Glycopeptide specificity of the secretory protein folding sensor UDP–glucose glycoprotein:glucosyltransferase

Sean C. Taylor^{1,2}, Pierre Thibault³, Daniel C. Tessier¹, John J.M. Bergeron² & David Y. Thomas^{1,2+}

¹Biochemistry Department, Faculty of Medicine, McGill University, Montreal, Quebec, Canada, ²Department of Anatomy and Cell Biology, McGill University, Montreal, Quebec, Canada, and ³Institute for Biological Sciences, National Research Council of Canada, Ottawa, Ontario, Canada

Secretory and membrane N-linked glycoproteins undergo folding and oligomeric assembly in the endoplasmic reticulum with the aid of a folding mechanism known as the calnexin cycle. UDP–glucose glycoprotein:glucosyltransferase (UGGT) is the sensor component of the calnexin cycle, which recognizes these glycoproteins when they are incompletely folded, and transfers a glucose residue from UDP–glucose to N-linked Man9-GlcNAc2 glycans. To determine how UGGT recognizes incompletely folded glycoproteins, we used purified enzyme to glucosylate a set of Man9-GlcNAc2 glycopeptide substrates *in vitro*, and determined quantitatively the glucose incorporation into each glycan by mass spectrometry. A ranked order of glycopeptide specificity was found that provides the criteria for the recognition of substrates by UGGT. The preference for amino-acid residues close to N-linked glycans provides criteria for the recognition of glycopeptide substrates by UGGT.

EMBO reports 4, 405–411 (2003)

doi:10.1038/sj.embor.embor797

INTRODUCTION

UDP–glucose glycoprotein:glucosyltransferase (UGGT) is an enzyme that resides in the endoplasmic reticulum (ER). UGGT recognizes incompletely folded glycoproteins with N-linked Man9-GlcNAc2 glycans, and transfers a glucose residue from UDP–glucose to produce the Glc1-Man9-GlcNAc2 glycan (for a review, see Parodi, 2000; Tessier *et al.*, 2000). This monoglucosylated glycan is recognized specifically by the lectin-like chaperones calnexin and calreticulin, which retain incompletely folded glycoproteins in the ER and recruit ERp57 to accelerate disulphide bond interchange for further folding (Zapun *et al.*, 1997, 1998; Schrag *et al.*, 2001).

¹Biochemistry Department, Faculty of Medicine, McGill University, McIntyre Medical Sciences Building, 3655 Promenade Sir William Osler, Montreal, Quebec, Canada H3G 1Y6

²Department of Anatomy and Cell Biology, McGill University, 3640 University Street, Montreal, Quebec, Canada H3A 2B2

³Institute for Biological Sciences, National Research Council of Canada, 100 Sussex Drive, Ottawa, Ontario, Canada K1A 0R6

*Corresponding author. Tel: +514 398 2973; Fax: +514 398 7384;

E-mail: david.thomas@mcgill.ca

Received 31 October 2002; revised 20 January 2003; accepted 21 January 2003

Fig. 1 | Components of the *in vitro* system and glycan composition of acid phosphatase, exo-1,3-β-D-glucanase and α-galactosidase expressed in the *Saccharomyces cerevisiae* strain DT111. (A–C) Matrix-assisted laser desorption/ionization–time-of-flight (MALDI–TOF) mass spectra of the glycoproteins acid phosphatase (AcP; A), α-galactosidase (α-Gal; B) and exo-1,3-β-D-glucanase (β-Glc; C). The numbers at the top of each peak correspond to the isotope-average protonated glycoprotein molecule. In (A) and (B), other, less abundant glycoforms are indicated. (D) Endo H digestions of AcP, α-Gal and β-Glc. Approximately 2 μg of each glycoprotein substrate was digested. Shifts of 19 kDa (AcP), 13 kDa (α-Gal) and 4 kDa (β-Glc) were consistent with occupancy of all the N-linked glycosylation sites in the three substrates. A gel stained with Coomassie blue is shown.

Several *in vitro* studies have shown that denatured glycoproteins are the preferred substrates for UGGT (Sousa & Parodi, 1995; Tessier *et al.*, 2000). It has been shown that denatured glycoproteins with a Man9-GlcNAc2 glycan are glucosylated preferentially over the Man8 or Man7 glycoforms (Sousa *et al.*, 1992), and the innermost GlcNAc residue is also required for recognition by UGGT (Sousa & Parodi, 1995). UGGT has also

been shown to monoglucosylate the glycan in the unfolded monomer of an RNase B heterodimer specifically, whereas it leaves the glycan in the folded monomer unmodified (Ritter & Helenius, 2000). This targeted monoglucosylation may direct the chaperones and folding enzymes of the calnexin cycle to the regions of incompletely folded glycoproteins that require further folding.

Fig. 2 | Quantification of UDP-glucose glycoprotein:glucosyltransferase monoglucosylation of glycopeptide substrates. (A) Monoglucosylation of α -galactosidase (α -Gal) tryptic glycopeptides. Approximately 1 μ g of α -Gal was chemically denatured, and was incubated with UDP-glucose glycoprotein: glucosyltransferase (UGGT) at a concentration of ~ 25 nM and UDP- 3 H-glucose at 2.5 μ M (19.5 Ci mmol $^{-1}$). Samples were then either electrophoretically separated by SDS-polyacrylamide gel electrophoresis (lane 1; α -Gal) or cleaved by trypsin, followed by SDS-PAGE (lane 2; labelled first, cut second (L1C2)). In lane 3 (cut first, labelled second (C1L2)), a tryptic digest of α -Gal was incubated with UGGT and 2.5 μ M UDP- 3 H]glucose (19.5 Ci mmol $^{-1}$). The gel was exposed for two days, and the resulting autoradiograph is shown. (B) Micro-liquid-chromatography-electrospray mass spectrometry analyses of UGGT-monoglucosylated tryptic glycopeptide T₄₄₄₋₄₆₆ from α -Gal (see Fig. 4 for reaction conditions). Two ion chromatograms are shown, representing the full mass spectral acquisition ($m/z = 400-2,000$) obtained with a low orifice voltage setting (40 V), together with a high orifice voltage scan (100 V) for $m/z = 366$. (Ba) Mass spectrum from a control tryptic digest of α -Gal with no incubation with UGGT, showing the Man9-GlcNAc2 ($m/z = 1,427$) and Man10-GlcNAc2 ($m/z = 1,481$) glycoforms that were used to calculate background levels of molar incorporation. (Bb) Mass spectrum obtained from the incubation of an α -Gal tryptic digest with UGGT and UDP-glucose, showing the Man9-GlcNAc2 ($m/z = 1,427$) and Glc1-Man9-GlcNAc2 ($m/z = 1,481$) glycoforms. Peak ratio (molar glucose incorporation) is the (area under the (Glc1 or Man1)-Man9-GlcNAc2 isotopic ion cluster divided by the sum of the areas under the (Glc1 or Man1)-Man9-GlcNAc2 isotopic ion cluster and the area under the Man9-GlcNAc2 isotopic ion cluster. All y axes indicate the ion intensity relative to the highest peak in each spectrum. Rel. int., relative ion intensity.

Fig. 3 | Micro-liquid-chromatography–electrospray-mass-spectrometry analyses of glycopeptides from acid phosphatase and α -galactosidase that are monoglucosylated by UDP–glucose glycoprotein:glucosyltransferase. Analyses were carried out as described for Fig. 2B. GT, Glu-C–trypsin combined digest; Rel. int., relative ion intensity; T, trypsin digest.

In contrast to previous reports, we show that, in a system using purified components, UGGT has distinct peptide-sequence preferences. We used defined Man₉-GlcNAc₂ glycopeptides and purified, recombinant rat liver UGGT, and carried out quantitative analysis by mass spectrometry of both the substrates and the products. Across a set of 24 glycopeptide substrates with diverse sequences, the activity of UGGT varied 20-fold.

RESULTS

Generation of glycopeptide substrates

The DT111 yeast strain (Tessier *et al.*, 2000) was used to express glycoproteins that carry Man₉-GlcNAc₂ glycans, which was confirmed for acid phosphatase (AcP), exo-1,3- β -D-glucanase (β -Glc) and α -galactosidase (α -Gal) by matrix-assisted laser desorption/ionization–time-of-flight (MALDI–TOF) mass spectrometry (Fig. 1A–C). However, in the case of AcP and α -Gal, some of these sites are partially occupied (glycoprotein peaks separated by 1,865 Da, see Fig. 1A,B). Endoglycosidase H (Endo H) digestion confirmed the results obtained by mass spectrometry (Fig. 1D).

Purified denatured Man₉-GlcNAc₂ glycoproteins from the yeast strain DT111 can be monoglucosylated using UDP-[³H]glucose and UGGT (Fig. 2A, lane 1). Subsequent proteolytic digestion of these glycoproteins with trypsin resulted in the production of monoglucosylated fragments (Fig. 2A, lane 2). These proteins can be denatured, trypsin digested and the resulting glycopeptides can be used as effective substrates for recombinant UGGT (Fig. 2A, lane 3). These

results provided the first evidence for the efficient recognition of glycopeptide substrates by UGGT, and allowed an analysis of the peptide features that are recognized by this enzyme. Mass spectrometry was used to detect and measure UGGT-catalysed glucose incorporation into glycopeptides (Figs. 2B and 3). The procedure is based on the ratio of the relative ion intensities of the Man₉-GlcNAc₂ to Glc1-Man₉-GlcNAc₂ forms of each glycopeptide. Hence, the unglucosylated Man₉-GlcNAc₂ form of each glycopeptide provides an internal control, from which the molar incorporation of glucose was calculated.

A ranked order of monoglucosylated glycopeptides

A total of 24 glycopeptide substrates from proteolytic digests of AcP, α -Gal and β -Glc were identified as substrates for UGGT. These monoglucosylated peptides were ranked on the basis of glucose incorporation (Fig. 4). The glycopeptides were then grouped as good (group 1), intermediate (group 2) and poor (group 3) substrates for UGGT. However, even the most poorly recognized glycopeptide (number 24) showed a detectable level of monoglucosylation. The reproducibility of this data was confirmed by comparing the levels of monoglucosylation between identical peptides from different digestion conditions (compare closed symbols in Fig. 4). Glycopeptides with two glycans were treated as special cases, and are shown in Fig. 5.

The amino-acid hydropathy of glycopeptides 1–14 in Fig. 4 were compared as described in Methods. The results of this comparison are shown in Fig. 6A.

	Rank	Protein	Peptide	Sequence	x	s.d.
Group 1	○ 1	AcP	T _{355–371}	NNLTAEVVPFMGNTFHR	0.883	0.002
	◇ 2	AcP	T _{101–118}	QFNGSLSFLNDDYEFFIR	0.811	0.005
	3	α-Gal	T _{444–446}	IGSLSPNAILNITVPAHGIAFYR	0.706	0.018
	□ 4	AcP	T _{248–270}	GLNLTSITDASTLFSWCAFEVNAK	0.660	0.068
	5	α-Gal	T _{369–385}	MNITLEEITFFDSNLGSK	0.597	0.035
	□ 6	AcP	GT _{248–266}	GLNLTSITDASTLFSWCAFE	0.501	0.008
Group 2	7	AcP	G _{385–406}	KFQCSNDTYVRYVINDAVVPIE	0.466	0.012
	8	AcP	GT _{176–200}	CHDTAQYFIDGLGDQFNITLQIVSE	0.421	0.064
	○ 9	AcP	G _{326–360}	IQDQKVWLSFTHDITDILNFLTTHAGTIDDKNNLTAE	0.394	0.015
	10	α-Gal	T _{246–277}	AAPMGQNAGVGGWNDLDNLEVGVGNLTDDDEEK	0.365	0.036
	11	β-Glc	T _{315–344}	VACEWGTGVILNESHWTVCGEFAAALITDCTK	0.322	0.044
	□ 12	AcP	G _{246–266}	NKGLNLTSITDASTLFSWCAFE	0.317	0.017
	◇ 13	AcP	GT _{101–114}	QFNGSLSFLNDDYE	0.308	0.083
Group 3	14	AcP	GT _{162–174}	NOTSFAMVTSNSK	0.141	0.061
	● 15	AcP	T _{308–319}	SVGSNLFNASVK	0.135	0.008
	16	α-Gal	T _{401–412}	VDNSTASAILGR	0.119	0.004
	● 17	AcP	GT _{308–319}	SVGSNLFNASVK	0.113	0.044
	18	α-Gal	T _{415–429}	TATGILYNATEQSYK	0.102	0.019
	■ 19	AcP	GT _{95–100}	LSNYTR	0.087	0.043
	■ 20	AcP	T _{95–1100}	LSNYTR	0.078	0.013
	◆ 21	AcP	T _{386–395}	FQCSNDTYVR	0.070	0.010
	○ 22	AcP	GT _{355–360}	NNLTAE	0.055	0.028
	◆ 23	AcP	GT _{386–395}	FQCSNDTYVR	0.054	0.016
	24	α-Gal	T _{169–176}	AMSDALNK	0.044	0.007

Fig. 4 | Ranked order of monoglucosylation of singly glycosylated tryptic glycopeptides. All reactions were performed at 37 °C for 4 h. Three separate reactions were performed in triplicate, containing 4 μM of the proteolytic digests from acid phosphatase (AcP), exo-1,3-β-D-glucanase (β-Glc) and α-galactosidase (α-Gal), ~58 nM UDP–glucose glycoprotein:glucosyltransferase (UGGT) and 1 mM UDP–glucose. The number of moles of glucose incorporated per glycopeptide were determined as in Fig. 2B. The average moles of glucose per mole of peptide (x) incorporated for each glycopeptide (for three experiments) and the corresponding standard deviation (s.d.) are shown. Symbols in front of the rank numbers indicate peptides that contain the same N-linked glycan. The letters in front of each peptide correspond to the digestion conditions used to produce that peptide (G, Glu-C ; GT, Glu-C–trypsin combination; T, trypsin). Branched symbols above the peptide sequence indicate sites of N-linked glycosylation. The amino acids enclosed in boxes indicate hydrophobic residues that may be recognized by UGGT, whereby the average hydropathy value indicates a hydrophobic patch for each tripeptide sequence. The shaded regions highlight the positions of hydrophobic amino acids, with respect to the glycan, for optimal recognition by UGGT.

Monoglucosylation of purified glycopeptides

The data shown for each experiment in Fig. 4 correspond to the monoglucosylation of mixtures of proteolytic digests of each glycoprotein with recombinant UGGT. To ensure that the identified ranked order of glucosylation was not a consequence of competition between the different glycopeptides in the mixture, the tryptic glycopeptides T_{248–270} (from group 1 in Fig. 4) and T_{386–395} (from group 3 of Fig. 4) from AcP were purified from the AcP digest by high-performance liquid chromatography (HPLC; Fig. 6B) and used as substrates. A sevenfold difference between the two glycopeptides in the incorporation of glucose was calculated, which was consistent with the ninefold difference observed between the corresponding peptides in the tryptic mixture (Fig. 4; compare peptides 4 and 21).

Rank	Protein	Peptide	Sequence	x	s.d.
1	AcP	GT _{449–466}	LTFYWDWNTTHYNASLLR	0.650	0.016
2	AcP	G _{449–467}	LTFYWDWNTTHYNASLLRQ	0.526	0.039
3	AcP	G _{427–448}	KRVAGTDFLKVCNVSSVSNSTE	0.130	0.010

Fig. 5 | Rank order of total glucosylation of doubly glycosylated tryptic glycopeptides. These peptides were part of the reaction mixture from acid phosphatase (AcP), as described for the experiments detailed in Fig. 4. The total moles of glucose incorporated per mole of glycopeptide are indicated as mean values (x) from three experiments carried out for each glycopeptide. The letters in front of each peptide correspond to the digestion conditions used to produce that peptide (G, Glu-C; GT, Glu-C–trypsin combination). Branched symbols above the peptide sequence indicate sites of N-linked glycosylation. See Fig. 4 for more details.

Fig. 6 | Hydropathy analysis of ranked glycopeptides and UDP–glucose glycoprotein–glucosyltransferase monoglucosylation of two purified tryptic glycopeptide substrates. (A) Hydropathy plots for monoglucosylated glycopeptides from acid phosphatase using the Kyte–Doolittle algorithm (see Methods). The x axis shows the position of each amino acid with respect to the N-linked glycan (position 0), where position 1 represents the amino acid immediately carboxy-terminal to the asparagine residue. The top panel shows the average Kyte–Doolittle value at each position for the group 1 glycopeptides from Fig. 4. The hydrophobic amino acids give positive values, and in a three-amino-acid window, a hydrophobic patch is defined as a positive average value for three consecutive amino acids. Note that the main difference between AcP_{101–118} and AcP_{101–114} is the hydrophobic patch downstream from residues 12–14. (B) Monoglucosylation of purified glycopeptides. AcP (280 μ g) was digested with trypsin. Purification of glycopeptides T_{248–270} and T_{386–395} of AcP was followed by micro-liquid-chromatography–electrospray-mass-spectrometry as described for Fig. 2B. Each purified glycopeptide was then monoglucosylated with recombinant UDP–glucose glycoprotein:glucosyltransferase (25 nM) in separate reactions containing 14, 28 or 56 μ M each purified glycopeptide and 1 mM UDP–glucose at pH 8 for 1.5 h at 37 °C. The molar incorporation of glucose (peak ratio) for each reaction was calculated as for Fig. 2B.

DISCUSSION

Previous studies have been unable to show that glycopeptides are efficiently recognized substrates for glucosylation by UGGT (Helenius, 2001). The engineered DT111 yeast strain (Tessier *et al.*, 2000) was used to produce homogeneous Man9-GlcNAc2-containing peptides with different amino-acid sequences, which were recognized by UGGT with a wide range of efficiencies (Fig. 4). Close inspection of the amino-acid sequences of the glycopeptides revealed two criteria for efficient recognition by UGGT: first, the presence of a polypeptide extension of ~12 amino acids from the N-linked glycan (Fig. 4). Second, hydrophobic patches, within both 1–3 and 6–14 amino acids, on the carboxy- or amino-terminal sides from the N-linked glycan (Fig. 4; shaded regions), as assessed using a number of hydrophobicity plots of the individual glycopeptides (Fig. 6A). Although there are clear peptide composition and sequence preferences for efficient UGGT substrates, no common structural elements that distinguish good and poor substrates were found

using a variety of algorithms to predict secondary structures (Appel *et al.*, 1994). Our data show that the amino-acid sequence of glycopeptides profoundly affects their recognition as substrates for UGGT. Residues close to the N-linked glycan seem to be the main determinants for recognition, and the presence of several proximal glycans does not seem to enhance recognition by UGGT (see Fig. 5, glycopeptide 3). Although our data predict that the recognition of incompletely folded glycoproteins by UGGT is affected by the sequence surrounding the glycan, we cannot exclude the possibility that these peptide-recognition elements could be provided by amino-acid residues brought into proximity of the glycans in folding intermediates. Therefore, for glycoproteins with multiple N-linked glycans, monoglucosylation of each glycan by UGGT would depend on the proximal peptide elements exposed during the folding process. This differential monoglucosylation could influence the behaviour of these glycoproteins in the calnexin cycle and in associated quality-control pathways in the ER.

METHODS

Mutant yeast glycoprotein substrates for recombinant rat UDP-glucose glycoprotein-glucosyltransferase. Uniform Man9-GlcNAc2 glycoprotein substrates were produced using the *Saccharomyces cerevisiae* mutant strain DT111, in which the *mnn1*, *mns1* and *och1* genes were deleted (Tessier et al., 2000).

The medium used for the production of AcP has been described previously (Tessier et al., 2000). The medium used for the production of β -Glc was prepared as described in Ramirez et al. (1989) with the following modifications: minimal medium SD-Leu was supplemented with 2% glucose and 0.15 M KCl, and was buffered with 0.16 M citrate/0.08 M phosphate, pH 5.8. α -Gal was expressed from the DT111 strain transformed with the *mel1* plasmid pMP550 (Post-Beittenmiller et al., 1984). Production of α -Gal was carried out in minimal medium SD-LEU-URA supplemented with 2% galactose, 1% raffinose and 0.15 M KCl.

Purification of Glycoproteins. Secreted glycoproteins expressed in the DT111 strain were prepared and purified on POROS 20HQ, as described previously (Tessier et al., 2000). AcP activity was monitored as described in Tessier et al. (2000). The hydrolysis of p-nitrophenyl α -D-galactopyranoside (Sigma; used at 1 mg ml⁻¹), which was carried out at 37 °C for 2–5 min in 25 mM sodium acetate, pH 5.0, was determined by measuring the optical density at a wavelength of 405 nm, and this was used to quantify α -Gal activity. The reaction was terminated by adding saturated Na₂CO₃ to one-half of the volume of the original reaction mixture. β -Glc was purified further using a 1-ml Phenyl Superose column (Pharmacia) equilibrated with 1 M ammonium sulphate, buffered with 20 mM sodium acetate, pH 5.0, on a BioCAD FPLC system (PerSeptive Biosystems). β -Glc was eluted with ~0.6 M ammonium sulphate. Measurement of β -Glc activity and termination of the reaction was carried out as described above for α -Gal. Protein concentrations were determined by measuring the extinction coefficient at a wavelength of 260 nm.

UDP-glucose glycoprotein-glucosyltransferase *in vitro* glucosylation assays. Recombinant rat liver UGGT was expressed, purified and characterized, as described previously (Tessier et al., 2000). For all monoglucosylation reactions, AcP, β -Glc and α -Gal were subjected to chemical denaturation, reduction and alkylation of any free sulphhydryls with iodoacetamide, as described previously (Zapun et al., 1997). Glucosylation reactions were buffered at pH 8 in 50 mM Tris-HCl and 5 mM CaCl₂. Monoglucosylation reactions and proteolytic digestion conditions were as described below and in the figure legends.

Proteolytic and Endo H digestion of glycoprotein substrates. Sequencing-grade endoproteinase Glu-C from *Staphylococcus aureus* (Roche) was used (mass ratio of 1:40, Glu-C:protein) in 50 mM ammonium carbonate, pH 7.8, for 18 h at 25 °C. Digestion with Glu-C was followed by evaporation to dryness and addition of 50 mM Tris-HCl, pH 8.0, 5 mM CaCl₂, with vortexing to redissolve the digested protein. Modified sequencing-grade trypsin from bovine pancreas (Roche) was used (mass ratio of 1:40, trypsin:protein) in 50 mM Tris-HCl, pH 8.0, 5 mM CaCl₂ for 18 h at 37 °C. Digestion with trypsin was followed by boiling the reaction mixture for 15 min to inactivate the trypsin, and monoglucosylation reactions were then performed in this mixture.

Cleavage of the glycan from 2 μ g of substrate glycoprotein was carried out using 5 μ l of Endo H (Boehringer Mannheim) in 40 μ l of 100 mM sodium acetate, pH 5.0, at 37 °C for 4 h.

Mass spectrometry. MALDI-TOF mass spectra were collected on a PerSeptive Biosystems Elite-STR spectrometer using the linear mode. The matrix, sinnapinic acid (Aldrich), was prepared as a 10 mg ml⁻¹ solution in acetonitrile/methanol/water (1:1:1, v/v/v) and used for all MALDI-TOF analyses. The dimeric, singly and doubly protonated ions of bovine serum albumin were used to obtain an external calibration of the mass spectrometer. All electrospray mass spectrometry (ESMS) experiments were conducted using a Q-TOF II (Micromass) hybrid quadrupole/time-of-flight instrument (Morris et al., 1996). A modular capillary liquid chromatography system (Micromass) was used for online micro-liquid-chromatography (μ LC)-ESMS experiments. Chromatographic separations were carried out using a 15 cm x 0.32 mm Pepmap C₁₈ capillary column (LC Packings) using a linear gradient elution of 5–95% acetonitrile (0.2% HCOOH) for 40 min with a flow rate of 3.5 μ l min⁻¹. Conventional mass spectra were obtained by operating the quadrupole in a radio-frequency-only mode, while a pusher electrode was pulsed (16-kHz frequency) to transfer all ions to the time-of-flight analyser. Mass spectra were acquired using stepped orifice voltage scanning similar to that described previously (Carr et al., 1993; Bateman et al., 1998). Tandem mass spectrometry experiments on glycopeptide peaks identified in preliminary μ LC-ESMS analyses were obtained in a subsequent sample injection for sequence analysis. Glycopeptide precursor ions were selected by the first quadrupole, while a pusher electrode was pulsed to transfer fragment ions formed in the RF-only hexapole cell to the time-of-flight analyser. Mass spectral resolution (50% FWHM definition) was typically 4,000–5,000. Scan durations of 1 s and 2 s were set for conventional and tandem MS acquisition modes, respectively. Collision activation was carried out using argon as a collision gas, with a 25-V offset between the direct current voltage of the entrance quadrupole and the RF-only hexapole cell. Data were acquired and processed in the Mass Lynx Window NT based data system (Micromass).

Primary sequence analysis of glycopeptides. The primary sequences of the 14 most highly ranked glycopeptides from Fig. 4 were analysed for a wide variety of characteristics using 42 of the available algorithms found in the ProtScale primary structure analysis tool on the ExPASy website (<http://ca.expasy.org/cgi-bin/protscale.pl>; Appel et al., 1994). For each algorithm, a window size of five was used with a relative weight of the window edges compared with the window centre of 10%, an exponential weight-variation model and no normalization of the scale. The numeric output of each algorithm was compared between the top 14 glycopeptides in Fig. 4. The hydropathy plots from four algorithms (Hopp & Woods, 1981; Kyte & Doolittle, 1982; Mohana Rao & Argos, 1986; Cowan & Whittaker, 1990) showed clear differences between the group 1 and group 2 glycopeptides, as shown by the Kyte–Doolittle plots in Fig. 6A. For the glycopeptides ranked 8–10 (Fig. 4), the amino-acid sequence N-terminal to the N-linked glycan was also analysed.

Supplementary information is available at *EMBO reports* online (<http://www.emboreports.org>)

ACKNOWLEDGEMENTS

The authors thank Micromass for the loan of a Q-TOF instrument, D. Krajcarski for technical assistance in the preliminary MALDI-TOF experiments and A. Migneault for assistance with the figures. This work was supported by grants from the Canadian Institutes of Health Research to D.Y.T. and J.J.M.B.

REFERENCES

- Appel, R.D., Bairoch, A. & Hochstrasser, D.F. (1994) A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. *Trends Biochem. Sci.*, **6**, 258–260.
- Bateman, K.P., White, R.L., Yaguchi, M. & Thibault, P. (1998) Characterization of protein glycoforms by capillary zone electrophoresis–nanoelectrospray mass spectrometry. *J. Chromatogr. A*, **794**, 327–344.
- Carr, S.A., Huddleston, M.J. & Bean, M.F. (1993) Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography–mass spectrometry. *Protein Sci.*, **2**, 183–196.
- Cowan, R. & Whittaker, R.G. (1990) Hydrophobicity indices for amino-acid residues as determined by high-performance liquid chromatography. *Peptide Res.*, **3**, 75–80.
- Helenius, A. (2001) Quality control in the secretory assembly line. *Phil. Trans. R. Soc. Lond. B*, **356**, 147–150.
- Hopp, T.P. & Woods, K.R. (1981) Prediction of protein antigenic determinants from amino-acid sequences. *Proc. Natl Acad. Sci. USA*, **78**, 3824–3828.
- Kyte, J. & Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105–132.
- Mohana Rao, J.K. & Argos, P. (1986) A conformational preference parameter to predict helices in integral membrane proteins. *Biochim. Biophys. Acta*, **869**, 197–214.
- Morris, H.R., Paxton, T., Dell, A., Langhorne, J., Berg, M., Bordoli, R.S., Hoyes, J. & Bateman, R.H. (1996) High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.*, **10**, 889–896.
- Parodi, A.J. (2000) Protein glucosylation and its role in protein folding. *Annu. Rev. Biochem.*, **69**, 69–93.
- Post-Beittenmiller, M.A., Hamilton, R.W. & Hopper, J.E. (1984) Regulation of basal and induced levels of the MEL1 transcript in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **4**, 1238–1245.
- Ramirez, M., Hernandez, L.M. & Larriba, G. (1989) A similar protein portion for two exoglucanases secreted by *Saccharomyces cerevisiae*. *Arch. Microbiol.*, **151**, 391–398.
- Ritter, C. & Helenius, A. (2000) Recognition of local glycoprotein misfolding by the ER folding sensor UDP–glucose:glycoprotein glucosyltransferase. *Nature Struct. Biol.*, **7**, 278–280.
- Schrag, J.D., Bergeron, J.J., Li, Y., Borisova, S., Thomas, D.Y. & Cygler, M. (2001) The structure of calnexin, an ER chaperone involved in quality control of protein folding. *Mol. Cell*, **8**, 633–644.
- Sousa, M. & 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.C., Ferrero-Garcia, M.A. & Parodi, A.J. (1992) Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP–Glc:glycoprotein glucosyltransferase. *Biochemistry*, **31**, 97–105.
- Tessier, D.C., Dignard, D., Zapun, A., Radomska-Pandya, A., Parodi, A.J., Bergeron, J.J. & Thomas, D.Y. (2000) Cloning and characterization of mammalian UDP–glucose glycoprotein: glucosyltransferase and the development of a specific substrate for this enzyme. *Glycobiology*, **10**, 403–412.
- Zapun, A., Petrescu, S.M., Rudd, P.M., Dwek, R.A., Thomas, D.Y. & Bergeron, J.J. (1997) Conformation-independent binding of monoglucosylated ribonuclease B to calnexin. *Cell*, **88**, 29–38.
- Zapun, A., Darby, N.J., Tessier, D.C., Michalak, M., Bergeron, J.J. & Thomas, D.Y. (1998) Enhanced catalysis of ribonuclease B folding by the interaction of calnexin or calreticulin with ERp57. *J. Biol. Chem.*, **273**, 6009–6012.