Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) of proteins are prevalent in both eukaryotes and prokaryotes. Although often poorly conserved in sequence, the amino acid content of IDPs and IDRs is actively regulated, and IDPs and IDRs serve functions such as connecting ordered domains, regulating translation, molecular recognition and signaling, and assisting in protein folding. Because of their inherent flexibility and lack of structure, IDPs and IDRs are susceptible to proteolytic cleavage in the competitive, extracellular milieu, and O-glycosylation – the attachment of a sugar moiety to the β-hydroxyl group of serine or threonine – is an important mechanism to protect against proteolysis in these regions. In fungi and yeasts in particular, most of the secreted IDPs and proteins exhibiting IDRs are O-mannosylated, but the evolutionary preference for this specific glycosylation pattern is not well understood. The present study uses glycopeptide synthesis and molecular dynamics (MD) simulations to reveal that O-mannosylation is the preferred glycan motif on fungal IDP sequences and reveals the biophysical reasons underpinning this observation, in turn suggesting an evolutionary selection for α-mannose as the preferred glycan for IDP/IDR stabilization in some eukaryotic systems.

O-Mannosylation is strongly preferred for proteolysis protection of a model fungal IDP. To investigate how glycan identity affects IDP proteolytic stability, we employed the naturally O-mannosylated linker from the Trichoderma reesei glycoside hydrolase family 7 cellobiohydrolase, TcCel7A, as a model. This enzyme is one of the most important industrial cellulases and its linker is a well-studied O-mannosylated IDP. The α-anomeric configuration was chosen since it is the...
only type reported so far in reducing terminal mannose residues of O-mannosylated proteins from fungi and yeasts. We used solid-state glycopeptide synthesis to produce four variants (Fig. 1), including the non-glycosylated linker, and measured the half-life to thermolysin degradation with MALDI-TOF MS (Fig. S1–S4). As shown in Table 1, all glycosylated variants improve proteolytic stability over the non-glycosylated linker, LNG, but the O-mannosylated linker (Lman) exhibits an striking 112-fold improvement over LNG, 16-fold proteolysis protection over the O-galactosylated linker (Lgal) and 3-fold over the O-glucosylated linker (Lglc). These results, obtained using a model IDP, align with our previous observation that O-mannosylation is able to restrict protein flexibility. That is, the relative stiffening of Lman was corroborated by its greater persistence length (Table S1†). Also, similar structures from T-REMD were clustered considering the Cα atoms with a root mean squared deviation cutoff of 1.5 Å (Fig. S7, Table S3†). The most populated clusters were found for Lman. Moreover, values of root-mean-square deviation relative to average structures computed for 10 ns trajectory blocks also indicate lower mobility of the Lman backbone (Table S3†). Small differences in protein backbone flexibility and concomitant large differences in resistance to proteolysis were also recently found for a structured protein with a single attached glycan, α-mannose or α-glucose. Chaffey et al. suggested that a chain of specific interactions between O-mannosyl and side chains of close residues may be propagating stiffening along the protein backbone. The similar behavior observed with IDPs suggests that the effects of α-mannose on protein stiffening may not be exclusive to a specific protein fold. From these observations, we further hypothesized that the observed differences in linker extension are caused by local interactions with the C2-hydroxyl group (2-OH) adjacent to the glycan–peptide bond, which is equatorial in α-glucose and α-galactose and axial in α-mannose. Fig. 3A shows the average number of hydrogen bonds (HBs) between the protein and each of the carbohydrate

<table>
<thead>
<tr>
<th>Variants</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNG</td>
<td>1.6</td>
<td>1.5</td>
<td>1.2</td>
<td>1.5 ± 0.2</td>
</tr>
<tr>
<td>Lman</td>
<td>163.9</td>
<td>196.4</td>
<td>130.5</td>
<td>163.6 ± 32.9</td>
</tr>
<tr>
<td>Lgal</td>
<td>8.5</td>
<td>8.8</td>
<td>12.1</td>
<td>9.8 ± 2.0</td>
</tr>
<tr>
<td>Lglc</td>
<td>53.3</td>
<td>62.0</td>
<td>45.9</td>
<td>53.7 ± 8.1</td>
</tr>
</tbody>
</table>

proteolysis than LNG. These results suggest that steric hindrance alone cannot fully explain proteolytic resistance, since the glycan moieties occupy roughly the same volume.

We subsequently examined how glycan chemistry affects protein flexibility, glycan orientation, specific interactions, and backbone torsional preferences in an attempt to explain the high proteolysis resistance imparted by O-mannosylation. Information about protein flexibility and extension were obtained from the free energy profiles, or potential of mean force (PMF), as a function of the end-to-end distance for all linkers (Fig. 2). Unlike Lgal and Lglc for which the PMFs are somewhat flat-bottomed and resemble that of the non-glycosylated linker LNG, the PMF for Lman is slightly narrower and shows a well-defined local minimum at larger distances (~3.0–3.5 nm). This indicates that Lman is, on average, stiffer and adopts more extended conformations than its counterparts. Further analyses reinforce the hypothesis that α-mannosylation is able to restrict protein flexibility. That is, the relative stiffening of Lman was corroborated by its greater persistence length (Table S1†). Also, similar structures from T-REMD were clustered considering the Cα atoms with a root mean squared deviation cutoff of 1.5 Å (Fig. S7, Table S3†). The most populated clusters were found for Lman. Moreover, values of root-mean-square deviation relative to average structures computed for 10 ns trajectory blocks also indicate lower mobility of the Lman backbone (Table S3†). Small differences in protein backbone flexibility and concomitant large differences in resistance to proteolysis were also recently found for a structured protein with a single attached glycan, α-mannose or α-glucose. Chaffey et al. suggested that a chain of specific interactions between O-mannosyl and side chains of close residues may be propagating stiffening along the protein backbone. The similar behavior observed with IDPs suggests that the effects of α-mannose on protein stiffening may not be exclusive to a specific protein fold. From these observations, we further hypothesized that the observed differences in linker extension are caused by local interactions with the C2-hydroxyl group (2-OH) adjacent to the glycan–peptide bond, which is equatorial in α-glucose and α-galactose and axial in α-mannose. Fig. 3A shows the average number of hydrogen bonds (HBs) between the protein and each of the carbohydrate

Fig. 1 The four linker models examined experimentally and computationally (left). Chair representations of α-mannose, α-galactose, and α-glucose are also depicted (right).
hydroxyl groups computed from the T-REMD simulations. The HBs between the 2-OH group and the peptide contribute significantly to the higher total number of HBs in Lgal and Lglc. Compared to Lman, this indicates that the equatorial configuration of 2-OH, the closest hydroxyl to the peptide chain, favors glycan–protein HBs.

Next, we show that orientation of the glycans relative to the peptide chain depends on the glycan chemistry and affects the conformational freedom of the glycosylated IDP. Fig. 3B shows the normalized distribution of the angle \( \theta \) between the normal to the plane of the sugar ring and the vector between \( \mathrm{Ca} \) and \( \mathrm{CB} \) of the threonine residues to which the glycan is attached. Values near 180° and 90° correspond, respectively, to conformations in which the plane of the rings are nearly parallel and perpendicular to the direction of the peptide chain (Fig. 3C). The shoulder at ~90° observed for Lgal and Lglc indicates that the glycans are more frequently oriented perpendicularly to the peptide chain than in Lman, and, therefore, exhibit smaller contact surface with the protein (Table S2†). This effect is associated to the pronounced glycan–protein HBs involving the equatorial 2-OH in Lgal and Lglc. The normalized angle distributions computed for the subset of molecular frames in which these specific interactions are absent (Fig. 3B, dashed lines) lack the characteristic shoulder in the 80–100° range, demonstrating that the C2 stereochemistry impacts the glycan conformation.

Taken together, the results presented thus far demonstrate that the 2-OH position affects glycan conformation and that protein dynamics differ depending on glycan chemistry. Next, why \( \alpha \)-mannosylation leads to more extended conformations and reduces protein flexibility requires an explanation. To this end, we examined how glycans affect the protein backbone conformational sampling at the residue level. Fig. 4A shows the Ramachandran plots for the LNG threonines, in which the protein backbone frequently visits all three major conformational regions. The R3 region corresponds to \( \alpha \)-helix like conformations, whereas R1 and R2 correspond to more extended conformations, such as those found in \( \beta \)-sheets and polyproline II structures. Although no persistent secondary structures were detected during the simulations, these results reflect the structural features of the linkers. We verified that attached glycans alter torsional sampling of the nearest amino acids, as seen elsewhere.25,26 For Lgal and Lglc, the same three regions are populated as in LNG, except that the peak in the R3 region occurs only every other residue because of the excluded volume of neighboring glycans (Fig. 4C and D, S8†). In contrast, the R2 region is predominantly favored in Lman for all glycosylated residues, suggesting that the relative rigidity of the \( \alpha \)-mannosylated linker results in part from a reduced local dihedral flexibility of the glycosylated residues imparted by \( \alpha \)-mannosylation (Fig. 4B). We suggest that perpendicularly oriented glycan rings in Lgal and Lglc allow for improved accommodation of neighboring glycan rings, favoring more compact conformations. Conversely, the preferred orientation
of α-mannose glycans hinders the mobility of the surrounding atoms in the peptide chain, thus revealing a direct relationship between glycan chemistry, orientation, and protein conformational freedom.

Variants decorated with O-mannobiosyl (L<sub>man</sub>) or O-galactobiosyl (L<sub>gal</sub>) were also simulated, as well as the linker with a putative natural decoration based on a previous experimental characterization (L<sub>man-h</sub>) (Fig. S9 and S10†). Our analyses suggest that the length of the glycan only slightly changes the dynamics of the protein when the chemistry of the 2-OH groups in the immediately attached glycosyl unit is preserved, reinforcing its importance (Fig. S11†).

Glycosylation pattern and protein primary sequence are correlated. Although less well studied, many secreted bacterial proteins are also O-glycosylated. For example, the multienzyme cellulosome from <i>Clostridium thermocellum</i> exhibits O-glycans on its linkers. Similarly, the thermostable enzyme CelA from <i>Caldicellulosiruptor bescii</i> has linkers of up to 70 amino acids rich in O-glycans. However, unlike the typical O-mannosylated linkers from eukaryotic proteins, these linkers exhibit mostly O-galactosylation, and are enriched in proline, relative to eukaryotic IDRs.

Aiming to understand why O-mannosylation is not prevalent in bacterial IDP and IDRs relative to their eukaryotic counterparts, we also studied a “PT linker”, which comprises a proline–threonine repeat sequence, and represents a fragment of glycosylated linkers found in bacterial cellulas. PT linker models were uniformly decorated with α-mannose (L<sub>PT-man</sub>), α-galactose (L<sub>PT-gal</sub>), and α-glucose (L<sub>PT-glc</sub>) (Fig. 5A).

It is well known that high proline content is generally found in disordered proteins and favors extended conformations of IDRs. Accordingly, the end-to-end distance PMF shows that the non-glycosylated PT linker favors extended conformations similarly to the glycosylated ThCel7A linker L<sub>man</sub> (Fig. 5B). Elongation and further stiffening of the linkers are observed upon glycosylation and is consistent with NMR spectroscopy data, which demonstrated that glycosylation of PT linkers dampens the dynamics. Interestingly, in the PT linkers, varying the glycan chemistry is not as impactful to the protein dynamics as in the eukaryotic linker cases. To understand this difference, we examined the correlation between protein dynamics and carbohydrate structuring proposed from the findings with the eukaryotic linker models. In the PT linkers, the presence of the equatorial 2-OH groups in galactosylated and glucosylated linkers does not increase the number of protein–glycan HB compared to L<sub>man</sub> nor favor perpendicular ring orientations, unlike L<sub>gal</sub> and L<sub>glc</sub> (Fig. S12†). Moreover, the Ramachandran plots of threonines are remarkably similar for the three glycosylated PT linkers (Fig. 6), and show the same preference for extended conformations as L<sub>man</sub> (R2 region). Together, these results predict that the C2 hydroxyl stereochemistry is unlikely to impact proline-rich IDPs. That may result from the loss of one of the HB sites in the protein backbone, since the backbone nitrogen atom is part of the pyrrolidine ring of proline residues.

T-REMD simulations of glycosylated tripeptides GTG were also performed to evaluate the effects of 2-OH configuration on glycan orientation and interactions without the influence of neighboring glycans and amino acids. A single glycan, α-mannose, α-galactose or α-glucose, was O-linked to the central threonine in the models T<sub>man</sub>, T<sub>gal</sub> and T<sub>glc</sub>, respectively (Fig. S9†). The parallel glycan-peptide backbone orientation is favored in the small model systems with α-O-mannosylation, T<sub>man</sub>, relative to other glycans (Fig. S13†).

---

<sup>†</sup> This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.
In the tripeptides $T_{\text{gal}}$ and $T_{\text{glc}}$, the equatorial configuration of 2-OH in $\alpha$-Gal and $\alpha$-Glc favors HB interactions with the peptide as in the $L_{\text{gal}}$ and $L_{\text{glc}}$ linkers. However, an excess of perpendicularly-oriented glycans relative to $L_{\text{man}}$ is not observed for these tripeptides, indicating that the local HB interactions between 2-OH and the peptide are not the only factor affecting glycan conformation. Instead, these results indicate that the glycans in $L_{\text{gal}}$ and $L_{\text{glc}}$ are primarily perpendicularly oriented because of the excluded volumes of neighboring glycans and amino acid side chains, and that the 2-OH—peptide HBs stabilize this glycan conformation. Thus, our results with the small tripeptides suggest that the primary sequence and the distribution of glycosylated residues along the peptide chain are important factors for carbohydrate orientation in these systems.

In summary, experimental comparisons of glycosylated and non-glycosylated IDPs show that O-mannosylation enhances protection against proteolysis by two orders of magnitude relative to the non-glycosylated parent IDP, followed by O-galactosylation (10-fold improved stability). Our results suggest that the resistance to proteolysis is an important driving force for the natural selection of $\alpha$-mannose as the main O-linked glycan motif decorating IDRs and IDPs in secreted eukaryotic proteins. Furthermore, these results demonstrate that the stereochemistry of C2 in the carbohydrate rings plays a key role on glycan orientation, which is correlated to protein flexibility and extension. Accordingly, the axial position of 2-OH in an $\alpha$-mannose glycan is related to the observed higher rigidity and extension of the studied IDR. While associating protein elongation with resistance to proteolysis is perhaps counterintuitive, protein stiffening can explain the remarkably higher stability of the O-mannosylated linker. That is, although we have not investigated the interactions between a protease and IDPs, we conjecture, in the light of the present findings, that increasing the peptide rigidity impairs binding to the catalytic site of a protease. This hypothesis is reinforced by the observation of a similar trend of glycan chemistry impacting resistance to proteolysis of a structured protein and its thermal stability, which is often linked to protein stiffening. Moreover, the effect of glycosylation on the average elongation of the studied IDR, as a protein linker, may be important to provide the optimum distance between the connected domains for protein function. Therefore, O-linked $\alpha$-mannose exhibits the unique ability of both extending the IDR while protecting it against proteolysis.

These results also suggest that the high content of proline residues, especially found in linkers from bacterial cellulases, avoids the need for $\alpha$-mannose for increased protection against proteolysis. This hypothesis will be tested in future experimental studies. We further suggest that the glycosylation pattern in eukaryotic IDRs co-evolved with the primary sequence. That is, the lower content of proline residues in IDPs and IDRs from fungi compared to bacteria is compensated by O-linked $\alpha$-mannosylation to guarantee optimal linker length, flexibility, and protection against proteolysis. Given the compelling alignment of experimental and computational results, we anticipate that our findings will be useful in the burgeoning field of glycoprotein engineering.

Conflicts of interest
There are no conflicts to declare.

Acknowledgements

ETP and MSS thank the Sao Paolo Research Foundation (Grant # 2016/04775-5 and 2013/08293-7). XG, YL, XW, PKC, and ZT thank the National Science Foundation CAREER Award (Grant CHE-1454925). MFC and GTB thank the US Department of Energy Efficiency and Renewable Energy (EERE) Bioenergy Technologies Office (Contract DE-AC36-08GO28308 with NREL). Computer time was provided by Extreme Science and Engineering Discovery Environment (XSEDE) allocation MCB-090159 at the San Diego Supercomputing and Texas Advanced Computing Centers and by the NREL Computational Sciences Center supported by the DOE Office of EERE (Contract DE-AC36-08GO28308).

References


18 S. Ahmad, V. Kumar, K. B. Ramanand and N. M. Rao, *Protein Sci.*, 2012, 21, 433–446.


