

N-glycosylation in sugarcane

Ivan G. Maia^{1,2} and Adilson Leite^{1*}

Abstract

The *N*-linked glycosylation of secretory and membrane proteins is the most complex posttranslational modification known to occur in eukaryotic cells. It has been shown to play critical roles in modulating protein function. Although this important biological process has been extensively studied in mammals, much less is known about this biosynthetic pathway in plants. The enzymes involved in plant *N*-glycan biosynthesis and processing are still not well defined and the mechanism of their genetic regulation is almost completely unknown. In this paper we describe our first attempt to understand the *N*-linked glycosylation mechanism in a plant species by using the data generated by the Sugarcane Expressed Sequence Tag (SUCEST) project. The SUCEST database was mined for sugarcane gene products potentially involved in the *N*-glycosylation pathway. This approach has led to the identification and functional assignment of 90 expressed sequence tag (EST) clusters sharing significant sequence similarity with the enzymes involved in *N*-glycan biosynthesis and processing. The ESTs identified were also analyzed to establish their relative abundance.

INTRODUCTION

In plants, as in other eukaryotes, most of the soluble and membrane bound proteins that are synthesized on polyribosomes associated with the endoplasmic reticulum (ER) are glycoproteins, including those proteins which will later be exported to the Golgi apparatus, lysosomes, plasma membrane or extracellular matrix. The glycans attached to glycoproteins contain a variety of sugar residues linked in linear or branched structures that can assume many different conformations. These glycans play a fundamental role in promoting correct protein folding and assembly and, as a consequence, enhance protein stability. They may also contain targeting information, or may be directly involved in protein recognition. The three main posttranslational modifications of proteins that involve carbohydrates are *N*- and *O*-linked glycosylation and the insertion of glycosyl phosphatidyl inositol anchors.

The precursor oligosaccharide that initiates *N*-glycosylation is assembled sugar by sugar onto a carrier lipid molecule called dolichol (Figure 1A), which is anchored in the ER membrane (Kornfeld and Kornfeld, 1985). The sugars are activated in the cytosol by the formation of nucleotide-sugar intermediates which then donate their sugar moieties to the membrane-bound lipid in an orderly sequence. The lipid-linked oligosaccharide is subsequently flipped from the cytosolic to the luminal face of the ER membrane. In the final step, an oligosaccharyl transferase catalyzes the transfer of the pre-formed oligosaccharide precursor, usually $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$, from the dolichyl pyrophosphate donor to specific asparagine residues (of the acceptor sequence Asn-X-Ser/Thr) of the nascent polypep-

tide chain (Hubbard and Ivatt, 1981). Immediately after the transfer, $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ undergoes trimming of the glucose (Glc) and some of the mannose (Man) residues, first in the ER and then in the Golgi apparatus (Figure 2A; for a review see Herscovics, 1999), giving rise to high-mannose-type *N*-glycans containing from five to nine mannose residues. Subsequent transfer to the resulting glycan of branching *N*-acetylglucosamine (GlcNAc) and additional sugars by Golgi glycosyltransferases to form hybrid and complex *N*-glycans also occurs (Figure 2A).

In plants, complex-type *N*-glycans are characterized by the presence of α -1,3-fucose (as opposed to α -1,6-linked fucose in mammals) and β -1,2-xylose residues respectively, linked to the proximal GlcNAc and core β -Man residues, and by the presence of β -1,2-GlcNAc residues linked to the α -Man units. Larger, complex-type plant glycans are rare and were recently identified as containing additional α -1,4-fucose and β -1,3-galactose residues linked to the terminal GlcNAc units. These modifications result in the Lewis antigens usually found on mammalian cell-surface glyco-conjugates (Lerouge *et al.*, 1998). In mammals the core structure is extended to contain penultimate galactose and terminal sialic acid residues.

After maturation in the ER and the Golgi apparatus, complex-type *N*-glycans can be further modified during the glycoprotein transport to their final destination. Many undergo removal of the terminal GlcNAc residues by the action of a *N*-acetylglucosaminidase present within the vacuole (Figure 2A). When *N*-glycans are fully accessible to processing enzymes, further *N*-glycosylation with the

¹Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, C.P. 6010, 13083-970 Campinas, SP, Brazil.

²Present address: Instituto de Biociências, Departamento de Genética, UNESP – Campus de Botucatu, 18618-000 Botucatu, SP, Brazil.

Send correspondence to Adilson Leite. E-mail aleite@unicamp.br.

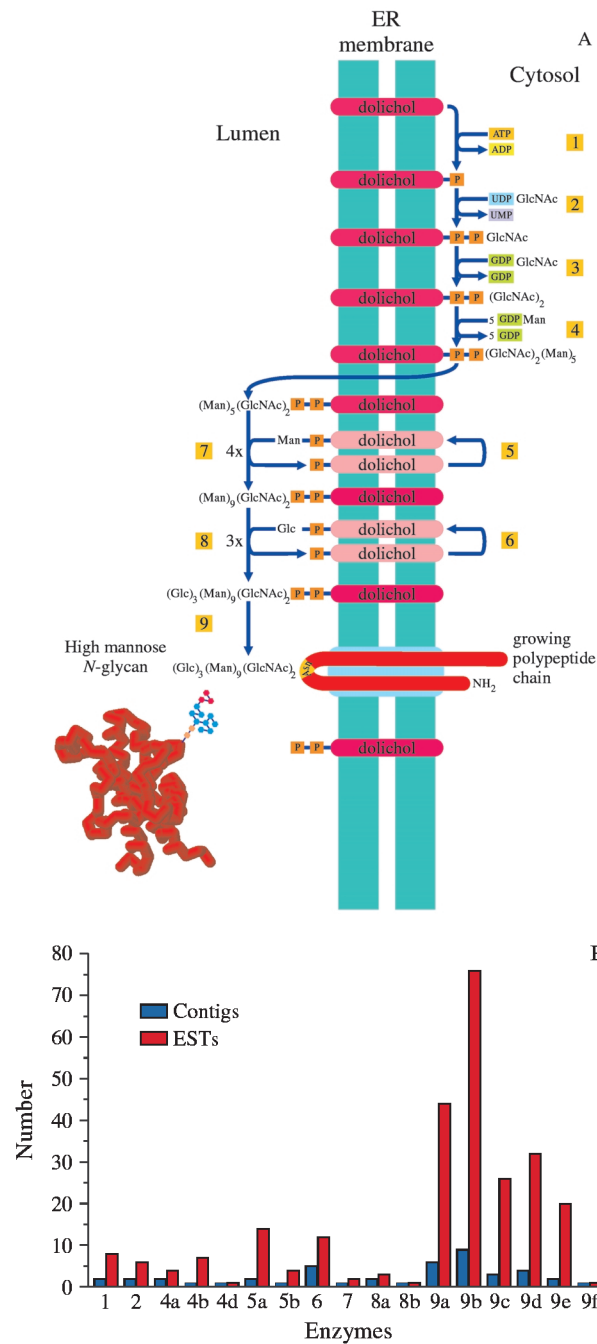


Figure 1 - Schematic representation of the N-linked dolichol-phosphate pathway (A). The enzymes are listed by numbers: 1, dolichol kinase (EC 2.7.1.108); 2, UDP-GlcNAc N-acetylglucosaminophosphotransferase (EC 2.7.8.15); 3, GDP-GlcNAc:GlcNAc beta-1,4-N-acetylglucosaminyltransferase (EC 2.4.1.-); 4a, glycolipid 4-beta-mannosyltransferase (EC 2.4.1.-); 4b, glycolipid 3-alpha-mannosyltransferase (EC 2.4.1.132); 4c, glycolipid 6-alpha-mannosyltransferase (EC: 2.4.1.-); 4d, glycolipid 2-alpha-mannosyltransferase (EC 2.4.1.131); 5, dolichyl-phosphate beta-D-mannosyltransferase (EC 2.4.1.83), polypeptide 1 (a), polypeptide 2 (b) and polypeptide 3 (c); 6, dolichyl-phosphate beta-glucosyltransferase (EC 2.4.1.117); 7, dolichyl-phosphate-mannose-glycolipid alpha-mannosyltransferase (EC 2.4.1.130); 8, dolichyl-phosphate-glucose-glycolipid-alpha-glycosyltransferase (EC 2.4.1.-) yeast alg6-like (a) or yeast alg8-like (b); 9, dolichyl-diphosphooligosaccharide-protein glycosyltransferase (EC 2.4.1.119), subunit 48K (a), ribophorin I (b), ribophorin II (c), yeast SST3-like (d), yeast DAD-1-like (e), yeast OST4-like (f). (B) Distribution of the sugarcane contigs (and corresponding ESTs) encoding the enzymes listed above.

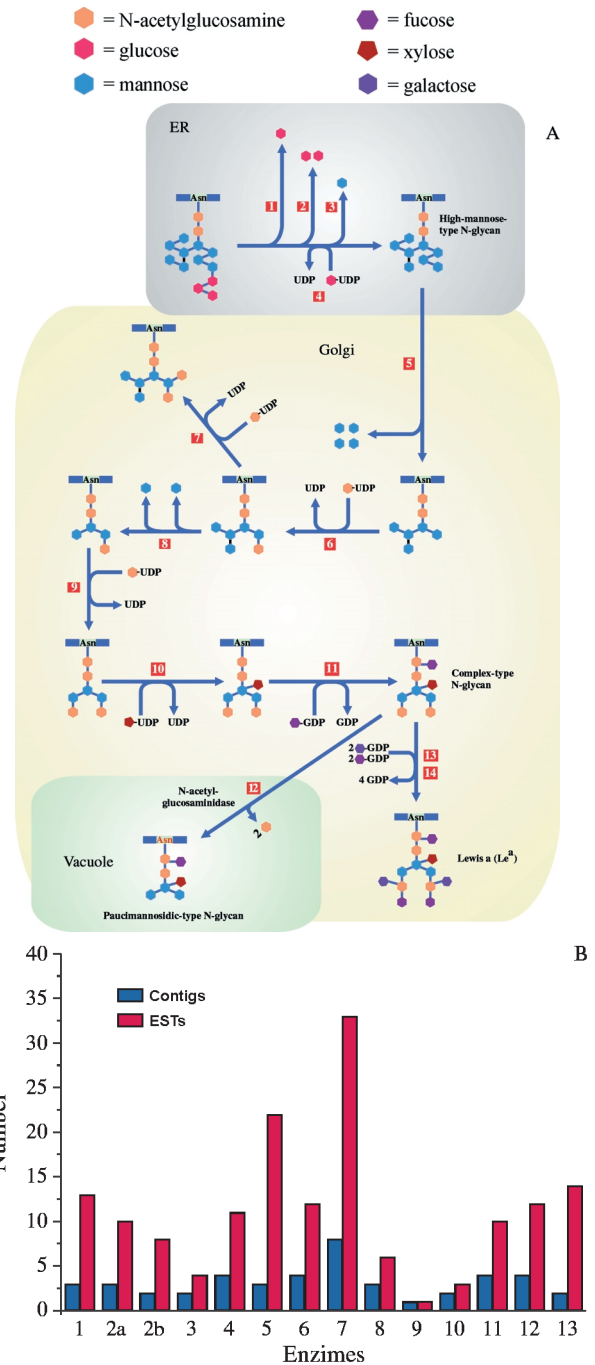


Figure 2 - Major pathway representing N-glycan processing in plants (A). The enzymes are indicated by numbers: 1, mannosyl-oligosaccharide glucosidase I (EC 3.2.1.106); 2, mannosyl-oligosaccharide glucosidase II, alpha subunit (a) and beta subunit (b); 3, ER mannosyl-oligosaccharide 1,2-alpha-mannosidase (EC 3.2.1.-); 4, UDP-glucose: glycoprotein glucosyltransferase (EC 2.4.1.-); 5, mannosyl-oligosaccharide 1,2-alpha-mannosidase (EC 3.2.1.113); 6, alpha-1,3-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase (EC 2.4.1.101); 7, beta-1,4-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase (EC 2.4.1.144); 8, mannosyl-oligosaccharide 1,3-1,6-alpha-mannosidase (EC 3.2.1.114); 9, alpha-1,6-mannosyl-glycoprotein beta-1,2-N-acetylglucosaminyltransferase (EC 2.4.1.143); 10, beta-1,2-xylosyltransferase (EC 2.4.1.-); 11, alpha-1,3-fucosyltransferase (EC 2.4.1.-); 12, alpha-N-acetylglucosaminidase; 13, beta-1,3-galactosyltransferase (EC 2.4.1.-); 14, alpha-1,4-fucosyltransferase (EC 2.4.1.-). (B) Distribution of the sugarcane contigs (and corresponding ESTs) encoding the enzymes listed above.

same type of modified *N*-glycans containing fucose and/or xylose residues but devoided of terminal GlcNAc residues can occur (Lerouge *et al.*, 1998).

Although the *N*-linked glycosylation mechanisms in mammalian and plant systems have been conserved during evolution, differences are observed in the final steps of oligosaccharide trimming and glycan modification in the Golgi apparatus. High mannose-type *N*-glycans in plants have structures identical to those found in other eukariotic cells, but plant complex *N*-linked glycans differ substantially (Lerouge *et al.*, 1998).

Our group has been using plants as bioreactors for the production of human therapeutic proteins (Leite *et al.*, 2000). Although many glycoproteins, such as those found in blood plasma, are of great pharmaceutical interest, their production in plants could lead to their inactivation as a consequence of the different patterns of host protein glycosylation (Chrispeels and Faye, 1996; Giddings *et al.*, 2000). Alternatively, prolonged human exposure to large quantities of immunogenic plant glycans can produce sensitization to these antigens. The use of plants as factories for the production of glycoproteins for therapeutic purposes depends on the suppression or 'humanization' of the glycosylation system of these plants. To apply these strategies, a precise knowledge of the glycosylation pathway in plants is needed.

In this paper we describe our first attempt to better understand the *N*-linked glycosylation mechanism in plants by mining the Sugarcane Expressed Sequence Tag (SUCEST) project database for sugarcane gene products potentially involved in the *N*-glycosylation pathway. The annotated expressed sequence tag (EST) sequences derived from diverse sugarcane cDNA libraries offer an attractive tool for future gene cloning and characterization.

MATERIAL AND METHODS

To identify sugarcane gene products sharing similarities with enzymes involved in *N*-glycosylation, homology-based searches against cluster consensus sequences (contigs obtained using the Phrap algorithm) in the SUCEST database were performed in-house using the basic local alignment search tool (BLAST) algorithm of Altschul *et al.* (1997). Contigs were also identified by keyword searches using a provisional functional assignment generated by automated annotation, the resulting hits being validated against existing homologous sequences in public databases. An estimate of the relative abundance of the identified putative genes was generated based on EST counts per corresponding contig.

RESULTS AND DISCUSSION

The SUCEST database was mined for sugarcane gene products potentially involved in the *N*-glycosylation pathway. In the first step, amino acid sequences of known glycosylation enzymes from different organisms were used

as queries for TBLASTN searches against the cluster consensi database of SUCEST. Further EST identification was obtained by keyword searches of a preexisting provisional functional assignment of the database. A reciprocal search against a non-redundant protein database was then performed to validate the matched contig sequences.

By doing so, we were able to identify 90 sugarcane EST clusters sharing significant sequence similarity to enzymes known to be implicated in *N*-glycan biosynthesis and processing. Of these, 50% showed similarity to enzymes of the *N*-linked dolichol-phosphate pathway (Figure 1A), and 50% matched several known processing glycosidases and glycosyltransferases of the secretory pathway (Figure 2A). In some cases, gene products with no homology to previously described enzymes in plants were identified, thus representing likely novel genes. For example, searching revealed gene products whose amino acid sequences were related to an ER-mannosidase originally encountered in mammals. In contrast, BLAST searches revealed no gene products in sugarcane with significant amino acid sequence similarity to GDP-GlcNAc:GlcNAc β -1,4-*N*-acetylglucosaminyltransferase and α -1,4-fucosyltransferase.

An estimate of the relative abundance of the identified genes was obtained by comparing the number of times ESTs were assigned to a particular contig. As shown in Figure 1B and 2B, the most abundant gene products corresponded to dolichyl-diphosphooligosaccharide-protein glycosyltransferase (represented by a 48 kDa subunit, and ribophorin I and II), β -1,4-mannosyl-glycoprotein β -1,4-*N*-acetylglucosaminyltransferase and mannosyl-oligosaccharide 1,2- α -mannosidase.

The *N*-glycosylation pathway involves a large number of catalytic enzymes that are relatively well conserved in different organisms. We have searched for genes encoding 23 enzymes implicated in *N*-glycan biosynthesis and processing and have found matches for 21 of them. The putative plant homologues identified within the generated set of sugarcane ESTs will contribute significantly to the understanding of this biosynthetic pathway in plants. This EST collection will enable more refined analysis to produce further insights into the respective molecular nature and expression profiles of the corresponding genes. Moreover, although the EST-based approach described here focused mainly on the identification of sugarcane genes, it can easily be extended to identify and map homologues in closely related plants such as maize and sorghum.

ACKNOWLEDGMENTS

We are grateful to FAPESP for the financial support (Processo 00/07420-5). I. G. Maia was the recipient of a FAPESP postdoctoral fellowship. A. Leite is in receipt of a research fellowship from CNPq (Processo 301937/88-5).

RESUMO

A *N*-glicosilação é uma das principais modificações pós-traducionais, sendo responsável por alterações na conformação, estabilidade e conseqüentemente na funcionalidade de proteínas em eucariotos. Com a finalidade de melhor compreender a via de *N*-glicosilação em plantas foi realizada uma prospecção no banco de seqüências expressas do projeto genoma da cana de açúcar (SUCEST). Foram identificadas noventa seqüências cujos produtos gênicos apresentam alto grau de similaridade com enzimas envolvidas na biossíntese e processamento de *N*-glicanos. Dos vinte e três genes da via de *N*-glicosilação previamente descritos em diferentes espécies, vinte e um foram detectados em cana de açúcar.

REFERENCES

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.
- Chrispeels, M.J. and Faye, L. (1996). The production of recombinant glycoproteins with defined non-immunogenic glycans. In: *Transgenic plants: a production system for industrial and pharmaceutical proteins* (Owen, M.R.L. and Pen, J., eds.). John Wiley & Sons Ltd., England, pp. 99-113.
- Giddings, G., Allison, G., Brooks, D. and Carter, A. (2000). Transgenic plants as factories for biopharmaceuticals. *Nat. Biotechnol.* 18: 1151-1155.
- Herscovics, A. (1999). Importance of glycosidases in mammalian glycoprotein biosynthesis. *Bioch. Biophys. Acta* 1473: 96-107.
- Hubbard, S.C. and Ivatt, R.J. (1981). Synthesis and processing of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* 50: 555-583.
- Kornfeld, R. and Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* 54: 631-664.
- Leite, A., Kemper, E.L., da Silva, M.J., Luchessi, A.D., Siloto, R.M.P., Bonaccorsi, E.D., El-Dorry, H. and Arruda, P. (2000). Expression of correctly processed human growth hormone in seeds of transgenic tobacco plants. *Mol. Breeding* 6: 47-53.
- Lerouge, P., Cabanes-Macheteau, M., Rayon, C., Fischette-Lainé, A.C., Gomord, V. and Faye, L. (1998). N-glycoprotein biosynthesis in plants: recent developments and future trends. *Plant Mol. Biol.* 38: 31-48.
- Staudacher, E., Altmann, F., Wilson, I.B.H. and Marz, L. (1999). Fucose in N-glycans: from plant to man. *Bioch. Biophys. Acta* 1473: 216-236.