The EMBO Journal vol.15 no.4 pp.705–713, 1996

A new stress protein: synthesis of *Schizosaccharomyces pombe* UDP-Glc:glycoprotein glucosyltransferase mRNA is induced by stress conditions but the enzyme is not essential for cell viability

Fabiana Fernández[1], Mehrdad Jannatipour[2], Ulf Hellman[3], Luis A.Rokeach[2] and Armando J.Parodi[1,4]

[1]Instituto de Investigaciones Bioquímicas Fundación Campomar, Antonio Machado 151, 1405 Buenos Aires, Argentina, [2]Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada and [3]Ludwig Institut for Cancer Research, Uppsala Branch, Box 595, S-751 24 Uppsala, Sweden

[4]Corresponding author

We have identified and begun the characterization of the gene encoding UDP-Glc:glycoprotein glucosyltransferase in *Schizosaccharomyces pombe*. This gene, here designated *gpt1*, codes for a polypeptide having a signal peptide of 18 amino acids followed by 1429 amino acids with no transmembrane domain, as expected for a soluble protein of the endoplasmic reticulum (ER). The C-terminal tetrapeptide PDEL most probably corresponds to a novel ER retention signal in this fission yeast. Synthesis of the corresponding mRNA was induced 2- to 9-fold by conditions known to affect glycoprotein folding in the ER (e.g. heat shock, culture in the presence of a Ca^{2+} ionophore, 2-mercaptoethanol or inhibitors of protein *N*-glycosylation such as tunicamycin or 2-deoxyglucose). This is the first evidence obtained *in vivo* that supports the proposed involvement of the enzyme in the quality control of glycoprotein folding in the ER. Thus far, the said involvement was inferred solely from the ability of the enzyme to glucosylate misfolded but not native glycoproteins in cell-free assays. The *gpt1* gene was disrupted and *gpt1*$^-$ cells were found to be viable. Moreover, no significant differences in the growth rate patterns at 18, 28 or 39°C or in cell morphology between *gpt1*$^+$ and *gpt1*$^-$ cells were observed, although they differed slightly in size.

Keywords: glycoprotein folding/glucosyltransferase/*Schizosaccharomyces pombe*/sequence

Introduction

The endoplasmic reticulum (ER) is the subcellular location where proteins entering the secretory pathway undergo many post-translational modifications such as removal of the signal peptide, *N*-glycosylation and coupling of protein structures to phosphatidylinostol-containing anchors. Initial steps in *N*-glycosylation have been known for some years, but their true biological significance has only begun to emerge recently. As was shown in 1972, *N*-glycosylation starts by transfer of an oligosaccharide ($Glc_3Man_9GlcNAc_2$ in most species) from a dolichol-P-P derivative to asparagine residues (Parodi *et al.*, 1972). As soon as it is transferred to the protein, processing of the oligosaccharide is initiated: the more external glucose unit is removed by glucosidase I, a membrane-bound α(1,2)glucosidase, whereas the remaining two glucose residues are excised by glucosidase II, an α(1,3)glucosidase only loosely attached to the inner membrane of the ER. Another processing reaction also occurring in the lumen of the ER is the removal of one or two α(1,2)-linked peripheral mannose residues (Kornfeld and Kornfeld, 1985).

Protein-linked, glucose-free oligosaccharides are then transiently re-glucosylated: $Glc_1Man_9GlcNAc_2$, $Glc_1Man_8GlcNAc_2$ and $Glc_1Man_7GlcNAc_2$ are formed from the unglucosylated species. The resulting compounds are immediately deglucosylated *in vivo* (Parodi *et al.*, 1983, 1984). The enzymes responsible for the so-called transient glucosylation of glycoproteins in the ER are the UDP-Glc:glycoprotein glucosyltransferase and glucosidase II (Trombetta *et al.*, 1989). Further processing of oligosaccharides may then occur in the Golgi apparatus (Kornfeld and Kornfeld, 1985).

Proteins entering the secretory pathway adopt their final tertiary (and in some cases also their quaternary) structures in the lumen of the ER. Although the information determining the tertiary structure of polypeptides is contained in their amino acid sequence, the *in vivo* folding of proteins is assisted by a battery of proteins designated chaperones. Chaperones preclude undesirable aggregation of folding proteins and prevent the newly synthesized proteins from following non-productive folding pathways. Proteins that fail to fold properly are retained in the ER (or in a ER–Golgi intermediate compartment), where they are proteolytically degraded (Lodish, 1988). This fact implies that a very stringent quality control of protein folding, able to discriminate between misfolded and properly folded species, operates in this subcellular location. A model for such quality control applicable to glycoproteins has been proposed recently (Sousa *et al.*, 1992; Hammond *et al.*, 1994; Helenius, 1994). According to this model, folding glycoproteins in the ER oscillate between monoglucosylated and unglucosylated states. Transition between these two oligosaccharide forms results from the opposite catalytic action of the UDP-Glc:glycoprotein glucosyltransferase and glucosidase II enzymes. Also playing a key role in the proposed mechanism is the membrane-bound chaperone calnexin. Due to its lectin-like activity selectively recognizing monoglucosylated oligosaccharides, calnexin is able to retain monoglucosylated polypeptides in the ER for as long as these are not properly folded (Ou *et al.*, 1993; Hammond *et al.*,1994). However, once they have attained their correct native conformations, glycoproteins become substrates for the glucosidase but not for the glucosyltransferase enzyme, which thus results

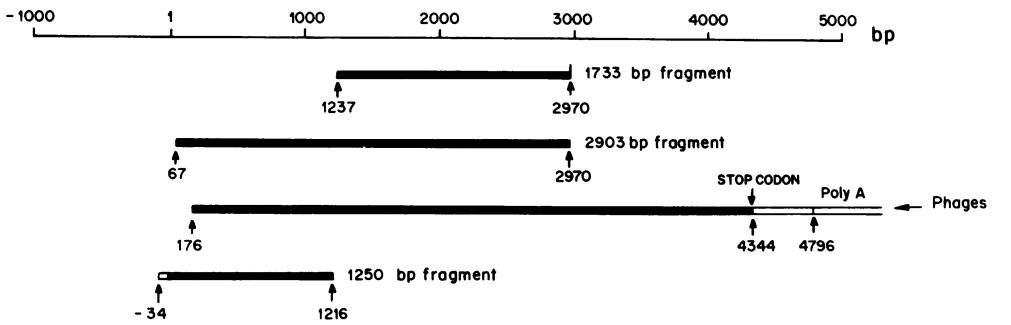

Fig. 1. Strategy used for cloning and sequencing *S.pombe gpt1* gene. Base +1 corresponds to the A in ATG of the Met initiator codon. Shaded areas correspond to the coding region. For details see Materials and methods and Results.

in their release from the calnexin anchor. Consequently, glycoproteins could then be amenable to be transported to the Golgi apparatus.

More recent reports on the interaction of calnexin with glycoproteins are somewhat contradictory. Zhang *et al.* (1995) showed that calnexin only binds class I major histocompatibility complex molecules that are monoglucosylated. Nevertheless, treatment of the glycoprotein–chaperone complex with endo-β-*N*-acetylglucosaminidase H resulted in the release of all oligosaccharides from the glycoprotein without dissociation of the complex. Similar observations were made by Ware *et al.* (1995) with respect to the association of calnexin with α1-antitrypsin. Both groups concluded that recognition of monoglucosylated oligosaccharide moieties by calnexin was only the first step in the formation of the calnexin–glycoprotein complex, whereas the oligosaccharide moieties would not play an important role after the onset of the interaction between calnexin and its ligand glycoproteins. On the other hand, Hebert *et al.* (1995), showed that addition of castanospermine (an inhibitor of glucosidase II) to cells precluded release of fully folded haemagglutinin molecules from calnexin, thus indicating that *in vivo* this chaperone remains bound to monoglucosylated oligosaccharides during the entire folding process.

A crucial element in the model of quality control of glycoprotein folding is the enzyme UDP-Glc:glycoprotein glucosyltransferase. Its involvement in the proposed model was based on its unique property, that of covalently glucosylating denatured but not native glycoproteins in cell-free assays (Trombetta *et al.*, 1989; Sousa *et al.*, 1992). As yet, however, no evidence gathered in experiments performed in intact cells supports the proposed role for this enzyme.

In order to gain a deeper insight into the function of UDP-Glc:glycoprotein glucosyltransferase in cellular metabolism, we have isolated its coding gene, *gpt1*, in the fission yeast *Schizosaccharomyces pombe*. We have chosen this organism because its subcellular structure, ER oligosaccharide processing reactions and involvement of molecular chaperones in ER folding closely resemble what has been described to occur in mammalian cells. In addition, gene disruption and genetic studies are readily carried out in this organism. Furthermore, we have examined the effect of disrupting this gene and studied its expression under conditions leading to the accumulation of misfolded glycoproteins in the ER.

Results

Cloning and sequencing the S.pombe UDP-Glc:glycoprotein glucosyltransferase gene

To obtain amino acid sequence information for the cloning of the *S.pombe* UDP-Glc:glycoprotein glucosyltransferase, the enzyme was purified up to the Mono Q ion-exchange chromatography step as previously reported (Fernández *et al.*, 1994), and resolved on SDS–PAGE. Following this, the band corresponding to the enzyme was digested with trypsin and the resulting peptides were subjected to amino acid microsequencing according to Hellman *et al.* (1995). The following sequences were obtained: peptide 1, FHCQDDIEDWR/K; peptide 2, DAWTVMQK; peptide 3, DLSSSFFDHYMS; peptide 4, RKPGFESADIMDE; peptide 5, LYSKNPLDYEVV; and peptide 6, NHVVPPLF. Peptides 5 and 6 appeared as a double sequence. In addition, we had previously reported the sequence of the N-terminus of the mature protein (AKPLDVKIAATFN; Fernández *et al.*, 1994). As a first step towards the isolation of the gene encoding UDP-Glc:glycoprotein glucosyltransferase, we undertook a PCR approach. Degenerate oligonucleotides were synthesized (sense 5′-TTYCAYTGYCARGAYGAYAT-3′), from peptide 1 and (antisense 5′-YTTYTGCATNACXGTCCA-3′) from peptide 2 and used as primers for PCR using genomic DNA or cDNA as templates. In both cases, a 1733 bp fragment was obtained, which was then cloned and sequenced (Figure 1). PCR with a sense primer derived from amino acids 5–12 of the N-terminus (5′-GAYGTNAARATHGCXGCXAC-3′) and the antisense primer derived from peptide 2 (see above) yielded a 2903 bp amplification product using either genomic DNA or cDNA as templates (Figure 1). Nucleotide analysis revealed that the smaller fragment (1733 bp) was contained in the larger one (2903 bp), thereby confirming that the peptide sequences obtained corresponded to the same protein.

The 1733 bp fragment was then used as probe in the screening of a cDNA library in the λZAPII express vector. Four independent phages were isolated and sequenced. All four clones were practically identical at their 3′ ends and contained the 452 bp 3′-untranslated region (UTR) and poly(A) tails. However, none of them extended to the coding region of the N-terminus of the mature protein. The longest insert spanned the 3′ last 2793 bp of the 2903 bp fragment (Figure 1). To isolate the sequences encoding the N-terminus and corresponding to the 5′ UTR of the mRNA, a PCR methodology was applied. A sense

primer derived from the λZAPII express vector and an antisense primer (5′-ATACCAAATGAATCCCTC-3′) annealing at 1150 bp from the 5′ end of the 2903 bp fragment were used to amplify the cDNA library. A 1250 bp fragment was synthesized by PCR and subsequently sequenced. In this way, 66 bp of coding sequences were determined in addition to the 2903 bp fragment. This region encodes an 18 residue long putative signal peptide and the first four amino acids that had been previously determined by amino acid microsequencing of the N-terminus of the mature protein. In order to sequence the 5′ UTR, the genomic DNA library was screened with a fragment spanning from 67 to 4467 bp. Three positive phages were isolated and sequenced up to –752 bp.

The gene coding for the *S.pombe* UDP-Glc:glycoprotein glucosyltransferase was designated *gpt1* and its GenBank accession number is U38417. Analysis of the deduced amino acid sequence is given below. Concerning the 5′ UTR, three TATAA boxes were found at –179 to –183, –477 to –481 and –491 to –495 bp. In addition, a decanucleotide (ATTCTGGAAT) with a sequence virtually identical to *Saccharomyces cerevisiae* and *S.pombe* heat shock gene promoters was found at –366 to –375 bp (Tuite *et al.*, 1988; Jannatipour and Rokeach, 1995).

Analysis of the amino acid sequence

The predicted amino acid sequence of the *S.pombe* UDP-Glc:glycoprotein glucosyltransferase is depicted in Figure 2. The peptides determined by amino acid microsequencing of the protein are indicated in the same figure. Of the peptides that had not been used for designing primers for PCR (peptides 3, 4, 5 and 6, see above), peptides 3 and 6 appeared to be coded in both the 1733 and 2903 bp fragments synthesized by PCR as well as in the phages. Peptide 5 was coded in both the 1250 and 2903 bp fragments and in the phages, and peptide 4 only in the last elements. Moreover, the entire microsequenced peptide containing the N-terminus of the mature protein (of which only amino acids 5–12 had been used for designing primers) appeared in the 1250 bp fragment. These results confirm, therefore, that the DNA-deduced protein sequence corresponds to that of the purified protein.

The first 18 amino acids have the characteristics of a signal peptide, rich in hydrophobic amino acids, with a predicted cleavage site between residues 18 and 19. This corresponds precisely to the site determined by amino acid microsequencing of the N-terminus of the mature protein (Fernández *et al.*, 1994). The calculated M_r of the predicted, mature polypeptide is 162 921 Da. This value agrees closely with that of the purified enzyme obtained from SDS–PAGE under reducing conditions (158 000 Da, Fernández *et al.*, 1994). The isoelectric point predicted from the amino acid sequence is 5.11. The deduced polypeptide displays five potential sites for *N*-glycosylation (Figure 2). At least one of them is functional *in vivo* as the enzyme is retained in concanavalin A–Sepharose columns from where it may be eluted with α-methyl-mannoside (Fernández *et al.*, 1994). No possible transmembrane domain was found. This was expected as the enzyme is a soluble protein of the ER (Trombetta *et al.*, 1989, 1991; Fernández *et al.*, 1994). The C-terminal tetrapeptide PDEL most likely represents a novel ER retention sequence for soluble proteins. The only *S.pombe* ER retention signal known to date is ADEL, which was found to occur in the molecular chaperone BiP (binding protein) (Pidoux and Amstrong, 1992).

The only other sequence of a UDP-Glc:glycoprotein glucosyltransferase known thus far, that from *Drosophila melanogaster*, was reported in a recent publication (Parker *et al.*, 1995). Comparison of the fly and the fission yeast enzyme sequences revealed that both enzymes have almost the same length and share extensive identity in their C-terminal portions (see Figure 2). In contrast, the N-terminal regions of the enzymes present scattered sequence similarities. Both *S.pombe* and *D.melanogaster* glucosyltransferases also show similarities with the *S.cerevisiae* Kre5 protein (not shown; for comparison of the *Drosophila* and *S.cerevisiae* proteins, see Parker *et al.*, 1995). The function of Kre5p is unknown, but it is somehow related to the synthesis of the cell wall glucan (Meaden *et al.*, 1990). Since this polysaccharide apparently is not a glycoprotein, it is unlikely that Kre5p carries the function of UDP-Glc:glycoprotein glucosyltransferase. In addition, experimental evidence strongly suggests that *S.cerevisiae* is the only eukaryote known to date to be devoid of UDP-Glc:glycoprotein glucosyltransferase activity (Fernández *et al.*, 1994).

Data obtained from GenBank showed that the C-terminal domains of *S.pombe* and *D.melanogaster* glucosyltransferases have significant similarities with conceptual translations of expressed sequence tags (ESTs) from cDNA libraries of *Caenorhabditis elegans*, rice and *Arabidopsis thaliana* (accession nos M75929, D24933 and T23006, respectively) (Figure 2B).

Expression of gpt1 mRNA under different growth conditions

As mentioned above, it has been proposed that the UDP-Glc:glycoprotein glucosyltransferase is involved in the quality control of glycoprotein folding in the ER (Sousa *et al.*, 1992; Hammond *et al.*, 1994; Helenius, 1994). This proposal was based exclusively on the fact that the enzyme glucosylates denatured but not native glycoproteins in cell-free assays (Trombetta *et al.*, 1989; Sousa *et al.*, 1992); however, no *in vivo* evidence supporting this proposition has been presented to date. It may be argued that if glucosyltransferase is indeed involved in the quality control of glycoprotein folding, environmental conditions leading to the accumulation of misfolded proteins should result in an increased expression of the mRNA coding for the enzyme. To investigate this point, haploid *S.pombe* cells were submitted to conditions that lead to the accumulation of misfolded proteins in the ER of several organisms, and that are known to induce transcription of ER chaperone genes (Normington *et al.*, 1989; Rose *et al.*, 1989; Pidoux and Amstrong, 1992; Cox *et al.*, 1993; Kohno *et al.*, 1993; Mori *et al.*, 1993; Partaledis and Berlin, 1993). RNA was extracted and analysed by Northern blotting. The intensity of the bands was normalized with respect to values obtained when the same filter was hybridized with an actin probe (not shown). A single message of ~5 kb was detected in control cells. As shown in Figure 3, a 5 min shift from 30 to 39°C resulted in a 2-fold induction of expression of *gpt1* mRNA. A 9-fold induction was observed after a 10 min shift. After a 60 min temperature shift, expression of *gpt1* mRNA was still 5-fold higher

Fig. 2. Comparison of amino acid sequences. (**A**) Deduced amino acid sequence of *S.pombe* glucosyltransferase (Sp), and alignment with the *D.melanogaster* (Dm) enzyme. For better comparison, the first 69 amino acids of the *Drosophila* enzyme are omitted. Identical residues are denoted by vertical dashes (|), highly conserved residues by colons (:), less conserved residues by single points (.) and spaces introduced for sequence alignment are designated by dashes (-). The arrow indicates the cleavage point of the signal peptide. Underlined sequences correspond to potential *N*-glycosylation sites. Peptides derived from amino acid microsequencing are within boxes. (**B**) Comparison of *S.pombe* (Sp) glucosyltransferase sequence with conceptual translations of expressed sequence tags (ESTs) from cDNA libraries of *C.elegans* (Ce), rice and *A.thaliana* (Ara). Identical residues are within boxes. Numbers refer to *S.pombe* extreme right amino acid residues.

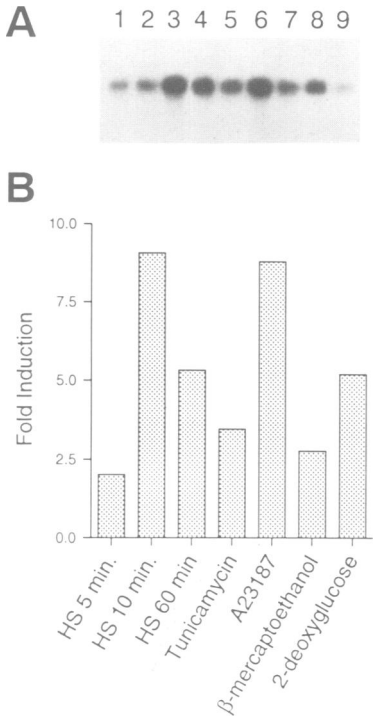

Fig. 3. Expression of *gpt1* mRNA under different physiological conditions. (**A**) Cells were grown at 30°C and shifted to 39°C for the indicated times. Alternatively, cells were grown for 2 h in the presence of tunicamycin, the Ca^{2+} ionophore A23187, 2-mercaptoethanol or 2-deoxyglucose. Total RNA was prepared and analysed by Northern blot. Lane 1, control cells grown at 30°C for 2 h; lanes 2–4 correspond to 5, 10 and 60 min shifts from 30 to 39°C, respectively; lane 5, 2 h in 10 μg/ml tunicamycin at 30°C; lane 6, 2 h in 10 μM A23187 at 30°C; lane 7, 2 h in 15 mM 2-mercaptoethanol at 30°C; lanes 8 and 9, 2 h in the presence or absence of 10 mM 2-deoxyglucose at 30°C, respectively. Growth conditions in both controls (lanes 1 and 8) were different (see Materials and methods). (**B**) For quantitation purposes, the relative intensity of each *gpt1* mRNA was determined by soft-LASER densitometric scanning. The intensity of the bands was normalized with respect to values obtained when the same filter was hybridized with an actin probe (not shown). For further details see Materials and methods and Results.

than basal levels. Folding of glycoproteins in the ER is, for many species, heavily dependent on the presence of saccharide moieties (Helenius, 1994). The expression of *gpt1* mRNA was also stimulated by inhibitors of glycosylation, such as tunicamycin or 2-deoxyglucose, as well as by the Ca^{2+} ionophore A23187 and the reducing agent 2-mercaptoethanol (see Figure 3). The presence of Ca^{2+} is required for proper folding in the ER, as the action of several chaperones is dependent on this cation. Moreover, addition of 2-mercaptoethanol to the medium interferes with the redox potential of the ER, and thus with the disulphide formation machinery of the cell. It may be concluded, therefore, that expression of *gpt1* mRNA is induced by conditions known to produce misfolding of glycoproteins in the ER. These results strongly support the involvement of the UDP-Glc:glycoprotein glucosyltransferase in the quality control of glycoproteins in the ER.

Disruption of the glucosyltransferase gene

To examine whether UDP-Glc:glycoprotein glucosyltransferase supports a function essential for *S.pombe* viability, we disrupted the *gpt1* gene (see Materials and methods and Figure 4A). A 966 bp segment of the *gpt1*-coding region was eliminated by endonuclease digestion and replaced with the *ura4* marker gene (Figure 4A). A linear restriction fragment containing the *ura4*-disrupted *gpt1* sequence was transformed into the Ura4$^-$ Sp13 diploid strain and Ura$^+$ cells were selected. To confirm that a copy of *gpt1* had been disrupted correctly by *ura4*, genomic DNA from both the parental and Ura$^+$ strains was digested with *Bgl*II and analysed by Southern blotting using a 1200 bp probe derived from the glucosyltransferase gene. The 5′-terminus of this probe coincided with that of the 2903 bp fragment mentioned above. A predicted 2.4 kb band was observed in all 15 Ura$^+$ clones analysed, in addition to the 0.8 and 1.2 kb wild-type bands (Figure 4B, lanes 1–3), whereas the 2.4 kb band is clearly absent in the parental strain (Figure 4B, lane 4). Similarly, an expected 2.4 kb band was observed in all 15 Ura$^+$ clones when a *ura4*-specific probe was used (Figure 4B, lanes 5–7). These results agree with what would be expected if homologous recombination in the *gpt1* gene had occurred. Heterozygous, *gpt1*-disrupted strains were sporulated, and random spore analysis showed that all of them were viable. Tetrad analysis further confirmed that *gpt1*$^-$ cells are viable (data not shown).

The glucosyltransferase activity in microsomes of *gpt1*$^+$ and *gpt1*$^-$ cells was assayed (Fernández *et al.*, 1994). In wild-type cells, protein-linked $Glc_1Man_9GlcNAc_2$, $Glc_1Man_8GlcNAc_2$ and $Glc_1Man_7GlcNAc_2$ were formed in the absence of denatured thyroglobulin (Figure 5A). Addition of this substrate to the incubation mixture stimulated the formation of the same compounds. No reaction products were formed with *gpt1*$^-$ microsomes either in the presence or absence of denatured thyroglobulin, thus confirming that the disrupted gene indeed corresponds to that encoding the UDP-Glc:glycoprotein glucosyltransferase and that a single gene codes for the enzyme (Figure 5A).

To investigate whether the growth rate of *gpt1*$^-$ cells was affected, the mutant and parental haploid strains were cultured at different temperatures. No major differences between the growth rates of *gpt1*$^+$ and *gpt1*$^-$ strains at 18, 28 and 39°C were found (Figure 5B). However, microscopic examination revealed that *gpt1*$^-$ cells are ~30% shorter than the parental cells (Figure 6).

Discussion

The primary sequence of *S.pombe* UDP-Glc:glycoprotein glucosyltransferase displays the characteristic features of soluble proteins of the ER: a leading signal peptide followed by a protein sequence devoid of transmembrane domains ending up in a novel *S.pombe* ER retention sequence (PDEL).

The other known sequence of a UDP-Glc:glycoprotein glucosyltransferase is that from *D.melanogaster*, which has been published recently (Parker *et al.*, 1995). What is striking when the fission yeast and fly sequences are compared is that, although both proteins have nearly the same size and their C-terminal regions have a high degree of identity, they differ vastly in their N-terminal portions. These structural features are reminiscent of the hsp70 family of proteins which includes the ER chaperone BiP: their N-termini, carrying the ATPase activity, have a

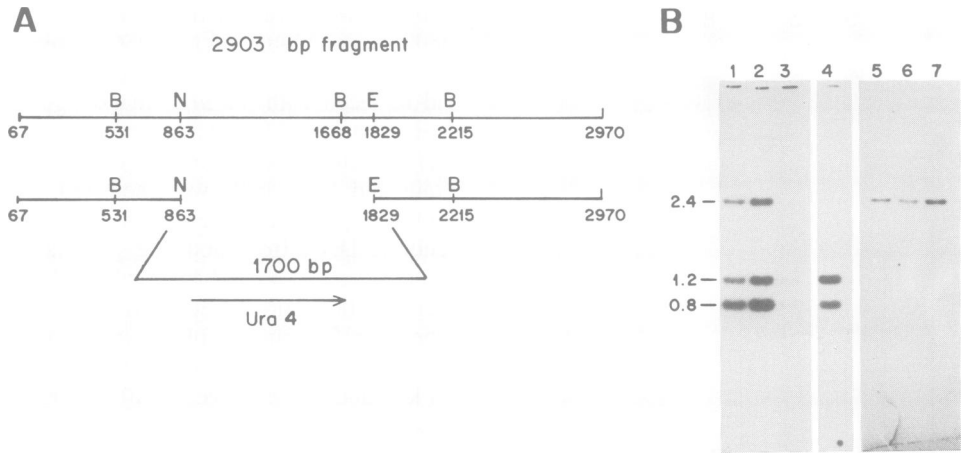

Fig. 4. Disruption of the *gpt1* gene. (**A**) Strategy used for disrupting the gene. The 2903 bp fragment (see Figure 1) was digested with *Nhe*I and *Eco*RV to delete the *gpt1* coding sequences between positions 863 and 1829 bp, and the 1700 bp *ura4* marker gene fragment was inserted. Digestion with *Xba*I and *Sac*I yielded a linear fragment containing 400 bp and 1000 bp of *gpt1*, 5′ and 3′ of the *ura4* gene, respectively. This fragment was used to transform the *ura4*$^-$ Sp13 *S.pombe* diploid strain and Ura$^+$ cells were selected. Only relevant restriction sites are shown as follows: B, *Bgl*II; E, *Eco*RV; and N, *Nhe*I. (**B**) Analysis of heterozygous cells carrying a disrupted *gpt1* allele. DNA was extracted from three Ura$^+$ isolates (lanes 1–3 and 5–7, which correspond to the same isolates, respectively) as well as from the parental strain (lane 4), and digested with *Bgl*II. Lanes 1–4 depict a Southern blot probed with a 1200 bp glucosyltransferase fragment extending from position 67 bp. Lanes 5–7 correspond to the hybridization with the *ura4* probe. Positions of marker bands in kb are indicated. For further details see Materials and methods and Results.

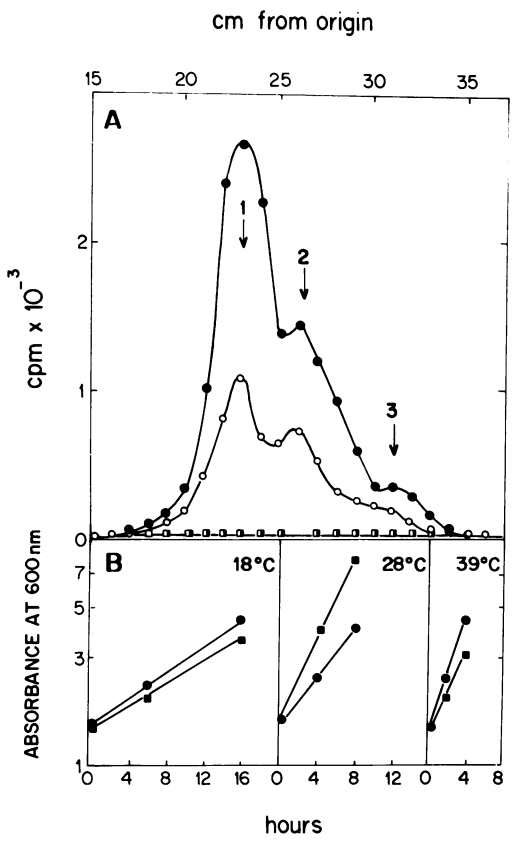

Fig. 5. Phenotypic characterization of *gpt1*$^-$ cells. (**A**) Enzymatic activity. Microsomes were prepared from haploid wild-type (circles) or *gpt1*$^-$ (squares) cells and incubated with UDP-[^{14}C]Glc in the presence (●—● or ■—■) or absence (○—○ or □—□) of denatured thyroglobulin. Total proteins were degraded with a protease and resulting glycopeptides were incubated with endo-β-*N*-acetylglucosaminidase H. Oligosaccharides were then run on paper chromatography in 1-propanol/nitromethane/water (5:2:4). Standards: 1, $Glc_1Man_9GlcNAc$; 2, $Glc_1Man_8GlcNAc$ and 3, $Glc_1Man_7GlcNAc$. (**B**) Growth of haploid *gpt1*$^+$ (●—●) or *gpt1*$^-$ (■—■) cells at the indicated temperatures. For further details see Materials and methods.

high degree of identity, whereas the C-terminal portions, containing the hydrophobic peptide binding sites, are much less conserved (Hendrick and Hartl, 1993). We have shown recently that, under physiological salt and pH conditions, glucosyltransferase binds hydrophobic peptides exposed in misfolded glycoproteins (Sousa and Parodi, 1995). Glucosyltransferase shares this property, therefore, with known chaperones, and it may well perform this function *in vivo*. Further, comparison of the fission yeast and fly sequences leads us to speculate that the conserved C-terminal portion of the enzyme carries the glucosyltransferase activity, which recognizes both the sugar nucleotide and the oligosaccharide, whereas the N-terminus, having to recognize a wide variety of different hydrophobic peptides, may be constituted by quite different amino acid sequences. This two-domain notion is supported by the fact that a limited but significant sequence similarity has been observed between some bacterial glycosyltransferases and the *D.melanogaster* glucosyltransferase precisely in their C-terminal portions (Parker *et al.*, 1995; similarity was also observed between *S.pombe* glucosyltransferase and the bacterial enzymes). The two-domain proposal is supported further by the fact that, as mentioned above, a significant similarity was found between the C-terminal portions of *S.pombe* and *D.melanogaster* glucosyltransferases and conceptual translations of ESTs from cDNA libraries from rice, *A.thaliana* and *C.elegans*. Concerning this last organism, an open reading frame (ORF; gene F48E3.3′) recently submitted to GenBank (accession no. U28735) probably corresponds to the gene encoding the UDP-Glc:glycoprotein glucosyltransferase. The conceptual translation of the ORF (that contains the EST mentioned above) corresponds to a protein having 1493 amino acids with the C-terminal tetrapeptide HTEL. Moreover, extensive identity between the C-terminal portions of the *S.pombe* and *D.melanogaster* glucosyltransferases and the *C.elegans* protein was found. The N-terminal domains of the proteins from these organisms only showed scattered similarity.

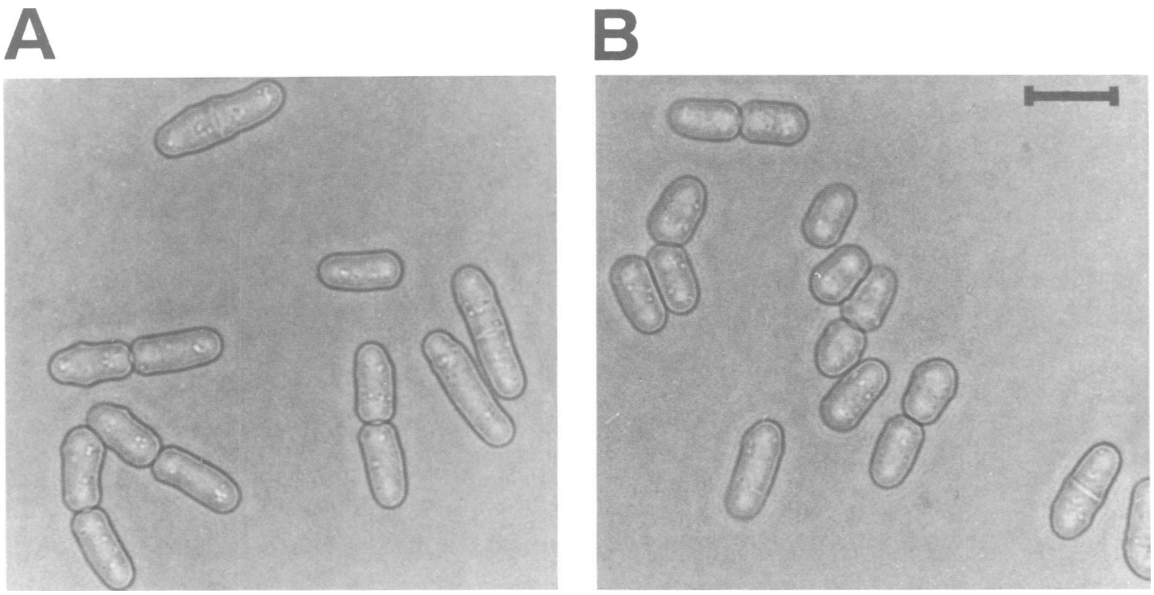

Fig. 6. Microscopy of (**A**) wild-type (*gpt1*$^+$) and (**B**) *gpt1*$^-$ cells. *Schizosaccharomyces pombe* cells were photographed using phase-contrast microscopy. Bar = 10 μm.

By pursuing further the analogy with hsp70 proteins, it may be speculated that part of the energy spent in the transfer of the glucose unit is also used to facilitate the release of the enzyme from the misfolded substrate, as is the case for chaperones upon ATP hydrolysis (Pelham, 1986).

Disruption of the glucosyltransferase gene did not cause any major change in the growth rate of *S.pombe* cells at 18, 28 and 39°C. In contrast, calnexin has been shown to be essential for *S.pombe* viability (Jannatipour and Rokeach, 1995). To reconcile this apparent incongruity, it may be speculated that formation of $Glc_1Man_9GlcNAc_2$ (the ligand binding to calnexin) by deglucosylation of the transferred oligosaccharide ($Glc_3Man_9GlcNAc_2$) is indeed sufficient to ensure binding to the chaperone, and that the glucosyltransferase activity only functions as a spare mechanism. In other words, the role of glucosyltransferase could be to facilitate calnexin recognition of partially folded/misfolded glycoproteins from which all the glucose units have been removed, prior to any interaction with this membrane-tethered chaperone. Likewise, the initial binding of substrates via their monoglucosylated oligosaccharides could be a 'safety', but non-essential, mechanism of recognition of partially folded/misfolded glycoproteins by calnexin. In this vein, it has been shown recently that mammalian calnexin can form stable and long-lived complexes with the T-cell receptor ε subunit lacking Asn-linked oligosaccharides as well as with a recombinant version of the multidrug resistance P-glycoprotein from which *N*-glycosylation sites had been removed by mutation (Loo and Clarke, 1994; Rajagopalan *et al.*, 1994). In addition, deletion of the transmembrane and cytoplasmic domains from the α subunit of the T-cell receptor almost completely eliminated calnexin binding, despite the fact that this truncated subunit still possessed its full complement of *N*-linked oligosaccharides (Margolese *et al.*, 1993). That formation of monoglucosylated oligosaccharides is not essential for cell survival is supported further by the fact that mutant mammalian cells devoid of glucosidase I, i.e. cells in which monoglucosylated oligosaccharides cannot be formed either by deglucosylation of the transferred oligosaccharide or by reglucosylation by the glucosyltransferase, have no alteration in their growth pattern (Ray *et al.*, 1991).

Stress conditions known to accumulate misfolded proteins in the ER of *S.pombe* and/or of other cell types (Normington *et al.*, 1989; Rose *et al.*, 1989; Pidoux and Amstrong, 1992; Cox *et al.*, 1993; Kohno *et al.*, 1993; Mori *et al.*, 1993; Partaledis and Berlin, 1993; e.g. heat shock, and growth in the presence of 2-mercaptoethanol, Ca^{2+} ionophore or inhibitors of protein glycosylation) were shown to induce synthesis of the mRNA encoding the UDP-Glc:glycoprotein glucosyltransferase. As this feature is shared by other chaperones, this result strongly supports the role proposed for the enzyme as a sensor of misfolded glycoproteins in the quality control of glycoprotein folding. The identification of the glucosyltransferase as a heat shock protein was reinforced by the fact that, as mentioned above, a highly conserved yeast heat shock gene promoter sequence was found at –366 bp of the 5′ UTR (Tuite *et al.*, 1988; Jannatipour and Rokeach, 1995). To date, the involvement of glucosyltransferase in such processes has been inferred solely from the known property of the enzyme, that of covalently tagging denatured but not native glycoproteins in cell-free assays.

Materials and methods

Bacterial and S.pombe strains

Escherichia coli strain XL1-Blue was used for all cloning procedures. *Escherichia coli* XL1-Blue MRF′ was used for screening of libraries constructed in the λZAP express vector (Stratagene). *Escherichia coli* strain XLOR (Stratagene) was used for plating excised phagemids. The *S.pombe* strain used was SP13 (h^+/h^-, *ade6*-M210/*ade6*-M216, *leu1*-32/*leu1*-32, *ura4*-D18/*ura4*-D18; Jannatipour and Rokeach, 1995).

Schizosaccharomyces pombe media

YE and minimal medium were prepared as described (Moreno *et al.*, 1991). The extremely low nitrogen (ELN) medium had the same

composition as the minimal medium, except that NH_4Cl was at 50 mg/l, and was used for sporulation of diploids.

DNA procedures

Standard DNA manipulations were carried out as described (Sambrook *et al.*, 1989). *Schizosaccharomyces pombe* genomic DNA was prepared as described before (Moreno *et al.*,1989). Construction of the *S.pombe* genomic library was previously described (Jannatipour and Rokeach, 1995). Briefly, the library was constructed by ligating partially *Sau*3A-digested, size-selected (5–7 kb) genomic DNA into *Bam*HI-digested λZAP express arms (Stratagene). The *S.pombe* cDNA library, constructed in λZAPII vector (Stratagene) was a gift from Dr David Beach (Cold Spring Harbor Laboratory, NY). Hybridization conditions were as described before (Sambrook *et al.*, 1989).

RNA procedures

Preparation of RNA was as previously described (Jannatipour and Rokeach, 1995). Briefly, exponentially growing cells (50 ml, minimal medium) were subjected to different treatments. Control cells were incubated at 30°C for 2 h. For heat shock, cell cultures were shifted from 30 to 39°C for 5, 10 and 60 min. Alternatively, cells were grown for 2 h at 30°C in the presence of either 10 μg/ml of tunicamycin, 10 μM A231187 or 15 mM 2-mercaptoethanol. In other treatments, cells were centrifuged, washed and resuspended in minimal medium minus glucose, supplemented with 3% glycerol and 2% lactic acid, in the presence or absence of 10 mM 2-deoxyglucose and incubation continued for a further 2 h. At the end of the incubation, cells were placed on ice, pelleted and total RNA was prepared as described (Moreno *et al.*, 1991). RNA samples (2.5 μg) were analysed by Northern blotting as previously described (Rokeach *et al.*, 1989). The cDNA probe used was derived from one of the λZAPII express clones isolated (see Results) and comprised from base 188 to the poly(A) tail. The gel was stained with ethidium bromide to verify that equivalent amounts of RNA were loaded.

Gene disruption

A 2903 bp fragment of *gpt1* cloned in the T vector (Novagen) was isolated after restriction with *Nhe*I and *Eco*RV. The corresponding sites are at 863 and 1829 bp from the ATG initiation codon of the *gpt1* gene. In this way, 966 bp were eliminated and replaced with a 1700 bp fragment containing the *ura4* gene. Restriction of the latter plasmid with *Xba*I and *Sal*I yielded a 3100 bp linear fragment containing the *ura4* gene flanked by 400 bp (5′) and 1000 bp (3′) of *gpt1* sequences. The *ura4*$^-$ Sp13 *S.pombe* diploid strain was transformed with this linear fragment and Ura$^+$ cells were selected. Genomic DNA from parental Sp13 and Ura$^+$ transformants was analysed by digestion with *Bgl*II followed by Southern blotting using a probe specific for the *ura4* gene and with a 1200 bp probe derived from the glucosyltransferase gene. The 5′ end of this probe coincided with that of the *gpt1* 2903 bp fragment. The glucosyltransferase-disrupted heterozygous cells were sporulated on ELN medium. The mixture of asci and vegetative cells was treated in 1 ml of 100 mM sodium acetate pH 5.5, containing 10 μl of β-glucuronidase (Sigma) at room temperature overnight. Digestion was examined microscopically and ~250 spores were washed and germinated onto media with or without uracil to determine the genotype of the progeny.

Enzymatic analysis

Schizosaccharomyces pombe microsomes were prepared from *gpt1*$^+$ and *gpt1*$^-$ cells and the UDP-Glc:glycoprotein glucosyltransferase activity assayed in the absence or presence of denatured thyroglobulin (Sigma) as described before (Fernández *et al.*, 1994).

Acknowledgements

This work was supported by the United States Public Health Service Grant GM 44500, by the Swedish Agency for Research Cooperation with Developing Countries (SAREC), by the University of Buenos Aires, and by the Medical Research Council of Canada (MRC). F.F. is a Doctoral Fellow of the National Research Council (Argentina), A.J.P. is a Career Investigator of the same institution, and L.A.R. is an MRC Scholar.

References

Cox,J.S., Shamu,C.E. and Walter,P. (1993) Transcriptional induction of genes encoding endoplasmic reticulum resident proteins requires a transmembrane protein kinase. *Cell*, **73**, 1197–1206.

Fernández,F., Trombetta,S.E., Hellman,U. and Parodi,A.J. (1994) Purification to homogeneity of UDP-Glc:glycoprotein glucosyltransferase from *Schizosaccharomyces pombe* and apparent absence of the enzyme from *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **269**, 30701–30706.

Hammond,C., Braakman,I. and Helenius,A. (1994) Role of *N*-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc. Natl Acad. Sci. USA*, **91**, 913–917.

Hebert,D.H., Foellmer,B. and Helenius,A. (1995) Glucose trimming and reglucosylation determine glycoprotein association with celnexin in the endoplasmic reticulum. *Cell*, **81**, 425–433.

Helenius,A. (1994) How *N*-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol. Biol. Cell.*, **5**, 253–265.

Hellman,U., Wernstedt,C., Gonez,J. and Heldin,C.-H. (1995) Improvement of an in-gel digestion procedure for the micropreparation of internal fragments for amino acid sequencing. *Anal. Biochem.*, **224**, 451–455.

Hendrick,J.P. and Hartl,F.U. (1993) Molecular chaperone functions of heat-shock proteins. *Annu. Rev. Biochem.*, **62**, 349–384.

Jannatipour,M. and Rokeach,L.A. (1995) The *Schizosaccharomyces pombe* homologue of the chaperone calnexin is essential for viability. *J. Biol. Chem.*, **270**, 4845–4853.

Kohno,K., Normington,K., Sambrook,J., Gething,M.J. and Mori,K. (1993) The promoter region of the yeast *KAR2* (BiP) gene contains a regulatory domain that responds to the presence of unfolded proteins in the endoplasmic reticulum. *Mol. Cell. Biol.*, **13**, 877–890.

Kornfeld,R. and Kornfeld,S. (1985) Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.*, **54**, 631–664.

Lodish,H.F. (1988) Transport of secretory and membrane glycoproteins from the rough endoplasmic reticulum to the Golgi. A rate-limiting step in protein maturation and secretion. *J. Biol. Chem.*, **263**, 2107–2110.

Loo,T.W. and Clarke,D.M. (1994) Prolonged association of temperature-sensitive mutants of human P-glycoprotein with calnexin during biogenesis. *J. Biol. Chem.*, **269**, 28683–28689.

Margolese,L., Waneck,G.L., Suzuki,C.K., Degen,E., Flavell,R.A. and Williams,D.B. (1993) Identification of the region on the class I histocompatibility molecule that interacts with the molecular chaperone p88 (calnexin, IP90). *J. Biol. Chem.*, **268**, 17959–17966.

Meaden,P., Hill,K., Wagner,J., Slipetz,D., Sommer,S.S. and Bussey,H. (1990) The yeast *KRE5* gene encodes a probable endoplasmic reticulum protein required for (1–6)-β-D-glucan synthesis and normal cell growth. *Mol. Cell. Biol.*, **10**, 3013–3019.

Moreno,S., Klar,A. and Nurse,P. (1991) Molecular genetic analysis of fission yeast *Schizosaccharomyces pombe*. *Methods Enzymol.*, **194**, 795–823.

Mori,K., Ma,W., Gething,M.J. and Sambrook,J. (1993) A transmembrane protein with a cd2$^+$/CDC28-related kinase activity is required for signaling from the ER to the nucleus. *Cell*, **74**, 743–756.

Normington,K., Kohno,K., Kozutsumi,Y., Gething,M.J. and Sambrook,J. (1989) *S.cerevisiae* encodes an essential protein homologous in sequence and function to mammalian BiP. *Cell*, **57**, 1223–1236.

Ou,W.-J., Cameron,P.H., Thomas,D.Y. and Bergeron,J.J.M. (1993) Association of folding intermediates of glycoproteins with calnexin during protein maturation. *Nature*, **364**, 771–776.

Parker,C.G., Fessler,L.I., Nelson,R.E. and Fessler,J.H. (1995) *Drosophila* UDP-Glc:glycoprotein glucosyltransferase: sequence and characterization of an enzyme that distinguishes between denatured and native proteins. *EMBO J.*, **14**, 1294–1303.

Parodi,A.J., Behrens,N.H., Leloir,L.F. and Carminatti,H. (1972) The role of polyprenol-bound saccharides as intermediates in glycoprotein synthesis in liver. *Proc. Natl Acad. Sci. USA*, **69**, 3268–3272.

Parodi,A.J., Lederkremer,G.Z. and Mendelzon,D.H. (1983) Transient glucosylation of protein bound $Man_9GlcNAc_2$, $Man_8GlcNAc_2$ and $Man_7GlcNAc_2$ in calf thyroid cells. A possible recognition signal in the processing of glycoproteins. *J. Biol. Chem.*, **258**, 8260–8265.

Parodi,A.J., Mendelzon,D.H., Lederkremer,G.Z. and Martín-Barrientos,J. (1984) Evidence that transient glucosylation of protein-linked $Man_9GlcNAc_2$, $Man_8GlcNAc_2$ and $Man_7GlcNAc_2$ occurs in rat liver and *Phaseolus vulgaris* cells. *J. Biol. Chem.*, **259**, 6351–6357.

Partaledis,J.A. and Berlin,V. (1993) The *FKB2* gene of *Saccharomyces cerevisiae* encoding the immunosuppresant-binding protein FKBP-13, is regulated in response to accumulation of unfolded proteins in the endoplasmic reticulum. *Proc. Natl Acad. Sci. USA*, **90**, 5450–5454.

Pelham,H.R.B. (1986) Speculations on the functions of the major heat shock and glucose-regulated proteins. *Cell*, **46**, 959–961.

Pidoux,A.L. and Amstrong,J. (1992) Analysis of the BiP gene and

identification of an ER retention signal in *Schizosaccharomyces pombe*. *EMBO J.*, **11**, 1583–1591.

Rajagopalan,S., Xu,Y. and Brenner,M.B. (1994) Retention of unassembled components of integral membrane proteins by calnexin. *Science*, **263**, 387–390.

Ray,M.K., Yang,J., Sundaram,S. and Stanley,P. (1991) A novel glycosylation phenotype expressed by Lec23, a Chinese hamster ovary mutant deficient in α-glucosidase I. *J. Biol. Chem.*, **266**, 22818–22825.

Rokeach,L.A., Jannatipour,M., Haselby,J.A. and Hoch,S.O. (1989) Primary structure of a human small nuclear ribonucleoprotein polypeptide as deduced by cDNA analysis. *J. Biol. Chem.*, **264**, 5024–5030.

Rose,M.D., Misra,L.M. and Vogel,J.P. (1989) *KAR2*, a karyogamy gene, is the yeast homolog of the mammalian BiP/GRP78 gene. *Cell*, **57**, 1211–1221.

Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sousa,M. and Parodi,A.J. (1995) The molecular basis for the recognition of misfolded glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *EMBO J.*, **14**, 4196–4203.

Sousa,M., Ferrero-García,M. and Parodi,A.J. (1992) Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *Biochemistry*, **31**, 97–105.

Trombetta,S., Bosch,M. and Parodi,A.J. (1989) Glucosylation of glycoproteins by mammalian, plant, fungal and trypanosomatid protozoa microsomal proteins. *Biochemistry*, **28**, 8108–8116.

Trombetta,S., Gañán,S. and Parodi,A.J. (1991) The UDP-Glc: glycoprotein glucosyltransferase is a soluble protein of the endoplasmic reticulum. *Glycobiology*, **1**, 155–161.

Tuite,M.F., Bossier,P. and Fitch,I.T. (1988) A highly conserved sequence in yeast heat shock gene promoters. *Nucleic Acids Res.*, **16**, 11845.

Ware,F.E., Vassilakos,A., Peterson,P.A., Jackson,M.R., Lehrman,M.A. and Williams,D.B. (1995) The molecular chaperone calnexin binds $Glc_1Man_9GlcNAc_2$ oligosaccharide as an initial step in recognizing unfolded glycoproteins. *J. Biol. Chem.*, **270**, 4697–4704.

Zhang,Q., Tector,M. and Salter,R.D. (1995) Calnexin recognizes carbohydrate and protein determinants of class I major histocompatibility complex molecules. *J. Biol. Chem.*, **270**, 3944–3948.

Received on August 30, 1995; revised on October 27, 1995