Overview of Nucleotide Sugar Transporter Gene Family Functions Across Multiple Species

Ariel Orellana1,2, Carol Moraga1, Macarena Araya1 and Adrian Moreno1,2

1 - Centro de Biotecnología Vegetal, Universidad Andres Bello, Av. República 217, Santiago, RM 837-0146, Chile
2 - FONDAP Center for Genome Regulation, Santiago, RM, Chile

Correspondence to Ariel Orellana: Centro de Biotecnología Vegetal, Universidad Andres Bello, Av. República 217, Santiago, RM 837-0146, Chile. aorellana@unab.cl

http://dx.doi.org/10.1016/j.jmb.2016.05.021

Edited by Thomas J. Smith

Abstract

Glycoproteins and glycolipids are crucial in a number of cellular processes, such as growth, development, and responses to external cues, among others. Polysaccharides, another class of sugar-containing molecules, also play important structural and signaling roles in the extracellular matrix. The additions of glycans to proteins and lipids, as well as polysaccharide synthesis, are processes that primarily occur in the Golgi apparatus, and the substrates used in this biosynthetic process are nucleotide sugars. These proteins, lipids, and polysaccharides are also modified by the addition of sulfate groups in the Golgi apparatus in a series of reactions where nucleotide sulfate is needed. The required nucleotide sugar substrates are mainly synthesized in the cytosol and transported into the Golgi apparatus by nucleotide sugar transporters (NSTs), which can additionally transport nucleotide sulfate. Due to the critical role of NSTs in eukaryotic organisms, any malfunction of these could change glycan and polysaccharide structures, thus affecting function and altering organism physiology. For example, mutations or deletion on NST genes lead to pathological conditions in humans or alter cell walls in plants. In recent years, many NSTs have been identified and functionally characterized, but several remain unanalyzed. This study examined existing information on functionally characterized NSTs and conducted a phylogenetic analysis of 257 NSTs predicted from nine animal and plant model species, as well as from protists and fungi. From this analysis, relationships between substrate specificity and the primary NST structure can be inferred, thereby advancing understandings of nucleotide sugar gene family functions across multiple species.

© 2016 Published by Elsevier Ltd.

Introduction

The addition of glycans to proteins and lipids is a fundamental process for many organisms. Another essential process is the cellular assembly of sugars into polysaccharides, the primary polymeric constituent of the extracellular matrix, especially in plants. Interestingly, plant cell walls account for most of the biomass on earth. These processes evidence the mechanisms developed by eukaryotic cells to enable proper macromolecule formation. In turn, these macromolecules (e.g., proteins, lipids, and polysaccharides) are responsible for a wide range of physiological functions.

Glycoconjugate synthesis primarily occurs in the Golgi apparatus and the endoplasmic reticulum (ER). These organelles contain glycosyltransferases, enzymes capable of transferring sugars to proteins, lipids, and polysaccharides. These macromolecules can also receive additional sulfate groups in the Golgi apparatus and ER through a process mediated by the sulfotransferases contained in these organelles. Glycosyltransferases and sulfotransferases use activated sugars and sulfates as substrates. These substrates include nucleoside diphosphate or monophosphate sugars, referred to as nucleotide sugars, and the nucleotide sulfate 3′-phosphoadenosine 5′-phosphosulfate (PAPS).

While both substrates (i.e., nucleotide sugars and PAPS) are present in the cytosol, these must be transferred into the ER and Golgi apparatus since the catalytic domain of most glycosyltransferases
and sulfotransferases faces the lumen of these organelles. Specifically, the molecular mass (~500–650 Da) and negative charge of phosphate or sulfate groups present in nucleotide sugars mean that these substrates cannot diffuse through membranes. Instead, translocation of these substrates depends on membrane proteins known as nucleotide sugar transporters (NSTs). Proper glycoconjugate synthesis and the addition of sulfate groups onto proteins, lipids, and polysaccharides depend on these transporters, and any alteration in the function of a given NST will disrupt normal organism functioning. In humans, disruptions can lead to varying pathological conditions \[1-4\]. In plant species, disruptions can alter the formation of the cell wall, which acts as an exoskeleton and contains sugars useful as lignocellulosic biofuels \[5\].

The first biochemical characterization of NSTs revealed these to be saturable in a \(K_m\) range of 1–10 \(\mu\)M, and additional evidence suggests that different transporters import distinct nucleotide sugars \[6,7\]. Biochemical analyses further showed that nucleotide sugar import concomitantly occurs with the export of a nucleoside monophosphate. Specifically, nucleoside diphosphate is produced upon the transfer of the nucleotide sugar into the acceptor. This diphosphate is then cleaved by a nucleoside diphosphatase into nucleoside monophosphate + \(P_i\) \[7,8\].

Biochemical features of the NSTs were made clearer through the reconstitution of these transporters into liposomes, an experimental approach that led to the development of a purification assay. Subsequent assays confirmed results obtained with Golgi apparatus-derived vesicles, including substrate specificities and that NSTs are low abundance proteins, making the identification of transporter-encoding genes through protein purification and sequencing a difficult task \[9–12\]. Genes encoding for NSTs were first cloned by complementing mutant sequencing a difficult task \[9–12\]. Genes encoding for NSTs were first cloned by complementing mutant sequence homology \[26,27\]. Indeed, sequence comparisons and assessments of NST-related sequences led to the identification of the human gene coding for the PAPS transporter \[28\], indicating that nucleotide sugars and nucleotide sulfate are transported by the same class of polytopic proteins. Analyses of the protein sequences predicted from the different cloned genes revealed common features among all NSTs. These transporters possess 6–10 transmembrane domains, range from 45 to 55 kDa in size, and form homodimers, the possible functional unit, in the Golgi apparatus \[10–12\].

As more NST sequences become available, it becomes possible to examine the primary sequence similarities of NSTs that share substrate specificity. This information is highly relevant when considering that numerous developing genome projects are releasing inaccurate annotations of new genes coding for NSTs. Initial analyses of the UDP-N-acetylgalactosamine transporters from \textit{Kluyveromyces lactis} and mammals showed that these two transporters were less similar than a group of three NSTs fromin mammals that transport UDP-N-acetylgalactosamine, UDP-galactose, and CMP-SA \[29\]. Therefore, substrate specificity predictions based on the primary sequence of an NST might be incorrect \[30\].

Functional characterizations of NSTs in different species have aided in clarifying the connection between the primary sequence and substrate specificity. In particular, Handford \textit{et al.} \[31\] performed a phylogenetic analysis of 26 NST sequences from different species and suggested a relationship between substrate specificity and the primary sequence. However, expanding on studies such as the one presented by Handford \textit{et al.} \[31\] has been difficult due to the limited number of characterized NSTs and limited availability of various nucleotide sugars.

**Functional assessment of cloned NSTs**

Functionally characterizing cloned NSTs is challenging since the biochemistry of membrane proteins is more complex than that of soluble proteins. Nevertheless, the expression of NST genes in heterologous systems, such as in isolated vesicles of the yeast \textit{Saccharomyces cerevisiae}, permits activity assessments. An advantage of a yeast-based system is that it only transports GDP-mannose and UDP-glucose into the ER and Golgi apparatus, therefore facilitating the detection of transport signals from other substrates. Despite this advantage, a potential drawback of this system is a likely restriction of nucleoside monophosphate, one of the substrates involved in the nucleotide sugar/nucleoside monophosphate exchange. This restriction can result from some nucleotide sugars being unused by the yeast. Consequently, the sugar moieties needed for the release and formation of nucleoside monophosphate would not be released. However, the endogenous transport of UDP-glucose and GDP-mannose suggests the likely existence of UMP and GMP pools in the Golgi lumen, thus allowing for the functional characterization of NSTs in yeast vesicles.

The study of NSTs via heterologous systems has revealed that substrate specificity varies from a small to larger number of nucleotide sugars \[21,24,25,32,33\]. Although informative, the reliance of transport assays on commercially limited, radiolabeled nucleotide sugars curbs the number of testable substrates. This problem is particularly relevant for species with a large

\[3151\]
repertoire of nucleotide sugars, which are usually not commercially available even in unlabeled forms.

An alternative method for measuring NST activity was recently described by Rautengarten et al. [34]. Specifically, activity was measured by heterologously expressing an NST gene in yeast, extracting the corresponding protein from the membrane, and incorporating this protein into liposomes loaded with nucleoside monophosphate. These proteoliposomes were then incubated with a large set of non-radiolabeled nucleotide sugars. The nucleotide sugars incorporated into the liposomes were detected by the combined use of high performance liquid chromatography and mass spectrometry. This non-radiolabeled procedure has already resulted in evaluations of 13 nucleotide sugars and facilitated broad substrate specificity assessments [34–36]. Several NSTs from the model plant Arabidopsis thaliana have been functionally characterized using this method, constructing a phylogenetic tree with six clades [34]. Although a recent development, non-radiolabeled analyses of NSTs are already evidencing associations between substrate specificity and genes within a clade. This knowledge could be used to provide more information on the clustering of genes coding for transporter clusters such as UDP-galactose/UDP-rhamnose, as well as for UDP-xylene or GDP-sugars [34–36].

## Linking the function of NSTs across species

Given the results from the NST characterizations in A. thaliana and of gene clusters exhibiting similar substrate specificities, we hypothesized that a correlation existed between the structure and function of NSTs and PAPS transporters from different species. To examine this postulation, genome databases were searched for model organisms such as Saccharomyces cerevisiae, Caenorhabditis elegans, Drosophila melanogaster, Danio rerio, Mus musculus, and Rattus norvegicus. The plants A. thaliana and Oryza sativa, as well as humans and protists, were also included.

The genes encoding for NSTs in humans are members of solute carrier family (SLC) 35, which has seven identified subfamilies [4], although only six have an associated function. Genes coding for NSTs have also been identified and, in some cases, functionally characterized in S. cerevisiae, D. melanogaster, C. elegans, and M. musculus. Among the protist species in which NSTs have been identified and characterized are Leishmania major [37,38], Giardia lamblia [39], Entamoeba histolytica [39–41], Trypanosoma brucei [42], Trypanosoma cruzi [43], and Toxoplasma gondii [44]. In turn, fungi with identified NSTs include Aspergillus fumigatus [45], Aspergillus nidulans [46], and Aspergillus niger [47]. Plants and algae produce significant amounts of polysaccharides in the Golgi apparatus, and several NSTs have been identified in Arabidopsis, rice, Vitis vinifera, and Volvox carteri [48,49]. In the model plant A. thaliana, there are more than 40 genes coding for nucleotide sugars [34]. Interestingly, one of the clades present in the Arabidopsis phylogenetic tree contains the triose phosphate translocator (TPT) gene family. These genes are responsible for mobilizing sugar precursors produced in plastids during photosynthesis and have been associated with the NST gene family [50].

Due to the increase in available data regarding functionally characterized NSTs (Table 1), this study was able to reexamine possible relationships between the primary sequence and substrate specificity. For this, information on 257 proteins identified as putative NSTs was collected (Table S1). Of these, 72 were already functionally evaluated (Table 1). A phylogenetic tree was constructed containing the NST coding genes of the aforementioned 257 proteins, which corresponded to 20 species (Fig. 1). Ten clades (A–J) were identified, and eight of these had at least one gene with a known function (Table 1).

Some SLC35 subfamily and A. thaliana phylogenetic tree clade [34] genes were separated into different clades by the present analysis, indicating that a new phylogenetic organization occurred when sequences from the 20 species were assembled in the tree. Every clade contained animal sequences; seven contained protist sequences, and six contained plant sequences. Yeast sequences were only found in four clades (Table 2).

To better understand the relationship between substrate specificity and the primary sequence, analysis of the phylogenetic tree focused on the distribution of the different sequences based on the reported use of the different classes of nucleotide sugars and nucleotide sulfate (UDP-sugars, GDP-sugars, CMP-SA, and PAPS). Since thiamine was recently reported as a substrate of the human SLC35F3 gene and since the TPT gene family is part of the NST gene family in plants, these were also considered.

### UDP-sugars

UDP-derived nucleotide sugars are the main substrates utilized during the biosynthesis of glycoconjugates or polysaccharides in eukaryotes, processes that primarily occur in the Golgi lumen. Other important reactions take place in the ER lumen, such as the UDP-glucose:glycoprotein glucosyltransferase-mediated glycosylation of unfolded N-glycoproteins containing an N-oligosaccharide composed of GlcNAc3Man9 [87]. This reaction is part of a quality control process that glycoproteins undergo in the ER, necessitating the transportation of UDP-glucose into the organelle.
<table>
<thead>
<tr>
<th>Uniprot accession number</th>
<th>Description (OS = organism species, GN = gene name)</th>
<th>Transport activity</th>
<th>References</th>
<th>Clade</th>
</tr>
</thead>
</table>
| P78382                   | CMP-SA transporter
(OS = Homo sapiens, GN = SLC35A1)                                                                   | CMP-SA             | [51]       | A     |
| Q61420                   | CMP-SA transporter
(OS = Mus musculus, GN = Slc35a1)                                                                    | CMP-SA             | [33]       | A     |
| P78381                   | UDP-galactose translocator
(OS = Homo sapiens, GN = SLC35A2)                                                                  | UDP-GalNAc, UDP-Gal | [52]       | A     |
| Q9Y2D2                   | UDP-N-acetylglucosamine transporter
(OS = Homo sapiens, GN = SLC35A3)                                                                 | UDP-GlcNAc         | [53]       | A     |
| Q9W4W6                   | CMP-SA/UDP-galactose transporter
(OS = Drosophila melanogaster, GN = Csat)                                                                   | UDP-GalNAc, UDP-Gal | [52,54]   | A     |
| Q9VMU8                   | CG14040
(OS = Drosophila melanogaster, GN = CG14040)                                                            | UDP-Gal            | [55]       | A     |
| O16658                   | NST family
(OS = Caenorhabditis elegans, GN = nstp-4)                                                            | UDP-GlcNAc         | [30]       | A     |
| Q93890                   | UDP-galactose/UDP-N-acetylglucosamine transporter
srf-3
(OS = Caenorhabditis elegans, GN = srf-3)                                                                 | UDP-GlcNAc, UDP-Gal | [56]       | A     |
| O02345                   | NST family
(OS = Caenorhabditis elegans, GN = nstp-5)                                                            | UDP-GlcNAc, UDP-Gal, UDP-Glc | [57]       | A     |
| Q8LGE9                   | CMP-SA transporter_1
(OS = Arabidopsis thaliana, GN = At5g41760)                                                           | CMP-SA             | [58,59]    | A     |
| Q654D9                   | CMP-SA transporter 1
(OS = Oryza sativa subp. Japonica, GN = CSTLP1)                                                        | CMP-SA             | [59]       | A     |
| Q8LES0                   | CMP-SA transporter 5
(OS = Arabidopsis thaliana, GN = At5g65000)                                                          | UDP-GlcNAc         | [60]       | A     |
| S8EVV0                   | UDP-galactose transporter subfamily protein
(OS = Toxoplasma gondii, GN = TGME49.267380)                                                               | UDP-GlcNAc, UDP-Gal | [44]       | A     |
| Q388K9                   | UDP-galactose transporter putative
(OS = Trypanosoma brucei (strain 927/4_GUTat10.1)
GN = Tb10.61.3010)                                                                                      | UDP-Gal, UDP-GlcNAc | [42]       | A     |
| Q381D0                   | CMP-SA transporter putative
(OS = Trypanosoma brucei (strain 927/4_GUTat10.1)
GN = Tb11.01.8280)                                                                                      | UDP-Gal, UDP-GlcNAc | [42]       | A     |
| C4LTQ4                   | UDP-N-acetylglucosamine transporter putative
(OS = Entamoeba histolytica
GN = EHI.050270)                                                                                           | UDP-Gal           | [40]       | A     |
| C4LZJ2                   | UDP-N-acetylglucosamine transporter putative
(OS = Entamoeba histolytica
GN = EHI.124580)                                                                                          | UDP-Glc           | [39]       | A     |
| Q8LY50                   | Putative thiamine transporter SLC35F3
(OS = Homo sapiens, GN = SLC35F3)                                                                       | Thiamine          | [61]       | B     |
| Q9M5A9                   | Glucose-6-phosphate/phosphate translocator 1 chloroplastic
(OS = Arabidopsis thaliana, GN = GPT1)                                                                       | Glucose 6-Phosphate, p phosphate | [62]       | C     |
| Q94B38                   | Glucose-6-phosphate/phosphate translocator 1 chloroplastic
(OS = Arabidopsis thaliana, GN = GPT2)                                                                       | Glucose 6-phosphate, p phosphate | [62]       | C     |
| Q8RXN3                   | Phosphoenolpyruvate/phosphate translocator 1 chloroplastic
(OS = Arabidopsis thaliana, GN = PPT1)                                                                       | Phosphoenolpyruvate, p phosphate | [63]       | C     |
| Q8H0T6                   | Phosphoenolpyruvate/phosphate translocator 2 chloroplastic
(OS = Arabidopsis thaliana, GN = PPT2)                                                                       | Phosphoenolpyruvate, p phosphate | [63]       | C     |

(continued on next page)
<table>
<thead>
<tr>
<th>Uniprot accession number</th>
<th>Description (OS = organism species, GN = gene name)</th>
<th>Transport activity</th>
<th>References</th>
<th>Clade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q9ZSR7</td>
<td>Triose phosphate/phosphate translocator TPT chloroplastic (OS = <em>Arabidopsis thaliana</em>, GN = TPT)</td>
<td>Triose Phosphate, phosphate</td>
<td>[64]</td>
<td>C</td>
</tr>
<tr>
<td>Q9LF61</td>
<td>Xylulose 5-phosphate/phosphate translocator (OS = <em>Arabidopsis thaliana</em>, GN = XPT)</td>
<td>Xylulose 5-phosphate, Phosphate</td>
<td>[65]</td>
<td>C</td>
</tr>
<tr>
<td>Q969S0</td>
<td>UDP-xylulose and UDP-N-acetylglucosamine transporter (OS = <em>Homo sapiens</em>, GN = SLC35B4)</td>
<td>UDP-Xyl, UDP-GlcNAc</td>
<td>[66]</td>
<td>E</td>
</tr>
<tr>
<td>Q9VAN9</td>
<td>CG14511</td>
<td>UDP-xylulose, UDP-GlcNAc, GDP-Fuc</td>
<td>[67]</td>
<td>E</td>
</tr>
<tr>
<td>P40004</td>
<td>UDP-N-acetylglucosamine transporter YEA4 (OS = <em>Saccharomyces cerevisiae</em> strain ATCC 205408/S288c, GN = YEA4)</td>
<td>UDP-GlcNAc</td>
<td>[68]</td>
<td>E</td>
</tr>
<tr>
<td>Q8TB61</td>
<td>Adenosine 3-phospho 5-phosphosulfate transporter 1 (OS = <em>Homo sapiens</em>, GN = SLC35B2)</td>
<td>PAPS</td>
<td>[28]</td>
<td>F</td>
</tr>
<tr>
<td>Q20787</td>
<td>Adenosine 3-phospho 5-phosphosulfate transporter 2 (OS = <em>Caenorhabditis elegans</em>, GN = pst-1)</td>
<td>PAPS</td>
<td>[69]</td>
<td>F</td>
</tr>
<tr>
<td>Q8MXJ9</td>
<td>Adenosine 3-phospho 5-phosphosulfate transporter 1 (OS = <em>Caenorhabditis elegans</em>, GN = pst-2)</td>
<td>PAPS</td>
<td>[69]</td>
<td>F</td>
</tr>
<tr>
<td>Q9VVD9</td>
<td>Adenosine 3-phospho 5-phosphosulfate transporter 2 (OS = <em>Drosophila melanogaster</em>, GN = Papst2)</td>
<td>PAPS</td>
<td>[70]</td>
<td>F</td>
</tr>
<tr>
<td>Q9VEI3</td>
<td>Adenosine 3-phospho 5-phosphosulfate transporter 1 (OS = <em>Drosophila melanogaster</em>, GN = sll)</td>
<td>PAPS</td>
<td>[28,71]</td>
<td>F</td>
</tr>
<tr>
<td>Q12520</td>
<td>UDP-galactose transporter homolog 1 (OS = <em>Saccharomyces cerevisiae</em> strain ATCC 205408/S288c, GN = HUT1)</td>
<td>UDP-Gal</td>
<td>[72]</td>
<td>F</td>
</tr>
<tr>
<td>Q64503</td>
<td>UDP-galactose/UDP-glucose transporter 1 (OS = <em>Arabidopsis thaliana</em>, GN = UTR1)</td>
<td>UDP-Glc</td>
<td>[27,73]</td>
<td>F</td>
</tr>
<tr>
<td>Q29Q28</td>
<td>UDP-galactose/UDP-glucose transporter 2 (OS = <em>Arabidopsis thaliana</em>, GN = UTR2)</td>
<td>UDP-Gal</td>
<td>[74]</td>
<td>F</td>
</tr>
<tr>
<td>Q9M9S6</td>
<td>UDP-galactose/UDP-glucose transporter 3 (OS = <em>Arabidopsis thaliana</em>, GN = UTR3)</td>
<td>UDP-Glc</td>
<td>[75]</td>
<td>F</td>
</tr>
<tr>
<td>Q69XD4</td>
<td>Oso06g0593100 protein (OS = <em>Oryza sativa</em> subsp. <em>Japonica</em>, GN = P0502H06.31)</td>
<td>UDP-Glc, UDP-Gal</td>
<td>[76]</td>
<td>F</td>
</tr>
<tr>
<td>Q91ZN5</td>
<td>Adenosine 3-phospho 5-phosphosulfate transporter 1 (OS = <em>Mus musculus</em>, GN = SLC35b2)</td>
<td>PAPS</td>
<td>[77]</td>
<td>F</td>
</tr>
<tr>
<td>Q92Q5</td>
<td>Adenosine 3-phospho 5-phosphosulfate transporter 2 (OS = <em>Mus musculus</em>, GN = SLC35b3)</td>
<td>PAPS</td>
<td>[77]</td>
<td>F</td>
</tr>
<tr>
<td>Q9H1N7</td>
<td>Adenosine 3-phospho 5-phosphosulfate transporter 2 (OS = <em>Homo sapiens</em>, GN = SLC35B3)</td>
<td>PAPS</td>
<td>[78]</td>
<td>F</td>
</tr>
<tr>
<td>B1N4Q4</td>
<td>UDP-galactose transporter-related protein 1 putative (OS = <em>Entamoeba histolytica</em>, GN = EHL.041050)</td>
<td>UDP-GlcNAc</td>
<td>[39]</td>
<td>F</td>
</tr>
<tr>
<td>V6TFL7</td>
<td>Transporter UAA family protein (OS = <em>Giardia intestinalis</em>, GN = DHA2.15483)</td>
<td>UDP-GlcNAc</td>
<td>[39]</td>
<td>F</td>
</tr>
<tr>
<td>Q76EJ3</td>
<td>UDP-N-acetylglucosamine/UDP-glucose/GDP-mannose transporter (OS = <em>Homo sapiens</em>, GN = SLC35D2)</td>
<td>UDP-GlcNAc</td>
<td>[80,81]</td>
<td>G</td>
</tr>
<tr>
<td>Q18779</td>
<td>UDP-sugar transporter sqv-7 (OS = <em>Caenorhabditis elegans</em>, GN = sqv-7)</td>
<td>UDP-Glc</td>
<td>[21]</td>
<td>G</td>
</tr>
<tr>
<td>Q94B65</td>
<td>Putative uncharacterized protein At4g31600 F28 M20.210</td>
<td>UDP-Gal, UDP-GlcNAc</td>
<td>[82]</td>
<td>G</td>
</tr>
<tr>
<td>Accession</td>
<td>Name</td>
<td>Function</td>
<td>Species</td>
<td>Gene Name</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>----------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>Q6K5Y4</td>
<td>Os02g0614100 protein</td>
<td></td>
<td>Arabidopsis thaliana, GN = F28 M20.210</td>
<td></td>
</tr>
<tr>
<td>Q9VHT4</td>
<td>GDP-fucose transporter 1</td>
<td></td>
<td>Oryza sativa subsp. Japonica, GN = OJ2055_H10.17</td>
<td></td>
</tr>
<tr>
<td>Q96A29</td>
<td>GDP-fucose transporter 1</td>
<td></td>
<td>Drosophila melanogaster, GN = nac</td>
<td></td>
</tr>
<tr>
<td>P40107</td>
<td>GDP-mannose transporter 1</td>
<td></td>
<td>Homo sapiens, GN = SLC35C1</td>
<td></td>
</tr>
<tr>
<td>Q941R4</td>
<td>GDP-mannose transporter GONST1</td>
<td></td>
<td>Arabidopsis thaliana, GN = GONST1</td>
<td></td>
</tr>
<tr>
<td>Q84L08</td>
<td>GDP-mannose transporter GONST4</td>
<td></td>
<td>Arabidopsis thaliana, GN = GONST4</td>
<td></td>
</tr>
<tr>
<td>D7U8E5</td>
<td>Putative uncharacterized protein</td>
<td></td>
<td>Vitis vinifera, GN = VIT.10s0092g00200</td>
<td></td>
</tr>
<tr>
<td>F6HH60</td>
<td>Putative uncharacterized protein</td>
<td></td>
<td>Vitis vinifera, GN = VIT.11s0016g04650</td>
<td></td>
</tr>
<tr>
<td>Q581R7</td>
<td>Lipophosphoglycan biosynthetic protein 2 putative</td>
<td></td>
<td>Trypanosoma brucei brucei (strain 927/4 GUTat10.1)</td>
<td></td>
</tr>
<tr>
<td>D8TXZ2</td>
<td>Putative uncharacterized protein invB</td>
<td></td>
<td>Volvox carteri, GN = invB</td>
<td></td>
</tr>
<tr>
<td>Q8IR6</td>
<td>Lipophosphoglycan biosynthetic protein</td>
<td></td>
<td>Leishmania major, GN = LPG2</td>
<td></td>
</tr>
<tr>
<td>Q69V70</td>
<td>Oms7g0581000 protein</td>
<td></td>
<td>Oryza sativa subsp. Japonica, GN = P0453G03.28</td>
<td></td>
</tr>
<tr>
<td>F4HS9</td>
<td>Nucleotide/sugar_transporter-like protein</td>
<td></td>
<td>Arabidopsis thaliana, GN = UXT1</td>
<td></td>
</tr>
<tr>
<td>Q5RXL8</td>
<td>Uncharacterized membrane protein At1g06890</td>
<td></td>
<td>Arabidopsis thaliana, GN = P0453G03.28</td>
<td></td>
</tr>
</tbody>
</table>

(continued on next page)
<table>
<thead>
<tr>
<th>Uniprot accession number</th>
<th>Description (OS = organism species, GN = gene name)</th>
<th>Transport activity</th>
<th>References</th>
<th>Clade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q8GUJ1</td>
<td>At2g30460 (OS = Arabidopsis thaliana, GN = UXT2)</td>
<td>UDP-Xyl</td>
<td>[35]</td>
<td>I</td>
</tr>
<tr>
<td>Q9LPU2</td>
<td>At1g21070 (OS = Arabidopsis thaliana, GN = T22111.10)</td>
<td>UDP-Rhm, UDP-Gal</td>
<td>[34]</td>
<td>I</td>
</tr>
<tr>
<td>Q9FIH5</td>
<td>At5g42420 (OS = Arabidopsis thaliana, GN = UXT2)</td>
<td>UDP-Rhm</td>
<td>[34]</td>
<td>I</td>
</tr>
<tr>
<td>Q8RW7</td>
<td>Bi-functional UDP-rhamnose/UDP-galactose transporter (OS = Arabidopsis thaliana, GN = URGT3)</td>
<td>UDP-Gal</td>
<td>[34]</td>
<td>I</td>
</tr>
<tr>
<td>Q9SZ96</td>
<td>At4g09810 (OS = Arabidopsis thaliana, GN = F17 A8.160)</td>
<td>UDP-Rhm, UDP-Gal</td>
<td>[34]</td>
<td>I</td>
</tr>
<tr>
<td>Q9FDZ5</td>
<td>At1g34020 (OS = Arabidopsis thaliana, GN = T15 K4.8)</td>
<td>UDP-Rhm, UDP-Gal</td>
<td>[34]</td>
<td>I</td>
</tr>
<tr>
<td>Q9SRE4</td>
<td>UDP-galactose transporter 2 (OS = Arabidopsis thaliana, GN = UDP-GALT2)</td>
<td>UDP-Gal</td>
<td>[34]</td>
<td>I</td>
</tr>
<tr>
<td>Q6ZJN0</td>
<td>Os08g0104900 protein (OS = Oryza sativa subsp. Japonica, GN = OJ1300_E01.7)</td>
<td>UDP-Glc, UDP-Gal</td>
<td>[76]</td>
<td>I</td>
</tr>
<tr>
<td>Q58S2Z4</td>
<td>Uncharacterized protein (OS = Trypanosoma brucei brucei (strain 927/4 GUTat10.1)</td>
<td>UDP-GlcNAc, UDP-GalNAc, GDP-Man</td>
<td>[42]</td>
<td>I</td>
</tr>
<tr>
<td>C5I6M0</td>
<td>UDP-galf transporter (OS = Neosartorya fumigata, GN = gflB)</td>
<td>UDP-Gal</td>
<td>[45]</td>
<td>I</td>
</tr>
<tr>
<td>Q7XKA0</td>
<td>OSJNBb0020J19.10 protein (OS = Oryza sativa subsp. Japonica, GN = OSJNBb0020J19.10)</td>
<td>UDP-Gal, UDP-Glc</td>
<td>[76]</td>
<td>I</td>
</tr>
</tbody>
</table>

* The transport of UDP-Galf was indirectly determined by measuring the incorporation of $[^{3}H]$UMP into Golgi apparatus vesicles isolated from yeast (S. cerevisiae) expressing gflB
Among different fungi, animal, and plant species, UDP-sugars are used in numerous reactions that lead to the formation of glycoproteins, glycolipids, and polysaccharides in the Golgi apparatus. Briefly, UDP-galactose and UDP-N-acetylglucosamine are essential to the remodeling of the N-linked oligosaccharide present in glycoproteins while these traverse through the Golgi apparatus [88]; UDP-N-acetylgalactosamine, UDP-glucuronic acid, and UDP-glucuronic acid are required for the synthesis of proteoglycans in animal cells [89]; and, in plants, UDP-galacturonic acid, UDP-xylene, UDP-arabinose, UDP-rhamnose, and UDP-apiose are fundamental to the biosynthesis of polysaccharides that compose the cell wall [90]. Currently, evidence exists for the transport of most UDP-sugars, with the exceptions of UDP-galacturonic acid, UDP-arabinose, and UDP-apiose.

Studies in yeast (e.g., S. cerevisiae and K. lactis) have shown that these species are capable of transporting UDP-glucose, UDP-galactose, and UDP-N-acetylglucosamine in the ER and Golgi apparatus [13,68,72,91]. Further functional analyses of NSTs have demonstrated that protists also transport these three nucleotide sugars [44]. It is likely that the first UDP-sugar transporters to emerge during evolution were those capable of transporting these three substrates.

The present phylogenetic analysis showed that five clades (A, E, F, G, and I) contained NSTs that transport UDP-sugars. The NSTs within all five clades presented UDP-glucose, UDP-galactose, and

![Phylogenetic tree of the SLC35/NST-TPT gene superfamily from model species.](image)
UDP-N-acetylglucosamine activities, which could indicate that transporters for these nucleotide sugars evolved to create NSTs for the remaining UDP-sugars. Analysis of substrate specificity for the NSTs present in these clades also revealed that some could transport two or more UDP-sugars. However, many NSTs retained the ability to transport either UDP-glucose, UDP-galactose, or UDP-N-acetylglucosamine, in addition to one or more other types of UDP-sugars. For example, some NSTs were found able to transport UDP-galactose and UDP-ribose, or UDP-N-acetylglucosamine and UDP-xylitol, or UDP-N-acetylglucosamine and UDP-N-acetylgalactosamine, or UDP-galactose and UDP-N-acetylgalactosamine and UDP-glucuronic acid. Gaining further knowledge on the substrate specificity of some transporters that accept multiple UDP-sugars is difficult due to differences in the number of nucleotide sugars utilized in activity assessments [24,25,52,54]. Recently, the functional characterization of NSTs was reevaluated using a larger number of nucleotide sugars [34]. Application of this information should result in better descriptions of NST substrate specificities.

**GDP-mannose/GDP-fucose**

Genes coding for NSTs involved in the transport of GDP-sugars were found exclusively in clade H; however, two branches were distinguished. One branch contained genes exclusively coding for GDP-fucose and GDP-mannose NSTs from plants and yeast. The other branch only contained genes from animal species. Yeast and plants use GDP-mannose for polysaccharide and glycolipid biosynthesis in the Golgi apparatus, and GDP-fucose is utilized for polysaccharide biosynthesis as well as protein fucosylation in plants. In contrast, there is no evidence for mannosylation reactions in the Golgi apparatus of animals, and, therefore, no need for GDP-mannose transport. This suggests that the GDP-mannose transporters, which are present in yeast and plants, could have evolved separately. The animal NSTs found in clade H are capable of transporting GDP-fucose, but whether these transporters are capable of transporting other GDP-sugars is a subject for future investigation.

**CMP-sialic acid**

While sialic acid is present in the glycoproteins and lipids of animals, it appears to be absent in yeasts and plants, likely due to a late evolutionary appearance. The substrate for all sialylation reactions is CMP-SA, a nucleotide sugar synthesized in the nucleus [92] that then migrates to the cytosol. CMP-SA incorporation into the Golgi apparatus occurs via a transporter, which, coincidentally, was one of the first to be biochemically characterized [93]. Cloning of a gene coding for a CMP-SA transporter was first achieved in mice by

<p>| Table 2. Information regarding the 10 clades formed in the phylogenetic tree by the assessed 257 protein sequences |</p>
<table>
<thead>
<tr>
<th>Clade</th>
<th>SLC35 component</th>
<th>Arabidopsis clade</th>
<th>Substrates</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>SLC35A1–5</td>
<td>III</td>
<td>CMP-SA, CMP-KDO?, UDP-Gal, UDP-GalNAc, UDP-GlcNAc</td>
<td>Hs, Rn, Mm, Dr, Dm, Ce, Os, At, Vv, Tc, Tb, Tg, Lm, Ld, Eh</td>
</tr>
<tr>
<td>B</td>
<td>SLC35F3–5</td>
<td>—</td>
<td>Thiamine</td>
<td>Hs, Rn, Mm, Dr, Ce, Tb</td>
</tr>
<tr>
<td>C</td>
<td>SLC35E4</td>
<td>IV</td>
<td>Triose phosphate/phosphate, Phosphoenolpyruvate/Phosphate, Glucose-6-phosphate/phosphate</td>
<td>Hs, Rn, Mm, Dr, Os, At</td>
</tr>
<tr>
<td>D</td>
<td>SLC35E2</td>
<td>—</td>
<td>N.D.</td>
<td>Hs, Rn, Mm</td>
</tr>
<tr>
<td>E</td>
<td>SLC35B4</td>
<td>II</td>
<td>UDP-Xyl, UDP-GlcNAc</td>
<td>Hs, Rn, Mm, Dr, Dm, Ce, Sc, Tc, Tb</td>
</tr>
<tr>
<td>F</td>
<td>SLC35B1–3</td>
<td>II</td>
<td>PAPS, UDP-Gal, UDP-Glc</td>
<td>Hs, Rn, Mm, Dr, Dm, Ce, Sc, At, Os, Tc, Tb, Lm, Eh, Gi</td>
</tr>
<tr>
<td>G</td>
<td>SLC35D1–3</td>
<td>III</td>
<td>UDP-Gal, UDP-Glc, UDP-GlcA, UDP-GlcNAc, UDP-GalNAc</td>
<td>Hs, Rn, Mn, Dr, Dm, Ce, At, Os, Tc</td>
</tr>
<tr>
<td>H</td>
<td>SLC35C1</td>
<td>III</td>
<td>GDP-Fuc, GDP-Man</td>
<td>Hs, Rn, Mm, Dr, Dm, Ce, Sc, Os, At, Vv, Vc, Tc, Tb, Lm</td>
</tr>
<tr>
<td>I</td>
<td>SLC35C2, SLC35E3, SLC35E4</td>
<td>I, V, VI</td>
<td>UDP-Gal, UDP-Glc, UDP-Rhm, UDP-Xyl, UDP-Gal</td>
<td>Hs, Rn, Mm, Dr, Ce, Sc, At, Os, Tc, Tb, Lm, Af, Anig, Anid</td>
</tr>
<tr>
<td>J</td>
<td>SLC35F1–2, SLC35F6</td>
<td>—</td>
<td>N.D.</td>
<td>Hs, Rn, Mm, Dr</td>
</tr>
</tbody>
</table>

“SLC35 component” refers to the family and subfamilies [4]. “Arabidopsis clade” refers to the clades identified in the Arabidopsis phylogenetic tree of the previously described NST/PTP family [34]. At: Aspergillus fumigatus; Anid: Aspergillus nidulans; Anig: Aspergillus niger; At: Arabidopsis thaliana; Ce: Caenorhabditis elegans; Dr: Drosophila melanogaster; Dr: Drosophila reici; Eh: Entamoeba histolytica; Gi: Giardia intestinalis; Hs: Homo sapiens; Ld: Leishmania donovani; Lm: Leishmania major; Mm: Mus musculus; Os: Oryza sativa; Rn: Rattus norvegicus; Sc: Saccharomyces cerevisiae; Tb: Trypanosoma brucei; Tc: Trypanosoma cruzi; Tg: Toxoplasma gondii; Vc: Volvox carteri; Vv: Vitis vinifera.
complementing a mutant Chinese hamster ovary cell line deficient in CMP-SA transporter activity [15]. This gene was then functionally characterized in yeast [33] and showed no activity toward UDP-galactose. Similar features were described for the human transporter, although no other nucleotide sugars were utilized in the assay.

CMP-SA transporters from humans and mice were found in clade A (Fig. 1). Clade A also contained NSTs from plants that are capable of transporting CMP-SA in vitro, even though they do not contain CMP-SA. Bakker et al. [58] hypothesized that since plants require CMP-KDO, a nucleotide sugar related to CMP-SA, for the biosynthesis of the pectin rhamnogalacturonan II, these orthologous genes may be involved in nucleotide sugar transport.

Clade A also contained NSTs from animals, plants, and protists that transport UDP-sugars, such as UDP-galactose, UDP-N-acetylgalactosamine, UDP-N-acetylglucosamine, and UDP-glucose, suggesting that CMP-SA transporters and their plant equivalents may have evolved from this branch.

3′-Phosphoadenosine 5′-phosphosulfate

PAPS is the donor molecule for all sulfate reactions in the Golgi apparatus and is transported from the cytosol into this organelle [94]. The identification of the first gene encoding for a PAPS transporter determined that NST gene family members are responsible for this activity. Genes coding for PAPS transporters have been identified and functionally characterized in humans, Drosophila, and C. elegans [28,69,71]. These genes were located in clade F (Fig. 1), a clade that also contained yeast, protist, and plant NSTs that transport UDP-glucose, UDP-galactose, and UDP-GlcNAc [27,72,74,75].

In animals, PAPS is the donor for the sulfation of proteins and proteoglycans. In contrast, yeast and plants have no sulfate in extracellular matrix polysaccharides, but there is evidence that sulfated peptides play a role in plant root development and the immune response [95,96]. In plants, the sulfation of these peptides is catalyzed by a sulfotransferase in the Golgi apparatus [97], necessitating the transport of PAPS into this organelle.

On the other hand, plant UDP-glucose/UDP-galactose transporters (UTr1 and UTr3) found in clade F are located in the ER and play a role in the unfolded protein response [73,75]. UDP-glucose:glycoprotein glucosyltransferase is a key enzyme in ER lumen quality control. This enzyme was described in Arabidopsis [98] and is an important component of the ER stress response in plants. Therefore, the import of UDP-glucose by UTr1 and UTr3 is consistent with a physiological role of these NSTs in ER homeostasis. Also present in clade F was a related C. elegans NST (CeHUT1) in the ER that is required to maintain ER homeostasis [99]. HUT1 from S. cerevisiae and Schizosaccharomyces pombe, grouped in clade F, also plays a role in the ER protein folding process [100], information that supports the importance of these ER-located NSTs in preserving homeostasis.

While additional research is needed to determine if UTr1 and UTr3 are able to also transport PAPS, in vitro analysis of PAPS transporters demonstrated no activity toward UDP-glucose when this was evaluated as a substrate. Although published data indicate that PAPST2 transporter from humans shows a low transport signal for UDP-glucose, it is still about twice that of the control for UDP-galactose [78]. Further experiments are necessary to confirm whether or not these clade F NSTs are capable of transporting UDP-galactose and/or UDP-glucose together with PAPS.

From an evolutionary perspective, sulfation is a process that emerged after the formation of glycoconjugates or polysaccharides. Given the proximity of genes encoding for UDP-glucose/UDP-galactose transporters, it is likely that PAPS transporters evolved from these NSTs.

Triose phosphates

The NST gene family in plants is closely related to the TPT gene family, a group of transporters that plays an important role during photosynthesis [50]. Clade C contained these plant genes, but this clade also contained human, rat, mouse, zebra fish, and Drosophila sequences from the SLC35E1 family. This observation indicates that these proteins may play different roles in animals and plants. However, additional analyses are needed to fully understand the function and physiological roles of these proteins in mammals.

Thiamine

Recently, the SLC35F3 gene was shown to code for a thiamine transporter [61]. This gene was found present in clade B and contained several sequence clusters from animals as well as T. cruzi. This group did not contain any plant or yeast sequences, suggesting a later evolution. Further characterization is needed to correlate this function with the NST gene family.

Relationship between the transport activity, glycan structure, and the biological role of NSTs

The discovery of mutations on genes coding for NSTs allowed establishing a correlation between the in vitro activity, the biological role, and the
connection between the transport activity and the glycan or polysaccharide that is/are affected in vivo by the impairment of a NST. Liu and Hirschberg [3] and Song [4] provided a comprehensive review on this topic; however, many aspects remain to be studied to get a better understanding of the biological role of NSTs in different species. Recent evidence related to this topic is described below:

SLC35A3 codes for a UDP-N-acetylg glucosamine transporter (Fig. 1, clade A) and a mutation causes autism spectrum disorder, epilepsy and arthrogryposis [101]. Studies in cell deficients in SLC35A3 showed defects in the synthesis of highly branched N-linked glycans in CHO and HeLa cells, whereas in MDCK cells, the content of keratan sulfate was reduced while no change was observed in heparan sulfate [102]; thus, a direct relationship between the activity of SLC35A3 and the glycans and polysaccharides that contain GlcNAc has been established.

An interesting observation came from the study of NSTs in plants, where nucleotide sugars are mostly used in the Golgi for the biosynthesis of cell wall polysaccharides (hemicellulosics and pectin). In A. thaliana, knock-out plants on URTG1 and URTG2 (UDP-galactose/UDP-rhamnose transporters) (Fig. 1, clade I) show some interesting phenotypes [34]. While in vitro both transport UDP-galactose and UDP-rhamnose with similar kinetics, the mutant on URTG1 exhibit less galactose in the cell wall from leaves whereas the levels of rhamnose were normal. Furthermore, the overexpression of URTG1 led to increased levels of galactose, likely in galactan, but no difference was observed in the levels of rhamnose. In contrast, mutant plants in URTG2 showed lower levels of rhamnose and galacturonic acid in seed mucilage, likely due to a decrease in rhamnogalacturonan I. Furthermore, no changes were observed in the content of galactose. URTG1 and URTG2 are paralogues with high levels of identity but different levels of expression; then, it is noteworthy that the alteration on each of them produces different consequences on the structure of polysaccharides. This suggests that their activity in vivo is regulated, such that depending when and where they are expressed, the consequences on the structure of polysaccharides formed will be different.

Another interesting example is related to the transport of UDP-xylene. Three paralogue genes (UXT1–3), capable of transporting this nucleotide sugar in vitro were recently characterized [35]. Mutants in UXT2 and UXT3 showed no phenotype and no change in the structure of glycans or polysaccharides were observed, suggesting that functional redundancy might exist. Interestingly, two allelic mutant lines in UXT1 exhibited a lower content of xylose in xylan, the most abundant hemicellulose in the secondary plant cell wall. Interestingly, no change was observed on the structure of xyloglucan, another abundant hemicellulose that contains significant amounts of xylose. Once again, it seems that UDP-xylose transported by UXT1 was mainly targeted to one type of polysaccharide. This finding provides additional support for the hypothesis that channeling of nucleotide sugar might occur depending on the presence of other partners involved in the biosynthesis of certain glycans or polysaccharides. These results suggest that NSTs are subjected to regulatory mechanisms, which determine their role in the synthesis of glycans and polysaccharides in vivo.

**Summary and perspectives**

This study collected data from 257 NST-predicted protein sequences, 72 of which corresponded to proteins with reported transport activity. These data were used to build a phylogenetic tree, which clustered all sequences within 10 clades. Some primary conclusions of tree analysis included the following: (1) CMP-SA was exclusively transported by proteins present in clade A. (2) GDP-sugars were transported by proteins present in clade H, which had two branches. One branch contained only animal sequences while the other branch contained yeast and plant sequences. (3) Transporters for UDP-sugars were spread among clades A, E, F, G, and I. UDP-glucose, UDP-galactose, and UDP-N-acetylg glucosamine were the predominant substrates, and the phylogenetic distribution of the transporters for these three sugars suggests that they may have evolved to create transporters for other UDP-sugars and CMP-SA. (4) Golgi-localized PAPS transporters were clustered in clade F, the same clade as UDP-glucose/UDP-galactose transporters located in the ER, and are likely involved in organelle homeostasis. This indicates that PAPS transporters also evolved from UDP-glucose/UDP-galactose transporters.

The presented phylogenetic tree could be a useful tool for identifying the potential roles of new NSTs in other species. Further assessments of uncharacterized NSTs, as well as reevaluations of some already characterized NSTs using a larger set of substrates, might improve understandings on the relationship between a primary NST sequence and substrate specificity.

An increasing number of studies are focused on determining substrate specificity for any given NST. This interest has already led to the development of transport assays for 13 different nucleotide sugars, thus accelerating the functional characterization of corresponding proteins. These advancements are enhancing the available knowledge of in vitro protein activities and, consequently, understandings of the in vivo physiological role played by these proteins. It is likely that more human pathologies associated with mutations in genes coding for NSTs will emerge, and knowing the functions of the NSTs involved is critical for creating possible disease therapies.
From a biotechnological perspective, there are numerous glycoproteins of pharmaceutical interest produced in heterologous systems. The structure of the glycan is often important for the function, and expression systems sometimes cannot properly produce glycan. Engineering these expression systems by adding the enzymes required to produce the correct glycan structure, such as glycosyltransferases or nucleotide sugar synthases, among others, could solve this issue. Likewise, the incorporation of NSTs with the correct substrate specificity should improve the production of glycoproteins in recombinant systems. Furthermore, using plants as a source for biofuels may require modifications of the recombinant systems. Furthermore, using plants as a source for biofuels may require modifications of the plant cell wall to alter the supply of substrates into the Golgi apparatus. In this context, understanding the function of NSTs in these and other processes is highly relevant. Comparing the sequences and relationships of these sequences to substrate specificity across species is a first step forward in predicting the function of many uncharacterized transporters.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2016.05.021.

Acknowledgments

Thanks to the Orellana’s laboratory for helpful discussions. Thanks to Carlos B. Hirschberg for the critical reading of this manuscript. This work has been supported by grants Fondecyt 1151335, Fondap-CRG 15090007, and Basal PB-16.

References


