reagent (Bio-Rad, Cat. No. 5000006).

SDS-PAGE western blotting and lectin blot

Serological assays were performed by resolving the protein sample in a sodium dodecylsulphate-polyacrylamide gel electrophoresis (SDS-PAGE), and transferred to a PVDF or Nitrocellulose membrane. Recombinant proteins were detected with a 1:5000 dilution of the primary mouse anti-c-Myc monoclonal antibody (GenScript, Cat. No. A00864) or 1:5000 dilution of the primary rabbit anti-GCSF polyclonal antibody (GeneTex, Cat. No. GTX31157), and 1:3000 dilution of the goat anti-mouse IgG HRP-conjugated secondary antibody (Bio-Rad, Cat. No. 170-6516) or 1:3000 dilution of the goat anti-rabbit IgG-conjugated secondary antibody (Bio-Rad, Cat. No. 1706515). In the case of lectin blots, a 1:1000 VVA-HRP conjugate (EY Laboratories, Cat. No. H-4601-1) was used. Blotted membranes were visualized with the enhanced chemiluminescence (ECL) detection kit (GE Healthcare, Cat. No. RPN2232), following the manufacturer recommendations, and imaged with the DNR Bio-Imaging System MicroChemi (RANCOM A/S, Birkerød, Denmark). Band analysis was performed using the TotalLab TL 100 software (Nonlinear Dynamics, Durham, NC).

Protein purification

Protein extract was transferred to a spin column, and anti-c-Myc beads were added to the column, following the manufacturer instructions (MBL International, Cat. No. 3305). The sample was incubated at 4 °C for 1 h, with gentle agitation end-over-end. The column was briefly centrifuged for 10 s, and the flow through was recovered. Three washes were performed by adding 200 µL of the washing buffer provided with the kit, and briefly centrifuging. Finally, 20 µL of elution peptide was added to the column, the sample was incubated for 5 min at 4 °C, and the purified protein was recovered by centrifuging the column for 10 s. Elution was repeated twice.

Analysis of *O*-glycans by mass spectometry

SDS-PAGE excised bands were Trypsin/Chymotrypsin (6.25 mg/L each) digested [45] at 30 °C overnight. The peptide digests were analyzed using an Easy-nLC 1000 nano-flow system with a 100 µm × 2 cm Acclaim C18 PepMap™ trap column and a 75 µm × 15 cm Acclaim C18 PepMap™ analytical (Thermo Scientific, MA, USA) coupled to a Q-Exactive™ Quadrupole Orbitrap mass spectrometer (Thermo Scientific, MA, USA). The flow rate was 300 nL min^{-1} and 10 µL of the protein digest was injected. 97% mobile phase A (LC/MS Optima water, 0.1% formic acid) was decreased to 90% over 3 min. Peptides were eluted with a linear gradient from 10 to 35% mobile phase B (LC/MS Optima acetonitrile 0.1% formic acid) over 21 min followed by 35–90% over 3 min and maintained for 8 min. The nanospray voltage was set at 2.1 kV, capillary temperature 275 °C, and S-lens RF level 55. The Q-Exactive was operated in top 10 data-dependent acquisition mode with a full scan mass range of 400–2000 *m/z* at 70,000 resolution, automatic gain control (AGC) of 1e6 and maximum injection time (IT) of 250 ms. The MS/MS scans were acquired at 17,500 resolution, AGC of 2e5, maximum IT of 50 ms, intensity threshold of 8e4, normalized collision energy of 27 and isolation window of 1.2 *m/z*. Unassigned, singly and >4 charged peptides were not selected for MS/MS and a 20 s dynamic exclusion was used. The Thermo .raw files were converted to mascot generic format using Proteowizard v2 [46] and the MS/MS scans were searched against the target/reverse human G-CSF amino acid sequence and the *N. benthamiana* proteome (Sol Genomics Network, accessed Jan 10th, 2015) using X! tandem [47] search algorithm operated from the SearchGUI v.2.35 [48] interface and processed in PeptideShaker v1.3.6 [49]. A 3 ppm precursor ion mass error and a 0.02 Da product ion error were used along with carbamidomethylation as a constant modification and oxidation of methionine, Hex(1) NAc(1) of Thr, HexNAc of Thr as variable modifications. A 1% FDR rate was used at the protein, peptide and peptide spectrum match level.

Statistical analysis

The statistical analyses were performed with the Minitab 18 software (Minitab Inc., PA, USA). A one-way analysis of variance (one-way ANOVA) was performed followed by Tukey test to find significance differences between means (statistical difference was defined as $p \leq 0.05$).

Results

1. Transient expression of G-CSF in *N. benthamiana* leaves.

To explore the feasibility of using *N. benthamiana* transient expression for the production of a mammalian glycoprotein, like G-CSF, several factors need to be explored, such as the accumulation levels of the recombinant protein, proper protein folding and native post-translational modifications. To address these issues, the native secretory human G-CSF (Sec-G-CSF) coding sequence, including its native secretory signal sequence, was used to build the corresponding expression cassettes,

in order to target the model protein to the secretory pathway (Fig. 1). Two other G-CSF variants were also built, which would target the recombinant protein to the cytosol and as ER-derived protein bodies. These variants were used as controls to determine protein subcellular localization, as well as to evaluate the impact of protein accumulation. The cytoplasmic variant (Cyt-G-CSF) corresponds to the mature amino acid sequence with an N-terminal methionine, allowing the accumulation of the protein in the cytosol. The protein body variant (Zera-G-CSF) was achieved by fusing the N-terminus of the mature G-CSF coding sequence (no signal peptide) to the γ-Zein ER-accumulating domain: Zera® peptide [38]. The different constructs were fused to an eYFP tag in order to track expression and proper subcellular localization of the recombinant protein. All of the G-CSF variant constructs-containing *Agrobacterium* strains were co-infiltrated with an *Agrobacterium* strain containing the p19 construct, a post-transcriptional gene silencing suppressor from *Cymbidium ringspot tombusvirus* (CymRSV) [40].

Expression of G-CSF:eYFP was first detected by confocal microscopy (Fig. 2a–l). Infiltrated *N. benthamiana* leaves were monitored from 2 to 8 dpi (days post-infiltration). Sec-G-CSF:eYFP was detected in the ER (endoplasmic reticulum), as a multiple dot-like pattern indicating a potential localization in the Golgi, as well as in the apoplast (Fig. 2a–c), suggesting localization in the secretory pathway. The cytoplasmic variant, Cyt-G-CSF:eYFP, was detected in the cytosol, and no signal was detected in the apoplast (Fig. 2d–f). Cytoplasmic eYFP (CyteYFP) was used as control for cytoplasmic accumulation pattern, also allowing the visualization of an apoplast with no presence of recombinant protein (Fig. 2g–i). Finally, the Zera-G-CSF:eYFP was found forming protein bodies due to the N-terminal fusion of the Zera® peptide (Fig. 2j–l).

Identity of the transiently expressed G-CSF:eYFP recombinant proteins was further corroborated by western blot (Fig. 2m, n). Sec-G-CSF:eYFP was detected as a monomeric form with an apparent molecular weight of 49 kDa, which corresponds to the mature G-CSF:eYFP protein (after signal peptide cleavage). The same molecular weight was observed with Cyt-G-CSF:eYFP, confirming signal peptide processing of the Sec-G-CSF:eYFP. Furthermore, multimers were observed with Cyt-G-CSF:eYFP. In the case of Zera-G-CSF:eYFP a weak band was detected when protein extraction was performed with PBST (PBS + Tween), but the Zera® fusion variant was properly extracted when the plant tissue sample was treated in denaturing and reducing conditions, using an SDS-DTT containing extraction buffer. Multimers were also observed with Zera-G-CSF:eYFP. eYFP cleavage was observed with all the different variants, and the cleaved eYFP band was corroborated using the CyteYFP expressing samples. It has to be noted that both Sec-G-CSF and Cyt-G-CSF were successfully extracted with PBST extraction buffer (Additional file 1: Fig. S3), but Zera-G-CSF could only be properly extracted with a reducing extraction buffer (Fig. 2; Additional file 2: Fig. S2).

G-CSF:eYFP accumulation was then assessed by densitometry analysis (Figs. 2n, 3a–d), with samples collected at 4, 6 and 8 dpi for each variant. We found that Sec-G-CSF reached an accumulation level of 17 mg/Kg F.W. (Fresh Weight), followed by Cyt-G-CSF:eYFP

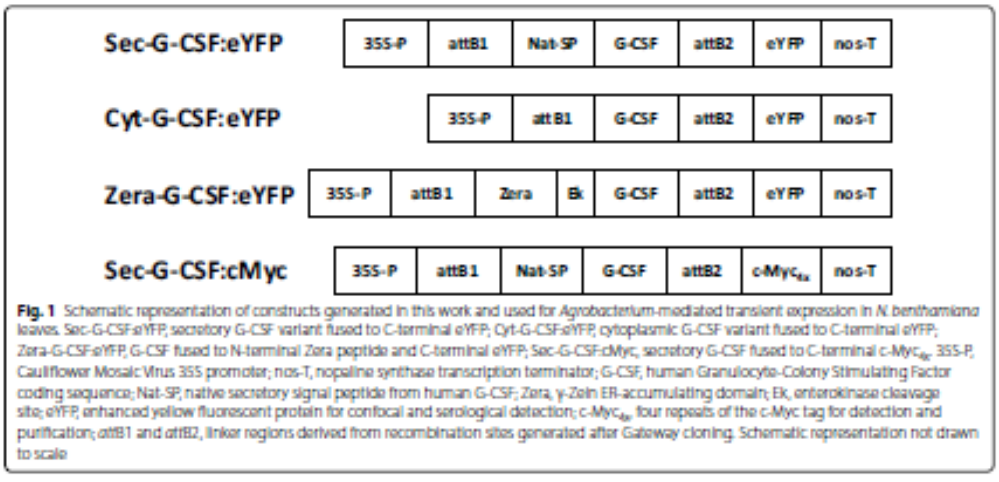


Fig. 1 Schematic representation of constructs generated in this work and used for *Agrobacterium*-mediated transient expression in *N. benthamiana* leaves. Sec-G-CSF:eYFP, secretory G-CSF variant fused to C-terminal eYFP; Cyt-G-CSF:eYFP, cytoplasmic G-CSF variant fused to C-terminal eYFP; Zera-G-CSF:eYFP, G-CSF fused to N-terminal Zera peptide and C-terminal eYFP; Sec-G-CSF:cMyc, secretory G-CSF fused to C-terminal c-Myc$_{4x}$. 35S-P, Cauliflower Mosaic Virus 35S promoter; nos-T, nopaline synthase transcription terminator; G-CSF, human Granulocyte-Colony Stimulating Factor coding sequence; Nat-SP, native secretory signal peptide from human G-CSF; Zera, γ-Zein ER-accumulating domain; Ek, enterokinase cleavage site; eYFP, enhanced yellow fluorescent protein for confocal and serological detection; c-Myc$_{4x}$, four repeats of the c-Myc tag for detection and purification; *attB1* and *attB2*, linker regions derived from recombination sites generated after Gateway cloning. Schematic representation not drawn to scale

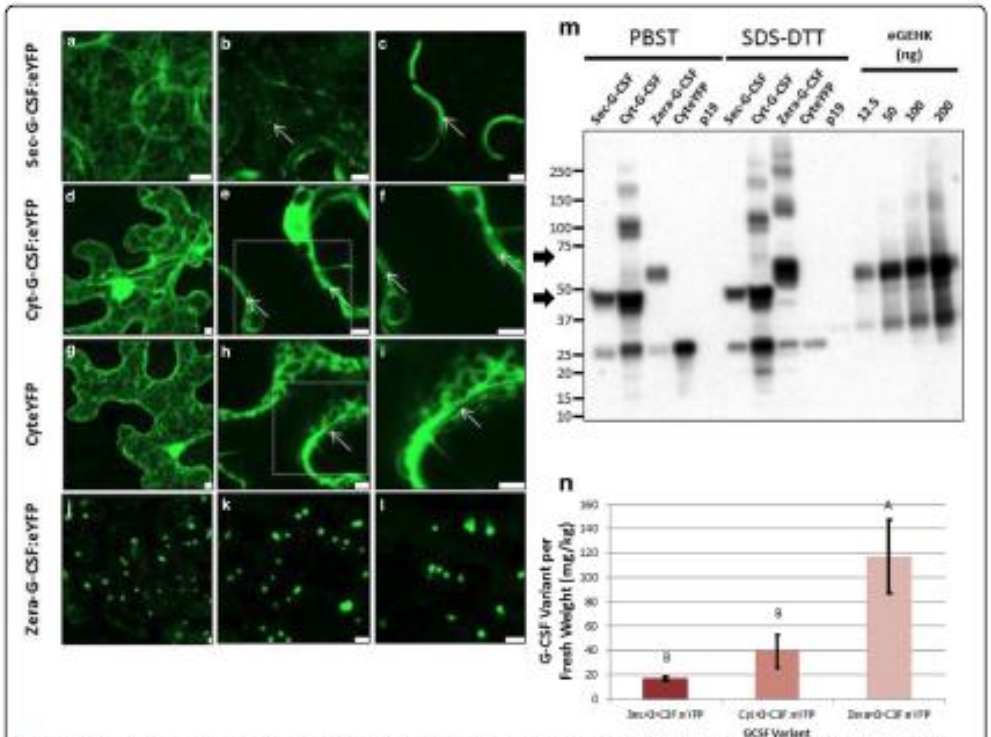


Fig. 2 Transient expression of G-CSF:øYFP variants in *N. benthamiana* plants. Expression of Sec-G-CSF:øYFP (**a–c**), Cyt-G-CSF:øYFP (**d–f**), CyteYFP (**g–i**), Zera-G-CSF:øYFP (**j–l**). White arrows indicate apoplast localization (**c**) or absence of signal in the apoplast (**e**, **f**, **h**, **i**). White arrow indicating Golgi localization (**b**). Inset white boxes (**e**, **f**) depict microscopic zoom in shown in **f** and **i**, respectively. Samples collected at 4 dpi. Size bar: 5 µm. **m** Western Blot detection of Sec-G-CSF:øYFP (Secretory variant, 49 kDa), Cyt-G-CSF:øYFP (Cytoplasmic and mature protein variant, 49 kDa), Zera-G-CSF:øYFP (Zera fused variant, 61 kDa), øYFP (Cytoplasmic øYFP, 29 kDa). All samples were treated with PBS + Tween (PBST) extraction buffer and reducing extraction buffer (SDS-DTT). **n** Band quantification of SDS-DTT treated samples. 20 µg TSP of PBS samples were loaded on the gel. 9 µL (equalized volume to the smallest PBS sample amount) of SDS-DTT samples was loaded on the gel. Black arrows denote monomeric G-CSF:øYFP variants. øGEHK: protein standard. Proteins were detected with GFP antibody. Samples collected at 6 dpi. Columns denoted with a different letter are significantly different ($p \leq 0.05$) using one-way ANOVA and followed by Tukey test. Error bars are standard deviation of the means

with 40 mg/Kg F.W., while Zera-G-CSF:eYFP reached 117 mg/Kg F.W., at 6 dpi (Figs. 2n, 3; Additional file 1: Fig. S3). We also observed that accumulation was stable at 4 and 6 dpi for Sec-G-CSF and Cyt-G-CSF, beginning to drop at 8 dpi, but not significantly (Fig. 3). There is no significant impact concerning efficiency of the extraction buffer of Sec-G-CSF (PBST or SDS-DTT) (Additional file 1: Fig. S3), but higher levels of accumulation were observed with Cyt-G-CSF when samples were treated with PBST, in comparison to SDS-DTT (Additional file 1: Fig. S3). Zera-G-CSF was the only variant that showed a significant increase in accumulation levels on the time-course, in comparison to the other two variants, reaching up to 135 mg/Kg F.W. at 8 dpi (Fig. 3), but as previously mentioned Zera-G-CSF could only be properly extracted under denaturing and reducing conditions using an extraction buffer containing SDS-DTT (Figs. 2, 3; Additional file 2: Fig. S2). This result was expected since previous studies indicate the usage of denaturing and reducing conditions for extraction of recombinant proteins fused to Zera [38, 50, 51]. The previous observations highlight the vulnerability of G-CSF to proteolytic degradation, and especially in the case of the secretory G-CSF.

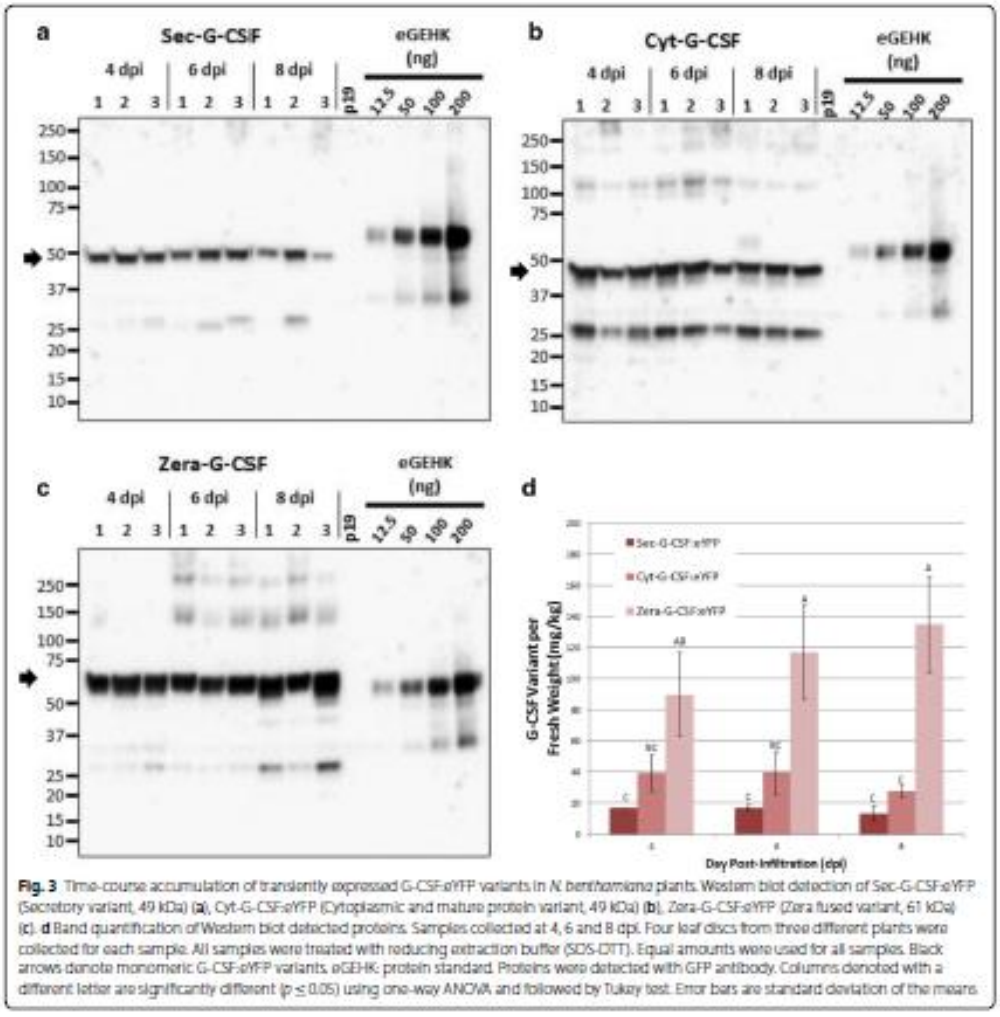


Fig. 3 Time-course accumulation of transiently expressed G-CSF:eYFP variants in *N. benthamiana* plants. Western blot detection of Sec-G-CSF:eYFP (Secretory variant, 49 kDa) (**a**), Cyt-G-CSF:eYFP (Cytoplasmic and mature protein variant, 49 kDa) (**b**), Zera-G-CSF:eYFP (Zera fused variant, 61 kDa) (**c**). **d** Band quantification of Western blot detected proteins. Samples collected at 4, 6 and 8 dpi. Four leaf discs from three different plants were collected for each sample. All samples were treated with reducing extraction buffer (SDS-DTT). Equal amounts were used for all samples. Black arrows denote monomeric G-CSF:eYFP variants. eGEHK: protein standard. Proteins were detected with GFP antibody. Columns denoted with a different letter are significantly different ($p \leq 0.05$) using one-way ANOVA and followed by Tukey test. Error bars are standard deviation of the means

2. Sec-G-CSF localization in the secretory pathway.

To determine the feasibility of producing a mucin-type *O*-glycosylated G-CSF recombinant protein, we then corroborated that the native secretory signal peptide used in this work was targeting G-CSF to the secretory pathway. For this purpose, we determined the co-localization of Sec-G-CSF with Rat sialyltransferase, a Golgi resident enzyme, fused with monomeric red fluorescent protein (STtmd:mRFP) as a Golgi marker, which would suggest trafficking of secretory G-CSF along the secretory network (Fig. 4). Sec-G-CSF:eYFP was observed to localize in the ER as previously observed (Fig. 2) but also localized in the Golgi together with STtmd:mRFP (Fig. 4a–c). Although only a few Golgi bodies were captured where Sec-G-CSF would co-localize with the Golgi marker, a previously tested secretory GFP (SecGFP) construct from the lab [52] was used as a positive control for co-localization in the Golgi, and a similar pattern was observed, being SecGFP also localized in the ER (Fig. 4e–g).

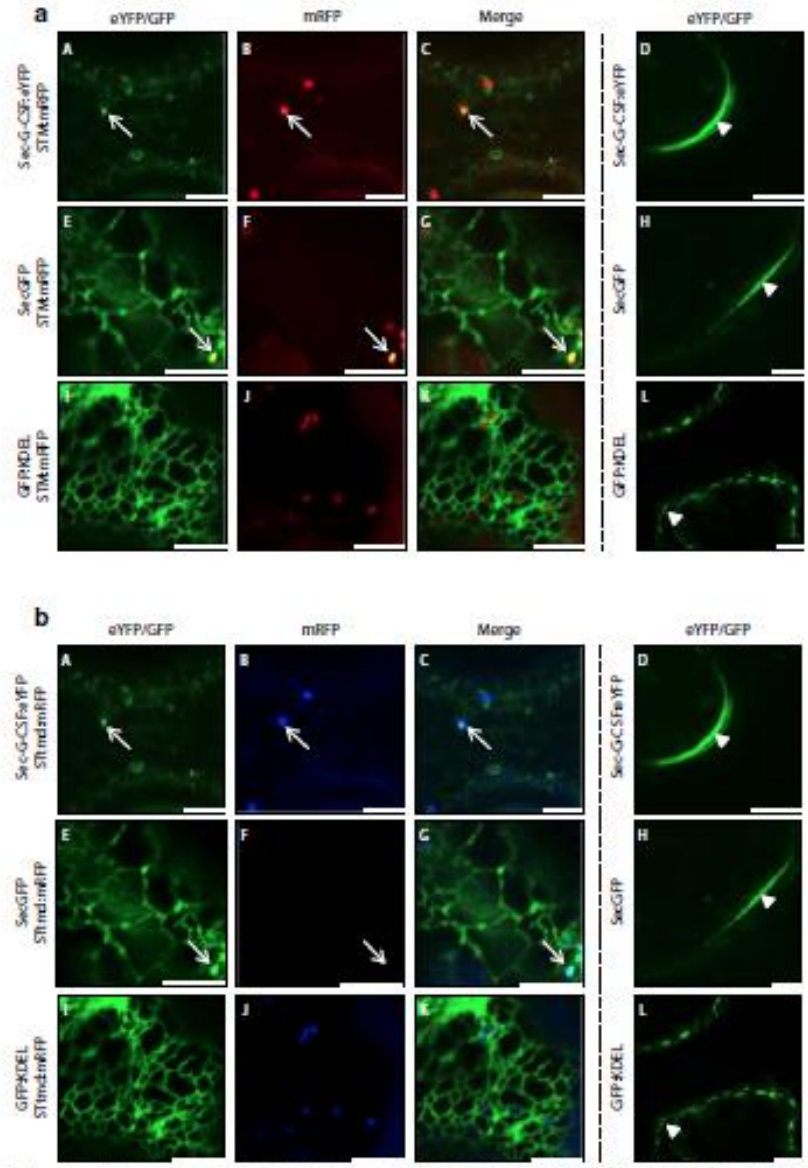


Fig. 4 **a**, **b** Golgi and apoplast localization of Sec-G-CSF in *N. benthamiana* plants. Sec-G-CSF:eYFP (A, C, D), SecGFP (E, G, H), GFP:KDEL (I, K, L). Sec-G-CSF, SecGFP and GFP:KDEL were co-infiltrated with STM:mRFP (B, C, F, G, J, K). White arrows depict Golgi localization (A, B, E, F) and colocalization (C, G) with STM:mRFP. White arrowheads (images on the right) depict localization in apoplast (D, H), or apoplast without presence of recombinant protein (J). *Agrobacterium* AGL-1 was carrying Sec-G-CSF, SecGFP, GFP:KDEL, STM:mRFP and p19 were delivered by EHA105. 4 dpi. Size bar: 5 µm. **a** shows the colors as originally displayed, while **b** shows a color palette adjusted for those who have difficulty viewing red and green colors

Thus, both Sec-G-CSF:eYFP and SecGFP were detected throughout the secretory pathway, in the ER, Golgi and in the apoplast (Fig. 4a–h). As a negative control for secretion, an ER-retained GFP (GFP:KDEL) construct was used, which would be directed to the secretory pathway but retrieved to the ER. The GFP:KDEL negative control was not found to co-localize with STtmd:mRFP, and it was not observed in the apoplast (Fig. 4i–k). The previous observations indicate that the native G-CSF secretory signal peptide indeed targets G-CSF to the secretory pathway. The same constructs were used to monitor apoplast signal, and only Sec-G-CSF:eYFP and SecGFP were observed in the apoplast, while GFP:KDEL was not, corroborating the previous observations (Fig. 4d, h, l). Although Sec-G-CSF was not exclusively observed in the Golgi, where *O*-glycosylation is traditionally referred to begin (Additional file 3: Fig. S1), the fact that the recombinant protein was observed in the apoplast suggested secretion of the protein to certain extent, thus trafficking from the ER to the apoplast. Moreover, some studies [53–56] have also suggested that the attachment of GalNAc to substrate proteins might also occur in regions of the ER (Additional file 3: Fig. S1), therefore we could consider to subject Sec-G-CSF to the *O*-glycosylation machinery.

3. *O*-Glycosylation of Sec-G-CSF.

Once we confirmed that Sec-G-CSF was targeted for secretion, the Sec-G-CSF coding sequence was co-expressed with the *O*-glycosylation machinery (glyco-machinery) comprised by the *Y. enterocolitica* UDP-GlcNAc/GalNAc 4-epimerase, the *C. elegans* UDP-GlcNAc/GalNAc transporter, and the human GalNAc-Transferase 2, to allow the mammalian-specific GalNAc-*O*-glycosylation of G-CSF. In this case, the eYFP tag was replaced by a C-terminal c-Myc$_{4x}$ tag (Sec-G-CSF:cMyc) for detection and purification purposes. Expression of Sec-G-CSF:cMyc was determined by western blot with a molecular weight of 28 kDa (Fig. 5a). Sec-G-CSF:cMyc reached accumulation levels of 49 mg/Kg F.W. and 31 mg/Kg F.W., at 4 and 6 dpi, respectively (Fig. 5b).

Once expression of Sec-G-CSF:cMyc was assessed, the secretory variant was then co-expressed with the *O*-glycosylation machinery. Sec-G-CSF:cMyc co-expressed with and without the *O*-glycosylation machinery was purified by c-Myc monoclonal antibody affinity chromatography, and analyzed via SDS-PAGE, c-Myc detection by western blot and GalNAc-*O*-glycosylation detection by lectin blotting (Fig. 5c). Sec-G-CSF:cMyc was detected both when co-expressed with or without the *O*-glycosylation machinery by SDS-PAGE and western blot using a c-Myc specific antibody (Fig. 5c). For detection of the *O*-glycosylated residue, the lectin *Vicia villosa* agglutinin conjugated with HRP (VVA-HRP), which is specific for GalNAc-*O*-T/S moieties, was used as probe (Fig. 5c). VVA-HRP reacted only with the Sec-G-CSF:cMyc co-expressed with the glycosylation machinery, but not with the Sec-G-CSF:cMyc expressed alone, indicating proper GalNAc-*O*-glycosylation of Sec-G-CSF:cMyc. To further corroborate the specific attachment of the GalNAc to the Thr-163 residue (Additional file 4: Fig. S4), Sec-G-CSF:cMyc co-expressed with and without the *O*-glycosylation machinery was purified and in-gel Trypsin/Chymotrypsin digested. A Trypsin/Chymotrypsin digestion system was used to generate a smaller peptide around the predicted *O*-glycosylation site that would be more amenable to MS/MS analysis. The resulting peptides were identified by nano LC–MS/MS and GalNAc attachment to the specific Thr-163 site was confirmed by the identification of an *N*-acetylhexosamine (HexNAc) derived from the peptide where the Thr-163 is localized, as indicated by the characteristic HexNAc product ion at *m/z* 204 (Additional file 4: Fig. S4; Additional file 5: Table S1). Nevertheless, Sec-G-CSF expressed alone could not be detected during the MS/MS analysis; moreover, analyses targeted to detect plant-specific post-translational modifications, such as proline hydroxylation and further plant-specific *O*-glycosylation, which are likely to occur during secretion of recombinant proteins, were not carried out in the present study. Thus, future studies will have to be carried out in order to fully characterize the recombinant protein.

Discussion

Accumulation has been a major focal point for the production of recombinant proteins in plants [4, 38, 57, 58]. Although it may be assumed that post-translational modifications, such as glycosylation which might have an impact on protein stability or activity [6, 11], would be produced by plants and other eukaryotic expression platforms, most studies concerning expression of glycoproteins do not address this issue [9, 11, 16, 32, 59–63]. In this study, we address the feasibility of introducing the mammalian-specific mucin-type *O*-glycosylation to the pharmaceutical glycoprotein human G-CSF, in *N. benthamiana* plants via transient expression.

G-CSF is commercially produced in *E. coli* and CHO cells, well-established expression platforms. The bacterial system is capable of producing the recombinant protein in its mature form, with an extra N-terminal Met, and the protein is not glycosylated [35, 64]. The mammalian system is capable of decorating the recombinant protein with Gal-GalNAc-*O*-glycosylation with up to two sialic acids, at the single Thr residue reported to be mucin-type *O*-glycosylated in humans [35, 65].

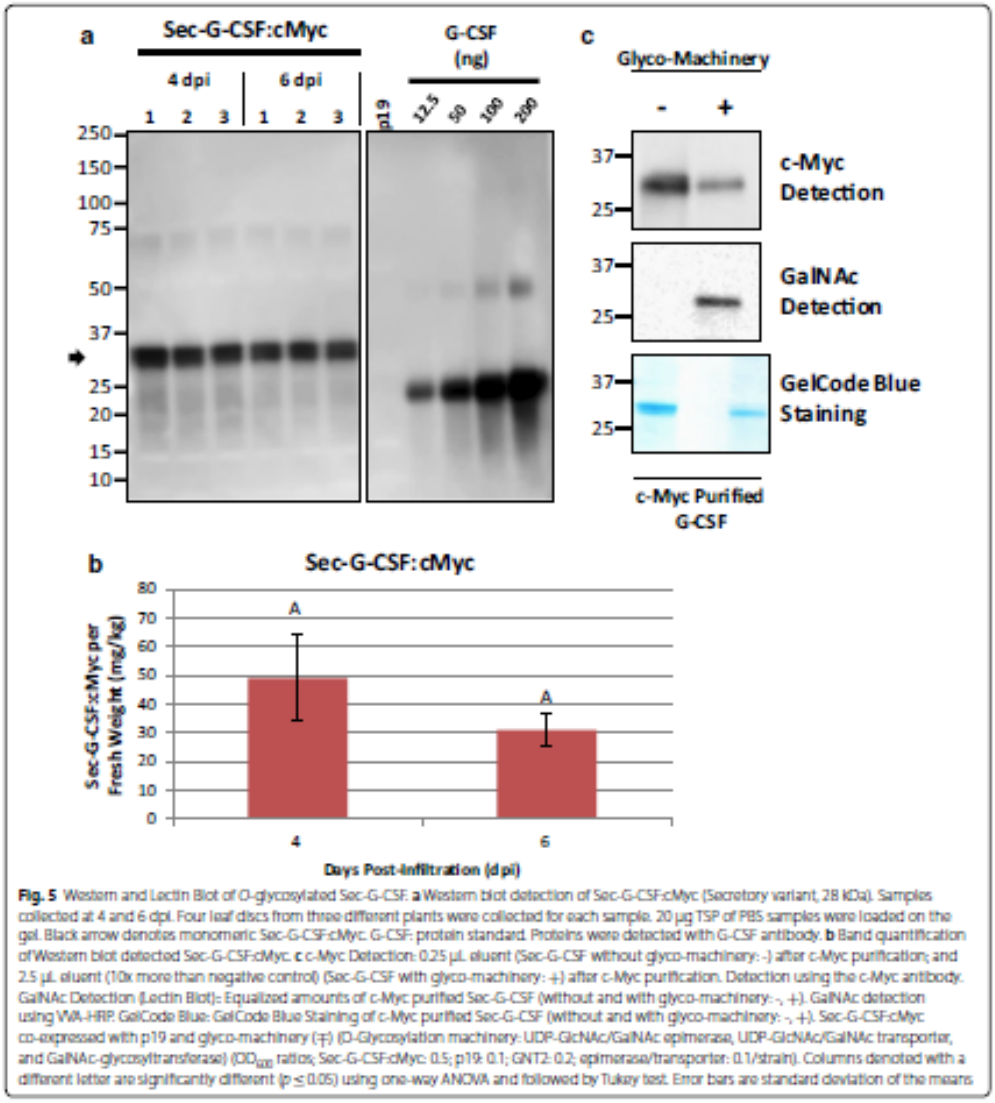


Fig. 5 Western and Lectin Blot of O-glycosylated Sec-G-CSF. **a** Western blot detection of Sec-G-CSF:cMyc (Secretory variant, 28 kDa). Samples collected at 4 and 6 dpi. Four leaf discs from three different plants were collected for each sample. 20 µg TSP of PBS samples were loaded on the gel. Black arrow denotes monomeric Sec-G-CSF:cMyc. G-CSF: protein standard. Proteins were detected with G-CSF antibody. **b** Band quantification of Western blot detected Sec-G-CSF:cMyc. **c** c-Myc Detection: 0.25 µL eluent (Sec-G-CSF without glyco-machinery: -) after c-Myc purification; and 2.5 µL eluent (10x more than negative control) (Sec-G-CSF with glyco-machinery: +) after c-Myc purification. Detection using the c-Myc antibody. GalNAc Detection (Lectin Blot): Equalized amounts of c-Myc purified Sec-G-CSF (without and with glyco-machinery: -, +). GalNAc detection using WA-HRP. GelCode Blue: GelCode Blue Staining of c-Myc purified Sec-G-CSF (without and with glyco-machinery: -, +). Sec-G-CSF:cMyc co-expressed with p19 and glyco-machinery (∓) (O-Glycosylation machinery: UDP-GlcNAc/GalNAc epimerase, UDP-GlcNAc/GalNAc transporter, and GalNAc-glycosyltransferase) (OD_{600} ratios; Sec-G-CSF:cMyc: 0.5; p19: 0.1; GNT2: 0.2; epimerase/transporter: 0.1/strain). Columns denoted with a different letter are significantly different ($p \leq 0.05$) using one-way ANOVA and followed by Tukey test. Error bars are standard deviation of the means

Studies addressing the glycosylation differences between these two commercially available G-CSF's have determined that lack of glycosylation has an impact on the protein's susceptibility to degradation, decreasing the biological activity of the cytokine, and thus an increased amount of the non-glycosylated form is required to obtain a similar biological effect as that obtained with the glycosylated form [33, 34, 37]. Other studies have also suggested that the lack of glycosylation might promote protein aggregation, which would hamper the biological activity of the protein [30, 66–68].

Due to the pharmaceutical relevance of this glycoprotein, several alternative production platforms have been considered. Among these platforms, yeasts have been explored due to the secretory system they possess, which is similar to that of higher order eukaryotes [29, 69]. Recombinant hG-CSF has been successfully expressed in *Saccharomyces cerevisiae* and *Pichia pastoris* [29, 31, 66], but recent analysis reported that the recombinant *P. pastoris*-produced protein was not mucin-type *O*-glycosylated. Instead, the *O*-glycosylation Thr site used in mammalian cells was mannose *O*-glycosylated [70]. In the case of the *S. cerevisiae*-derived recombinant protein, glycosylation was apparently absent [66]. Furthermore, the yeast-derived recombinant protein was found mainly as multimers, which were not biologically active and had to be denatured and re-natured to recover proper protein folding and regain biological activity. The issue with multimerization is believed to result from lack of proper protein folding, as it has also been observed when G-CSF was expressed using the filamentous fungus *Aspergillus niger*, where glycosylation could not be determined [67].

Recently, G-CSF production has been achieved in tobacco and rice suspension cells [59, 62] and tobacco plants [60], targeting the protein to the secretory pathway; ER-retained in BY-2 cells [61]; in lettuce chloroplasts [32]; and transiently in *N. benthamiana* [63]. These studies do not report multimerization as it was observed in the case of the yeast and *A. niger* reports, and the protein was proved to be biologically active. However, glycosylation was not addressed in any of the previous studies.

In the present study, we have successfully expressed recombinant GalNAc-*O*-glycosylated G-CSF in *N. benthamiana* plants. Mammalian-specific *O*-glycosylation of the plant-derived G-CSF was achieved by targeting the recombinant protein to the secretory pathway, using its native secretory signal peptide and co-expressing it with a GalNAc-*O*-glycosylation machinery [6]. The recombinant secretory G-CSF (Sec-G-CSF) was confirmed to be GalNAc-*O*-glycosylated via lectin blot. Further mass spectrometry analysis corroborated the proper GalNAc-*O*-glycosylation at the single Thr residue reported to be mucin-type *O*-glycosylated in the native protein. Although plant-specific glycosylation is known to occur on proteins targeted to the secretory pathway, more specifically *O*-glycosylation at hydroxylated prolines, analyses in the present study were not addressed to characterize such post-translational modifications. Thus, further analytical studies will be required to fully characterize the glycosylation profile of the recombinant protein, as well as to determine glycosylation ratios. Nevertheless, corroboration of proper GalNAc-*O*-glycosylation of the Sec-G-CSF variant highlights the feasibility to further elongate the GalNAc-*O*-glycosylation moiety and imitate native mucin-type *O*-glycosylation. This objective would require the co-expression of additional glycosyltransferases and other proteins, as it has been demonstrated by recent studies where the GalNAc-*O*-glycosylated sites of recombinant proteins transiently expressed in *N. benthamiana* plants, were further elongated with Galactose and sialic acid [14–16, 71, 72].

In relation to multimerization, Sec-G-CSF did not seem to multimerize, in contrast to the other two variants targeted to the cytoplasm or expressed as ER-derived protein bodies (Cyt-G-CSF and Zera-G-CSF, respectively) (Figs. 2, 3; Additional file 3: Fig. S3). As reported previously by other studies where recombinant G-CSF was expressed in yeast and *A. niger* [29, 31, 66, 67], multimerization had a negative impact on biological activity and it has been suggested that it is caused by the lack of proper protein folding, by escaping the ER processing [67]. The fact that Sec-G-CSF was glycosylated does not only support the evidence that the native secretory signal peptide did indeed target the recombinant protein to the secretory pathway, since the GalNAc-Transferase is known to locate and exert its activity in the Golgi apparatus or in subregions of the ER (Additional file 3: Fig. S1) [15, 49–52], but localization of G-CSF in the secretory pathway also suggests proper protein processing by the ER, which together with further glycosylation, might decrease the nature of oligomerization of G-CSF, as suggested by other studies [30, 66–68]. Lack of multimerization and proper glycosylation are positively related to biological activity in this case, indicating the potential of the expression platform used in this project for the production of this glycoprotein. Nevertheless, proper studies related to the impact of glycosylation and ER processing to multimerization and biological activity will have to be carried out in future work.

It must also be noticed that the accumulation levels obtained with the secretory G-CSF were the lowest in comparison to the other two variants (Cyt-G-CSF and Zera-G-CSF). Co-expression with the glycosylation machinery decreased the accumulation levels of Sec-G-CSF:cMyc even more (approximately 10–30 × decrease was observed during the experiments, data not shown) (Fig. 5). Such decrease was expected since the expression of several recombinant proteins and their respective markers would cause an exhaustive usage of the translation machinery, therefore it would be advisable to generate a single multiple expression cassette containing only the required transgenes, assuring at the same time that all genes get expressed in the same cells [16]. Likewise, future subcellular relocalization of the recombinant G-CSF together with glycosyltransferases, as well as developing stable transgenic lines with the corresponding glycosylation machinery, should also be considered as an

attempt to increase accumulation levels, while allowing the production of mucin-type *O*-glycosylated G-CSF in plants [4].

Conclusions

In conclusion, we have successfully expressed GalNAc-*O*-glycosylated G-CSF in *N. benthamiana* plants via agroinfiltration, offering an alternative system for the production of this pharmaceutical protein with its native mammalian-specific post-translational modification. Further studies would be required to explore elongation of the glycosylation moiety and approaches to increasing accumulation of this recombinant glycoprotein.

Additional files

Additional file 1: Figure S3. Impact of extraction buffer on G-CSF. A Western Blot detection of Sec-G-CSF-sYFP and Cyt-G-CSF-sYFP extracted with $PBST^{0.1\ \%}$ (gels on the left) or under reducing and denaturing extraction conditions (SDS-DTT) (gels on the right). **B** Band quantification of Western blot detected proteins. Samples collected at 4, 6 and 8 dpi. Four leaf discs from different leaves were collected from each biological sample. 20 µL TSP of $PBST^{0.1\ \%}$ treated sample or equivalent volume of SDS-DTT treated sample were loaded on the gel. Black arrows denote monomeric G-CSF-sYFP variant. p19: negative control. αGEHK: protein standard. Proteins were detected with GFP antibody. Band quantification of Western blot detected proteins (Graph). Columns denoted with a different letter are significantly different ($p \leq 0.05$) using one-way ANOVA and followed by Tukey test. Error bars are standard deviation of the means.

Additional file 2: Figure S2. Secretory and Cytoplasmic G-CSF can be extracted with a PBS-based extraction buffer, but Zera-G-CSF can only be extracted under denaturing and reducing conditions. Western Blot detection of G-CSF-sYFP variants transiently expressed in *N. benthamiana* plants. Samples collected at 4 dpi. Upper panel: Plant 1; Lower panel: Plant 2. Sec-G-CSF-sYFP (Secretory variant, 49 kDa), Cyt-G-CSF-sYFP (Cytoplasmic and mature protein variant, 49 kDa), Zera-G-CSF-sYFP (Zera fused variant, 61 kDa), sYFP (Cytoplasmic sYFP, 29 kDa), p19 negative control. Four leaf discs from two different plants were collected for each sample. 50 µg TSP of PBST samples or equivalent volume of SDS-DTT samples were loaded on the gel. Black arrows denote Zera-G-CSF-sYFP. Gray arrows denote Sec-GCSF-sYFP and Cyt-G-CSF-sYFP. Red asterisks denotes a faint band corresponding to Zera-G-CSF-sYFP extracted with PBST extraction buffer and its corresponding proper extraction with SDS-DTT extraction buffer. Proteins were detected with anti GFP.

Additional file 3: Figure S1. Compartmentalization of GalNAc attachment to target proteins in the secretory pathway. GalNAc-Transferases are traditionally referred to be localized in the Golgi Apparatus (Cis-Golgi), but recent studies suggest also ER localization, thus proposing attachment of GalNAc in subregions of ER and proximal Golgi compartment. GalNAc-T, GalNAc transferase. ER, Endoplasmic Reticulum. ERGIC, Intermediate ER-Golgi Compartment. Grey box, GalNAc.

Additional file 4: Figure S4. MS/MS identification of glycosylated Sec-G-CSF-cMyc-derived peptide. A Schematic illustration of released peptide after Trypsin/Chymotrypsin in gel digestion of c-Myc purified Sec-G-CSF-cMyc expressed alone or co-expressed with the *O*-glycosylation machinery. Native glycosylation site (Thr-163) is denoted by underlining. **B** Extracted ion chromatograms of predicted QQMEELGMAPALQPTQGAMPAFASAF peptide derived from Sec-G-CSF-cMyc expressed alone (upper panel), not being detected, or co-expressed with the *O*-glycosylation machinery (lower panel), where it was detected. **C** MS/MS of c-Myc purified and Trypsin/Chymotrypsin in gel digested Sec-G-CSF-cMyc-derived peptide (QQMEELGMAPALQPTQGAMPAFASAF) co-expressed with the *O*-glycosylation machinery, showing identified *b*- and *y*-ions. Modified Thr-163 is marked in red. Complete list of detected product ions is shown in Supplementary Table S1.

Additional file 5: Table S1. *y*- and *b*- product ions detected in GalNAc-*O*-glycosylated QQMEELGMAPALQPTQGAMPAFASAF derived peptide and associated mass errors.

Authors' contributions

IAR-A, RM and GAC conceived, designed the experiments. IAR-A performed all the experiments. JBR designed, performed and interpreted MS/MS analysis. IAR-A, JBR, SGL, RM and GAC helped to write, improve and review the latest version of this manuscript. All authors read and approved the final manuscript.

Author details

[1] School of Engineering and Sciences, Tecnológico de Monterrey, Campus Monterrey, Av. Eugenio Garza Sada 2501 Sur, C.P. 64849 Monterrey, NL, Mexico. [2] Arizona State University, Phoenix, AZ 85004-4467, USA. [3] Agriculture and Agri-Food Canada, London, ON, Canada. [4] Department of Biology, University of Western Ontario, London, ON, Canada.

Acknowledgements

Authors would like to thank Centro de Biotecnología-FEMSA and Centro de Agrobiotecnología at Tecnológico de Monterrey, Mexico, in particular the Research Group of Bioprocessing and Synthetic Biology (Dr. Marco Antonio Rito-Palomares); and the Center for Protein Identification and Development (CIDPRO) (Dr. Sergio Román Othón Serna-Saldívar) for infrastructure and economic support. Special thanks are also given to CONACyT for Grant (265404/227702-351384) to Israel A. Ramírez-Alanís. We also would like to especially thank Hong Zhu and Angelo Kaldis for technical assistance, Reza Saberianfar for providing confocal microscopy expert assistance, and the AAFC foreign research participant program. We also thank Sasha Daskalova for providing the glycosylation related vectors. We thank Dr. Tsuyoshi Nakagawa (Shimane University) for providing Gateway binary vectors that contain the *bar* gene, which was identified by Meiji Seika Kaisha, Ltd.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

Not applicable.

Publisher's Note

Received: 3 April 2018 Accepted: 19 October 2018
Published online: 06 November 2018

References

1. Daniell H, Singh ND, Mason H, Streatfield SJ. Plant-made vaccine antigens and biopharmaceuticals. Trends Plant Sci. 2009;14:669–79.
2. Sharma AK, Sharma MK. Plants as bioreactors: recent developments and emerging opportunities. Biotechnol Adv. 2009;27:811–32.
3. Sack M, Hofbauer A, Fischer R, Stoger E. The increasing value of plant-made proteins. Curr Opin Biotechnol. 2015;32:163–70.

4. Dicker M, Schoberer J, Vavra U, Strasser R. Subcellular targeting of proteins involved in modification of plant N- and O-glycosylation. In: Castilho A, editor. Methods mol. biol. New York: Humana Press; 2015. p. 249–67.
5. Faye L, Boulaflous A, Benchabane M, Gomord V, Michaud D. Protein modifications in the plant secretory pathway: current status and practical implications in molecular pharming. Vaccine. 2005;23:1770–8.
6. Daskalova SM, Radder JE, Cichacz ZA, Olsen SH, Tsaprailis G, Mason H, et al. 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 Biotechnol. 2010;10:62.
7. Bardor M, Faveeuw C, Fitchette AC, Gilbert D, Galas L, Trottein F, et al. Immunoreactivity in mammals of two typical plant glyco-epitopes, core α(1,3)-fucose and core xylose. Glycobiology. 2003;13:427–34.
8. Gomord V, Fitchette AC, Menu-Bouaouiche L, Saint-Jore-Dupas C, Plasson C, Michaud D, et al. Plant-specific glycosylation patterns in the context of therapeutic protein production. Plant Biotechnol J. 2010;8:564–87.
9. Jin C, Altmann F, Strasser R, Mach L, Schähs M, Kunert R, et al. A plant-derived human monoclonal antibody induces an anti-carbohydrate immune response in rabbits. Glycobiology. 2008;18:235–41.
10. Yang Z, Drew DP, Jørgensen B, Mandel U, Bach SS, Lavery SB, et al. Engineering mammalian mucin-type O-glycosylation in plants. J Biol Chem. 2012;287:11911–23.
11. Vukušić K, Šikić S, Balen B. Recombinant therapeutic proteins produced in plants: towards engineering of human-type O- and N-glycosylation. Period Biol. 2016;118:75–90.
12. Khan AH, Bayat H, Rajabibazl M, Sabri S, Rahimpour A. Humanizing glycosylation pathways in eukaryotic expression systems. World J Microbiol Biotechnol. 2017;33:1–12.
13. Castilho A, Neumann L, Gattinger P, Strasser R, Vorauer-Uhl K, Sterovsky T, et al. Generation of biologically active multi-sialylated recombinant human EPOFc in plants. PLoS ONE. 2013;8:1–13.
14. Dicker M, Tschofen M, Maresch D, Konig J, Juarez P, Orzaez D, et al. Transient glyco-engineering to produce recombinant IgA1 with defined N- and O-glycans in plants. Front Plant Sci. 2016;7:18.
15. Strasser R. Challenges in O-glycan engineering of plants. Front Plant Sci. 2012;3:1–6.
16. Strasser R. Engineering of human-type O-glycosylation in *Nicotiana benthamiana* plants. Bioengineered. 2013;4:191–6.
17. Yang Z, Bennett EP, Jørgensen B, Drew DP, Arigi E, Mandel U, et al. Toward stable genetic engineering of human O-glycosylation in plants. Plant Physiol. 2012;160:450–63.
18. Houston AC, Stevens LA, Cour V. Pharmacokinetics of glycosylated recombinant human granulocyte colony-stimulating factor (lenograstim) in healthy male volunteers. Br J Clin Pharmacol. 1999;47:279–84.
19. Reeves G. Overview of use of G-CSF and GM-CSF in the treatment of acute radiation injury. Health Phys. 2014;106:699–703.
20. Chavez-Tapia NC, Mendiola-Pastrana I, Ornelas-Arroyo V, Noroña-Herrera C, Vidaña-Perez D, Delgado-Sanchez G, et al. Granulocyte-colony stimulating factor for acute-on-chronic liver failure: systematic review and meta-analysis. Ann Hepatol. 2015;14(5):631–41.
21. Hermesh T, Moran TM, Jaln D, López CB. Granulocyte colony-stimulating factor protects mice during respiratory virus infections. PLoS ONE. 2012;7:2–9.
22. Barreda DR, Hanington PC, Belosevic M. Regulation of myeloid development and function by colony stimulating factors. Dev Comp Immunol. 2004;28:509–54.
23. Bönig H, Silbermann S, Weller S, Kirschke R, Körholz D, Janssen G, et al. Glycosylated vs non-glycosylated granulocyte colony-stimulating factor (G-CSF)—results of a prospective randomised monocentre study. Bone Marrow Transplant. 2001;28:259–64.
24. Dale DC. Colony-stimulating factors for the management of neutropenia in cancer patients. Drugs. 2002;62:1–15.
25. Duncan N, Hewetson M, Atra A, Dick G, Pinkerton R. An economic evaluation of the use of granulocyte colony-stimulating factor after bone marrow transplantation in children. Pharmacoeconomics. 1997;11:169–74.
26. Barnes G, Pathak A, Schwartzberg L. Pharmacoeconomics of granulocyte colony-stimulating factor: a critical review. Adv Ther. 2014;31:683–95.
27. Whyte S, Cooper KL, Stevenson MD, Madan J, Akehurst R. Cost-effectiveness of granulocyte colony-stimulating factor prophylaxis for febrile neutropenia in breast cancer in the United Kingdom. Value Health. 2011;14:465–74.
28. Mire-Sluis AR, Das RG, Thorpe R. The international standard for granulocyte colony stimulating factor (G-CSF). Evaluation in an international collaborative study. Participants of the collaborative study. J Immunol Methods. 1995;179:117–26.
29. Suk Yang D, Soon Bae C, Lee J. Production of recombinant human granulocyte-colony-stimulating factor in high cell density yeast cultures. Biotechnol Lett. 1997;19:655–9.
30. Bae CS, Yang DS, Lee J, Park YH. Improved process for production of recombinant yeast-derived monomeric human G-CSF. Appl Microbiol Biotechnol. 1999;52:338–44.
31. Lasnik MA, Porekar VG, Stalc A. Human granulocyte colony stimulating factor (hG-CSF) expressed by methylotrophic yeast pichia pastoris. Pflugers Arch. 2001;442:R184–6.
32. Sharifi Tabar M, Akbar Habashi A, Momari HR. Human granulocyte colony-stimulating factor (hg-csf) expression in plastids of *Lactuca sativa*. Iran Biomed J. 2013;17:158–64.
33. Carter CRD, Keeble JR, Thorpe R. Human serum inactivates non-glycosylated but not glycosylated granulocyte colony stimulating factor by a protease dependent mechanism: significance of carbohydrates on the glycosylated molecule. Biologicals. 2004;32:37–47.
34. Carter CRD, Whitmore KM, Thorpe R. The significance of carbohydrates on G-CSF: differential sensitivity of G-CSFs to human neutrophil elastase degradation. J Leukoc Biol. 2004;75:515–22.
35. Höglund M. Glycosylated and non-glycosylated recombinant human granulocyte colony-stimulating factor (rhG-CSF)—what is the difference? Med Oncol. 1998;15:229–33.
36. Mattii L, Azzarà A, Fazzi R, Carulli G, Chimenti M, Cicconi N, et al. Glycosylated or non-glycosylated G-CSF differently influence human granulocyte functions through RhoA. Leuk Res. 2005;29:1285–92.
37. Martin-Christin F. Clinical report Granulocyte colony stimulating factors: how different are they? How to make a decision? Anticancer Drugs. 2001;12:185–91.
38. Alvarez ML, Topal E, Martin F, Cardineau GA. Higher accumulation of F1-V fusion recombinant protein in plants after induction of protein body formation. Plant Mol Biol. 2010;72:75–89.
39. Nakamura S, Mano S, Tanaka Y, Ohnishi M, Nakamori C, Araki M, et al. Gateway binary vectors with the bialaphos resistance gene, bar, as a selection marker for plant transformation. Biosci Biotechnol Biochem. 2010;74:1315–9.
40. Silhavy D, Molnár A, Lucioli A, Szittya G, Hornyik C, Tavazza M, et al. A viral protein suppresses RNA silencing and binds silencing-generated, 21- to 25-nucleotide double-stranded RNAs. EMBO J. 2002;21:3070–80.
41. Conley AJ, Joensuu JJ, Menassa R, Brandle JE. Induction of protein body formation in plant leaves by elastin-like polypeptide fusions. BMC Biol. 2009;7:48.
42. Latijnhouwers M, Hawes C, Carvalho C, Oparka K, Gillingham AK, Boevink P. An Arabidopsis GRIP domain protein localizes to the trans-Golgi and binds the small GTPase ARL1. Plant J. 2005;44:459–70.
43. Kapila J, De Rycke R, Van Montagu M, Angenon G. An Agrobacterium-mediated transient gene expression system for intact leaves. Plant Sci. 2005;122:101–8.
44. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976;72:248–54.
45. Finnie C. Plant proteomics. Plant Proteomics. 2007;1072:1–253.
46. Keller D, Chambers M, Burke R, Agus D, Mallick P. ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics. 2008;24:2534–6.
47. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. Bioinformatics. 2004;20:1466–7.
48. Vaudel M, Barsnes H, Berven FS, Sickmann A, Martens L. SearchGUI: an open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. Proteomics. 2011;11:996–9.
49. Vaudel M, Burkhart JM, Zahedi RP, Oveland E, Berven FS, Sickmann A, et al. PeptideShaker enables reanalysis of MS-derived proteomics data sets. Nat Biotechnol. 2015;33:22–4.
50. Torrent M, Llompart B, Lasserre-Ramassamy S, Llop-Tous I, Bastida M, Marzabal P, et al. Eukaryotic protein production in designed storage organelles. BMC Biol. 2009;7:5.

51. Joseph M, Ludevid MD, Torrent M, Rofidal V, Tauzin M, Rossignol M. Proteomic characterisation of endoplasmic reticulum-derived protein bodies in tobacco leaves. BMC Plant Biol. 2012;12:36.
52. Saberianfar R, Joensuu JJ, Conley AJ, Menassa R. Protein body formation in leaves of *Nicotiana benthamiana*: a concentration-dependent mechanism influenced by the presence of fusion tags. Plant Biotechnol J. 2015;13:927–37.
53. Stanley P. Golgi glycosylation. Cold Spring Harb Perspect Biol. 2011;3:1–13.
54. Gill DJ, Chia J, Senewiratne J, Bard F. Regulation of O-glycosylation through Golgi-to-ER relocation of initiation enzymes. J Cell Biol. 2010;189:843–58.
55. Van den Steen P, Rudd PM, Dwek RA, Opdenakker G. Concepts and principles of O-linked glycosylation. Crit Rev Biochem Mol Biol. 1998;33:151–208.
56. Bennett EP, Mandel U, Clausen H, Gerken TA, Fritz TA, Tabak LA. Control of mucin-type O-glycosylation: a classification of the polypeptide GalNAc-transferase gene family. Glycobiology. 2012;22:736–56.
57. Ahmad A, Pereira EO, Conley AJ, Richman AS, Menassa R. Green biofactories: recombinant protein production in plants. Recent Pat Biotechnol. 2010;4:242–59.
58. Alvarez ML, Pinyerd HL, Topal E, Cardineau GA. P19-dependent and P19-independent reversion of F1-V gene silencing in tomato. Plant Mol Biol. 2008;68:61–79.
59. Hong S-Y, Kwon T-H, Lee J-H, Jang Y-S, Yang M-S. Production of biologically active hG-CSF by transgenic plant cell suspension culture. Enzyme Microb Technol. 2002;30:763–7.
60. Tabar MS, Solouki M, Tohidfar M, Sadeghizadeh M. Expression of human granulocyte-colony stimulating factor (hG-CSF) gene in tobacco (*Nicotiana tobacum*). Aust J Crop Sci. 2012;6:135–40.
61. Nair NR, Chidambareswaren M, Manjula S. Enhanced heterologous expression of biologically active human granulocyte colony stimulating factor in transgenic tobacco BY-2 cells by localization to endoplasmic reticulum. Mol Biotechnol. 2014;56:849–62.
62. Hong S-Y, Kwon T-H, Jang Y-S, Kim S-H, Yang M-S. Production of bioactive human granulocyte-colony stimulating factor in transgenic rice cell suspension cultures. Protein Expr Purif. 2006;47:68–73.
63. Zvonova AS, Petrovskaya LE, Rodina AV, Frolova OY, Ivanov PA, Shingarova LN, et al. Production of biologically active human myelocytokines in plants. Biochem Biokhimiia. 2009;74:1187–94.
64. Perez-Perez J, Martinez-Caja C, Barbero J, Gutierrez J. DnaK/DnaJ supplementation improves the periplasmic production of human granulocyte-colony-stimulating factor in *Escherichia coli*. Biochem Biophys Res Commun. 1995;210(2):524–32.
65. Oh-eda M, Hasegawa M, Hattori K, Kuboniwa H, Kojima T, Orita T, et al. O-linked sugar chain of human granulocyte-colony stimulating factor protects it against polymerization and denaturation allowing it to retain its biological activity. J Biol Chem. 1990;265:11432–5.
66. Bae CS, Yang DS, Chang KR, Seong BL, Lee J. Enhanced secretion of human granulocyte colony stimulating factor directed by a novel hybrid fusion peptide from recombinant *Saccharomyces cerevisiae* at high cell concentration. Biotechnol Bioeng. 1998;57:600–9.
67. Kraševec N, Milunovic T, Lasnik MA, Lukančič I, Komel R, Porekar VG. Human granulocyte colony stimulating factor (G-CSF) produced in the filamentous fungus aspergillus niger. Acta Chim Slov. 2014;61:709–17.
68. Raso SW, Abel J, Barnes JM, Maloney KM, Pipes G, Treuheit MJ, et al. Aggregation of granulocyte-colony stimulating factor in vitro involves a conformationally altered monomeric state. Protein Sci. 2005;14:2246–57.
69. Böer E, Steinborn G, Kunze G, Gellissen G. Yeast expression platforms. Appl Microbiol Biotechnol. 2007;77:513–23.
70. Gong B, Bumina I, Lynaugh H, Li H. O-linked glycosylation analysis of recombinant human granulocyte colony-stimulating factor produced in glycoengineered *Pichia pastoris* by liquid chromatography and mass spectrometry. J Chromatogr B Anal Technol Biomed Life Sci. 2014;945–946:135–40.
71. Dicker M, Maresch D, Strasser R. Glyco-engineering for the production of recombinant IgA1 with distinct mucin-type O-glycans in plants. Bioengineered. 2016;5979:484–9.
72. Castilho A, Neumann L, Daskalova S, Mason HS, Steinkellner H, Altmann F, et al. Engineering of sialylated mucin-type O-glycosylation in plants. J Biol Chem. 2012;287:36518–26.

Chapter 13

• Curriculum Vitae

PERSONAL INFORMATION

Name: Israel Alfonso Ramírez Alanis

Born November 2nd, 1983, in Monterrey, Mexico.

e-mail: israel.ramirezalanis@gmail.com

EDUCATION

Degree: ***PhD in Biotechnology***

University: Instituto Tecnológico de Estudios Superiores de Monterrey (ITESM), Monterrey Campus in December 2018.

Accepted in the PhD program in Biotechnology in January, 2013

Degree: ***Master in Science in Biotechnology***

University: Instituto Tecnológico de Estudios Superiores de Monterrey (ITESM), Monterrey Campus

Degree: ***Bachelor of Science in Chemistry Bacteriology Parasitology***

University: Universidad Autónoma de Nuevo León (UANL), Faculty of Biological Sciences

RESEARCH PROJECTS

ITESM (Monterrey, Mx), Agriculture and Agri-Food Canada (AAFC, London, Ca)	
November 2012 - Today	**PhD Student in Biotechnology:** - **Advisor: PhD Guy Albert Cardineau, PhD Sergio Serna-Saldivar** - Molecular farming: Production of pharmaceutical and nutritional recombinant proteins in plants. - Research project: "Engineering mammalian-specific posttranslational modifications in plant-made nutraceuticals and pharmaceuticals: phosphorylation and sialylated Mucin-type O-glycosylation as a challenge."
Ludwig-Maximillians-Universität, Munich, Germany.	

March 2011- October 2012	**Department of Genetics:** - Advisor: PhD Thomas Lahaye - Synthetic biology: Modulating specificity of custom designed Transcription Activator-Like Effectors (TALEs) based on Golden Gate cloning and Gateway technology.
Instituto Tecnológico de Estudios Superiores de Monterrey (ITESM), Monterrey.	
January 2009 to February 2011	**Master Student in Industrial Biotechnology:** – Advisor: PhD Guy Albert Cardineau – Expression of recombinant proteins with pharmaceutical properties using plants as production platforms. – Increased accumulation of the Recombinant Human Granulocyte-Colony Stimulating Factor (rhG-CSF) in NT1 cells, after protein body formation and gene silencing through RNAi suppression. (Work used as Thesis to obtain the Master in Science Degree)
University of Würzburg, Germany.	
October 2005 to July 2006	**Undergraduate Research Project in the Department of Microbiology (Biozentrum):** – Advisor: PhD Jürgen Kreft – Investigations on *Listeria monocytogenes* virulence; determination of importance of two genes implicated in the response of the bacterium against oxidative stress; – Molecular analysis of the expression of some genes implicated in the oxidative stress response. (Work used as Thesis to obtain the degree of B.Sc.).

CONFERENCE

Biology Graduate Research Forum, Western University, London, Canada.	
October 2016	**Poster :** – Engineering Plant *O*-Glycosylation.
Plant Biotech 2016, Queen´s University, Kingston, Canada.	
June 2016	**Poster :** – Plant-derived *O*-glycosylated pharmaceuticals.
Congreso de Investigación y Desarrollo Tec, ITESM, Monterrey, Mexico.	
January 2010	**Speaker :** – Design of the recombinant human Granulocyte-Colony Stimulating Factor (rhG-CSF) gene for its expression in NT1 cells.

This document was typed in using Microsoft Word by Israel Alfonso Ramírez-Alanis