

## Analytical Glycobiology

# Analysis of site-specific *N*-glycan remodeling in the endoplasmic reticulum and the Golgi

Ivan Hang<sup>2,5</sup>, Chia-wei Lin<sup>2,5</sup>, Oliver C Grant<sup>4</sup>, Susanna Fleurkens<sup>2</sup>, Thomas K Villiger<sup>3</sup>, Miroslav Soos<sup>3</sup>, Massimo Morbidelli<sup>3</sup>, Robert J Woods<sup>4</sup>, Robert Gauss<sup>1,2</sup>, and Markus Aebi<sup>1,2</sup>

<sup>2</sup>Institute of Microbiology, Department of Biology, <sup>3</sup>Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology, ETH Zurich, CH-8093 Zurich, Switzerland, and <sup>4</sup>Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, USA

<sup>1</sup>To whom correspondence should be addressed: e-mail: markus.aebi@micro.biol.ethz.ch (M.A.); robert.gauss@micro.biol.ethz.ch (R.G.)

<sup>5</sup>These authors contributed equally to this work.

Received 12 June 2015; Revised 14 July 2015; Accepted 27 July 2015

### Abstract

The hallmark of *N*-linked protein glycosylation is the generation of diverse glycan structures in the secretory pathway. Dynamic, non-template-driven processes of *N*-glycan remodeling in the endoplasmic reticulum and the Golgi provide the cellular setting for structural diversity. We applied newly developed mass spectrometry-based analytics to quantify site-specific *N*-glycan remodeling of the model protein Pdi1p expressed in insect cells. Molecular dynamics simulation, mutational analysis, kinetic studies of *in vitro* processing events and glycan flux analysis supported the defining role of the protein in *N*-glycan processing.

**Key words:** endoplasmic reticulum glycosylation, Golgi, Golgi glycosylation, site-specific glycosylation

### Introduction

*N*-Linked glycosylation is an abundant posttranslational modification present on secretory proteins in all domains of life (Dell et al. 2010). In eukaryotes, attachment and subsequent modification of *N*-glycans affect the folding of glycoproteins and regulate their secretion, e.g. by providing signals to the endoplasmic reticulum (ER) quality control machinery (Shental-Bechor and Levy 2009; Aebi et al. 2010). On the final glycoprotein product, *N*-glycans act as ligands that mediate cell–cell interactions of cell surface glycoproteins and fine-tune the function of mature glycoproteins (van Kooyk and Rabinovich 2008).

Biosynthesis of *N*-glycoproteins starts with the stepwise assembly of a lipid-linked oligosaccharide precursor at the ER membrane. Oligosaccharyl transferase then transfers the G3M9Gn2 carbohydrate (G: glucose, M: mannose, Gn: *N*-acetylglucosamine) from the isoprenoid lipid carrier onto the asparagine side-chain of the N-X-S/T consensus sequence of polypeptides in the ER lumen (Kelleher and Gilmore 2006). Initial modifications of the *N*-glycan in the ER involve

successive removal of glucose and  $\alpha$ -1,2-mannosyl residues to generate oligomannose-type glycans (Hebert et al. 2005; Mast and Moremen 2006). It is in the Golgi where glycosyl hydrolases (GHs) and transferases (GTs) generate the heterogeneous population of carbohydrate structures presented on mature glycoproteins (Hua et al. 2012; Moremen et al. 2012).

GHs, GTs, and nucleotide sugar transporters represent the “hardware” of the Golgi remodeling pathway that converts oligomannose glycans into hybrid- and complex-type oligosaccharides (Varki 1998). This network of glycan modifying enzymes can vary between different species, tissues, and cell types and generates variable *N*-glycan structures. Glycoproteins expressed in mammalian cells generally carry an array of complex-type oligosaccharides decorated with terminal galactose and sialic acid residues (North et al. 2010). In contrast, the glycosylation machinery of insect cells produces a variety of paucimannose-type products with  $\alpha$ 1,6- or  $\alpha$ 1,3-linked fucose residues attached to the core Gn (Shi and Jarvis 2007; Hosokawa et al. 2010).

N-Glycan diversity is the result of the dynamic and competitive nature of N-glycan synthesis in the Golgi (Stanley 2011). Golgi GTs and GHs usually have different substrate specificities and act sequentially, but they may also compete for substrates (Moremen et al. 2012). Analysis of site-specific N-glycosylation of Sindbis virus glycoproteins, carboxypeptidase Y, and invertase provided the first evidence that processing of N-glycans is determined by the physical accessibility of the carbohydrates (Hsieh et al. 1983a, 1983b; Trimble et al. 1983; Hubbard 1988). Subsequently, site-specific N-glycosylation has been found on different proteins Thy1, lysosomal  $\alpha$ -mannosidase, lactoferrin, and HIV envelope glycoproteins (Parekh et al. 1987; Heikinheimo et al. 2003; Faid et al. 2006; Go et al. 2008; Hua et al. 2012; Nagae and Yamaguchi 2012) and alterations in the polypeptide sequence can yield different N-glycan structures (Yu et al. 2013).

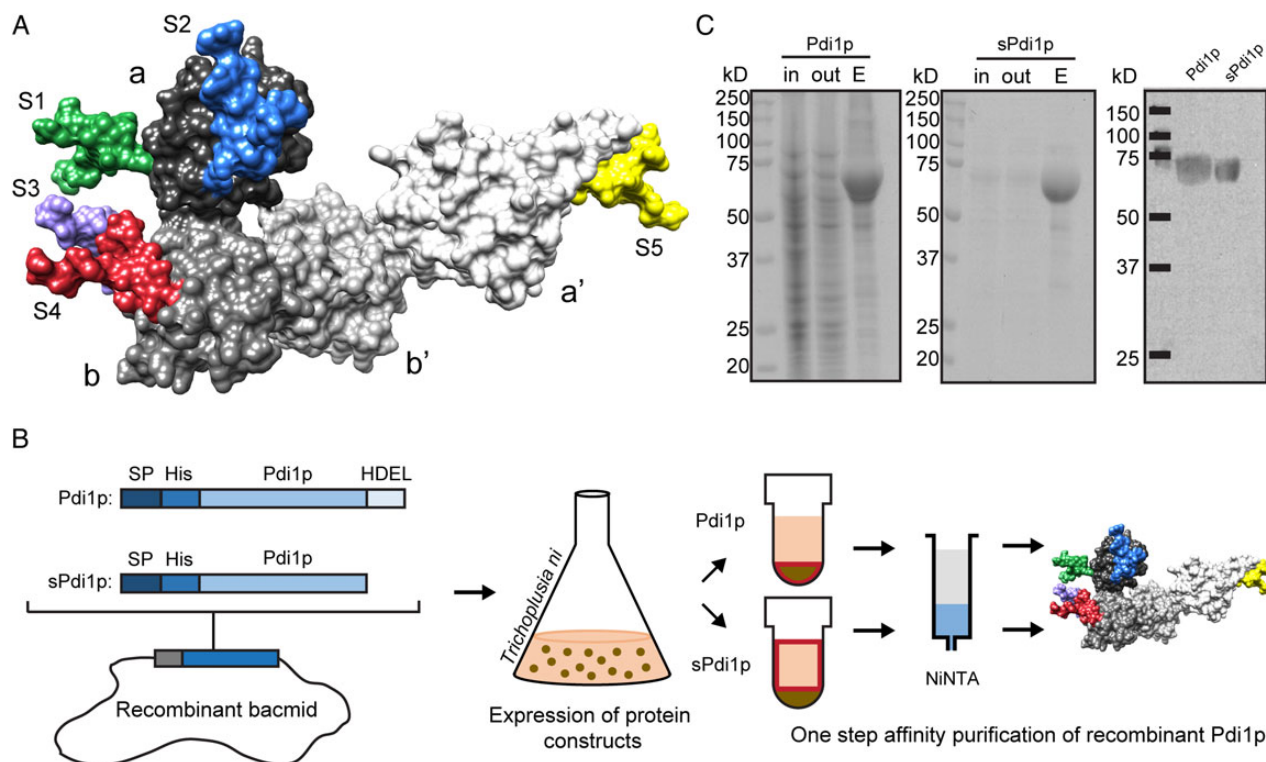
To study in detail the factors that affect site-specific N-glycosylation, we expressed a model protein, yeast Pdi1p, in insect cells and employed newly developed mass spectrometry-based analytical techniques to quantify site-specific N-glycan structures produced in the ER and the Golgi. Combining molecular dynamics simulations, mutational analyses, and in vitro assays we define the intramolecular interactions of glycans and the protein surface that are major determinants for carbohydrate modification. Implementing glycan flux analysis, we identified the rate-limiting processing steps for each glycosite. Our studies confirm the importance of the protein structure in the pathway of N-glycan processing.

## Results

### Establishing an MS-based workflow to analyze site-specific glycosylation

ER-resident protein disulfide isomerase 1 (Pdi1p) from *Saccharomyces cerevisiae* is composed of four thioredoxin-like domains (a, b, b' and a') (Wilkinson and Gilbert 2004; Tian et al. 2006). Unlike mammalian PDIs, yeast Pdi1p has five N-glycosylation sites (S1–S5) that are localized in three of the four domains (Figure 1A). Tryptic peptides containing these glycosites were used in our study (Table I). A C-terminal HDEL sequence retained the protein in the ER and early Golgi, while disruption of this sequence resulted in the secretion of the protein into the medium (sPdi1p), allowing us to analyze ER- and Golgi-located processing separately (Figure 1B). Pdi1p proteins were expressed in *Trichoplusia ni* insect cells using the baculovirus expression system and purified by Ni-NTA affinity chromatography via an N-terminal His<sub>10</sub>-tag. SDS-PAGE and immunoblot analysis verified that the purity was suitable for the following MS measurements (Figure 1C).

To assess site-specific glycan heterogeneity, we employed the filter-assisted sample preparation (FASP) method (Wisniewski et al. 2009). Peptides were analyzed on a nanoLC-LTQ Velos Orbitrap operated in scheduled data-dependent acquisition mode, one MS scan followed by 10 HCD MS/MS scans. Based on the unique fragmentation of glycopeptides by the HCD method, we developed the ExtractMgf algorithm



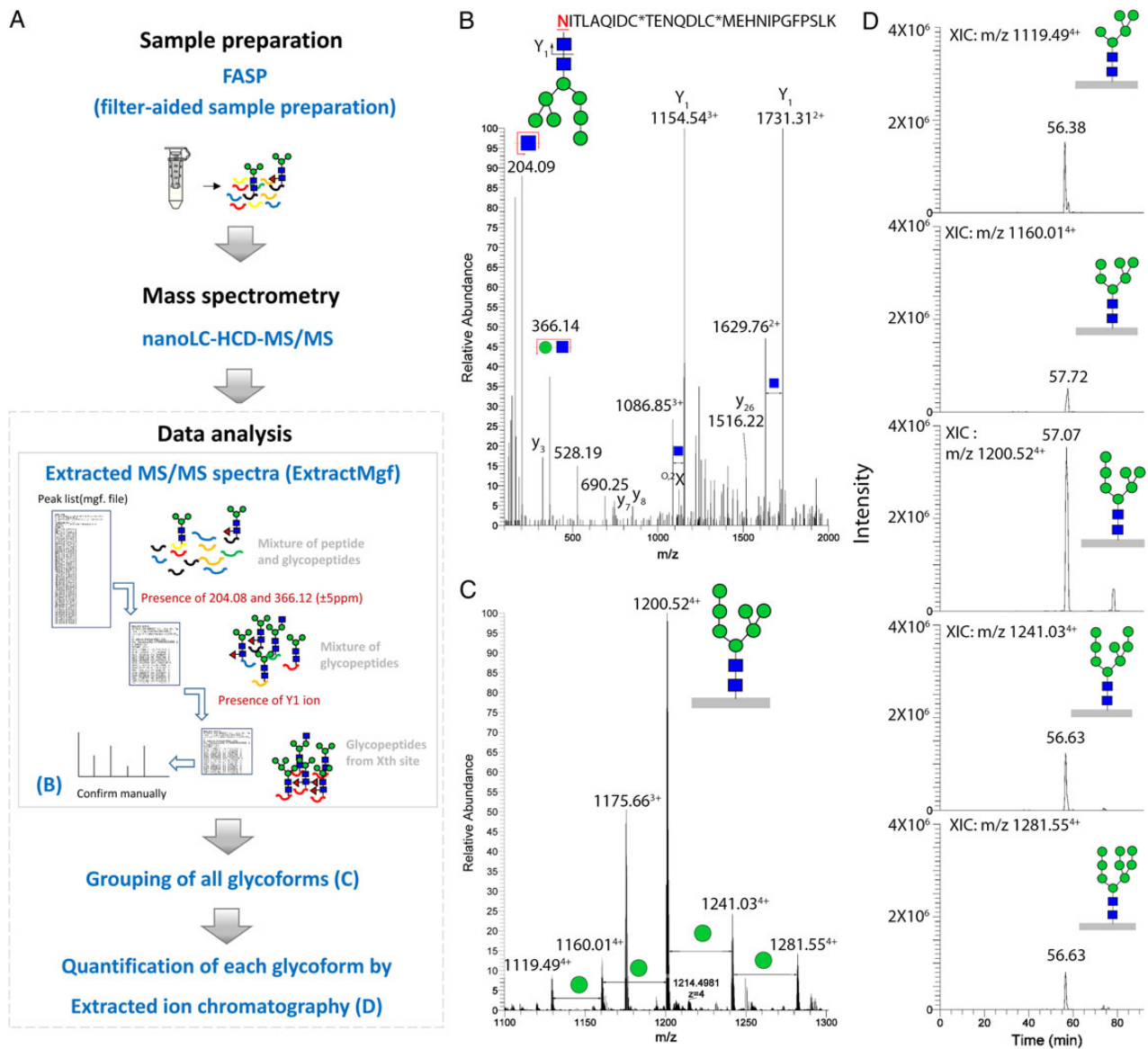
**Fig. 1.** Purification of recombinant Pdi1p and sPdi1p from insect cells. (A) Model of glycosylated Pdi1p. M9Gn2 N-glycans were modelled onto the five glycosites of the crystal structure of Pdi1p (PDB: 2B5E; [www.glycam.org](http://www.glycam.org)). Glycans are depicted in colors: S1 green, S2 blue, S3 purple, S4 red and S5 yellow. Different domains of Pdi1p are in different shades of gray. (B) Schematic representation of Pdi1p and sPdi1p purification. *Trichoplusia ni* cells were infected with recombinant viruses carrying expression copies of ER retained Pdi1p or secreted Pdi1p (sPdi1p). Pdi1p was purified via NiNTA chromatography from cell lysates; sPdi1p was isolated from culture supernatants. ER retention of full-length sPdi1p was disrupted by the presence of two additional amino acids at the protein's C-terminus (HDELLE) (Raykhel et al. 2007). SP: signal peptide; His: His<sub>10</sub> tag, HDEL: ER retrieval signal. (C) Purification of Pdi1p and sPdi1p. Samples of input (in), flow-through (out) and eluted (E) fractions were analyzed by SDS-PAGE and stained with Coomassie (left/middle), or analyzed by immunoblot using anti-His5 antibodies (right).

**Table 1.** Glycosylation site of yeast Pdi1p

Site	Sequence	[M+H] <sup>+</sup>
1	NITLAQIDC*TENQDLC*MEHNIPGFPSLK	3258.52
2	NSDVNNSIDYEGPR	1579.70
3	QSQPAVAVVADLPAYLANETFVTPVIVQSGK	3212.71
4	IDADFNATFYSMANK	1707.78
5	LAPTYQELADTYANATSDVLIAK	2468.25

A list of the five tryptic peptides containing a glycosylation sequon monitored in this study. C\* represents carbamidomethylated cysteine.

to speed up manual analyses of MS/MS spectra, as exemplified for the S1 glycopeptide (Segu and Mechref 2010) (Figure 2A). From 3648 measured spectra, the glycan oxonium ions [HexNAc]<sup>+</sup> 204.09 and [HexNAc + Hex]<sup>+</sup> 366.14 were used to identify the MS/MS scans of all glycopeptides from the peak list. Next, the Y1 ion corresponding to *m/z* of the peptide plus one HexNAc was used to identify the peptide backbone. In the case of S1, 1731.3034 corresponding to [S1 + HexNAc + 2H]<sup>2+</sup> was used to sort out all MS/MS spectra from that glycopeptide. After two sorting runs, the remaining 20 spectra were verified manually. Since <sup>0,2</sup>X ring cleavage on the single



**Fig. 2.** Overall workflow for glycopeptide analysis by mass spectrometry. (A) Purified proteins were processed by FASP. The mixture of peptides and glycopeptides was analyzed by LC-HCD mass spectrometry. Raw data were transformed into a peak list and processed by ExtractMgf. For relative quantification, XIC of each glycoform was plotted and its corresponding peak area was integrated. The relative abundance of each form was calculated. Symbols represent monosaccharides: mannose (circle), *N*-acetylglucosamine (square) and fucose (triangle) (B) One MS/MS spectrum at *m/z* 1200.54 (*z* = 4) was assigned to the first glycosylation site of Pdi1p containing M7Gn2 glycoform. Glycan oxonium ions and doubly/triply charged Y1 at *m/z* 1154.54 (*z* = 3)/1731.31 (*z* = 2) are indicated. The nomenclature of peptide fragment ions and glycan fragmentation ions was described previously (Roepstorff and Fohlman 1984; Domon et al. 1990; Dell et al. 1994). (C) The overall glycosylation profile of S1 obtained by grouping of MS spectra. Corresponding ions are indicated. (D) XIC of each glycoform sharing the same peptide backbone. See Supplementary data, Figure S2 and Table S2 for quantification of RNaseB glycans. (B–D) The structures of glycan isomers were assigned according to the biosynthesis pathway of insect cells shown in Figure 7A. This figure is available in black and white in print and in color at *Glycobiology* online.

remaining HexNAc was a frequent event and resulted in neutral loss of 120 Da from the Y1 ion (Figure 2B; Supplementary data, Figure S1), we used the triple peaks, Y1,  $^{0,2}X$  and [peptide + H] $^{+}$ , to manually confirm the identity of the Y1 ion in HCD spectra. One example of an S1 MS/MS spectrum of the M7Gn2 is shown in Figure 2B.

Since glycopeptides containing the same peptide backbone co-elute when applied to reverse phase liquid chromatography, we identified the different glycan structures of one site by grouping MS spectra based on the presence of Y1 ions observed in the original MS/MS spectra. In the example given, the overall glycosylation profile of S1 ranged from M5Gn2 to M9Gn2 (Figure 2C). Finally, we quantified the amount of each glycan structure by its extracted ion chromatogram (Figure 2D). We verified the accuracy of the method by analyzing bovine RNase B that carries a known mixture of oligomannose glycans on a single glycosite (Supplementary data, Figure S2).

### Pdi1p glycans are differentially processed

In an initial experiment, we analyzed the N-linked glycans of Pdi1p retained in the ER and early Golgi. As expected, we found oligomannose structures; yet, we did not observe a defined structure per site, but rather a mixture of glycans processed to varying degrees (Figure 2D). The processing of N-glycans was significantly different for the five sites analyzed: while S2, S3 and S5 mainly carried M7Gn2 to M5Gn2 glycans, S4 contained a higher fraction of M9Gn2 and M8Gn2 glycans (Figure 3A and B). S1 showed an intermediate processing pattern with M7Gn2 as the most prevalent structure. By comparing the glycan profiles for each site using Euclidean distance and hierarchical clustering analysis, we confirmed that S4 showed the most distinct glycan pattern (Figure 3C). These results indicated that the location of an N-glycan on Pdi1p was an essential parameter influencing the processing of the glycan by mannosidases of the ER and early Golgi.

### Interactions with the protein surface reduce accessibility of the S4 glycan

In order to test whether glycan–protein interactions influence oligosaccharide processing, we generated a three-dimensional model of a truncated version of Pdi1p composed of the a and b domain and with an M9Gn2 glycan attached to S4. To remove any bias from the initial glycan orientation, an explicitly solvated molecular dynamics simulation was performed for 0.5  $\mu$ s. Throughout the simulation, the B-branch of the S4 glycan remained in close contact with the surface of the b-domain, while the A- and C-branch formed interactions with the a-domain (Figure 4A; Supplementary data, Movie). Once formed, the interactions between the S4 glycan (located on the b-domain) and the a-domain remained stable over the course of the simulation timescale. The contacts formed with the a-domain reduced the accessibility of the S4 glycan, which correlated with the attenuated processing of the S4 glycan by mannosidases (Figure 4B). Notably, the interactions formed by the S4-glycan with the surface of the a-domain altered the relative orientation and dynamics of the a- and b-domains while the unglycosylated protein fluctuated about the crystal structure conformation (Figure 4C).

To confirm these *in silico* predictions, we designed truncated variants of Pdi1p, each containing the glycosylated b-domain with an N-terminal His<sub>10</sub>-tag and a C-terminal HDEL sequence (Figure 5A). The different proteins were expressed, purified, and the peptides were analyzed by the ExtractMgf workflow (Figure 5B and C). The glycan profiles on S4 of full-length Pdi1p, the abb'- and the ab-domain glycoprotein were almost identical. However, when we removed the a-domain, the distribution of glycoforms changed and shifted to

smaller, more processed glycans (Figure 5C; Supplementary data, Figure S3), resulting in a different glycosylation pattern (Figure 5D). These findings supported the hypothesis that the interaction of a glycan with the surface of the covalently linked protein altered the accessibility of the glycan to the processing machinery.

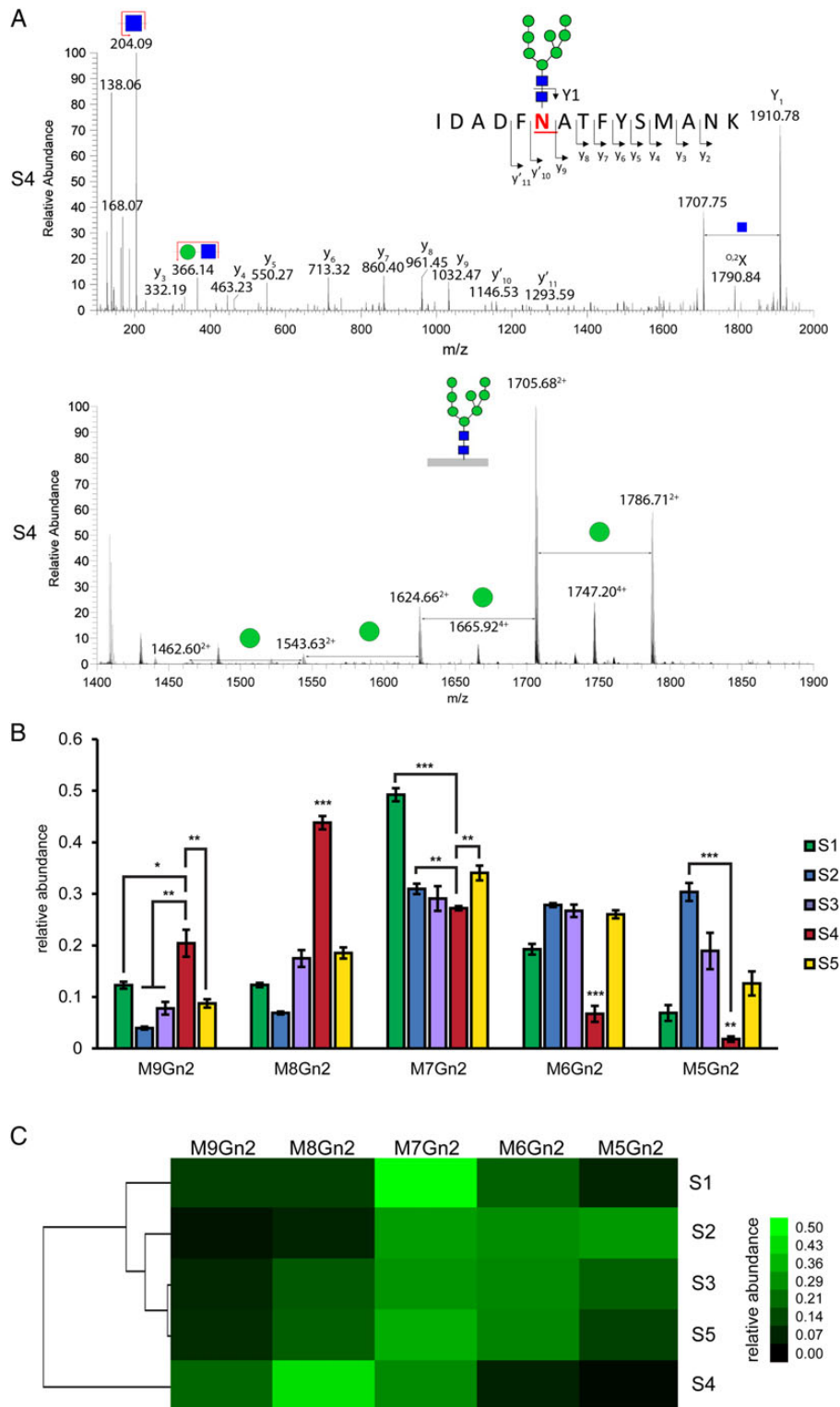
### N-Glycans on Pdi1p represent different substrates to Mns1p

According to this hypothesis, the five N-linked glycans should display different processing kinetics by a given enzyme. We addressed this by an *in vitro* experiment and analyzed the processing of the glycans of Pdi1p by ER  $\alpha$ -mannosidase from *S. cerevisiae*, Mns1p. Mns1p recognizes the terminal mannose residues of both, the B- and C-branch of an M9Gn2 glycan and specifically hydrolyzes the  $\alpha$ -1,2-glycosidic linkage of the terminal mannose of the B-branch (Jelinek-Kelly et al. 1985; Ziegler and Trimble 1991; Vallée et al. 2000). Pdi1p carrying M9Gn2 glycans was obtained from insect cells that were incubated with the  $\alpha$ -mannosidase inhibitor kifunensine during protein production. MS analysis confirmed that Pdi1p was homogeneously glycosylated with M9Gn2 N-glycans on all five sites (Figure 6A; Supplementary data, Figure S4). Mns1p was expressed in *T. ni* cells and purified by glutathione affinity chromatography (Supplementary data, Figure S4A). *In vitro* assays were performed with a 100 $\times$  molar excess of glycosylated Pdi1p (Karaveg and Moremen 2005). Aliquots were taken at different time points, and site-specific glycosylation was determined by quantitative MS analysis. Three minutes after initiating the reaction 84, 80, 62, and 54% of M9Gn2 were processed to M8Gn2 on S2, S3, S5, and S1, respectively, whereas the conversion of M9Gn2 on S4 was only 37% (Figure 6B). After 10 min, all but S4 glycan showed >90% processing; conversion to M8Gn2 was completed for all five glycans after 60 min. We concluded that interaction of a glycan with the protein surface competed with enzyme binding and, thus, reduced enzymatic turnover of the glycan: the more favorable this interaction the slower the processing.

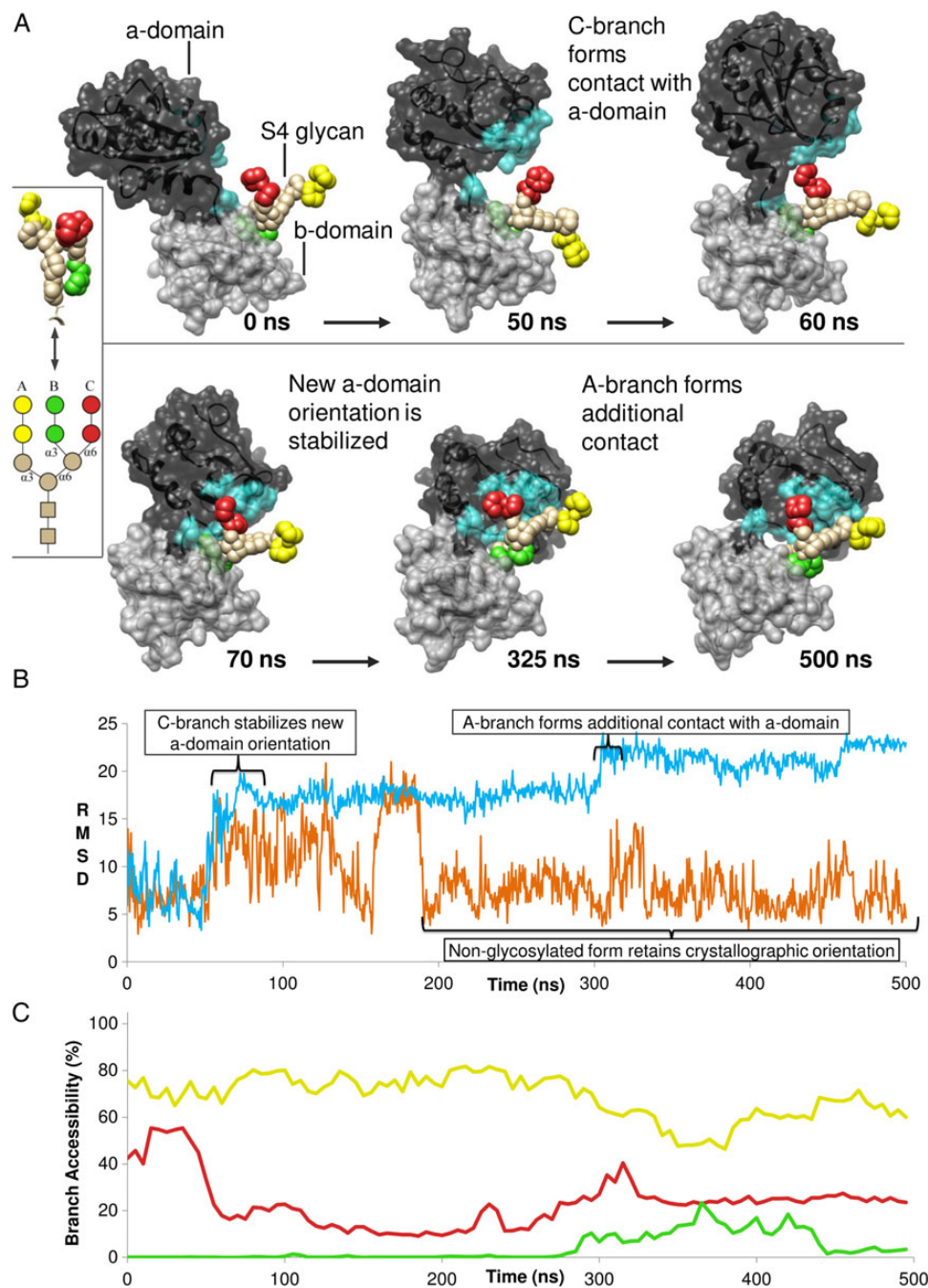
### N-Glycans of sPdi1p are processed differently in the Golgi

In the absence of the HDEL ER retrieval sequence, His<sub>10</sub>-tagged sPdi1p was secreted into the medium, allowing us to analyze the complete glycan-processing pathway. Employing the Extractmfg workflow, we found glycan structures reflecting the whole glycosylation machinery of *T. ni* cells, yet, glycans on the five glycosites of Pdi1p were processed differently (Figure 7A and B). S1, S2 and S3 glycans showed a similar distribution of oligomannose, paucimannose-type and complex-type structures. However, the ratio of di-fucosylated glycans was significantly higher on S2 and S3 when compared with S1 and oligomannose glycans were almost absent on S5. In contrast to these glycosites, S4 maintained a high ratio of oligomannose oligosaccharides and paucimannose-type glycans that were neither fucosylated nor decorated with a second Gn. Interestingly, one of the abundant glycoforms on S4 was GnM4Gn2, an intermediate product of  $\alpha$ -mannosidase II processing (Figure 7B; Supplementary data, Figure S5). Hierarchical clustering analysis confirmed that the glycan profiles of S2 and S3, and S1 and S5, respectively, were closely related while the S4 glycan profile was different to all the other sites (Figure 7C). These data confirmed that for the remodeling of N-glycans in the Golgi, the localization of the glycan on the protein surface was a major factor that determined both the structure of an oligosaccharide and the relative abundance of this structure.





**Fig. 3.** Pdi1p retained in the ER and early Golgi displays site-specific glycan profiles. **(A)** Analysis of M8Gn2 on S4. The MS/MS spectrum of  $m/z$  1705.68 ( $z=2$ ) was obtained as described in Figure 2. The peptide was identified by the Y1 ion (S4 peptide plus HexNAc at  $m/z$  1910.78). The peptide sequence is displayed together with the attached M8Gn2 glycan. The underlined *N* represents the *N*-glycosite and  $y'$  represents  $y$  ion without a glycan. **(B)** Site-specific *N*-glycan profile of Pdi1p. Pdi1p was expressed and purified and site-specific glycan profiles were analyzed. The relative abundance of the different glycoforms at a given site (S1 to S5) is shown. Data represent the mean values of glycan ratios of four independent experiments with error bars indicating the standard deviation. *P*-values were calculated by paired Student's *t*-test: \*  $P < 0.05$ ; \*\*  $P < 0.01$ ; \*\*\*  $P < 0.001$ . **(C)** Similarity representation of five Pdi1p *N*-glycosylation sites based on the relative glycoform distribution. Similarity between site-specific profiles was calculated using the Euclidean distance and the dendrogram was obtained by Centroid Linkage Clustering. The color scale represents the relative glycoform abundance from Figure 3B. See Supplementary data, Figure S3 and Table S3 for spectra and raw data. This figure is available in black and white in print and in color at *Glycobiology* online.

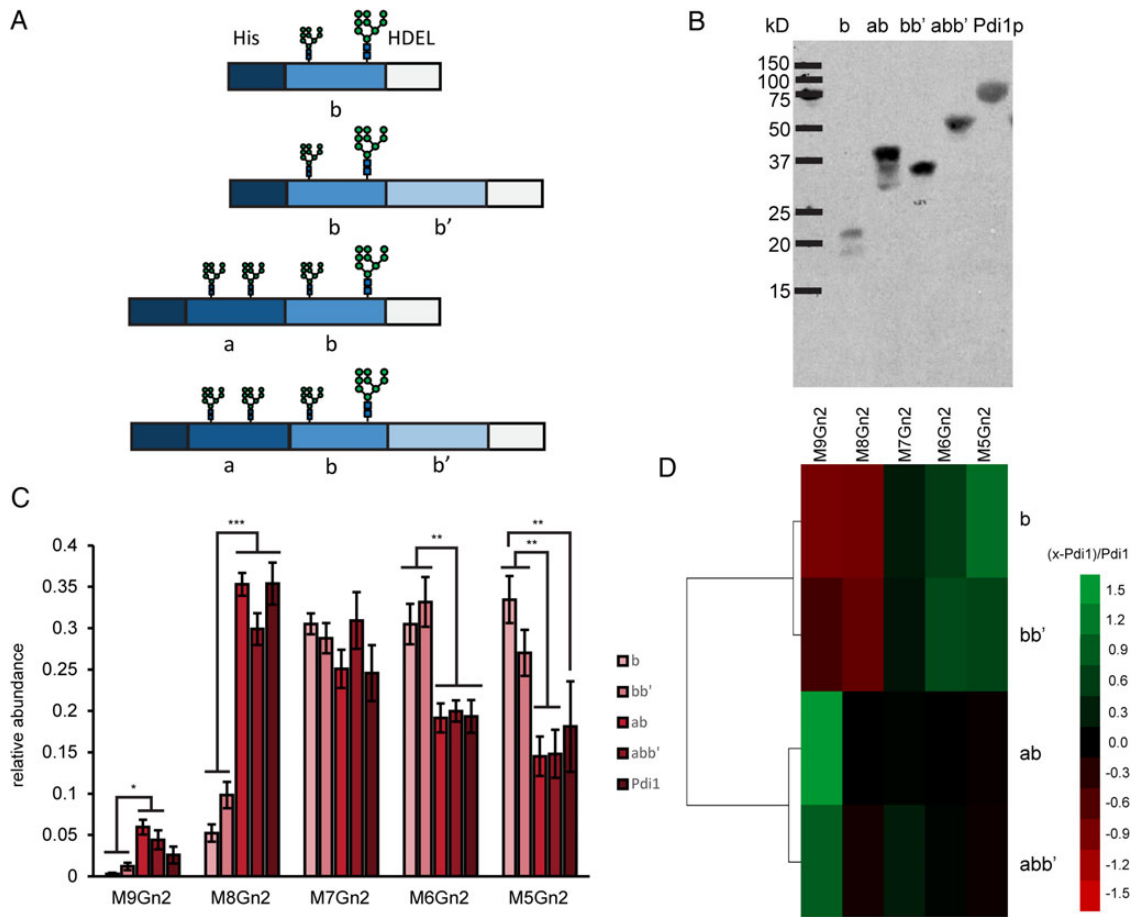


**Fig. 4.** Molecular dynamics simulation of glycosylated Pdi1p-ab. **(A)** Six snapshots from the molecular dynamics simulation of the S4 glycosylated ab-domain variant of Pdi1p, showing the contacts formed between the S4 glycan and the a-domain (blue surface). The b-domains of each snapshot are aligned to the crystal structure coordinates. The glycan is shown as van-der-Waals spheres, with the A-branch in yellow, the B-branch in green and the C-branch in red. **(B)** A plot of the RMSD of the a-domain obtained over the course of the simulation relative to the crystal structure orientation (blue line: RMSD of the a-domain with S4 glycosylated, orange line: RMSD of a-domain without S4 glycosylation). **(C)** The relative solvent accessibility of the non-reducing terminal disaccharides of the A- (yellow), B- (green) and C-branches (red) of the S4 glycan over the course of the simulation.

Consequently, the substrate properties of the five N-linked glycans of sPdi1p differ for a given processing enzyme.

To quantify the *in vivo* activity of different processing enzymes on the five N-glycans, we performed glycan flux analysis using a stoichiometric matrix to describe the connectivity of the network (see Materials and Methods for a detailed description). We calculated the

substrate fluxes at each enzyme employing the assumption that the secreted Pdi1p represented a quasi-steady state of the glycan processing in the cell. Conversion for each reaction was normalized to the incoming fluxes (Figure 7D). This analysis revealed that conversion for enzymes acting early in the pathway was generally higher than those for enzymes acting late in the processing. Substrate conversion by



**Fig. 5.** S4 glycan processing of Pdi1p is improved when the a-domain is removed. **(A)** Schematic representation of the Pdi1p variants used. His: N-terminal His<sub>10</sub> tag, HDEL: C-terminal HDEL ER retention signal. **(B)** Pdi1p variants were expressed and purified. Elution fractions were analyzed by SDS-PAGE and immunoblot using anti-His<sub>4</sub> antibodies. **(C)** Relative abundance of glycoforms on S4 of purified full-length and truncated Pdi1p, indicated in different colors. Data represent the mean values of glycan ratios of four independent experiments with error bars indicating the standard deviation \*  $P < 0.05$ ; \*\*  $P < 0.01$ ; \*\*\*  $P < 0.001$ . **(D)** Similarity representation of Pdi1p variants (indicated at the right) based on the difference in S4 glycoform abundance (given in Figure 6F) to full-length Pdi1p. The color scale represents values calculated by the equation:  $(x - \text{Pdi1}) / \text{Pdi1}$  (x is the average glycoform abundance of a specific variant; Pdi1p is the average glycoform abundance of full-length Pdi1p). Similarity between site-specific profiles was calculated as described in Figure 3C. See Supplementary data, Figure S4 and Table S4 for spectra and raw data. This figure is available in black and white in print and in color at *Glycobiology* online.

$\alpha$ -mannosidase I and II was significantly lower for the S4 glycan than for all the other glycans. On the other hand, the S2 and S3 glycans were more preferred substrates for  $\alpha$ 1,6- and  $\alpha$ 1,3-fucosyl transferases than the oligosaccharides on S1 and S5. Thus, altered conversion results in site-specific processing of the N-glycans. Importantly, this flux analysis revealed that only a subset of Golgi processing enzymes was sensitive to the localization of the N-glycan.

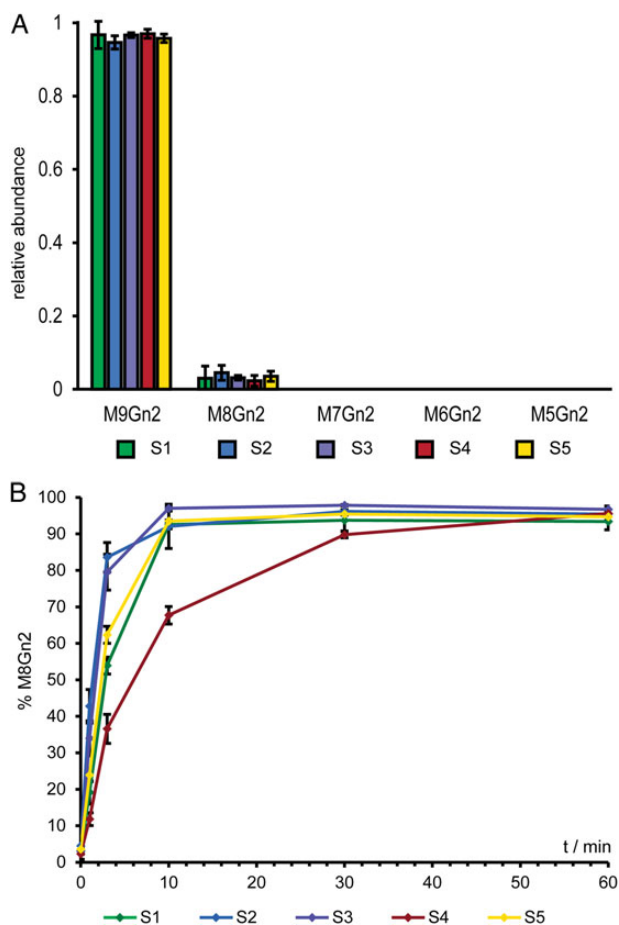
#### Site-specific processing profile of N-glycans on sPdi1p is dependent upon the protein structure and the Golgi processing machinery

We next analyzed whether protein structure also affect site-specific processing of Pdi1p in the Golgi. Again, we took advantage of the modular structure of Pdi1p and designed truncated variants as in Figure 4D, yet, without the C-terminal HDEL sequence. Constructs were expressed in *T. ni* cells and purified from the culturing media (Figure 8A). After MS-analysis and data processing, we observed that the glycans of the b- and bb'-domain constructs were more processed when compared with full-length sPdi1p and the variants containing the a-domain. The S4 glycan profiles on both b- and bb'-domain constructs contained

significantly lower amounts of oligomannose-type N-glycans, while the paucimannose-type structure M3Gn2 was more abundant (Figure 8B). Conversion analysis revealed that removal of the a-domain made the S4 glycan a more preferred substrate for  $\alpha$ -mannosidases I and II and N-acetylglucosaminidase (Figure 8C). We also noted changes in the processing pattern of the other N-glycans with respect to altered protein structure (data not shown).

#### Expression of an additional glycosyltransferase increases site-specific glycan diversity

Microheterogeneity of N-glycan structures on sPdi1p is the result of incomplete conversion by glycan-processing enzymes. In addition, site-specific processing by a subset of these enzymes resulted in site-specific glycan profiles. It is evident that increasing the number of processing steps in such a system will elevate the probability of site-specific N-glycan processing. We therefore expressed  $\beta$ 1,4-galactosyltransferase (GALT-1) from *Caenorhabditis elegans* in insect cells (Figure 9A). GALT-1 acts late in the pathway and adds a  $\beta$ 1,4-galactose to the  $\alpha$ 1,6-linked core fucose residue (Titz et al. 2009). We analyzed site-specific glycosylation profiles of sPdi1p purified from cells co-infected



**Fig. 6.** The S4 is processed more slowly by Mns1p in vitro. **(A)** N-Glycan analysis of M9Gn2-Pdi1p. *Trichoplusia ni* cells were incubated with kifunensine and infected with virus containing an expression copy of GST-tagged Pdi1p. GST-Pdi1p was affinity purified and the GST tag was removed. N-Glycan profiles were determined by MS analysis. Data represent the mean values of relative glycan abundances of three independent experiments with error bars indicating the standard deviation of mean. **(B)** M9Gn2-Pdi1p was incubated with Mns1p. At the indicated time points, aliquots were TCA precipitated and analyzed by MS. Mean values for the relative abundance of M8Gn2 at the different glycosites from three independent experiments were plotted with error bars representing the standard deviation of mean. See Supplementary data, Figure S5 and Table S5 for spectra and raw data.

with GALT-1 recombinant viruses (Figure 9B). Indeed, we observed site-specific processing: in contrast to S2, S3 and S5 oligosaccharides, GALT-1 did not process the S1 glycan, even though all of the sites presented suitable GALT-1 substrates. Thus, introducing an additional late-acting enzyme in the pathway allowed us to differentiate the glycans of Pdi1p further; the glycan profile of S1 deviated from the profile of S5 (Figure 9C). Taken together, the location of the N-linked glycan on a protein in a given setting of glycan-processing enzymes in the secretory pathway defines its processing.

## Discussion

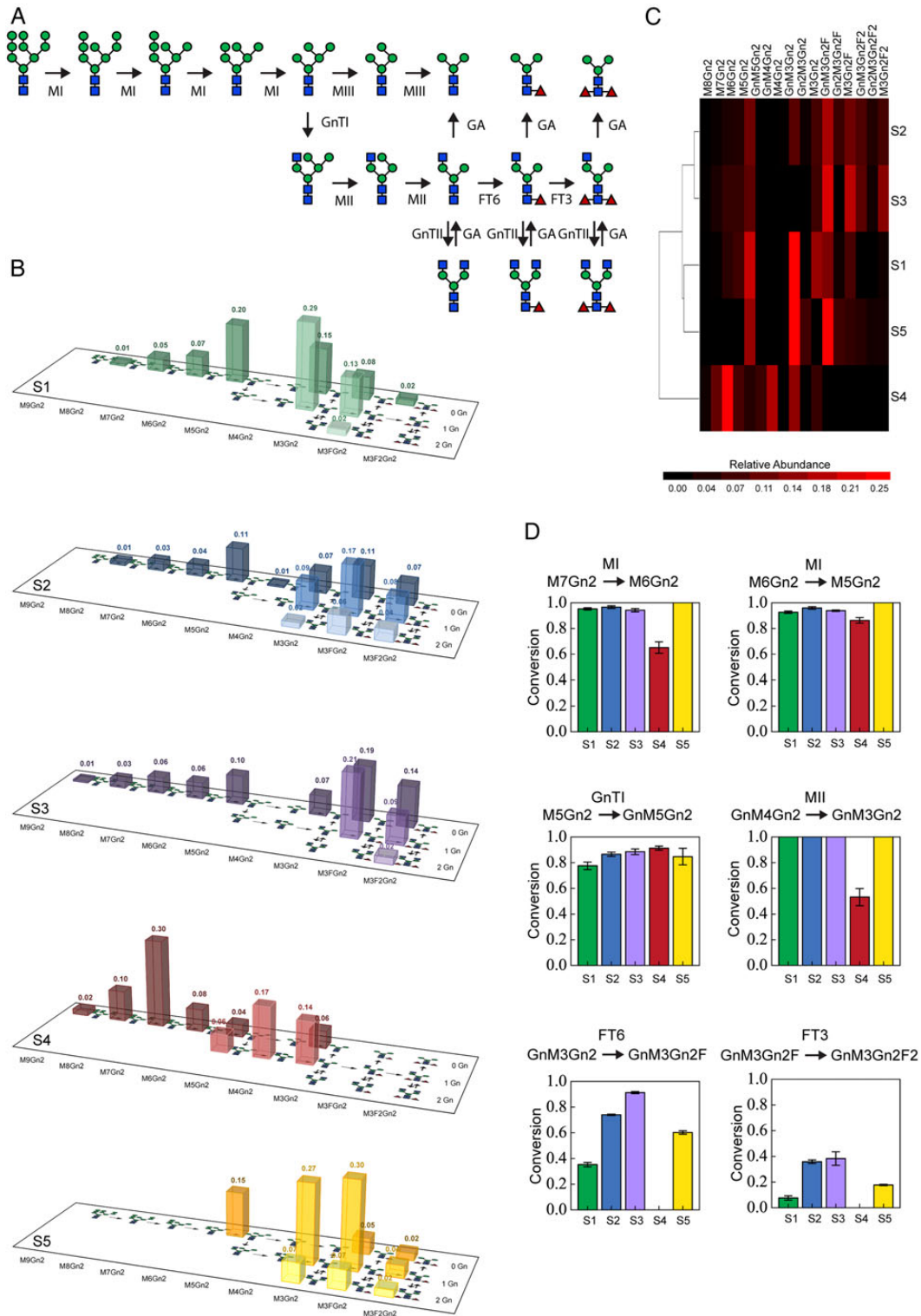
In this study, we exploited baculovirus-driven expression in insect cells to produce model glycoproteins that were retained in the ER and early Golgi (Pdi1p), or secreted into the medium (sPdi1p), respectively (Figure 1). Since yeast Pdi1p carries five N-glycans distributed over

the molecule, we could observe remodeling of individual N-linked oligosaccharides at different locations on the same protein while they were presented to the identical cellular glycan-processing machinery for the same amount of time. In addition, the protein has a modular composition of four thioredoxin-like domains that fold independently; hence, we were able to express truncated variants of Pdi1p lacking one or multiple domains without disturbing the structure of the remaining protein (Kemink et al. 1996, 1999). This set-up allowed us to assess the influence of the protein on glycan remodeling while keeping the cellular environment constant.

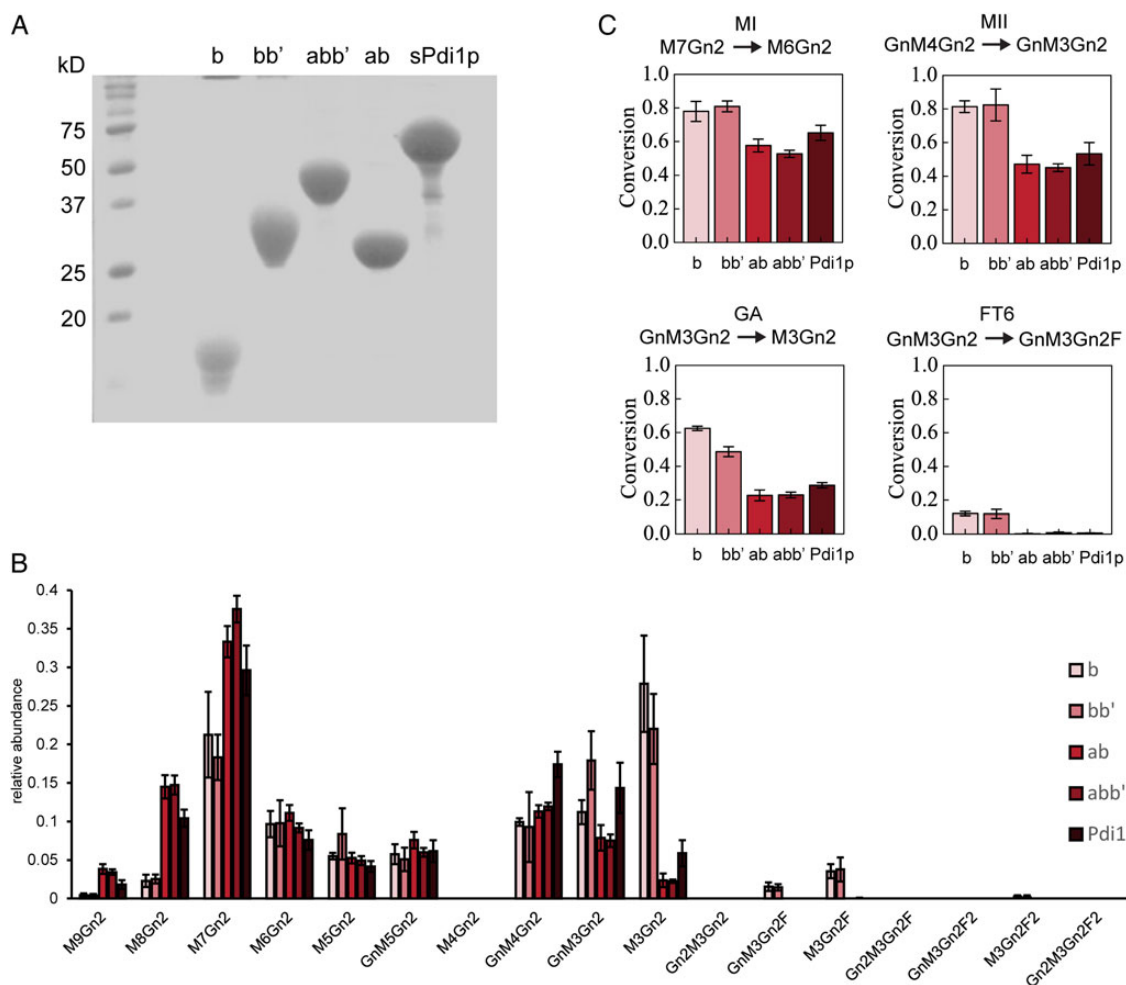
Mammalian glycans contain sialic acid residues that are easily lost during ionization in mass spectrometric analyses; consequently, it is difficult to quantify carbohydrates on mammalian glycoproteins (Powell and Harvey 1996). In contrast, insect cells produce fucosylated paucimannose structures without galactose or sialic acid, allowing us to quantify oligosaccharide structures linked to specific glycosites. To assess site-specific glycan composition, we developed an analytic workflow based on mass spectrometry operated in HCD fragmentation mode (Figure 2). Detection of Y1,  $O^2X$  and peptide peaks are frequently reported in CID MS/MS of glycopeptides by Q-TOF or TOF-TOF MS, but this peak triplet has not been exploited for identification in the HCD fragmentation method (Krokhin et al. 2004; Sparbier et al. 2007). Our data demonstrated that using the peak triplet and characteristic peptide fragment ions is a suitable tool for glycopeptide identification without further tandem mass spectrometry. Moreover, we developed and evaluated a reliable quantification method of each glycoform that allows to partly automate and speed up data analysis.

Although the glycan-processing machinery of insect cells used in this study is less complex than in mammalian cells, it implements the same underlying principles. According to the current view, a glycoprotein may transit through the Golgi too quickly to be processed completely by every possible enzyme (Stanley 2011; Moremen et al. 2012). In addition, many GHs and GTs compete for the same oligosaccharide substrate producing additional heterogeneity of glycan structures (Schachter 1991; Geisler and Jarvis 2012). Earlier studies suggested that the local protein surface influences enzyme accessibility to individual glycans (Hsieh et al. 1983a, 1983b; Trimble et al. 1983; Hubbard 1988); however, these studies did not provide direct evidence to support this hypothesis. Analyzing the glycan profiles of Pdi1p and sPdi1p, we found that the five glycosites displayed individual carbohydrate profiles throughout the secretory pathway (Figures 3 and 7). In particular, the S4 glycan was less processed in both, the ER and Golgi compartments relative to the other glycans. Molecular dynamics simulations showed that the terminal mannose residues of an M9Gn2 glycan on S4 made long-lasting contacts with specific amino acid residues of the a- and b-domain (Figure 4A). These interactions might reduce enzyme accessibility and, thus, processing of the S4 glycan. Along this line, ER mannosidase processed in vitro the S4 glycan more slowly than its counterparts on the other glycosites. Consequently, when we removed the a-domain from the protein in vivo, the glycan on S4 became more accessible and the resulting glycan profile shifted to more processed glycan structures (Figures 4 and 8). In addition, implementing glycan flux analysis, we identified the rate-limiting steps for each glycosite. Interestingly, it is only a subset of processing enzymes that are kinetically controlled (thereby generating microheterogeneity of the glycan structures) and some of these enzymes are affected by the location of the N-glycan on the protein (resulting in site-specific processing). Taken together, our data show that the structure of a glycoprotein can control the processing of some N-linked oligosaccharides at different stages of the glycosylation pathway.





**Fig. 7.** Secreted Pdi1p shows site-specific glycan profiles. **(A)** N-Glycan-processing pathway of *T. ni* showing glycan structures and enzymes (MI, MII, MIII:  $\alpha$ -mannosidase I, II and III; GnTI, GnTII: N-acetylglucosamine transferase I and II; GA: N-acetylglucosaminidase; FT6, FT3:  $\alpha$ 1,6- and  $\alpha$ 1,3-fucosyl transferase). **(B)** Site-specific glycan profiles of sPdi1p. Data represent the average values calculated from four independent experiments. Relative abundances for each glycoform are depicted above the corresponding column placed at the left of graphic structure representation. The processing pathway from (A) is depicted at the base of the charts. See Supplementary data, Table S6 for the complete data set including all glycoforms detected. **(C)** Pdi1p N-glycosylation sites arranged based on the similarity of their glycoprofiles. The similarity between site-specific glycan profiles and different glycoforms was calculated as described in Figure 3C. **(D)** Conversions for selected enzymes for each glycosite (S1 to S5). Site-specific enzyme conversion was obtained by glycan flux analysis using the relative abundances of each glycoform from (B). Error bars indicate standard deviation of mean values from four experiments. See Supplementary data, Figure S6 and Table S6 for spectra, a complete set of conversions for every enzyme and raw data. (B–D) The structures of glycan isomers were assigned according to the biosynthesis pathway of insect cells shown in (A). This figure is available in black and white in print and in color at *Glycobiology* online.



**Fig. 8.** The glycan profile of secreted Pdi1p is altered by changing the protein structure. (A) Secreted Pdi1p and truncated versions were expressed in insect cells and purified as described in Figure 1. Eluted fractions were analyzed by SDS-PAGE and stained with Coomassie (left/middle). ER retention of sPdi1p was disrupted by the presence of two additional amino acids at the protein's C-terminus (HDELLE) (Raykhel et al. 2007); truncated versions do not contain a HDEL sequence. (B) Relative glycan abundances on S4 of purified full-length sPdi1p and secreted b-, bb'-, ab-, abb'-variants shown in different colors. Data represent the mean values of glycan ratios of three independent experiments with error bars indicating standard deviation of mean values. \*  $P < 0.05$ ; \*\*  $P < 0.01$ ; \*\*\*  $P < 0.001$ . (C) Conversions of S4 N-glycan on sPdi1p variants were calculated as in Figure 6D. Conversions for selected enzymes were plotted for each variant. Error bars indicate standard deviation of mean values from four experiments. This figure is available in black and white in print and in color at *Glycobiology* online.

Site-specific N-glycans can affect the function of glycoproteins, e.g. immunoglobulins (Niwa et al. 2004; Scallion et al. 2007; Sazinsky et al. 2008; Ferrara et al. 2011), but for most glycoproteins, this has not been addressed experimentally. It is evident that the quantitative analysis of site-specific glycans will be essential for future research in this direction. The MS-based, quantitative glycoproteomics method presented in this report will become a valuable tool to define more precisely the role of the carbohydrate in glycoprotein function. Similarly, the mathematical modelling of glycan processing based on quantitative glycoproteomics data will be essential to experimentally approach and understand the function of the Golgi and will affect the biotechnological production of glycoproteins.

## Materials and methods

### Reagents

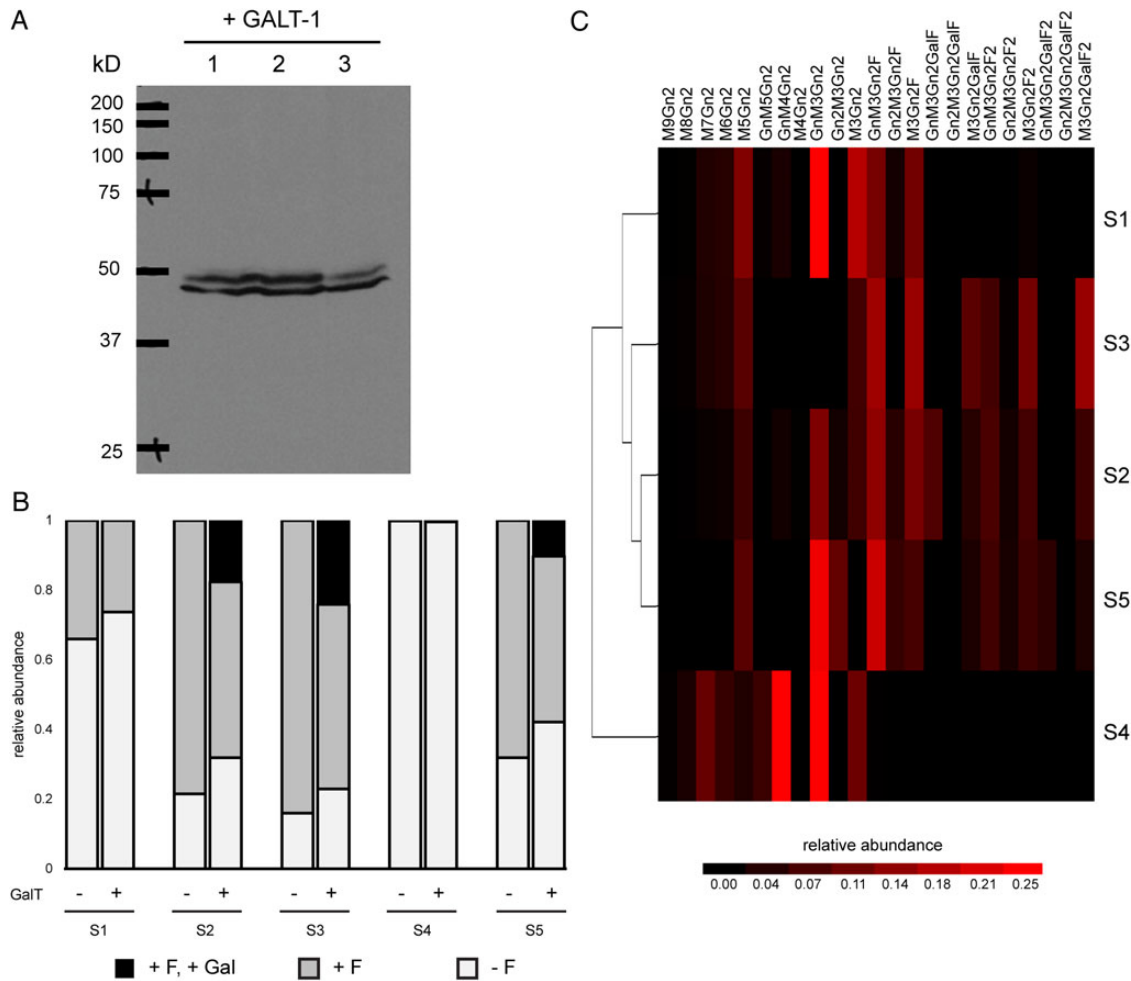
The monoclonal anti-His antibody was purchased from Qiagen AG (Cat. No. 34670). Monoclonal anti-GST antibody was purchased

from Rockland (Cat. No. 600-101-200). *Spodoptera frugiperda* Sf21 and *Trichoplusia ni* Hi5 cells were obtained from K. Locher and cultivated in Sf-900 II SFM medium (Invitrogen, Cat.No 10902104).

### Construction of plasmids for baculovirus-based protein expression

Plasmids used in this study are summarized in Supplementary data, Table S1. Plasmids were constructed using standard cloning protocols. Open reading frames of all constructs were confirmed by nucleotide sequencing.

pRG105: The plasmid pRG105 was constructed to allow for baculovirus-based expression of secreted proteins fused to an N-terminal affinity tag. gp67 secretion signal peptide followed by GST was PCR-amplified from pAcSecG2T baculovirus transfer vector (Fischer Scientific, Cat. no. BDB554797) using primers RG395 (5'-TGG GCG CGC ATG CTA CTA GTA AAT CAG TCA CAC-3') and RG396 (5'-CCC AAG CTT TTA CTC GAG CTG



**Fig. 9.** Altering the processing machinery changes the glycan profile of secreted Pdi1p. (A) GALT-1 expression levels were confirmed by immunoblotting using anti-FLAG antibody in three independent experiments. (B) sPdi1p was purified from cells infected with either an unmodified bacmid (–GalT) or a recombinant bacmid carrying an expression copy of GALT-1 (+GalT). Relative glycan abundances without core fucoses (–F), with core fucoses but without galactose (+F), and with galactose attached to the core fucose (+F, +Gal) were calculated for each site. The data represent mean values from three independent experiments. See Supplementary data, Table S7 for data on individual glycan structures. (C) Similarity of the five glycosites was calculated from relative glycan abundances as described in Figure 3C. See Supplementary data, Figure S7 for glycan profiles of S1, S2 and S3, and immunoblot of GALT-1 expression, and Supplementary data, Table S7 for raw data. This figure is available in black and white in print and in color at *Glycobiology* online.

CAG AGG CCT GAG CTC GGA TCC ACG CGG AAC CAGA-3'). The resulting PCR fragment was cloned into the BssHIII and HindIII restriction sites of pRG74 (Gauss et al. 2011) to generate pRG100. Next pRG100 was used as a template to introduce His<sub>10</sub>-tag following a cleavable signal sequence by PCR using RG394 (5'-AAG TGG TTC GCA TCC TCG GTTT-3') and RG398 (5'-CCC GAG CTC CAG GGG CCC CTG GAA CAG AAC TTC CAG ATG GTG ATG GTG ATG GTG ATG GTG ATG GTG GGT ACC CGC AAA GGC AGA ATG CGC-3') primers producing pRG101. Finally, the *PDI1* gene (without the signal peptide) was excised from pRG84 (Gauss et al. 2011) by digestion with SacI and XhoI and introduced into pRG101 generating pRG105. ER retention of sPdi1p is disrupted by the presence of two additional amino acids at the protein's C-terminus (HDELLE).

pRG143, pRG144, pRG145, pRG146: Truncated versions of the *PDI1* gene were PCR-amplified from pRG84 using the following primer pairs: RG476 (5'-CCG AGC TCA CTA GTC CGG CTG TCG

CCG TTG TTG CTG ATC-3') and RG477 (5'-CCC TGC AGC TCG AGG TAG GGC AAG GCT TCC ACT TGC AACC-3') (b-domain); RG475 (5'- CCG AGC TCA CTA GTG AGG CTG TGG CCC CTG AAG ACT CC-3') and RG477 (ab-domains); RG475 and RG478 (5'- CCC TGC AGC TCG AGA CCT TTC AAG AAG TCC TTA ACC AAAG-3') (abb'-domains); RG476 and RG478 (bb'-domains). To allow for expression of ER-retained proteins, the respective PCR products were cloned into the SacI and XhoI restriction sites of pRG85 upstream of the HDEL sequence thus replacing the *PDI1* gene and yielding pRG143 (b), pRG144 (ab), pRG145 (abb') and pRG146 (bb').

pRG135, pRG136, pRG137, pRG138: To allow for expression and secretion of truncated Pdi1p variants, the coding sequences of pRG143, pRG144, pRG145, and pRG146 were excised by SacI and XhoI. Fragments were ligated into the respective restriction sites of pRG105 downstream of the gp67 signal sequence yielding pRG135 (b), pRG136 (ab), pRG137 (abb') and pRG138 (bb').

## Expression and purification of recombinant proteins in insect cells

Standard procedures were used to generate recombinant baculoviruses in Sf21 cells. Recombinant viruses were collected after 72 h. Viruses were amplified by infecting Sf21 cell cultures with virus solution in a 1 : 10 and 1 : 100 (v/v) ratio. Expression of intracellular or secreted recombinant proteins was monitored by SDS-PAGE and immunoblotting after each virus amplification step. For protein expression, *T. ni* Hi5 cells were diluted to 1 million cells per milliliter in shaker flasks and infected 1 : 100 (v/v) with recombinant viruses. For the analysis of site-specific processing of sPdi1p by GALT-1, sPdi1p was expressed in Hi5 cells co-infected with GALT-1 recombinant viruses.

*Trichoplusia ni* Hi5 cells were diluted to 1 million cells per milliliter in shaker flasks and infected 1 : 100 (v/v) with recombinant viruses. Infected cells were incubated at 27°C for 48 h. To purify His<sub>10</sub>-tagged ER retained proteins, cells were pelleted by centrifugation (600 rcf, 5 min), washed with 1× PBS and lysed with lysis buffer [1% Triton X-100, 1× Protease Inhibitor cocktail (Roche, 11873580001) in 1× PBS]. The lysates were centrifuged at 3500 rcf for 10 min and the soluble fraction was filtered using PES-membrane 0.2 μm filters (TPP). For the purification of secreted proteins, the culture supernatant was cleared by centrifugation at 3500 rcf for 10 min and filtered. Affinity purification was performed on gravity flow columns filled with 1 mL NiNTA beads (Protino, 745400.100) equilibrated either with 10 column volumes (CVs) of lysis buffer or medium. The bound fraction was washed with 15 CV washing buffer (30 mM imidazole in 1× PBS) and proteins were eluted with 4 CV of elution buffer (250 mM imidazole in 1× PBS). Eluted fractions were concentrated on Amicon Ultra-4 Centrifugal Filter Devices. For MS analysis, samples were prepared as described in the “Sample preparation” paragraph. Proteins that were stored after buffer exchange to 1× PBS and kept frozen at -20°C.

To purify GST-tagged M9Gn2-Pdi1p, Hi5 cells were infected with recombinant virus in the presence of 10 μM kifunensine (Sigma-Aldrich, K1140-1MG) and incubated for 48 h. Mns1p was purified from Hi5 cells 48 h after infection with recombinant virus. Following cell lysis, proteins were bound to Glutathione Sepharose 4B matrix (GE Healthcare, 17-0756-01). Bound proteins were washed with 15 CV of 1× PBS and subsequently with 15 CV of PreScission Protease (PSP; VWR, 27-0843-01) activity buffer (150 mM NaCl, 1 mM EDTA, 1 mM DTT in 50 mM Tris, pH 7.0). On column cleavage was performed by incubating bound proteins with 20 U of PSP in 1 mL of PSP activity buffer for 16 h. Released proteins were collected by gravity flow. Sample was concentrated and the buffer was exchanged to Mns1p activity buffer (150 mM NaCl, 5 mM CaCl<sub>2</sub> in 20 mM MES, pH 7.0) (Karaveg and Moremen 2005).

## In vitro Mns1p activity assay

Fifty micromolars of Pdi1p were mixed with 0.5 μM Mns1p in total volume of 85 μL and incubated at 37°C. Aliquots of 17 μL were taken at different time points. Each aliquot was mixed with 3 μL of 100% TCA, vortexed, and incubated on ice for 5 min. After centrifugation at 20,000 × g and 4°C for 5 min, the pellet was washed three times with 500 μL of ice cold acetone. The pellet was air dried and stored at -20°C. For MS analysis, protein pellets were resuspended in 50 μL of 8 M urea in 50 mM ammonium bicarbonate (pH 8.5).

## Sample preparation and glycopeptide analysis by nano-HPLC-HCD-MS/MS

Purified proteins were loaded onto a filter device (Microcon YM-30, Millipore) and washed three times with water. Usually, 50 μg was

reduced by 50 mM dithiothreitol in 50 mM ammonium bicarbonate buffer (pH 8.5) at 37°C for 1 h, following by alkylation with 65 mM iodoacetamide at 37°C in the dark for 1 h. After four washing steps with ammonium bicarbonate buffer, proteins were digested by trypsin (weight ratio 50 : 1, Promega, Cat. No. V5111) at 37°C for 16 h. Digested peptides and glycopeptides were collected by centrifugation and dried in a speedvac. Samples were desalted by Zip-Tip C18 (Millipore) prior to nanoLC-MS/MS analysis. Samples were analyzed on a calibrated LTQ-Orbitrap Velos mass spectrometer (Thermo Fischer Scientific) coupled to an Eksigent-Nano-HPLC system (Eksigent Technologies). Peptides were resuspended in 2.5% acetonitrile with 0.1% formic acid and loaded onto a self-made tip column (75 μm × 80 mm) packed with reverse phase C18 material (AQ, 3 μm 200 Å, Bischoff GmbH). Peptides were eluted with a flow rate of 200 nL per minute by a gradient from 3 to 30% of solution B (99.9% ACN, 0.1% FA) applied for 22 min, 50% B applied for 25 min, 97% B applied for 27 min. One scan cycle comprised of a full-scan MS survey spectrum, followed by up to 10 sequential HCD MS/MS on the most intense signals above a threshold of 2000. Full-scan MS spectra (800–2000 *m/z*) were acquired in the FT-Orbitrap at a resolution of 60,000 at 400 *m/z*, while HCD MS/MS spectra were recorded in the FT-Orbitrap at a resolution of 15,000 at 400 *m/z*. HCD was performed with a target value of 1e5 and stepped collision energy rolling from 35, 40 and 45 V was applied. AGC target values were 5e5 for full FTMS. For all experiments, dynamic exclusion was used with a single repeat count, 15 s repeat duration, and 60 s exclusion duration.

## Identification and quantification of different glycoforms sharing the same peptide backbone

MS and MS/MS data were processed into the Mascot generic format (mgf) file. Extractmgf was written in Perl to perform the following steps: glycopeptides were initially extracted from mgf file by identifying glycan oxonium ions, [HexNAc]<sup>+</sup> 204.09 and [HexNAc + Hex]<sup>+</sup> 366.12. Glycopeptides corresponding to each site were further obtained by sorting out the corresponding Y1 ion individually from extracted mgf file. All MS/MS spectra were confirmed manually. Here, XCalibur 2.2 sp1.48 was used for data processing. For quantification, extracted ion chromatography (XIC) of each glycoform was plotted by its individual *m/z* with the mass tolerance of 5 ppm. Peak area was defined manually and integrated by the program. The relative amount of each glycoform sharing same peptide backbone was calculated as following equation:

$$\text{Relative amount of each glycoform (\%)} = \frac{\text{Peak area of each glycoform}}{\text{Sum of peak area of all glycoforms}} \times 100\%.$$

## Molecular modelling

### Three-dimensional structure preparation

Initial starting coordinates for the a- and b- subdomains of Pdi1p were taken from PDB: 2B5E. Models were generated for the unglycosylated domains and for a glycoform with M9Gn2 at site 4 (S4) using the online glycoprotein builder available on GLYCAM-Web (Kirschner et al. 2008; Woods Group, 2005–2014). Using tleap (Case et al. 2014), the structures were placed in a truncated octahedron of TIP3P (Jorgensen et al. 1983) water with an 8 Å distance buffer from the solute to the edge of the periodic box and 0.5 Å spacing between solute and water molecules. Prior to solvation, all waters of crystallization were removed, and the systems were neutralized with sodium counter-ions.



### Energy minimization and molecular dynamics simulations

Energy minimization (10,000 steps of steepest descent, followed by 10,000 steps of conjugate gradient) was performed in a stepwise fashion with initially a  $5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  restraint applied to the solute (protein and glycan) heavy atoms. This was followed by minimization with restraints only on the glycan heavy atoms, before a final minimization with no restraints on any atoms. Minimizations were performed with sander.MPI in AMBER14 (Case et al. 2014). Molecular dynamics simulations were performed with the CUDA implementation of PMEMD (Gotz et al. 2012; Salomon-Ferrer et al. 2013) in AMBER14 using constant pressure (nPT) conditions. A Berendsen barostat with a time constant of 1 ps was employed for pressure regulation, while a Langevin thermostat with a collision frequency of  $2 \text{ ps}^{-1}$  was employed for temperature regulation. A non-bonded interaction cutoff of  $8 \text{ \AA}$  was employed. Long-range electrostatics were treated with the particle-mesh Ewald method (Darden et al. 1993). Covalent bonds involving hydrogen were constrained with the SHAKE algorithm allowing a time step of 2 fs (Ryckaert et al. 1977). Cartesian restraints were applied to the protein C $\alpha$  atoms ( $5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ ) during a 100 ps heating stage from 5 to 300 K, which was followed by a 100 ps equilibration phase. All restraints were then removed for a 1 ns equilibration phase prior to a 500 ns production phase.

### Data analysis

Root mean squared deviation (RMSD) of the *a*-domain: The cpptraj (Roe and Cheatham 2013) module was used to extract 1000 snapshots from each simulation at 500 ps intervals, these frames were aligned on the b-domain of the crystal structure, and the RMSD computed for the *a*-domain. Glycan branch accessibility: NACCESS (Hubbard and Thornton 1993) was employed with a  $6 \text{ \AA}$  radius probe to compute the accessibility of the non-reducing terminal disaccharides in the A-, B- and C-branches of the glycan over 100 snapshots from the simulation; the approximate end-to-end distance of the disaccharide being  $12 \text{ \AA}$ . The values were normalized relative to the accessibility of a Man $\alpha$ 1-2Man $\alpha$  disaccharide.

### Glycan flux analysis of *T. ni* Golgi glycosylation network

Glycan flux analysis is based on the principle of metabolic flux analysis (MFA) and thus describes a steady state of the system. The glycosylation network is regarded as an enclosed biological system, in which the glycan structures are connected by fluxes  $v$ . The network is characterized by the stoichiometric matrix  $S$  with the objective to balance fluxes across glycan structures. The stoichiometric matrix is an  $n$  (number of glycan structures) times  $m$  (number of fluxes) matrix with the following logic:

$$\begin{aligned} S(i, j) < 0 & \text{ glycan structure } i \text{ is a reactant in reaction } j, \\ S(i, j) = 0 & \text{ glycan structure } i \text{ does not appear in reaction } j, \\ S(i, j) > 0 & \text{ glycan structure } i \text{ is a product in reaction } j. \end{aligned}$$

The change of glycan concentrations  $c$  over time can be written as the product of the stoichiometric matrix and the fluxes:

$$\frac{dc(t)}{dt} = Sv(t). \quad (1)$$

In MFA, a pseudo steady-state assumption is made for cell internal metabolites, stating that internal species change much faster than the measured external ones (Antoniewicz 2013). In glycan flux analysis, this assumption is also valid considering the reported Golgi residence time of 20–40 min which is much shorter than the culture duration

and any observed changes in glycosylation (Jimenez del Val et al. 2011). Therefore, equation (1) can be written as

$$Sv = 0. \quad (2)$$

The stoichiometric matrix  $S$  can be split into a Golgi internal stoichiometric matrix  $S_i$  and a stoichiometric matrix of the exchange fluxes  $S_e$  that relates external with internal glycan structures. Equation (2) can be split accordingly:

$$S_i v_i + S_e v_e = 0. \quad (3)$$

Since in our case,  $S_i$  is a square matrix with full rank, the inverse  $S_i^{\#}$  was used to solve equation (3):

$$v_i = -(S_i)^{\#} S_e v_e. \quad (4)$$

The measured glycans are considered as constant fluxes  $v_e$  out of the Golgi apparatus and were used to calculate the fluxes  $v_i$  inside the network. M9Gn2 was assumed as the only structure entering the late ER and Golgi apparatus (Krambeck et al. 2009). Large glycosylation networks often contain cycles and therefore more reactions than species which leads to so-called underdetermined systems with multiple possible flux distributions. However, in our case there are as many measured glycan structures as there are internal fluxes and hence  $S_i$  is invertible leading to a unique solution of equation (4). It is worth noting that enzymatic reactions are in principle reversible, yet only the net fluxes can be derived from expressed glycan patterns. Based on the computed fluxes, the conversion for every enzymatic step was calculated by normalizing the effluxes with respect to the influxes over each glycan species. Small differences early in the glycosylation process such as a different mannosidase activity can significantly change the entire glycosylation profile even if all reaction further downstream are equally active on each site. This is mainly due to a changing availability of substrates for each reaction, yet the enzymes are not more or less occupied regarding all other glycoproteins which are simultaneously present in the Golgi apparatus. As a consequence, the flux conversion of each individual step in the pathway can be used to quantitatively compare the enzymatic preference towards site-specific processing.

### Supplementary data

Supplementary data for this article are available online at <http://glycob.oxfordjournals.org/>.

### Acknowledgements

We thank Emma Marriott for excellent technical assistance. We thank Tsung-Hsien Pu and Prof. Kay-Hooi Khoo of the Core Facility for Protein structural Analysis, Academia Sinica, Taipei, Taiwan for their help in the early stage of the software development. We thank the Functional Genomics Centre Zurich for help with the mass spectrometry analysis.

### Funding

R.J.W. thank the National Institutes of Health (P41GM103390, R01GM094919) (EUREKA), as well as the Science Foundation of Ireland (08/IN.1/B2070) for support. This work was supported by ETH Research Grant No. 07 10-1 to M.A. and R.G. and Swiss National Science Foundation Grant No. 31003A\_127098 to M.A.

### Conflict of interest statement

None declared.

## Abbreviations

CVs, column volumes; ER, endoplasmic reticulum; FASP, filter-assisted sample preparation; GALT-1,  $\beta$ 1,4-galactosyltransferase; GHs, glycosyl hydrolases; GTs, glycosyl transferases; MFA, metabolic flux analysis; mgf, Mascot generic format; PSP, PreScission Protease; RMSD, root mean squared deviation; XIC, extracted ion chromatography.

## References

- Aebi M, Bernasconi R, Clerc S, Molinari M. 2010. N-Glycan structures: Recognition and processing in the ER. *Trends Biochem Sci.* 35:74–82.
- Antoniewicz MR. 2013. Dynamic metabolic flux analysis – Tools for probing transient states of metabolic networks. *Curr Opin Biotechnol.* 24:973–978.
- Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, Cheatham TE, Darden TA, Duke RE, Gohlke H, et al. 2014. *AMBER 14*. San Francisco: University of California.
- Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: An N- log (N) method for Ewald sums in large systems. *J Chem Phys.* 98:10089.
- Dell A, Galadari A, Sastre F, Hitchen P. 2010. Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *Int J Microbiol.* 2010:148178.
- Dell A, Reason AJ, Khoo KH, Panico M, McDowell RA, Morris HR. 1994. Mass spectrometry of carbohydrate-containing biopolymers. *Methods Enzymol.* 230:108–132.
- Domon B, Vath JE, Costello CE. 1990. Analysis of derivatized ceramides and neutral glycosphingolipids by high-performance tandem mass spectrometry. *Anal Biochem.* 184:151–164.
- Faid V, Evjen G, Tollersrud OK, Michalski JC, Morelle W. 2006. Site-specific glycosylation analysis of the bovine lysosomal alpha-mannosidase. *Glycobiology.* 16:440–461.
- Ferrara C, Grau S, Jager C, Sondermann P, Brunker P, Waldhauer I, Hennig M, Ruf A, Rufer AC, Stihle M, et al. 2011. Unique carbohydrate-carbohydrate interactions are required for high affinity binding between Fc $\gamma$ RIII and antibodies lacking core fucose. *Proc Natl Acad Sci USA.* 108:12669–12674.
- Gauss R, Kanehara K, Carvalho P, Ng DT, Aebi M. 2011. A complex of Pdi1p and the mannosidase Htm1p initiates clearance of unfolded glycoproteins from the endoplasmic reticulum. *Mol Cell.* 42:782–793.
- Geisler C, Jarvis DL. 2012. Substrate specificities and intracellular distributions of three N-glycan processing enzymes functioning at a key branch point in the insect N-glycosylation pathway. *J Biol Chem.* 287:7084–7097.
- Go EP, Irungu J, Zhang Y, Dalpathado DS, Liao HX, Sutherland LL, Alam SM, Haynes BF, Desaire H. 2008. Glycosylation site-specific analysis of HIV envelope proteins (JR-FL and CON-S) reveals major differences in glycosylation site occupancy, glycoform profiles, and antigenic epitopes' accessibility. *J Proteome Res.* 7:1660–1674.
- Gotz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. 2012. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *J Chem Theory Comput.* 8:1542–1555.
- Hebert DN, Garman SC, Molinari M. 2005. The glycan code of the endoplasmic reticulum: Asparagine-linked carbohydrates as protein maturation and quality-control tags. *Trends Cell Biol.* 15:364–370.
- Heikinheimo P, Helland R, Leiros HK, Leiros I, Karlsen S, Evjen G, Ravelli R, Schoehn G, Ruigrok R, Tollersrud OK, et al. 2003. The structure of bovine lysosomal alpha-mannosidase suggests a novel mechanism for low-pH activation. *J Mol Biol.* 327:631–644.
- Hosokawa N, Tremblay LO, Sleno B, Kamiya Y, Wada I, Nagata K, Kato K, Herscovics A. 2010. EDEM1 accelerates the trimming of alpha1,2-linked mannose on the C branch of N-glycans. *Glycobiology.* 20:567–575.
- Hsieh P, Rosner MR, Robbins PW. 1983a. Host-dependent variation of asparagine-linked oligosaccharides at individual glycosylation sites of Sindbis virus glycoproteins. *J Biol Chem.* 258:2548–2554.
- Hsieh P, Rosner MR, Robbins PW. 1983b. Selective cleavage by endo-beta-N-acetylglucosaminidase H at individual glycosylation sites of Sindbis virion envelope glycoproteins. *J Biol Chem.* 258:2555–2561.
- Hua S, Nwosu CC, Strum JS, Seipert RR, An HJ, Zivkovic AM, German JB, Lebrilla CB. 2012. Site-specific protein glycosylation analysis with glycan isomer differentiation. *Anal Bioanal Chem.* 403:1291–1302.
- Hubbard SC. 1988. Regulation of glycosylation. The influence of protein structure on N-linked oligosaccharide processing. *J Biol Chem.* 263:19303–19317.
- Hubbard SJ, Thornton JM. 1993. "NACCESS", Computer Program. Department of Biochemistry and Molecular Biology, University College London
- Jelinek-Kelly S, Akiyama T, Saunier B, Tkacz JS, Herscovics A. 1985. Characterization of a specific alpha-mannosidase involved in oligosaccharide processing in *Saccharomyces cerevisiae*. *J Biol Chem.* 260:2253–2257.
- Jimenez del Val I, Nagy JM, Kontoravdi C. 2011. A dynamic mathematical model for monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing Golgi apparatus. *Biotechnol Progr.* 27:1730–1743.
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J Chem Phys.* 79:926–935.
- Karaveg K, Moremen KW. 2005. Energetics of substrate binding and catalysis by class 1 (glycosylhydrolase family 47) alpha-mannosidases involved in N-glycan processing and endoplasmic reticulum quality control. *J Biol Chem.* 280:29837–29848.
- Kelleher DJ, Gilmore R. 2006. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology.* 16:47R–62R.
- Kemmink J, Darby NJ, Dijkstra K, Nilges M, Creighton TE. 1996. Structure determination of the N-terminal thioredoxin-like domain of protein disulfide isomerase using multidimensional heteronuclear <sup>13</sup>C/<sup>15</sup>N NMR spectroscopy. *Biochemistry.* 35:7684–7691.
- Kemmink J, Dijkstra K, Mariani M, Scheek RM, Penka E, Nilges M, Darby NJ. 1999. The structure in solution of the b domain of protein disulfide isomerase. *J Biomol NMR.* 13:357–368.
- Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, Foley BL, Woods RJ. 2008. GLYCAM06: A generalizable biomolecular force field. *Carbohydrates. J Comput Chem.* 29:622–655.
- Krambeck FJ, Bennun SV, Narang S, Choi S, Yarema KJ, Betenbaugh MJ. 2009. A mathematical model to derive N-glycan structures and cellular enzyme activities from mass spectrometric data. *Glycobiology.* 19:1163–1175.
- Krokhin O, Ens W, Standing KG, Wilkins J, Perreault H. 2004. Site-specific N-glycosylation analysis: Matrix-assisted laser desorption/ionization quadrupole-quadrupole time-of-flight tandem mass spectral signatures for recognition and identification of glycopeptides. *Rapid Commun Mass Spectrom.* 18:2020–2030.
- Mast SW, Moremen KW. 2006. Family 47 alpha-mannosidases in N-glycan processing. *Methods Enzymol.* 415:31–46.
- Moremen KW, Tiemeyer M, Nairn AV. 2012. Vertebrate protein glycosylation: Diversity, synthesis and function. *Nat Rev Mol Cell Biol.* 13:448–462.
- Nagae M, Yamaguchi Y. 2012. Function and 3D structure of the N-glycans on glycoproteins. *Int J Mol Sci.* 13:8398–8429.
- Niwa R, Shoji-Hosaka E, Sakurada M, Shinkawa T, Uchida K, Nakamura K, Matsushima K, Ueda R, Hanai N, Shitara K. 2004. Defucosylated chimeric anti-CC chemokine receptor 4 IgG1 with enhanced antibody-dependent cellular cytotoxicity shows potent therapeutic activity to T-cell leukemia and lymphoma. *Cancer Res.* 64:2127–2133.
- North SJ, Huang HH, Sundaram S, Jang-Lee J, Etienne AT, Trollope A, Chalabi S, Dell A, Stanley P, Haslam SM. 2010. Glycomics profiling of Chinese hamster ovary cell glycosylation mutants reveals N-glycans of a novel size and complexity. *J Biol Chem.* 285:5759–5775.
- Parekh RB, Tse AG, Dwek RA, Williams AF, Rademacher TW. 1987. Tissue-specific N-glycosylation, site-specific oligosaccharide patterns and lentil lectin recognition of rat Thy-1. *EMBO J.* 6:1233–1244.
- Powell AK, Harvey DJ. 1996. Stabilization of sialic acids in N-linked oligosaccharides and gangliosides for analysis by positive ion matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom.* 10:1027–1032.
- Raykhel I, Alanen H, Salo K, Jurvansuu J, Nguyen VD, Latva-Ranta M, Ruddock L. 2007. A molecular specificity code for the three mammalian KDEL receptors. *J Cell Biol.* 179:1193–1204.

- Roe DR, Cheatham TE. 2013. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput.* 9:3084–3095.
- Roepstorff P, Fohlman J. 1984. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom.* 11:601.
- Ryckaert J-P, Ciccotti G, Berendsen HJC. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J Comput Phys.* 23:327–341.
- Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. 2013. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J Chem Theory Comput.* 9:3878–3888.
- Sazinsky SL, Ott RG, Silver NW, Tidor B, Ravetch JV, Wittrup KD. 2008. Aglycosylated immunoglobulin G1 variants productively engage activating Fc receptors. *Proc Natl Acad Sci USA.* 105:20167–20172.
- Scallon BJ, Tam SH, McCarthy SG, Cai AN, Raju TS. 2007. Higher levels of sialylated Fc glycans in immunoglobulin G molecules can adversely impact functionality. *Mol Immunol.* 44:1524–1534.
- Schachter H. 1991. The 'yellow brick road' to branched complex N-glycans. *Glycobiology.* 1:453–461.
- Segu ZM, Mechref Y. 2010. Characterizing protein glycosylation sites through higher-energy C-trap dissociation. *Rapid Commun Mass Spectrom.* 24:1217–1225.
- Shental-Bechor D, Levy Y. 2009. Folding of glycoproteins: Toward understanding the biophysics of the glycosylation code. *Curr Opin Struct Biol.* 19:524–533.
- Shi X, Jarvis DL. 2007. Protein N-glycosylation in the baculovirus-insect cell system. *Curr Drug Targets.* 8:1116–1125.
- Sparbier K, Asperger A, Resemann A, Kessler I, Koch S, Wenzel T, Stein G, Vorweg L, Suckau D, Kostrzewa M. 2007. Analysis of glycoproteins in human serum by means of glycospecific magnetic bead separation and LC-MALDI-TOF/TOF analysis with automated glycopeptide detection. *J Biomol Tech.* 18:252–258.
- Stanley P. 2011. Golgi glycosylation. *Cold Spring Harb Perspect Biol.* 3:1–13.
- Tian G, Xiang S, Noiva R, Lennarz WJ, Schindelin H. 2006. The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. *Cell.* 124:61–73.
- Titz A, Butschi A, Henrissat B, Fan YY, Hennet T, Razzazi-Fazeli E, Hengartner MO, Wilson IB, Kunzler M, Aebi M. 2009. Molecular basis for galactosylation of core fucose residues in invertebrates: Identification of *caenorhabditis elegans* N-glycan core alpha1,6-fucoside beta1,4-galactosyltransferase GALT-1 as a member of a novel glycosyltransferase family. *J Biol Chem.* 284:36223–36233.
- Trimble RB, Maley F, Chu FK. 1983. GlycoProtein biosynthesis in yeast. Protein conformation affects processing of high mannose oligosaccharides on carboxypeptidase Y and invertase. *J Biol Chem.* 258:2562–2567.
- Vallée F, Lipari F, Yip P, Sleno B, Herscovics A, Howell PL. 2000. Crystal structure of a class I alpha1,2-mannosidase involved in N-glycan processing and endoplasmic reticulum quality control. *EMBO J.* 19:581–588.
- van Kooyk Y, Rabinovich GA. 2008. Protein-glycan interactions in the control of innate and adaptive immune responses. *Nat Immunol.* 9:593–601.
- Varki A. 1998. Factors controlling the glycosylation potential of the Golgi apparatus. *Trends Cell Biol.* 8:34–40.
- Wilkinson B, Gilbert HF. 2004. Protein disulfide isomerase. *Biochim Biophys Acta.* 1699:35–44.
- Wisniewski JR, Zougman A, Nagaraj N, Mann M. 2009. Universal sample preparation method for proteome analysis. *Nat Methods.* 6:359–362.
- Woods Group. (2005–2014). *GLYCAM Web*. Athens, GA: Complex Carbohydrate Research Center, University of Georgia (<http://www.glycam.com>).
- Yu X, Baruah K, Harvey DJ, Vasiljevic S, Alonzi DS, Song BD, Higgins MK, Bowden TA, Scanlan CN, Crispin M. 2013. Engineering hydrophobic protein-carbohydrate interactions to fine-tune monoclonal antibodies. *J Am Chem Soc.* 135:9723–9732.
- Ziegler FD, Trimble RB. 1991. Glycoprotein biosynthesis in yeast: Purification and characterization of the endoplasmic reticulum Man9 processing alpha-mannosidase. *Glycobiology.* 1:605–614.