ARTICLE

Received 14 Aug 2013 | Accepted 12 Dec 2013 | Published 17 Jan 2014

DOI: 10.1038/ncomms4099

Sequence-based protein stabilization in the absence of glycosylation

Nikki Y. Tan¹, Ulla-Maja Bailey¹, M Fairuz Jamaluddin¹, S Halimah Binte Mahmud¹, Suresh C. Raman¹
& Benjamin L. Schulz¹

Asparagine-linked *N*-glycosylation is a common modification of proteins that promotes productive protein folding and increases protein stability. Although *N*-glycosylation is important for glycoprotein folding, the precise sites of glycosylation are often not conserved between protein homologues. Here we show that, in *Saccharomyces cerevisiae*, proteins upregulated during sporulation under nutrient deprivation have few *N*-glycosylation sequons and in their place tend to contain clusters of like-charged amino-acid residues. Incorporation of such sequences complements loss of *in vivo* protein function in the absence of glycosylation. Targeted point mutation to create such sequence stretches at glycosylation sequons in model glycoproteins increases *in vitro* protein stability and activity. A dependence on glycosylation for protein stability or activity can therefore be rescued with a small number of local point mutations, providing evolutionary flexibility in the precise location of *N*-glycans, allowing protein expression under nutrient-limiting conditions, and improving recombinant protein production.

¹School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland 4072, Australia. Correspondence and requests for materials should be addressed to B.L.S. (email: b.schulz@uq.edu.au).

The presence of *N*-glycans assists protein folding, increases protein solubility, reduces aggregation and limits proteolysis^{1–5}, while their specific structures affect glycoprotein function in development, cancer, inflammation and infection^{3,6}. In eukaryotes, *N*-glycans are added to selected asparagine residues in nascent polypeptides co- and post-translocationally in the lumen of the endoplasmic reticulum (ER) by oligosaccharyl-transferase (OTase)^{7–10}. *N*-glycosylation by OTase occurs before substrate protein folding, as only flexible polypeptide substrate can bind at its active site¹¹. The presence of Ser or Thr at the +2 position markedly increases glycosylation efficiency of asparagines, due to specific interactions with the peptide-binding site of OTase¹¹, and such sequences are termed glycosylation sequons (N-x-S/T; x ≠ P). Asparagines in secondary structural elements tend to be inefficiently glycosylated^{12,13}, as their reduced flexibility lowers their affinity to the OTase active site¹¹. *N*-glycosylation therefore often occurs between secondary structural elements of folded proteins¹⁴. This allows *N*-glycans to optimally assist productive protein folding by their hydrophilic bulk lowering the entropy of the protein unfolded state and acting as nucleation sites for local folding of nascent glycopolypeptide chains². *N*-glycans can also increase protein stability by making direct contacts with amino acids in the folded protein^{15,16}, increasing overall protein solubility¹⁷, shielding exposed hydrophobic surfaces¹⁸ and limiting proteolysis¹⁹. In the ER, *N*-glycans act as signals to locally recruit ER chaperones to assist in glycoprotein oxidative folding and are important in glycoprotein quality control and degradation^{2,20–22}. Although *N*-glycans are important for protein folding and function, the precise sites of glycosylation within a protein are not generally conserved in homologues.

Here, we study the extent to which peculiar amino-acid sequences can functionally replace *N*-glycosylation sites. We show that clusters of like-charged amino acids in place of *N*-glycosylation sequons can impart local protein stability and improve protein function in the absence of glycosylation.

Results

Sequence analysis of yeast cell wall *N*-glycoproteins. In the absence of a nitrogen source and in the presence of a non-fermentable carbon source, bakers' yeast *S. cerevisiae* ceases vegetative growth and enters sporulation²³. This results in large changes to many aspects of yeast biology including remodelling of the cell wall. Key enzymes involved in this process are the GAS family of glucanases/transglycosidases. Gas1p is a major cell-wall protein during vegetative growth while its functionally equivalent paralogue Gas2p is predominantly expressed during sporulation^{24,25}. Gas1p and Gas2p share 45% sequence identity, conserved secondary structure and identical disulphide bond connectivity^{26,27}. However, Gas1p has 10 *N*-glycosylation sequons while Gas2p has only 1. This prompted us to test if other sporulation-induced proteins also had few glycosylation sequons. This analysis showed that glycoproteins upregulated in sporulation did have fewer glycosylation sequons than their paralogues expressed during vegetative growth (Fig. 1a). The nutrient limitations that induce sporulation may provide a selective pressure for reducing flux through the energetically expensive *N*-glycosylation pathway, resulting in loss of *N*-glycosylation sequons from sporulation-specific proteins.

Given the importance of *N*-glycosylation for protein folding, we investigated if the equivalent positions of *N*-glycans in paralogues that lacked a sequon at that particular site had compensatory amino-acid sequences. For this, we used a set of experimentally validated yeast glycosylation sites^{13,28–30}. Sequence analysis of positions in paralogues corresponding to

these sites showed that they had more acidic amino acids compared with both validated glycosylation sites and the average amino-acid composition of the whole protein set (Fig. 1b–d). This enrichment was only significant in the immediate vicinity of the residues corresponding to the location of the glycosylation sequon (Fig. 1e–h), although the high sequence diversity in this analysis precluded the identification of more specific patterns.

Functional complementation of glycosylation sites in yeast.

To test if these clusters of acidic amino acids served to provide local intrinsic stability to the protein even in the absence of *N*-glycosylation, we expressed and purified the model glycoproteins yeast Gas1p and Gas2p as secreted MBP fusions in *E. coli* and *S. cerevisiae*. Given Gas1p has many more *N*-glycosylation sequons than Gas2p, we predicted that Gas2p would be more stable than Gas1p when expressed without *N*-glycosylation in *E. coli*. Circular dichroism (CD) spectroscopy of *E. coli*-expressed protein was consistent with MBP-Gas2 containing more helical secondary structure than MBP-Gas1, while the native glycosylated forms of these *S. cerevisiae*-expressed proteins showed equivalent CD spectra (Fig. 2a,b). MS analysis of the disulphide bond connectivity by non-reducing tryptic peptide mapping showed that both *E. coli*-expressed MBP-Gas1 and MBP-Gas2 had incorrect connectivity compared with native protein expressed in yeast (Supplementary Tables 1 and 2)^{26,27}. To enable comparison with native glycosylated MBP-Gas1 and MBP-Gas2, we created disulphide shuffled forms of these *S. cerevisiae*-expressed proteins. CD spectra of these disulphide shuffled forms of MBP-Gas1 and MBP-Gas2 were similar to their native forms (Fig. 2a,b). This is consistent with the increased stability and secondary structure in nonglycosylated MBP-Gas2 compared with nonglycosylated MBP-Gas1 being local effects that influence secondary structure formation. That is, the absence of *N*-glycosylation at the 10 glycosylation sites in Gas1p strongly affected local formation of secondary structural elements, whereas this affect was not as strong in Gas2p, which with only a single glycosylation site does not natively rely on *N*-glycosylation for stability to the extent of Gas1p.

We next tested if the Gas2p sequences corresponding to Gas1p glycosylation sites were functionally equivalent *in vivo*. Yeast cells with deletion of *GAS1* showed reduced growth compared with wild-type cells, which was complemented by overexpression of native Gas1p (Fig. 2c). Gas1p is glycosylated at Asn40 (NN₄₀GS)¹³ and expression of variant Gas1-N40Q lacking this glycosylation site did not complement loss of *GAS1* (Fig. 2c). While Asn-Gln site-directed mutagenesis is routinely used to test the importance of particular glycosylation sites in glycoprotein folding and function, the conformational preferences of Gln are very different to those of Asn, while His is more similar to Asn³¹. Expression of variant Gas1-N40H only partially complemented loss of *GAS1* (Fig. 2c). This lack of complementation indicated that glycosylation at Asn40 was required for Gas1p function in the context of the native sequence. However, replacing this stretch of Gas1p with the corresponding sequence from Gas2p (39ESGE₄₂; Supplementary Fig. 1a) restored growth, as did point mutation to insert proximal Asp residues (Fig. 2c), while retaining the conserved Gly41 (Supplementary Fig. 1b). The dependency of *in vivo* Gas1p function on glycosylation at Asn40 could therefore be rescued by a small number of compensatory point mutations introducing local clusters of negatively charged residues.

Stabilization of unglycosylated human interferon- β . Animal glycoproteins are not generally required to be expressed under such severe nutrient limitations as during yeast sporulation.

Figure 1 | Equivalent sequences to glycosylation sites in yeast cell-wall proteins. (a) Glycosylation sequon density is higher in glycoprotein paralogues upregulated in vegetative growth than in those upregulated in sporulation. Paralogues of black squares, Gas1p; white squares, Mkc1p; black circles, Exg1p; black triangles, Plb1p; white triangles, Ecm33p; and white diamonds, Crh1p. Horizontal bars show mean for each group. * $P = 0.01$, 2-sided Mann-Whitney test. (b) Weblogo representation of aligned experimentally validated glycosylation sites and (c) corresponding sequences in homologues lacking glycosylation sites. (d) Proportion of hydrophilic, hydrophobic, acidic and basic amino acids in validated sites (black), corresponding sequences lacking sequons (white) and in the entire protein sequences of all homologues analysed (grey). * $P < 0.005$, Fisher's exact test. Position specific proportion of (e) hydrophilic, (f) hydrophobic, (g) acidic and (h) basic amino acids for validated glycosylation sites in yeast cell-wall proteins (black squares) and corresponding sequences in paralogs lacking sites (white squares). Horizontal lines indicate mean values of the entire protein sequence of all homologues analysed (full) and the values at which $P = 0.001$ by Fisher's exact test compared with this mean for a given position (dashed).

However, human glycoproteins are often expressed for biotechnological purposes in bacteria that lack glycosylation systems. Such a protein is the human glycoprotein hormone interferon- β (IFN β)³². The single *N*-glycosylation site in human IFN β is not strictly conserved in orthologues and paralogues (Fig. 3a, Supplementary Fig. 2), and in these proteins is often replaced by clusters of acidic amino acids similar to those observed in place of *N*-glycosylation sites in yeast proteins (Fig. 1, Supplementary Fig. 1). To test if such a sequence would increase IFN β stability in the absence of glycosylation, we expressed and purified IFN β as wild type and glycosylation site variant (Asn101Asp) forms from *E. coli* and refolded these proteins *in vitro*. Guanidinium-induced unfolding equilibria showed that at physiologically relevant pH 7.5 this point mutation increased stability ($\Delta\Delta G^0 = -9 \text{ kJ mol}^{-1}$, Fig. 3b). In contrast, at pH 5.0 the $\Delta\Delta G^0$ was negligible (Fig. 3c). We propose that this is consistent with stabilization of the IFN β_{N101D} variant at pH 7.5 through reduced entropy of the unfolded state due to charge-charge repulsion between Asp101 and Glu102 favouring a local turn or extended structure similar to the structure of Asn101 and Glu102 in folded native IFN β ¹⁸, or stabilization of the native state^{33,34}. At pH 5 there is a lower

probability that both these residues will be simultaneously deprotonated (negatively charged) in a given protein molecule, consistent with the lack of stabilization in the IFN β_{N101D} variant. The overall higher stability for both protein forms at pH 5.0 than at pH 7.5 is likely due to increased distance from the predicted pI of IFN β (~ 9.0). This confirmed that local clusters of negatively charged amino acids could increase protein stability in the absence of glycosylation in homologues of human IFN β . To confirm that stable IFN β_{N101D} retained native activity, we performed a cell culture growth inhibition assay (Fig. 3d). Stable IFN β_{N101D} showed significantly higher specific activity than IFN β wild type (Fig. 3d), correlating increased thermodynamic stability with improved protein function in this assay. This suggested that it is possible for IFN proteins to have either *N*-glycosylation sites or sequence stretches that can play a similar role in protein folding and stability.

Stabilization at an inefficiently glycosylated sequon. Asparagines in sequons generally show macroheterogeneity with not all of the protein pool glycosylated, and almost all sequons are

Figure 2 | Structure and function of Gas1p and Gas2p glycosylation site variants. (a) CD spectra of MBP-Gas1 and (b) MBP-Gas2 produced in *E. coli* (dashed black), or *S. cerevisiae* native (solid black) or disulphide-shuffled (solid gray). Mean residue ellipticity is shown. (c) Growth of yeast with wild-type Gas1p (wild type, pGAS1-NNGS) and Asn40 glycosylation site variants. The native Asn40 sequon is underlined. Point mutations in variants are bold.

modified to some extent⁹. We therefore tested whether essentially unmodified sequons could also be subject to sequence-based stabilization. For this purpose we studied the paralogous yeast proteins Ost3p and Ost6p, which are accessory subunits of OTase^{7,13}. The ER luminal domain of Ost3p has a single sequon at Asn33, which is not detectably glycosylated³⁵ (Fig. 4a and Supplementary Fig. 3), while Ost6p has no sequons and contains a cluster of acidic amino acids at the position corresponding to the Ost3p sequon (Supplementary Fig. 4). The ER luminal domain of Ost6p is more easily expressed than that of Ost3p in *E. coli*³⁰ and shows activity in an *in vitro* peptide-binding assay when Ost3p does not³⁶. To test if the higher *in vitro* activity of Ost6p relative to Ost3p correlated with increased thermodynamic stability of Ost6p, we performed guanidinium-induced equilibrium unfolding of Ost6 and Ost3 purified from *E. coli*. This showed that Ost6 ($\Delta G^0 = -79 \pm 10 \text{ kJ mol}^{-1}$) was more stable than Ost3 ($\Delta G^0 = -28 \pm 5 \text{ kJ mol}^{-1}$) (Fig. 4b,f), consistent with the presence of clusters of like-charged amino acids in place of an unused glycosylation sequon increasing thermodynamic stability

Figure 3 | Stability and activity of human IFNβ variants. (a) Alignment of human IFNβ and a selected orthologue and paralogue. The IFNβ glycosylation sequon is underlined and corresponding clusters of acidic amino acids are in bold. (b) Guanidinium induced equilibrium unfolding of IFNβ wild-type and IFNβ_{N101D} variant as measured by intrinsic tryptophan fluorescence and fitted to the two-state model of protein folding at pH 7.5 and at (c) pH 5.0. Solid line, IFNβ wild type; dashed line, IFNβ_{N101D}. (d) Activity of IFNβ_{wt}, IFNβ_{N101D} and IFNβ_{com} (Commercial nonglycosylated protein) in a cell culture growth inhibition assay. Values show mean of percent cell survival relative to untreated cells, $n > 8$; Error bars show s.e.m.; *P = 0.002, **P = 10⁻⁷, student's t-test.

and activity. To test this in the context of Ost6p, we introduced a glycosylation sequon in Ost6 at the equivalent position to the Ost3 sequon, and tested *in vitro* stability and peptide-binding activity. This variant of Ost6 with a glycosylation sequon had both reduced stability compared with wild-type Ost6 (Ost6_{D38N} $\Delta \Delta G^0 = 20 \text{ kJ mol}^{-1}$) (Fig. 4f,g), and essentially eliminated *in vitro* peptide-binding activity (Fig. 4l,m). We next tested whether the stability and *in vitro* activity of Ost3 could be increased by targeted point mutation at the unused sequon in Ost3. We first introduced an acidic Asp in place of Asn33, which increased stability compared with wild-type protein (Ost3_{N33D} $\Delta \Delta G^0 = -16 \text{ kJ mol}^{-1}$) (Fig. 4c), but did not result in a detectable increase in *in vitro* peptide binding (Fig. 4h,i). However, Ost3p orthologues lack sequons at this position and

Figure 4 | Stability and function of yeast Ost3 and Ost6 variants.

(a) Alignment of Ost3 and Ost6. The Ost3 glycosylation sequon is underlined and corresponding cluster of acidic amino acids in Ost6 are in bold. Guanidinium induced equilibrium unfolding of (b) Ost3, (c) Ost3_{N33D}, (d) Ost3_{N33D, K34D}, (e) Ost3_{N33K}, (f) Ost6 and (g) Ost6_{D38N}, as measured by intrinsic tryptophan fluorescence and fitted to the two-state model of protein folding. Peptide binding to Ost3 wild type (black) and Ost3_{N33D} (purple) by peptides (h) LVIWINGDK (m/z of 1,057.6⁺) and (i) FFYSNNGSQFYIR (m/z of 1,642.8⁺), to Ost3 wild type (black) and Ost3_{N33K} (red) by peptides (j) LVIWINGDK (m/z of 1,057.6⁺) and (k) FFYSNNGSQFYIR (m/z of 1,642.8⁺), and to Ost6 wild type (black)³⁶ and Ost6_{D38N} (blue) by (l) LVIWINGDK (m/z of 1,057.6⁺) and (m) AGTLFLVDLIK (m/z of 1,189.7⁺). Values show mean of triplicates, error bars show s.e.m.

instead typically possess clusters of basic amino acids, while Ost6p orthologues have acidic residues at this position (Supplementary Fig. 4). This difference may depend on Ost3p- and Ost6-specific charge distributions on a neighbouring α -helix (Supplementary Fig. 5). We therefore tested two additional variants with clusters of either acidic or basic like-charge. While Ost3_{N33D, K34D} (30KLADDSPPK₃₈) was less stable ($\Delta G^0 = 6$ kJ mol⁻¹) (Fig. 4d), Ost3_{N33K} (30KLAKKSPKK₃₈) was more stable than Ost3_{N33D} and wild-type Ost3 ($\Delta G^0 = -27$ kJ mol⁻¹) (Fig. 4e). Optimal stabilization was therefore achieved through minimal mutations providing a cluster of like-charged residues. This stable Ost3_{N33K} showed increased *in vitro* binding of several peptides (Fig. 4j,k). These peptides contained aromatic amino acids, complementary to Phe102 in the peptide-binding groove of Ost3, a key determinant of transient nascent polypeptide tethering by Ost3 crucial to its function in modulating OTase activity *in vivo*^{30,36}. The additional thermodynamic stability of Ost3_{N33K} therefore enabled its function to be investigated in this *in vitro* assay.

Discussion

We show that targeted introduction of like-charged amino acids at *N*-glycosylation sequons is a generally applicable technique for increasing the stability of normally *N*-glycosylated proteins when expressed from more efficient expression systems such as *E. coli*. The targeted nature of this approach also limits any unanticipated effects on protein structure or function of non-targeted mutagenesis. The presence of a sequon rather than a compensatory sequence at a specific location in a given protein may also depend on its protein context. For example, experimental^{1,16} and computational³⁷ characterization of variants of CD2 adhesion domain from human (which is glycosylated) and rat (which is not glycosylated) has shown that gain of glycosylation mutations in Rat CD2 do not immediately increase protein thermodynamic stability. However, acquisition of two additional mutations proximal to the newly glycosylated Asn introduces specific glycan-protein interactions, which lead to the formation of high-occupancy hydrogen bonds and stabilize the glycosylated protein. Alternatively, while loss of glycosylation mutations in human CD2 destabilizes the protein, two additional nearby mutations rescue stability and function³⁸. Experimentally, Asn-Gln site-directed mutagenesis is commonly used as a proxy for loss of glycosylation at a given site. Such a mutation may show if glycosylation at that site is important for protein folding, stability or activity, in the context of an otherwise unchanged protein sequence. However, our bioinformatics, *in vitro* and *in vivo* results here show that such a requirement could in general be easily rescued by evolutionary acquisition of a small number of stabilizing point mutations. The exact reasons why particular glycoproteins have *N*-glycosylation sites at specific locations are undoubtedly complex, depending not only on their affect on global protein stability but also on evolutionary history and the many roles of *N*-glycans in influencing protein function.

Methods

Sequence analyses. Families of yeast glycoproteins with members expressed predominantly during vegetative growth or sporulation were identified from expression data²³ and UniProt BLAST with a cutoff of e-4. Sequence analysis was performed based on a set of experimentally validated occupied glycosylation sites in *S. cerevisiae* cell-wall glycoproteins (28 sites in 13 glycoproteins)^{13,28,29}. Homologues to these from *S. cerevisiae* were identified with UniProt BLAST with a cutoff of e-4, and each of these sets of homologous proteins and glycoproteins were aligned with ClustalW. For each experimentally validated glycosylation site, aligned sequences of homologous proteins were identified that did not contain a glycosylation sequon at that position (41 sequences). The characteristics of the amino-acid sequences flanking these sites (glycosylation sites, and corresponding sequences in homologues lacking glycosylation sites) were determined through a non-gapped alignment visualized with WebLogo³⁹ and amino-acid composition

(hydrophilic, hydrophobic, acidic or basic). pI of the amino-acid sequences was determined with ProtParam⁴⁰.

Cloning and mutagenesis. DNA encoding the ER luminal domain of *S. cerevisiae* Gas2p as predicted in the UniProt Knowledgebase (position 25 to 531 of Gas2p/Q06135/YLR343W) was cloned into the pMAL-p2x vector, using PCR primers incorporating *Bam*HI and *Hind*III restriction sites, to give pMAL-GAS2. DNA encoding the GAS1 genomic locus (GAS1/P22146/YMR307W including 1 kb up and downstream) was cloned into the pRS415 vector to give pRS415-GAS1. Blunt ended DNA encoding mature MBP-Gas1 or MBP-Gas2 was amplified using PCR primers incorporating a C-terminal His-tag, phosphorylated with T4 polynucleotide kinase, and ligated into blunt ended DNA amplified from pRS415-GAS1 excluding DNA encoding mature GAS1 (ref. 41). This produced pRS415-MBP-GAS1-His and pRS415-MBP-GAS2-His, which encoded MBP-Gas1-His or MBP-Gas2-His with the native Gas1p signal sequence under the control of the native GAS1 promoter and terminator. DNA encoding the mature protein sequence of IFN β as predicted in the UniProt Knowledgebase (position 22 to 187 of P01574) with Cys17 changed to Ser³² and with an N-terminal decaHis-tag, optimized for *E. coli* protein expression was synthesized and cloned into the pJexpress404 expression vector (DNA 2.0). DNA sequence encoding a C-terminal FLAG-tag was inserted in *S. cerevisiae* OST3 on plasmid YEp352 (ref. 42) using site-directed mutagenesis⁴¹ to create pOST3-FLAG. Point mutation of GAS1, IFN β and OST3 were performed by site-directed mutagenesis⁴¹.

Cell culture and protein purification. BY4741 (MATa *his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0*) yeast cells and the *Agas1* mutant derivative thereof (Open Biosystems) was transformed with pRS415, pRS415-GAS1 or variants using standard protocols. Cells were diluted to equal densities at OD_{600 nm}, serially 10-fold diluted, equal volumes spotted to minimal media agar plates and incubated at 30°C. TOP10 *E. coli* cells carrying the plasmids encoding IFN β , Ost6 (pHis10-OST6L⁴³), Ost3 (pHis10-OST3L⁴³) or variants thereof were grown in LB media, and cells carrying pMAL-Gas1 (ref. 36) or pMAL-Gas2 were grown in LB-glucose media. Media was supplemented with 100 mg l⁻¹ ampicillin. Protein expression was induced by addition of IPTG to 1 mM at an OD_{600 nm} of 0.7. After incubation for 4 h at 37°C the cells were collected by centrifugation. MBP-Gas1 and MBP-Gas2 were purified as previously described³⁶, except that protein was eluted with 10 mM maltose. Yeast cells carrying pRS415-MBP-GAS1-His or pRS415-MBP-GAS2-His were grown in minimal media at 30°C with shaking and collected at an OD_{600 nm} of 1.0. Cells were resuspended in 50 mM phosphate buffer pH 7.5, 150 mM NaCl, 1 $\times$ protease inhibitor cocktail (Roche) and 1 mM PMSF, and lysed by French Press. Proteins were purified using TALON resin (Clontech) and eluted in 50 mM phosphate buffer pH 7.5, 150 mM NaCl and 250 mM imidazole. The TALON resin eluate was then applied to amylose-agarose beads and purified as previously described³⁶, except that protein was eluted with 10 mM maltose. Disulphide-shuffled protein was produced by reduction in 6 M guanidinium chloride, 50 mM phosphate buffer pH 7.5 with 10 mM DTT, buffer exchange into 1 M guanidinium chloride, 50 mM phosphate buffer pH 7.5 with 0.1 mM CuSO₄ and incubation at 25°C for 2 h. Protein was then buffer exchanged into 20 mM phosphate buffer pH 7.5, 200 mM NaCl and 1 mM EDTA. Collected cells expressing IFN β variants were resuspended in 50 mM phosphate buffer pH 7.5, 150 mM NaCl, 1 $\times$ protease inhibitor cocktail (Roche) and 1 mM PMSF, and lysed by French Press. Inclusion bodies were recovered by sequential centrifugation at 18,000 rcf. The pellets were washed with 1 ml 50 mM phosphate buffer pH 7.5 and 150 mM NaCl and re-centrifuged three times. The washed pellets containing inclusion bodies were resolubilized in 6 M guanidinium chloride, 50 mM phosphate buffer pH 7.5 and 1 mM β -mercaptoethanol and incubated at 37°C for 16 h. His-tagged IFN β was purified using TALON resin (Clontech) and eluted in 6 M guanidinium chloride, 50 mM phosphate buffer pH 5.0 and 1 mM β -mercaptoethanol. The pH was adjusted to pH 7.5 with other constituents remaining constant, DTT was added to 20 mM and the protein was incubated at 37°C for 1 h. Protein was buffer exchanged with PD-10 columns (GE Healthcare) into 50 mM phosphate buffer pH 7.5 and 6 M guanidinium chloride. Cysteines were oxidized to cystine by addition of 0.1 mM CuSO₄ and incubation at 25°C for 2 h. Oxidation was quenched by addition of 1 mM EDTA, and oxidized IFN β was buffer exchanged with PD-10 columns into 50 mM phosphate buffer pH 7.5 and 25% glycerol at 4°C. Cells expressing N-terminally His-tagged Ost6, Ost3 or variants thereof were resuspended in ice cold 50 mM Potassium Phosphate, pH 7.5, 150 mM NaCl, 10 mM Imidazole, 1 $\times$ complete mini EDTA free protease inhibitor (Roche) and 1 mM PMSF, and lysed by French Press. Protein was purified from clarified whole-cell extract using TALON resin and eluted with 200 mM imidazole. Proteins were buffer exchanged with PD-10 columns into 50 mM potassium phosphate buffer pH 7.0 and 1 mM EDTA. Yeast with pOST3 or pOST3-FLAG were grown in minimal media at 30°C and collected at mid log phase. Proteins were denatured in SDS sample buffer, incubated for 1 h at 37°C with 1 $\times$ G5 buffer in the presence or absence of Endo H¹³, separated by SDS-PAGE, transferred to nitrocellulose membrane and probed with an α -FLAG antibody.

In vitro protein analysis. CD spectra were obtained using a Jasco J-710 spectrometer (Jasco, Japan) as described⁴⁴. Free energies of stabilization (ΔG^0)

were determined based on the two-state model of protein folding⁴⁵ as described³⁰. To determine the equilibrium transitions, intrinsic tryptophan fluorescence of proteins was measured in different concentrations of guanidinium chloride in 50 mM potassium phosphate pH 7.5 or 5.0, and 150 mM NaCl at 25°C. ΔG^0 and cooperativity were determined from a non-linear least-squares 6 parameter fit to the data points⁴⁵, with uncertainty estimates from the corresponding six-parameter fits. Ost3 and Ost6 peptide-binding affinity chromatography was performed as described³⁶. Briefly, tryptic peptides from MBP-Gas1p were applied to peptide-affinity chromatography columns containing MBP-Ost3, MBP-Ost3_{N33K}, MBP-Ost6 or MBP-Ost6_{D38N} bound to amylose-agarose resin with or without pre-reduction with DTT. Peptides were eluted with physiological buffer, fractions collected, and peptides in the load and elution fractions were detected with MALDI-TOF-MS. Peptide retention was determined by the relative abundance of each peptide in wash fractions compared with the load, in oxidized relative to reduced. Reduction of the CxxC motif in Ost6p causes structural changes only at the peptide-binding groove³⁰, confirming that retained peptides bind at this groove rather than through other non-specific interactions³⁶.

Cell culture growth inhibition assay. A total of 4×10^3 Human MCF-7 breast tumour cells were cultured in 200 μ l of phenol-free media containing 5% charcoal stripped calf serum in 96-well plates. Equal amounts of either human recombinant IFN β (ProSpec CYT-234), IFN- β _{WT}, IFN- β _{N101D} or no protein were added to the cultures with retinoic acid (Sigma) and incubated for 4 days as described⁴⁶. Cells were fixed with a final concentration of 3.3% (wt/vol) TCA and stained with Sulforhodamine B (Sigma R2625) as described⁴⁷, after which the OD was measured at 490 nm. A control plate to determine initial seeding intensity was included where cells were fixed 2–3 h after plating. Absorbance with this plate was set as 0% growth, and growth obtained with untreated cells was taken as 100%.

References

- Hanson, S. R. *et al.* The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proc. Natl. Acad. Sci. USA* **106**, 3131–3136 (2009).
- Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73**, 1019–1049 (2004).
- Ohtsubo, K. & Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855–867 (2006).
- Rudd, P. M. *et al.* Glycoforms modify the dynamic stability and functional activity of an enzyme. *Biochemistry* **33**, 17–22 (1994).
- Wormald, M. R. & Dwek, R. A. Glycoproteins: glycan presentation and protein-fold stability. *Structure* **7**, R155–R160 (1999).
- Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell. Biol.* **13**, 448–462 (2012).
- Kelleher, D. J. & Gilmore, R. An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* **16**, 47R–62R (2006).
- Mohorko, E., Glockshuber, R. & Aebi, M. Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. *J. Inher. Metab. Dis.* **34**, 869–878 (2011).
- Schulz, B. L. in *Glycosylation*. (ed Petrescu, S.) 21–40 (Intech, 2012).
- Schwarz, F. & Aebi, M. Mechanisms and principles of N-linked protein glycosylation. *Curr. Opin. Struc. Biol.* **21**, 576–582 (2011).
- Lizak, C., Gerber, S., Numao, S., Aebi, M. & Locher, K. P. X-ray structure of a bacterial oligosaccharyltransferase. *Nature* **474**, 350–355 (2011).
- Bailey, U. M., Jamaluddin, M. F. B. & Schulz, B. L. Analysis of congenital disorder of glycosylation-Id in a yeast model system shows diverse site-specific under-glycosylation of glycoproteins. *J. Proteome Res.* **11**, 5376–5383 (2012).
- 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* **8**, 357–364 (2009).
- Petrescu, A. J., Milac, A. L., Petrescu, S. M., Dwek, R. A. & Wormald, M. R. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **14**, 103–114 (2004).
- Slynko, V. *et al.* NMR Structure Determination of a Segmentally Labeled Glycoprotein Using In Vitro Glycosylation. *J. Am. Chem. Soc.* **131**, 1274–1281 (2009).
- Culyba, E. K. *et al.* Protein native-state stabilization by placing aromatic side chains in N-glycosylated reverse turns. *Science* **311**, 571–575 (2011).
- Lizak, C., Fan, Y. Y., Weber, T. C. & Aebi, M. N-Linked Glycosylation of Antibody Fragments in *Escherichia coli*. *Bioconjug. Chem.* **22**, 488–496 (2011).
- Karpusas, M. *et al.* The crystal structure of human interferon β at 2.2-Å resolution. *Proc. Natl. Acad. Sci. USA* **94**, 11813–11818 (1997).
- Chen, M. M. *et al.* Perturbing the folding energy landscape of the bacterial immunity protein Im7 by site-specific N-linked glycosylation. *Proc. Natl. Acad. Sci. USA* **107**, 22528–22533 (2010).
- Aebi, M., Bemasconi, R., Clerc, S. & Molinari, M. N-glycan structures: recognition and processing in the ER. *Trends. Biochem. Sci.* **35**, 74–82 (2010).

21. Pearse, B. R. & Hebert, D. N. Lectin chaperones help direct the maturation of glycoproteins in the endoplasmic reticulum. *Biochem. Biophys. Acta* **1803**, 684–693 (2010).
22. Roth, J. *et al.* Protein N-glycosylation, protein folding, and protein quality control. *Mol. Cells* **30**, 497–506 (2010).
23. Chu, S. *et al.* The Transcriptional Program of Sporulation in Budding Yeast. *Science* **282**, 699–705 (1998).
24. Ragni, E., Fontaine, T., Gissi, C., Latgè, J. P. & Popolo, L. The Gas family of proteins of *Saccharomyces cerevisiae*: characterization and evolutionary analysis. *Yeast* **24**, 297–308 (2007).
25. Rolli, E. *et al.* Expression, stability, and replacement of glucan-remodeling enzymes during developmental transitions in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **22**, 1585–1598 (2011).
26. Popolo, L. *et al.* Disulfide bond structure and domain organization of yeast beta(1,3)-glucanoglucosyltransferases involved in cell wall biogenesis. *J. Biol. Chem.* **283**, 18553–18565 (2008).
27. Hurtado-Guerrero, R. *et al.* Molecular mechanisms of yeast cell wall glucan remodeling. *J. Biol. Chem.* **284**, 8461–8469 (2009).
28. Izquierdo, L. *et al.* Distinct donor and acceptor specificities of Trypanosoma brucei oligosaccharyltransferases. *EMBO J* **28**, 2650–2661 (2009).
29. Nasab, F. P., Schulz, B. L., Gamarro, F., Parodi, A. J. & Aebi, M. All in One: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **19**, 3758–3768 (2008).
30. Schulz, B. L. *et al.* Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc. Natl. Acad. Sci. USA* **106**, 11061–11066 (2009).
31. Hovmöller, S., Zhou, T. & Ohlson, T. Conformations of amino acids in proteins. *Acta. Crystallogr. D. Biol. Crystallogr.* **58**, 768–776 (2002).
32. Mark, D. F., Lu, S. D., Creasey, A. A., Yamamoto, R. & Lin, L. S. Site-specific mutagenesis of the human fibroblast interferon gene. *Proc. Natl. Acad. Sci. USA* **81**, 5662–5666 (1984).
33. Loladze, V. V. & Makhatadze, G. I. Energetics of charge-charge interactions between residues adjacent in sequence. *Proteins* **79**, 3494–3499 (2011).
34. Schweiker, K. L. & Makhatadze, G. I. Protein stabilization by the rational design of surface charge-charge interactions. *Methods Mol. Biol.* **490**, 261–283 (2009).
35. Kelleher, D. J. & Gilmore, R. The *Saccharomyces cerevisiae* oligosaccharyltransferase is a protein complex composed of Wbp1p, Swp1p, and four additional polypeptides. *J. Biol. Chem.* **269**, 12908–12917 (1994).
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.* **20**, 849–855 (2011).
37. Wang, X. Y., Ji, C. G. & Zhang, J. Z. Exploring the molecular mechanism of stabilization of the adhesion domains of human CD2 by N-glycosylation. *J. Phys. Chem. B* **116**, 11570–11577 (2012).
38. Wyss, D. F. *et al.* Conformation and function of the N-linked glycan in the adhesion domain of human CD2. *Science* **269**, 1273–1278 (1995).
39. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
40. Gasteiger, E. *et al.* in *The Proteomics Handbook*. (ed Walker, J. M.) 571–607 (Humana Press, 2005).
41. Imai, Y., Matsushima, Y., Sugimura, T. & Terada, M. A simple and rapid method for generating a deletion by PCR. *Nucleic Acids Res.* **19**, 2785 (1991).
42. Spirig, U., Bodmer, D., Wacker, M., Burda, P. & Aebi, M. The 3.4-kDa Ost4 protein is required for the assembly of two distinct oligosaccharyltransferase complexes in yeast. *Glycobiology* **15**, 1396–1406 (2005).
43. Mohd Yusuf, S. N., Bailey, U. M., Tan, N. Y. J., Jamaluddin, M. F. B. & Schulz, B. L. Mixed disulfide formation in vitro between a glycoprotein substrate and yeast oligosaccharyltransferase subunits Ost3p and Ost6p. *Biochem. Biophys. Res. Commun.* **432**, 438–443 (2013).
44. Zalucki, Y. M., Jones, C. E., Ng, P. S., Schulz, B. L. & Jennings, M. P. Signal sequence non-optimal codons are required for the correct folding of mature maltose binding protein. *Biochim. Biophys. Acta* **1798**, 1244–1249 (2010).
45. Pace, C. N. Measuring and increasing protein stability. *Trends Biotechnol.* **8**, 93–98 (1990).
46. Kolla, V., Lindner, D. J., Xiao, W., Borden, E. C. & Kalvakolanu, D. V. Modulation of interferon (IFN)-inducible gene expression by retinoic acid. Up-regulation of STAT1 protein in IFN-unresponsive cells. *J. Biol. Chem.* **271**, 10508–10514 (1996).
47. Vichai, V. & Kirtikara, K. Sulforhodamine B colorimetric assay for cytotoxicity screening. *Nat. Protoc.* **1**, 1112–1116 (2006).

Acknowledgements

This work was supported by National Health and Medical Research Council Project Grant 631615 and National Health and Medical Research Council Career Development Fellowship APP1031542 to B.L.S.

Author contributions

N.Y.T., S.C.R. and B.L.S. performed sequence analyses. N.Y.T., S.H.B. and B.L.S. characterized the proteins. U.M.B. performed the growth inhibition assay. M.F.J. performed the peptide affinity chromatography assay. U.M.B. and B.L.S. wrote the manuscript with contributions from all other authors.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Tan, N. Y. *et al.* Sequence-based protein stabilization in the absence of glycosylation. *Nat. Commun.* 5:3099 doi: 10.1038/ncomms4099 (2014).