Analysis of congenital disorder of glycosylation-Id in a yeast model system shows diverse site-specific under-glycosylation of glycoproteins.

Ulla-Maja Bailey[1], Muhammad Fairuz Jamaluddin[1] and Benjamin L. Schulz[1,2]

1 School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia.

2 To whom correspondence should be addressed: Benjamin L. Schulz, School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia. Phone: +61 7 3365 4875 Fax: +61 7 3365 4273 Email: b.schulz@uq.edu.au

Abbreviations used: CDG, congenital disorder of glycosylation; OTase, oligosaccharyltransferase.

Key words: ALG; *alg3*; N-Glycosylation; Oligosaccharyltransferase; mass spectrometry; Glycosylation occupancy; PNGase F; CDG.

Acknowledgements of financial support: We acknowledge the support of NHMRC project grant 631615 and CDF APP1031542

Abstract

Asparagine-linked glycosylation is a common post-translational modification of proteins in eukaryotes. Mutations in the human ALG3 gene cause changed levels and altered glycan structures on mature glycoproteins, and are the cause of a severe Congenital Disorder of Glycosylation (CDG-Id). Diverse glycoproteins are also underglycosylated in *Saccharomyces cerevisae alg3* mutants. Here we analysed site-specific glycosylation occupancy in this yeast model system using peptide-N-glycosidase F to label glycosylation sites with an asparagine-aspartate conversion that creates a new endoproteinase AspN cleavage site, followed by proteolytic digestion, and detection of peptides and glycopeptides by LC-ESI-MS/MS. We used this analytical method to identify and measure site-specific glycosylation occupancy in *alg3* mutant and wild type yeast strains. We found decreased site-specific N-glycosylation occupancy in the *alg3* knock-out strain preferentially at Asn-Xaa-Ser sequences located in secondary structural elements, features previously associated with poor glycosylation efficiency. Furthermore, we identified 26 previously experimentally unverified glycosylation sites. Our results provide insights into the underlying mechanisms of disease in CDG-1d and our methodology will be useful in site-specific glycosylation analysis in many model systems and clinical applications.

Introduction

Protein N-glycosylation is a common and essential co- and post-translocational modification of membrane and secreted proteins that in eukaryotes is catalysed by the multiprotein complex enzyme oligosaccharyltransferase (OTase) [1]. The oligosaccharide $Glc_3Man_9GlcNAc_2$ is assembled on the lipid carrier dolicholpyrophosphate and transferred *en bloc* by OTase

onto selected asparagines, primarily in N-glycosylation sequons (NX(S/T); X≠P) as nascent polypeptide is translocated into the endoplasmic reticulum (ER) [1, 2]. Further extension and modification of the glycan structures on mature proteins occurs in the Golgi [3-5]. N-glycans assist protein folding in the ER, and the specific glycan structures on mature glycoproteins modulate protein function in developmental programs, during immune responses, in cell-cell interactions and in many other cellular processes [6-8].

A number of the genes that encode for enzymes in the N-glycosylation pathway have been shown to be involved in human health and disease [6, 9]. Many studies have revealed that alterations in glycan expression are associated as causative or incidental factors in relatively rare Congenital Disorders of Glycosylation (CDG), widespread acquired diseases and cancer [6, 7, 9-11]. CDGs are conceptually divided into CDG-I, where the genetic defect is in the ER in the N-glycosylation pathway before addition of glycan to protein, and CDG-II, where the defect is in the Golgi. Thus, CDG-I results in inefficient glycosylation of asparagines, while CDG-II results in changes to the glycan structures on mature glycoproteins [6].

Underglycosylation is the underlying cause of disease in CDG-I, resulting in an increase in misfolded glycoproteins and alterations in their functions through reduced glycosylation of particular asparagines. However, with over half of all proteins predicted to be glycosylated, the sequence and structural diversity of protein acceptor sites is extraordinarily large [12]. It is currently not clear which particular subset of asparagines are underglycosylated in CDG-I. However, a study using targeted SRM-MS showed that transferrin isolated from the blood of CDG-I patients had lower glycosylation occupancy specifically at its second N-glycosylation site [13]. Additionally, several familial mutations in the homologous OTase subunits MagT1 and Tusc3 have been linked to a form of CDG-I leading to non-syndromic mental retardation [14-16].

Studies using a manipulable yeast genetic system showed that the yeast homologues of these proteins are required for efficient site-specific N-glycosylation of selected asparagines [17, 18]. Other CDGs have also been extensively studied in yeast. *Saccharomyces cerevisiae alg* mutants *(alg,* asparagine-linked glycosylation) are defective in the biosynthesis of the dolichol-linked $Glc_3Man_9GlcNAc_2$,[19, 20] and the mutations in the human *alg* homologs causing different types of CDG can be functionally studied by heterologous expression in yeast [6]. The ALG3 gene is highly conserved in eukaryotes, and its mutation also results in a similar phenotype in *Arabidobsis thalina* [21, 22]. Knock-out mutants of these *ALG* genes in *S. cerevisiae* are generally viable, and have also been used to characterize the pathway of N-glycosylation in the ER [19].

Glycosylation occupancy can be measured semi-quantitatively on many proteins using MS-based methods [17, 18, 23-25]. Traditionally in these approaches in yeast, Endo H is used as a deglycosylation agent followed by protease treatment and MS analysis to identify glycosylated and non-glycosylated peptides. Enrichment of glycoproteins rather than glycopeptides allows detection of both glycosylated and non-glycosylated versions of the same peptide. This allows direct relative quantification of site-specific glycosylation occupancy, although entails some reduction in the enrichment efficiency obtainable through glycopeptide enrichment [25]. Endo H cleaves all N-glycans of wild type yeast, which are all high-mannose structures, and conveniently leaves previously glycosylated asparagines tagged with a single *N*-acetylglucosamine (GlcNAc)[17]. Glycan release with Endo H therefore provides a clear distinction between the glycosylated and unglycosylated versions of the same peptide with a Δmass of 203.08 Da. However, in *S. cerevisiae Δalg3* mutants the incomplete precursor glycan $Man_5GlcNAc_2$-PP-Dol accumulates and Endo H resistant oligosaccharides are transferred to proteins [26].

In this study we chose a yeast *Δalg3* strain as a model system to develop an MS-based method that can measure the occupancy of specific glycosylation sites in different proteins in situations where Endo H cannot be used.

Experimental procedures

Yeast Strains

Yeast strains used were BY4741 (MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) and the *Δalg3* mutant derivative thereof (Open Biosystems). Cells were grown to mid log phase at 30˚C in YPD (2% Bacto peptone, 1% yeast extract, and 2% glucose).

Cell Wall Protein Sample Preparation

Triplicates of 50 ml of cells grown to mid log phase were harvested and lysed with agitation for 1 hr at 4˚C using glass beads (Sigma) in 50 mM Tris HCl, pH 7.5, 1 mM PMSF and 1 x Complete protease inhibitor cocktail (Roche). Based on previously reported methods [17, 27], covalently linked cell wall material was pelleted by centrifugation at 18000 rcf for 1 min and washed 4 times with ice-cold 50 mM Tris-HCl pH 7.5. The pellet was resuspended in 50 mM Tris-HCl pH 8.0, 2 % SDS, 7 M Urea and 2 M Thiourea. Cysteines were reduced and alkylated by addition of dithiothretiol to 10 mM and incubation at 30˚C with agitation for 30 min, followed by addition of acrylamide to 50mM and further incubation with agitation at 30˚C for 1 hr. Non-covalently linked proteins were removed by washing the 18000 rcf pellet five times with 1 ml 50mM Tris-HCl, pH 8, 2 % SDS, 7 M urea and 2 M thiourea, followed by five washes with 2 % SDS. The pellet was resuspended in 100 µl 1% NP40 and 1x G7 buffer (New England Biolabs) and aliquoted into 50 µl followed by addition of 500 units of PNGase F (New England Biolabs) and incubated at 37˚C with agitation for 16 h. The negative control

was incubated in the same buffer without PNGase F enzyme. The cell wall proteins were pelleted at 18000 rcf and washed 5 times in 50 mM Tris-HCl pH 8.0. The final cell wall pellet, equivalent to 12.5 ml of cell culture, was resuspended in 50 mM NH_4HCO_3, and proteins were digested with either trypsin 4 µg/ml (Sigma) or AspN 1 µg/ml (Promega) at 37˚C for 16 h with agitation.

Mass Spectrometry and data analysis

Peptides were desalted using C18 ZipTips (Millipore) and analysed by LC-ESI-MS/MS using a Prominence nanoLC system (Shimadzu) and TripleTof 5600 mass spectrometer with a Nanospray III interface (AB SCIEX). Approximately 2 µg peptides were desalted on an Agilent C18 trap (300 Å pore size, 5 µm particle size, 0.3 mm i.d. x 5 mm) at a flow rate of 30 µl/min for 3 min, and then separated on a Vydac EVEREST reversed-phase C18 HPLC column (300 Å pore size, 5 µm particle size, 150 µm i.d. x 150 mm) at a flow rate of 1 µl/min. Peptides were separated with a gradient of 10-60 % buffer B over 45 min, with buffer A (1 % acetonitrile and 0.1 % formic acid) and buffer B (80 % acetonitrile with 0.1 % formic acid). Gas and voltage setting were adjusted as required. An MS TOF scan from *m/z* of 350-1800 was performed for 0.5 s followed by information dependent acquisition of MS/MS with automated CE selection of the top 20 peptides from *m/z* of 40-1800 for 0.05 s per spectrum. Peptides were identified using ProteinPilot (AB SCIEX), searching the LudwigNR database (downloaded from http://apcf.edu.au as at 27 January 2012; 16,818,973 sequences; 5,891,363,821 residues) with standard settings: Sample type, identification; Cysteine alkylation, acrylamide; Instrument, TripleTof 5600; Species, *S. cerevisiae* with common contaminants; ID focus, biological modifications; Enzyme, trypsin or AspN; Search effort, thorough ID. False discovery rate analysis using ProteinPilot was performed on all searches.

Peptides identified with greater than 99 % confidence and with a local false discovery rate of less than 1 % were included for further analysis, and MS/MS fragmentation spectra were manually inspected. Extracted ion chromatograms were obtained using PeakView 1.1. Relative quantification of site-specific glycosylation occupancy was determined for each deamidated (deglycosylated) / unmodified (non-glycosylated) tryptic peptide pair without considering variable response factors of the unmodified and deamidated peptides. Occupancy was defined as the ratio of the extracted ion intensities of the deamidated peptide ion to the sum of the intensities of the deamidated and unmodified peptide ions [17].

Results

Identification and relative quantification of N-glycosylation site occupancy in cell wall proteins from wild type and Δalg3 yeast cells using PNGase F treatment followed by trypsin digestion. We asked whether glycosylation occupancy in the yeast *Δalg3* strain showed site-specific alterations. To efficiently detect multiple glycosylation sites, enrichment of glycoproteins is required [17, 28]. Enrichment of yeast cell wall proteins ensures detection of both glycosylated and non-glycosylated forms of proteins since the covalent linkage of the proteins to the yeast polysaccharide cell wall matrix does not rely directly on N-glycans. Previous methods for analysis of site-specific N-glycosylation occupancy in yeast have used glycan release with Endo H [17, 18, 23, 24]. However, the glycan structures on mature glycoproteins in the yeast *Δalg3* mutant are altered, with the mutant glycan structures engendering the glycoproteins resistant to deglycosylation with Endo H [29]. We therefore deglycosylated cell wall enriched glycoproteins from *Δalg3* and wild type yeast strains with the general deglycosylating enzyme PNGase F, leaving previously glycosylated asparagines “tagged” with a deamidation (m/z ~ 1

Da). Deglycosylated covalently attached cell wall proteins were digested with trypsin, detected by LC-ESI-MS/MS and identified with ProteinPilot (AB SCIEX). We identified peptides containing 19 NX(S/T) sequons as either previously glycosylated peptides modified with deamidation, or non-glycosylated unmodified peptides. This gave a total of 29 peptides (> 99 % confidence) from 12 different proteins (Table 1, Supporting Information). We measured the relative glycosylation occupancy [17] at these 19 sites by measurement of deamidated-peptide / peptide pairs (Table 2, and Fig. 1 A and C). We found 10 N-glycosylation sites that showed decreased levels of glycosylation occupancy in *Δalg3* compared to wild type. Ecm33p_328 showed the most severe decrease in glycosylation occupancy at 37 % (Table 2 and Fig. 1 A-D). Some glycosylation sites showed only mild under-glycosylation, such as Crh1p_177 that showed 96 % occupancy in the *Δalg3* mutant (Table 2). We found six sites that were completely (100 %) glycosylated in both *Δalg3* and wild type cells, and three sites were found to be completely non-glycosylated in both wild type and mutant yeast cells (Table 2). Two glycosylation sites were mildly under-glycosylated in wild type cells, Ecm33p_328 (Table 2 and Fig. 1 A and B) and Pry3p_101, and were both found to be further under-glycosylated in the *Δalg3* strain (Table 2). In summary, we found 16 glycosylation sites that were occupied in wild type cells, 10 of which were under-glycosylated in the *Δalg3* strain.

Spontaneous deamidation of NG sequences has been reported to occur during standard proteomic sample preparation [30-32]. It was therefore important to confirm that the deamidation events we detected were due to PNGase F catalysed glycan release, rather than spontaneous chemical deamidation of non-glycosylated asparagines. Two of the detected glycosylation sites, Crh2p_310 and Gas1p_40, have sequons with the sequence NGS/T

(Table 1), and would therefore be prone to spontaneous deamidation. We did not detect any deamidated peptide corresponding to Crh2p_310 in negative control samples not deglycosylated with PNGase F. In the negative control of *Δalg3* cells we did detect a weak MS peak corresponding to ions of deamidated Gas1p_40 peptide, indicating that some portion of the unglycosylated form of this peptide underwent spontaneous deamidation during sample preparation (data not shown). However, no such peak was detected in the wild type negative control, indicating that this site is completely glycosylated in wild type, and is partially under-glycosylated in *Δalg3* cells. The intensity of the deamidated Gas1p_40 peptide in the negative control was very low and was neglected in calculation of relative glycosylation occupancy at this site, leading to a slight underestimation of the level of under-glycosylation at Gas1p_40 in the *Δalg3* strain (Table 2). No detectable spontaneous deamidation of asparagines in other sequons were detected in any of the negative controls.

Identification of N-Glycosylation sites in yeast cell wall proteins with PNGase F treatment followed by AspN digestion.

Deglycosylation with PNGase F and digestion with trypsin allowed detection and relative quantification of site occupancy in *Δalg3* cells. However, the mass shift upon deamidation is ~1Da, which in some cases can be difficult to differentiate from the normal isotopic distribution of non-deamidated peptides in the case that the deamidated and non-deamidated versions of that peptide are not separated by C18 LC prior to MS detection. This is in contrast to the ready differentiation of previously glycosylated asparagines tagged with a Δmass of 203.08 Da in the case of Endo H glycan release. This ambiguity has the potential to make measurement of site-specific N-glycan occupancy difficult. To overcome these difficulties and to further identify specific N-glycosylation sites showing under-glycosylation in *Δalg3* yeast

cells, we therefore treated the cell-wall glycoproteins with PNGase F followed by digestion with the protease AspN and detection with LC-ESI-MS/MS. Deamidation of a previously glycosylated asparagine converts the amino acid to an aspartic acid that will be recognized as a cleavage site for AspN. Consequently AspN will cleave at deamidated asparagines that were glycosylated prior to treatment with PNGase F. This will allow clear differentiation of previously glycosylated asparagines, in contrast to the difficulties of PNGase F/trypsin as described above. Using this approach we identified 49 peptides (> 99 % confidence) that had been cleaved at a deamidated asparagine in a sequon (glycosylated) or contained an unmodified sequon (non-glycosylated) (Table 3, Supporting Information). The 49 peptides represented 42 NX(S/T) sequons in 14 different proteins (Table 3). 37 of these sequons were identified as glycosylated. We found seven sequons that were present in unmodified peptides (Table 3). Three of these unmodified sequons (Ecm33p_328, Gas3p_350 and Crh2p_96) were only found in *Δalg3* samples and were also detected as peptides cleaved at the corresponding deamidated asparagine in both wild type and *Δalg3* cells, indicating partial glycosylation at these three sites in the Δ*alg3* mutant. Four sequons were identified exclusively as non-glycosylated. Two of these (Ecm33p_209 and Crh1p_177) were only detected from *Δalg3* cells, indicating that these sites are under-glycosylated in the Δ*alg3* mutant strain (Table 3). Crh1p_177 was also identified in tryptic peptides both as a (deamidated) glycopeptide and a non-modified peptide (Table 1), showing some under-glycosylation at this site (Table 2). In total we identified five glycosylation sites that were under-glycosylated in the Δ*alg3* mutant strain using PNGase F followed by AspN treatment. In addition, we identified two sequons in non-modified peptides, Ccw14p_87 and Tos1p_417, present in both wild type and *Δalg3* positive and negative controls (no PNGase F treatment), suggesting that these two sequons are not glycosylated in wild type or the Δ*alg3* mutant strain

(Table 3). AspN cleavage due to spontaneous deamidation of an asparagine in an NGS sequon was detected at minor levels at Gas1p_40 in the *Δalg3* negative control, indicating that a portion of this site that was not glycosylated in these cells underwent spontaneous chemical deamidation. Relative quantification of glycosylation site occupancy was not performed using AspN peptides due to the inability to directly compare the deamidated and unmodified forms of the same peptide.

Characteristics of sequons that showed decreased glycosylation occupancy in Δalg3 cells.

N-Glycosylation is a general co-translational modification affecting many glycosylation sites on many different proteins within the cell. Only ~ 70 % of NX(S/T) sequons are glycosylated in proteins that are translocated into the ER [17, 18, 23, 24]. However, the sequons that were normally glycosylated in wild type cells were not evenly affected in the *Δalg3* mutant strain – some sites remained efficiently glycosylated while others showed severe under-glycosylation. We compared the characteristics of glycosylation sites that were underglycosylated and sites that showed no change in glycosylation occupancy in *Δalg3* cells (Table 2 and 4). We found that sequons that were efficiently glycosylated in *Δalg3* cells were more likely to be located in loops of folded proteins (not in secondary structural elements) and contain threonine in the +2 position (Fig 2).

Identification of novel N-glycosylation sites in yeast cell wall proteins.

In *S. cerevisae* there are 196 N-glycosylation sequons that have been directly experimentally shown to be occupied (UniProt) [17, 33]. The method we describe here used cell wall enriched proteins that contain a limited number of proteins compared to a recently described method using filter-aided-sample-preparation (FASP) from whole cell extracts enriched for

glycoproteins with ConA [33]. Nonetheless, using our approach we found in total 54 different occupied glycosylation sites (Table 1 and 3), 26 of which have not been reported previously (Table 5), raising the total number of known occupied N-glycosylation sites in *S. cerevisae* to 222.

Discussion

We developed an LC-MS based analytical method to identify N-glycosylation sites and successfully determined the relative glycosylation occupancy at 19 sites using the yeast *Δalg3* strain as a model. Independent of the yeast strain studied, deglycosylation with PNGase F followed by AspN digestion increased the number of glycosylation sites that could be detected in a complex protein sample at one time, when compared with trypsin digest. It appears that proteins with multiple N-glycosylation sites become more accessible to digest with AspN and MS analysis after treatment with PNGase F compared with Endo H. Of the 26 new sites we found in this study, 22 were identified using the combination of PNGase F and AspN. In total we could detect 54 different occupied glycosylation sites. The approach we describe here is therefore complementary to standard techniques, and increases the number of N-glycosylation sites that can be detected at the same time on many different proteins.

While most glycosylation sequons were efficiently glycosylated in wild type cells, the altered glycan structure in the Δ*alg3* strain reduces the general glycosylation activity of OTase, leading to underglycosylation [26, 29]. However, glycosylation sites were not uniformly affected in Δ*alg3* cells, with the subset of sites with lower affinity for OTase most strongly affected. We found 13 specific glycosylation sites that were under-glycosylated in the Δ*alg3* strain

compared to wild type cells (Table 4). The features of these glycosylation sites suggest that they are the subset of normally modified sites that are more difficult for OTase to glycosylate. For instance, these under-glycosylated sites are enriched in serine at the +2 position (Fig 2A). NxS sequons are glycosylated ~40 times less efficiently than NxT sequons,[34] because of specific recognition of threonine at the +2 position by the peptide acceptor binding site of OTase [35]. Under-glycosylated sites in the Δ*alg3* strain were also likely to be present in secondary structural elements (helices or sheets), whereas efficiently glycosylated sites were more likely present in flexible loops (Fig 2B). Protein folding competes with glycosylation by OTase because the peptide acceptor binding site of OTase requires a flexible stretch of peptide for tight binding [35]. As secondary structure forms early in protein folding, glycosylation sites in secondary structural elements will have reduced affinity to OTase. The subset of glycosylation sites both present in flexible loops and with threonine at the +2 position were efficiently glycosylated (Fig 2C), suggesting a combination of these factors influences peptide acceptor binding to OTase, and hence glycosylation efficiency.

Incorporation of either of the homologous proteins Ost3p or Ost6p into OTase results in enzyme isoforms with different protein substrate specificities at the level of individual glycosylation sites [17]. Six of the 13 sites that show under-glycosylation in *Δalg3* cells have been shown to be under-glycosylated in yeast cells with double knockout of these homologous OTase subunits (*Δost3*/*Δost6*) [17]. All six of these sites were shown to be rescued by overexpression of Ost3p, although two sites could be rescued by over-expressing of either Ost3p or Ost6p [17]. Furthermore, one glycosylation site, Gas1p_253, was shown to require Ost6p for complete glycosylation occupancy [17]. Interestingly, we found that glycosylation occupancy of Gas1p_253 was not affected in Δ*alg3* cells (Table 2), which express both

Ost3p- and Ost6p-containing OTase isoforms. This emphasizes the importance of the protein substrate specificities of OTase isoforms in defining the characteristics of an efficiently glycosylated asparagine.

Our results may be complicated by the ability of many proteins to fold correctly even without glycosylation at many sites, as long as a certain critical level of glycosylation is present. This will mean that if glycosylation is strictly required at a particular asparagine for correct protein folding, then that asparagine will appear to always be glycosylated, even though most of the nascent polypeptide that is not modified will be degraded by the quality control systems of the ER [2]. In addition, glycosylation efficiency may depend on factors such as local folding of the target protein close to the sequon and the protein substrate specificities of the particular OTase isoforms present [17, 36]. However, in combination this and previous studies have detected almost every sequon in both glycosylated and non-glycosylated forms, suggesting that these effects are relatively minor. Therefore, it is reasonable to imply that sites that are easier to glycosylate would be expected to show full occupancy and sites that are more difficult to glycosylate would be found under-glycosylated.

Conclusion

Mutations in the human ALG3 gene cause a rare form of CDG (CDG-Id) [37]. Reduced OTase activity due to the altered structure of the lipid-linked glycan substrate in this disorder is likely to affect N-glycosylation occupancy in a site-specific manner, similar to the patterns we observe here in a *S. cerevisiae* model system. We predict that glycosylation sites in NxS sequons and those located in secondary structural elements will be the most likely to show reduced glycosylation in CDG-Id. Many different diseases and health conditions in addition to

the CDG spectrum are caused by or associated with defects in glycosylation and consequently decreased levels of N-glycosylation occupancy at many different glycosylation sites in many different proteins. In conjunction with appropriate glycoprotein or glycopeptide enrichment strategies, the method we describe here could potentially be further developed to study human N-glycosylation site occupancy using clinical tissue, serum or saliva protein samples, leading to discovery of new biomarkers for CDG and other disorders and improved diagnostics.

References

1. Kelleher, D. J.; Gilmore, R., An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **2006,** 16, (4), 47R-62R.

2. Schulz, B. L., Beyond the sequon: sites of N-glycosylation. In *Glycosylation*, InTech: **2012**; ISBN 978-953-51-0771-2.

3. Helenius, A.; Aebi, M., Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem* **2004,** 73, 1019-49.

4. Chen, W.; Helenius, A., Role of Ribosome and Translocon Complex during Folding of Influenza Hemagglutinin in the Endoplasmic Reticulum of Living Cells. *Mol Biol Cell* **2000,** 11, (2), 765-72.

5. Whitley, P.; Nilsson, I. M.; von Heijne, G., A nascent secretory protein may traverse the ribosome/ER translocase complex as an extended chain. *J Biol Chem* **1996,** 271, 6241-44.

6. Haeuptle, M. A.; Hennet, T., Congenital disorders of glycosylation: an update on defects affecting the biosynthesis of dolichol-linked oligosaccharides. *Hum Mutat* **2009,** 30, (12), 1628-41.

7. Janik, M. E.; Litynska, A.; Vereecken, P., Cell migration-the role of integrin glycosylation. *Biochim Biophys Acta* **2010,** 1800, (6), 545-55.

8. Sato, S.; St-Pierre, C.; Bhaumik, P.; Nieminen, J., Galectins in innate immunity: dual functions of host soluble beta-galactoside-binding lectins as damage-associated molecular patterns (DAMPs) and as receptors for pathogen-associated molecular patterns (PAMPs). *Immunol Rev* **2009,** 230, (1), 172-87.

9. Bova, G. S.; Carter, B. S.; Bussemakers, M. J. G.; Emi, M.; Fujiwara, Y.; Kyprianou, N.; Jacobs, S. C.; Robinson, J. C.; Epstein, J. I.; Walsh, P. C.; Isaacs, W. B., Homozygous Deletion and Frequent Allelic Loss of Chromosome 8p22 Loci in Human Prostate Cancer. *Cancer Res* **1993,** 53, (17), 3869-73.

10. Honma, K.; Iwao-Koizumi, K.; Takeshita, F.; Yamamoto, Y.; Yoshida, T.; Nishio, K.; Nagahara, S.; Kato, K.; Ochiya, T., RPN2 gene confers docetaxel resistance in breast cancer. *Nat Med* **2008,** 14, (9), 939-48.

11. Pils, D.; Horak, P.; Gleiss, A.; Sax, C.; Fabjani, G.; Moebus, V. J.; Zielinski, C.; Reinthaller, A.; Zeillinger, R.; Krainer, M., Five genes from chromosomal band 8p22 are significantly down-regulated in ovarian carcinoma. *Cancer* **2005,** 104, (11), 2417-29.

12. Apweiler, R.; Hermjakob, H.; Sharon, N., On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* **1999,** 1473, (1), 4-8.

13. Hülsmeier, A. J.; Paesold-Burda, P.; Hennet, T., N-glycosylation site occupancy in serum glycoproteins using multiple reaction monitoring liquid chromatography mass spectrometry. *Mol Cell Proteomics* **2007,** 6, (12), 2132-8.

14. Garshasbi, M.; Hadavi, V.; Habibi, H.; Kahrizi, K.; Kariminejad, R.; Behjati, F.; Tzschach, A.; Najmabadi, H.; Ropers, H. H.; Kuss, A. W., A defect in the TUSC3 gene is

associated with autosomal recessive mental retardation. *Am J Hum Genet* **2008,** 82, (5), 1158-64.

15. Mohorko, E.; Glockshuber, R.; Aebi, M., Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. *J Inherit Metab Dis* **2011,** 34, (4), 869-78.

16. Molinari, F.; Foulquier, F.; Tarpey, P. S.; Morelle, W.; Boissel, S.; Teague, J.; Edkins, S.; Futreal, P. A.; Stratton, M. R.; Turner, G.; Matthijs, G.; Gecz, J.; Munnich, A.; Colleaux, L., Oligosaccharyltransferase-subunit mutations in non-syndromic mental retardation. *Am J Hum Genet* **2008,** 82, (5), 1150-7.

17. Schulz, B. L.; Aebi, M., Analysis of Glycosylation Site Occupancy Reveals a Role for Ost3p and Ost6p in Site-specific N-Glycosylation Efficiency. *Mol Cell Proteomics* **2009,** 8, (2), 357-364.

18. Schulz, B. L.; Stirnimann, C. U.; Grimshaw, J. P. A.; Brozzo, M. S.; Fritsch, F.; Mohorko, E.; Capitani, G.; Glockshuber, R.; Grütter, M. G.; Aebi, M., Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc Natl Acad Sci U S A* **2009,** 106, (27), 11061-6.

19. Burda, P.; Aebi, M., The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta* **1999,** 1426, (2), 239-57.

20. Runge, K. W.; Robbins, P. W., A new yeast mutation in the glucosylation steps of the asparagine-linked glycosylation pathway. Formation of a novel asparagine-linked oligosaccharide containing two glucose residues. *J Biol Chem* **1986,** 261, (33), 15582-90.

21. Henquet, M.; Eigenhuijsen, J.; Spiegel, H.; Schreuder, M.; van Duijn, E.; Cordewener, J.; Depicker, A.; van der Krol, A.; Bosch, D., Characterization of the single-chain Fv-Fc antibody MBP10 produced in Arabidopsis alg3 mutant seeds. *Transgenic Res* **2011,** 20, (5), 1033-42.

22. Henquet, M.; Lehle, L.; Schreuder, M.; Rouwendal, G.; Molthoff, J.; Helsper, J.; van der Krol, A.; Bosch, D., Identification of the gene encoding the alpha1,3-mannosyltransferase (ALG3) in Arabidopsis and characterization of downstream n-glycan processing. *Plant Cell* **2008,** 20, (6), 1652-64.

23. Izquiedro, L.; Schulz, B. L.; Rodrigues, J. A.; Güther, M. L.; Procter, J. B.; Barton, G. J.; Aebi, M.; Ferguson, M. A., Distinct donor and acceptor specificities of Trypanosoma brucei oligosaccharyltransferases. *EMBO J* **2009,** 28, (17), 2650-61.

24. Nasab, F.; Schulz, B.; Gamarro, F.; Parodi, A.; Aebi, M., All in one: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in Saccharomyces cerevisiae. *Mol Biol Cell* **2008,** 19, (9), 3758-68.

25. Zhang, H.; Li, X. J.; Martin, D. B.; Aebersold, R., Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* **2003,** 21, (6), 660-6.

26. Huffaker, T. C.; Robbins, P. W., Yeast mutants deficient in protein glycosylation. *Proc Natl Acad Sci U S A* **1983,** 80, (24), 7466-70.

27. Yin, Q. Y.; de Groot, P. W.; Dekker, H. L.; de Jong, L.; Klis, F. M.; de Koster, C. G., Comprehensive proteomic analysis of Saccharomyces cerevisiae cell walls: identification of proteins covalently attached via glycosylphosphatidylinositol remnants or mild alkali-sensitive linkages. *J Biol Chem* **2005,** 280, (21), 20894-901.

28. Schulz, B. L.; White, J. C.; Punyadeera, C., Saliva proteome research: current status and future outlook. *Crit Rev Biotech* **2012**.

29. Aebi, M.; Gassenhuber, J.; Domdey, H.; te Heesen, S., Cloning and characterization of the ALG3 gene of Saccharomyces cerevisiae. *Glycobiology* **1996,** 6, (4), 439-44.

30. Krokhin, O. V.; Antonovici, M.; Ens, W.; Wildins, J. A.; Standing, K. G., Deamidation of -Asn-Gly- sequences during sample preparation for proteomics: Consequences for MALDI and HPLC-MALDI analysis. *Anal Chem* **2006,** 78, (18), 6645-50.

31. Robinson, N. E.; Robinson, Z. W.; Robinson, B. R.; Robinson, A. L.; Robinson, J. A.; Robinson, M. L.; Robinson, A. B., Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. *J Pept Res* **2004,** 63, (5), 426-36.

32. Palmisano, G.; Melo-Braga, M. N.; Engholm-Keller, K.; Parker, B. L.; Larsen, M. R., Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *J Proteome Res* **2012,** 11, (3), 1949-57.

33. Breidenbach, M. A.; Palaniappan, K. K.; Pitcher, A. A.; Bertozzi, C. R., Mapping yeast N-glycosites with isotopically recoded glycans. *Mol Cell Proteomics* **2012,** 11, (6), M111.015339.

34. Kasturi, L.; Eshleman, J. R.; Wunner, W. H.; Shakin-Eshleman, S. H., The Hydroxy Amino Acid in an Asn-X-Ser/Thr Sequon Can Influence N-Linked Core Glycosylation Efficiency and the Level of Expression of a Cell Surface Glycoprotein. *J. Biol. Chem.* **1995,** 270, (24), 14756-61.

35. Lizak, C.; Gerber, S.; Numao, S.; Aebi, M.; Locher, K. P., X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **2011,** 474, (7351), 350-5.

36. Jamaluddin, M. F. B.; Bailey, U. M.; Tan, N. Y. J.; Stark, A. P.; Schulz, B. L., Polypeptide binding specificities of Saccharomyces cerevisiae oligosaccharyltransferase accessory proteins Ost3p and Ost6p. *Protein Sci* **2011,** 20, (5), 849-55.

37. Sun, L.; Eklund, E. A.; Chung, W. K.; Wang, C.; Cohen, J.; Freeze, H. H., Congenital disorder of glycosylation id presenting with hyperinsulinemic hypoglycemia and islet cell hyperplasia. *J Clin Endocrinol Metab* **2005,** 90, (7), 4371-5.

Tables

Table 1. *Peptides containing N-glycosylation sequons identified by mass spectrometry.* Proteins covalently linked to the yeast polysaccharide cell wall were prepared, glycans were released by peptide-N-glycosidase F, and proteins were digested with trypsin, detected by LC-ESI-MS/MS, and identified by ProteinPilot.

Protein name	Peptide position	Sequence	Asn	*m/z*	*z*	ΔMass
Gas1p	35-47	FFYSN**N**GSQFYIR	40	548.5878	3	-0.0001
Gas1p	35-47	FFYSNNGSQFYIR	40	548.2642	3	0.0132
Gas1p	48-77	GVAYQADTA**N**ETSGSTVNDPLANYESCSR	57	1031.453	3	0.0056
Gas1p	48-77	GVAYQADTANETSGSTVNDPLANYESCSR	57	1031.123	3	-0.0014
Gas1p	91-105	VYAI**N**TTLDHSECMK	95	599.6099	3	-0.0042
Gas1p	253-271	**N**LSIPVFFSEYGCNEVTPR	253	1122.536	2	0.0008
Gas5p	50-73	GVDYQPGGSS**N**LTDPLADASVCDR	60	837.0438	3	0.0022
Gas5p	50-73	GVDYQPGGSSNLTDPLADASVCDR	60	836.7169	3	0.0055
Pst1p	51-65	CDTLVG**N**LTIGGGLK	57	766.9019	2	-0.0025
Pst1p	51-65	CDTLVGNLTIGGGLK	57	766.4099	2	-0.0023
Ecm33p	318-335	GGANFDSSSS**N**FSCNALK	328	939.4022	2	0.0002
Ecm33p	318-335	GGANFDSSSSNFSCNALK	328	938.909	2	-0.0023
Pry3p	100-120	Y**N**YSNPGFSESTGHFTQVVWK	101	817.042	3	0.0019
Pry3p	100-120	YNYSNPGFSESTGHFTQVVWK	101	816.7147	3	0.0039
Yjl171cp	219-240	**N**SSSIGYYDLPAIWLLNDHIAR	219	840.4252	3	0.0044

Yjl171cp	219-240	NSSSIGYYDLPAIWLLNDHIAR	219	840.0991	3	0.0102
Plb2p	47-60	**N**ASGLSTAETDWLK	47	747.3575	2	-0.0042
Plb2p	47-60	NASGLSTAETDWLK	47	746.8659	2	-0.0032
Plb2p	185-200	SIVNPGGS**N**LTYTIER	193	574.6266	3	-0.0051
Plb2p	362-382	YVNNLSQDDDDDIAIYAANPFK	365	796.0468	3	0.0061
Plb2p	491-509	**N**LTDLEYIPPLVVYIPNTK	491	1102.099	2	0.0019
Gas3p	268-288	L**N**STFEDAVIPLIFSEYGCNK	269	811.3958	3	0.0039
Gas3p	268-288	LNSTFEDAVIPLIFSEYGCNK	269	811.0667	3	0.0004
Crh1p	175-186	FH**N**YTLDWAMDK	177	770.8482	2	0.0002
Crh1p	175-186	FHNYTLDWAMDK	177	771.3389	2	-0.0023
Crh2p	23-59	ATFC**N**ATQACPEDKPCCSQYGECGT GQYCLNNCDVR	28	1098.453	4	0.0215
Crh2p	310-326	**N**GTSAYVYTSSSEFLAK	310	609.288	3	0.0003
Tos1p	412-421	AAVIFNSSDK	417	526.271	2	-0.0071
Yil169cp	508-532	GEGVLAVDPTETNATPIPVVGYTGK	520	829.1041	3	0.0157

Deamidated asparagines are bold. Unglycosylated sequons are underlined. All methionines are oxidized and all cysteines are alkylated. All peptides listed were identified with > 99 % confidence.

Table 2. *Site-specific N-glycosylation occupancy in Δalg3 yeast strain.* The relative N-glycosylation occupancy at a given site was determined from the abundance of deamidated-modified and unmodified versions of the same sequon-containing peptide as measured by LC-ESI-MS.

Glycosylation site	Strain	
	Wild type	*Δalg3*
Gas1p_95	1.00±0.00*	1.00±0.00
Gas1p_253	1.00±0.00	1.00±0.00
Plb2p_193	1.00±0.00	1.00±0.00
Plb2p_491	1.00±0.00	1.00±0.00
Crh2p_28	1.00±0.00	1.00±0.00
Crh2p_310	1.00±0.00	1.00±0.00
Gas5p_60	1.00±0.00	0.97±0.05
Crh1p_177	1.00±0.00	0.96±0.01
Gas1p_40	1.00±0.00	0.94±0.05
Gas3p_269	1.00±0.00	0.92±0.04
Pry3p_101	0.96±0.02	0.85±0.02
Pst1p_57	1.00±0.00	0.76±0.12
Plb2p_47	1.00±0.00	0.66±0.17
Gas1p_57	1.00±0.00	0.57±0.11
Yjl171cp_219	1.00±0.00	0.51±0.18
Ecm33_328	0.99±0.01	0.37±0.08
Plb2_365	0.00±0.00	0.00±0.00
Tos1p_417	0.00±0.00	0.00±0.00

Yil169cp_520	0.00±0.00	0.00±0.00

*, Values are mean of biological triplicates. Error is SD.

Table 3. *Peptides containing N-glycosylation sequons identified by mass spectrometry.* Proteins covalently linked to the yeast polysaccharide cell wall were prepared, glycans were released by peptide-N-glycosidase F, and proteins were digested with AspN, detected by LC-ESI-MS/MS, and identified by ProteinPilot.

Protein name	Peptide Position	Peptide Sequence	Asn	*m/z*	*z*	ΔMass
Gas1p	23-39	D*DVPAIEVVGNKFFYSN.**N**GS	40	957.4678	2	0.0003
Gas1p	40-53	**N**GSQFYIRGVAYQA	40	525.5922	3	0.0021
Gas1p	77-94	DIPYLKKLNTNVIRVYAI.**N**TT	95	711.7537	3	0.0035
Gas5p	150-165	DNVLGFFAGNEVINSV.**N**TT	166	847.9219	2	-0.0020
Gas5p	166-179	**N**TTNTATYVKAVVR	166	769.9114	2	-0.0017
Gas5p	319-343	DFENLKNEYSKVSNPEGNGGYSTSN.**N**YS	344	917.4109	3	0.0007
Ecm33p	45-55	DKISGCSTIVG.**N**LT	56	575.7883	2	-0.0080
Ecm33p	82-91	**N**SSSLSSFSA	82	494.2102	2	-0.0135
Ecm33p	202-213	DNLVWAN<u>NIT</u>LR#	209	714.8843	2	0.0020
Ecm33p	214-226	DVNSISFGSLQTV.**N**AS	227	683.8457	2	-0.0007
Ecm33p	241-257	**N**LTQLSKVGQSLSIVSN	241	894.9885	2	-0.0006
Ecm33p	258-266	DELSKAAFS.**N**LT	267	484.2396	2	-0.0012
Ecm33p	279-286	**N**NTQLKVI	279	465.7654	2	-0.0020
Ecm33p	287-303	DGFNKVQTVGGAIEVTG.**N**FS	304	846.4332	2	-0.0007
Ecm33p	323-345	DSSSS<u>NFS</u>CNALKKLQSNGAIQG#	328	809.7242	3	0.0011
Ecm33p	328-345	**N**FSCNALKKLQSNGAIQG	328	655.6658	3	-0.0031
Pry3p	91-100	DAWYGEISKY.**N**YS	101	616.2849	2	-0.0005

Yjl171cp	286-299	DLGTGIQSYGYITR.**N**TT	300	772.392	2	0.0017
Ccw14p	73-91	DAAYSAFKSSCSEQNASLG	87	669.6309	3	0.0022
Plb1p	81-91	DTSLLSTLFGS.**N**SS	92	570.7927	2	-0.0001
Plb1p	265-276	DGRYPGTTVINL.**N**AT	277	653.3411	2	-0.0049
Plb1p	492-512	DLEYIPPLIVYIPNSRHSFNG.**N**QS	513	815.4273	3	0.0064
Plb2p	83-93	DTSLLSTLFSS.**N**SS	94	585.7971	2	-0.0019
Plb2p	162-170	**N**WTSVQEIV	162	538.7665	2	-0.0001
Plb2p	493-514	DLEYIPPLVVYIPNTKHSFNG.**N**QS	515	806.0919	3	0.0063
Plb2p	567-579	**N**ATLPPECTKCFA	567	769.3527	2	-0.0045
Plb2p	615-629	DGIPITALLGSSTSG.**N**TT	630	694.8669	2	-0.0003
Gas3p	201-211	**N**RSIPVGYSAA	201	568.2903	2	-0.0009
Gas3p	191-200	DMKQYISKHA.**N**RS	201	610.8044	2	-0.0078
Gas3p	339-349	D*DFVNLESQLK.**N**VS	350	654.3276	2	0.0001
Gas3p	350-363	**N**VSLPTTKESEISS	350	746.8725	2	0.0000
Gas3p	339-363	D*DFVNLESQLKNVSLPTTKESEISS#	350	927.4709	3	0.0145
Gas3p	370-384	DNSAITNIYSGFGTN.**N**FT	385	787.36	2	0.0000
Gas3p	414-421	DYAVPTTF.**N**YT	422	457.219	2	0.0005
Crh1p	117-129	**N**GTGIVSSFYLQS	117	687.3319	2	-0.0017
Crh1p	173-180	DKFHNYTL#	177	519.2554	2	-0.0016
Crh1p	192-200	DGESVRVLS.**N**TS	201	481.2486	2	-0.0051
Crh2p	82-95	DYSSKLGNANTFLG.**N**VS	96	743.8612	2	-0.0021
Crh2p	82-100	DYSSKLGNANTFLGNVSEA#	96	993.9658	2	-0.0160
Crh2p	175-189	DLETAQTNFYWESVL.**N**YT	190	908.4202	2	-0.0105
Crh2p	223-232	DGVVGRTLYK.**N**ET	233	554.3093	2	-0.0044

Crh2p	237-252	**N**ATTQKYQYPQTPSKV	237	927.9626	2	-0.0052
Crh2p	253-260	DISIWPGG.**N**ST	261	422.7143	2	0.0013
Crh2p	282-296	DISNPGYYYAIVNEV.**N**IT	297	858.9086	2	-0.0016
Crh2p	310-326	**N**GTSAYVYTSSSEFLAK	310	913.4259	2	-0.0045
Tos1p	222-235	DWSRGSYFVPGSTS.**N**CT	236	773.3513	2	-0.0016
Tos1p	407-419	DGTLKAAVIFNSS	417	661.8506	2	-0.0011
Pir3p	231-244	**N**STLSMSLSKGILT	231	726.8838	2	-0.0011
Pst1p	215-227	DVHSVSFANLQKI.**N**SS	228	729.3972	2	0.0124

Deamidated asparagines are bold. Unglycosylated sequons are underlined. *, missed cleavage. #, Peptides containing unglycosylated sequons identified in *Δalg3* only. All methionines are oxidized and all cysteines are alkylated. All peptides listed were identified with > 99 % confidence.

Table 4. *Summary of N-glycosylation sites that are underglycosylated in Δalg3*

Glycosylation site	Phenotype	Identified with
Gas1p_40	mild	Trypsin
Gas1p_57	strong	Trypsin
Gas5p_60	mild	Trypsin
Pst1p_57	medium	Trypsin
Ecm33_209	n.a.	AspN
Ecm33_328	strong	Trypsin/AspN
Pry3p_101	medium	Trypsin
Yjl171cp_219	strong	Trypsin
Plb2_47	strong	Trypsin
Gas3p_269	mild	Trypsin
Gas3p_350	n.a.	AspN
Crh1p_177	mild	Trypsin/AspN
Crh2p_96	n.a.	AspN

n.a. not applicable

Table 5. *Novel N-glycosylation sites detected using PNGase F.*

Uniprot ID	Protein name	Occupied N-glycosylation sites
Q08193	Gas5p	Asn 166
P38248	Ecm33p	Asn 56, 82, 227, 241, 304
P47033	Pry3p	Asn 101
P46992	Yjl171cp	Asn 219, 300
P39105	Plb1p	Asn 92, 277
Q03674	Plb2p	Asn 47, 94, 162, 515, 567
Q03655	Gas3p	Asn 269, 385, 422
P53301	Crh1p	Asn 201
P32623	Crh2p	Asn 190, 237, 261, 297
Q03180	Pir3p	Asn 231
Q03180	Pst1p	Asn 57

Figure legends

Figure 1. *LC-ESI-MS/MS identification and relative quantificatiotion of glycosylation occupancy at Asn^{328} of Ecm33p from wild type and Δalg3 cells.* Extracted ion chromatograms of the $[M+2H]^{2+}$ ions corresponding to deamidated peptide at an *m/z* of 939.40 (dashed line) and unmodified peptide at an *m/z* of 938.91 (solid line) from wild type (A) and (B) and *Δalg3* (C) and (D). The glycoprotein enriched cell wall proteins were treated with PNGase F followed by trypsin (A and C), or with trypsin (B and D). The ratio of the area under the deamidated peptide to the total of deamidated and unmodified peptide gave a relative occupancy of 0.98 in wild type (A) and 0.37 in *Δalg* (C) in this single measurement. The scale of the intensity is the same in A and B. MS/MS spectra and ProteinPilot matched fragmentation of deamidated peptide (E) and unmodified peptide (F) are shown.

Figure 2: *Characteristics of glycosylation sites that are underglycosylated in Δalg3 cells.* Proportion of glycosylation sites that are present in (A) loops, (B) NxT sequons or (C) NxT sequons in loops, that show no change in occupancy (black) or that are underglycosylated (white) in *Δalg3* cells compared with wild type cells. * $p<0.05$, ** $p<0.01$, Fisher's exact test.

Figure 1

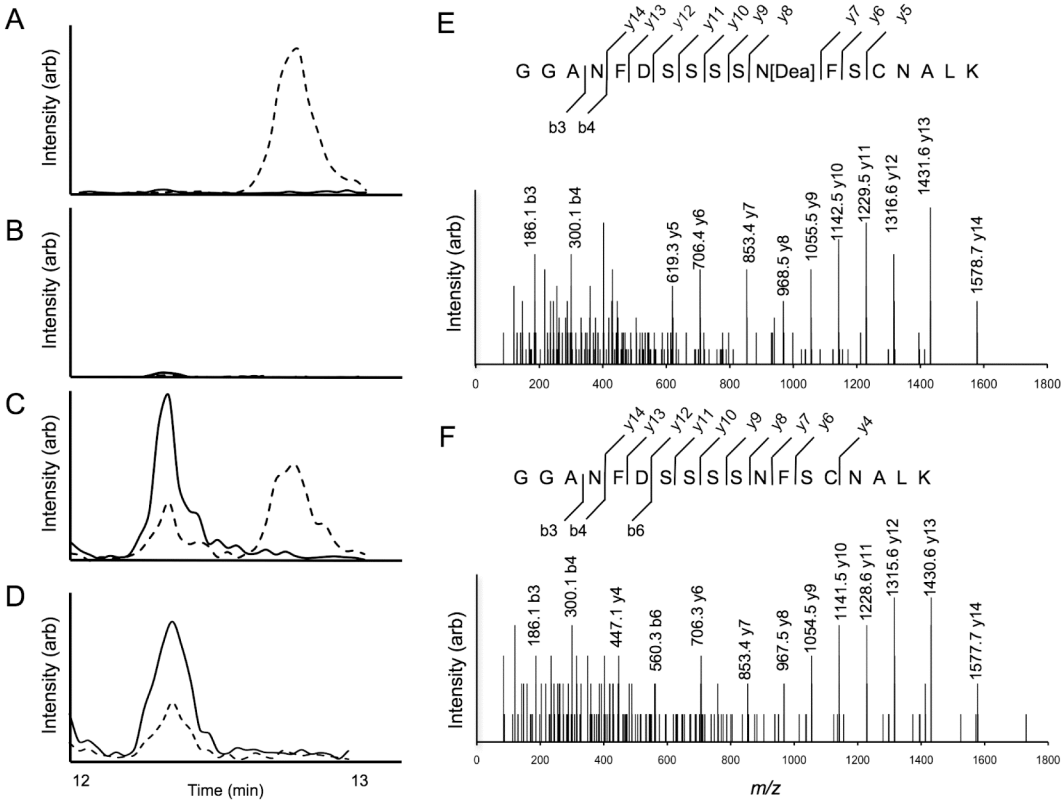
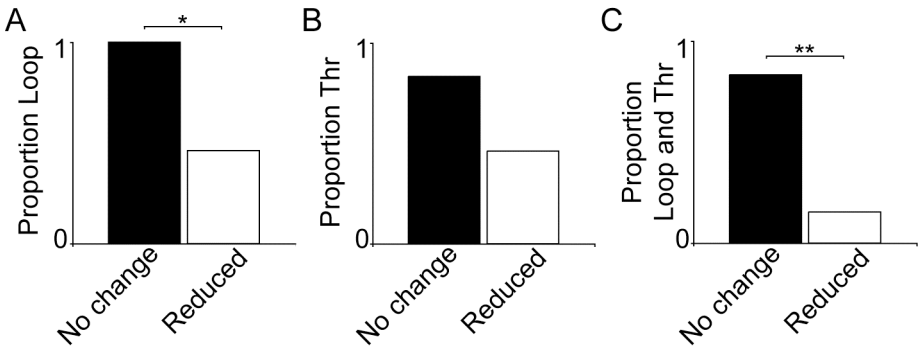

A
B
C
D
Intensity (arb)
12
13
Time (min)
E
F
G G A N F D S S S S N[Dea] F S C N A L K
G G A N F D S S S S N F S C N A L K
b3 b4
b6
y14 y13 y12 y11 y10 y9 y8 y7 y6 y5 y4
186.1 b3
300.1 b4
447.1 y4
560.3 b6
619.3 y5
706.4 y6
706.3 y6
853.4 y7
853.4 y7
968.5 y8
967.5 y8
1055.5 y9
1054.5 y9
1141.5 y10
1142.5 y10
1228.6 y11
1229.5 y11
1315.6 y12
1316.6 y12
1430.6 y13
1431.6 y13
1577.7 y14
1578.7 y14
0 200 400 600 800 1000 1200 1400 1600 1800
m/z

Figure 2


A
*
Proportion Loop
1
0
No change
Reduced
B
Proportion Thr
1
0
No change
Reduced
C
**
Proportion Loop and Thr
1
0
No change
Reduced

Graphical Abstract