Discrimination of epimeric glycans and glycopeptides using ion-mobility mass spectrometry: towards a comprehensive carbohydrate sequencing strategy

P. Both,[†,1] A. P. Green,[†,1] C. Gray,[†,1] R. Šardzík,[1] J. Voglmeir,[1] C. Fontana,[2] M. Austeri,[1] M. Rejzek,[3] D. Richardson,[1] R. A. Field,[3] G. Widmalm,[2] S. L. Flitsch [*,1] and C. E. Eyers [*,1]

[†]These authors contributed equally

[1]School of Chemistry & Manchester Institute of Biotechnology, The University of Manchester, 131 Princess Street, Manchester, M1 7DN, United Kingdom

[2]Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden

[3]Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, United Kingdom

Abstract

Mass Spectrometry (MS) is the primary analytical technique used to characterise the complex oligosaccharide structures which decorate cell surfaces. The relatively limited number of monomeric glycan building blocks found in mammals can be assembled in numerous ways to produce a diverse set of higher-order structures. A number of these monosaccharide units are simple epimers of one another and when combined produce diastereomers of glycoconjugates that cannot be distinguished by existing MS-based techniques. Structure elucidation frequently relies on assumptions that biosynthetic pathways are highly conserved. Here we show that biosynthetic enzymes can display unexpected promiscuity. We demonstrate that human UDP-GalNAc: peptide α-*N*-acetylgalactosaminyltransferase 2 (pp-α-GanT2) is able to utilise UDP-GlcNAc as well as its natural donor (UDP-GalNAc) as a substrate, leading to the synthesis of epimeric glycopeptides *in vitro.* We describe the use of ion-mobility MS (IM-MS) for the separation of these isobaric structures which contain multiple asymmetric centres and differ only in their configuration at a single position. Significantly, collision-induced dissociation (CID) of these glycopeptides followed by IM-MS provides a method for the distinction and identification of epimeric glycans that is independent of polypeptide sequence. Finally, the combination of CID and IM-MS performed on a series of isobaric disaccharides demonstrates that the nature of both the reducing and non-reducing glycans can be determined, and supports the future application of this technique in the sequencing of complex oligosaccharides. The results presented clearly demonstrate the potential of IM-MS to overcome a major limitation currently associated with glycomic and glycoproteomic analysis by providing a platform for the distinction of epimeric structures.

Introduction

It is well established that protein glycosylation plays a critical role in the regulation of a diverse range of biological processes.[1] To improve our understanding of these complex mechanisms, it is necessary to perform detailed studies regarding carbohydrate structure-function relationships. However, the characterisation of glycan and glycopeptide structures from biological samples represents a significant challenge due to the microheterogeneity and diversity of these molecules. Although a surprisingly limited pool of monomeric glycan building blocks are found in mammals, they can potentially be combined to produce a large variety of linear and branched oligosaccharide structures which may differ in glycan sequence, regiochemistry and stereochemistry.[2] Mass spectrometry (MS) has played a pivotal role in the characterisation of these structures due to its high sensitivity and minimal sample requirements. As a result of significant advances in MS-based

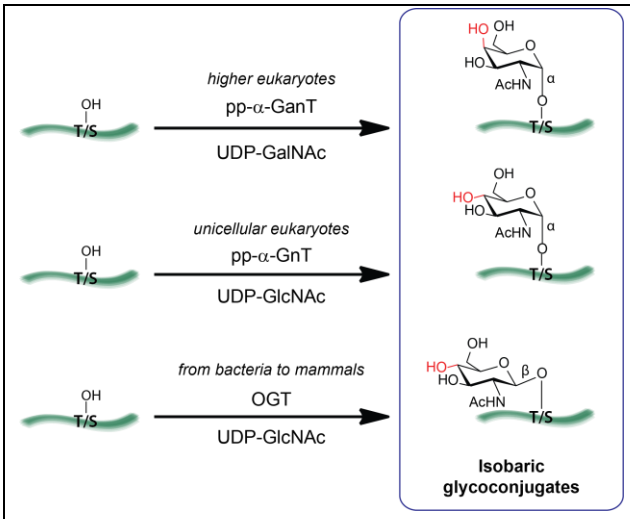


Figure 1 | Examples of common isobaric glycoconjugates and families of enzymes involved in their biosynthesis. (*top*) pp-α-GanT family of enzymes mediate the transfer of an α-linked GalNAc linked to the hydroxyl groups of serine and threonine residues in proteins. This modification is common in higher eukaryotes (including humans) and represents the first step in the biosynthesis of mucin-type *O*-glycans. (*middle*) In unicellular eukaryotes, mucin-type *O*-glycans have been found initiated with an α-linked GlcNAc residue generated by a pp-α-GnT family of enzymes. (*bottom*) β-linked GlcNAc is a common post-translational modification (PTM) observed in many organisms and is especially prevalent among multicellular eukaryotes. The attachment of this residue is mediated by O-GlcNAc Transferase (OGT). β-linked GlcNAc is also transferred to the Notch epidermal growth factor (EGF) repeats by extracellular OGT (EOGT)[36]. The differentiation of epimeric glycopeptides by mass spectrometry alone has not been reported previously.

techniques, impressive structural information including peptide/protein glycosylation site and regiochemistry of glycosidic linkages can now be obtained from complex mixtures of glycoconjugates from biological samples.[3][4]5][6] However, a major limitation that remains unresolved is the distinction of isobaric epimers of glycoconjugates (Fig. 1) which can occur as a result of alternative configuration at the anomeric linkages or due to the presence of epimeric glycan monomers such as the three hexoses (Hex) glucose (Glc), galactose (Gal) and mannose (Man) or the two HexNAcs *N*-acetylglucosamine (GlcNAc) and *N*-acetylgalactosamine (GalNAc). This is of particular importance as these five monosaccharides are amongst the most abundant building blocks present in mammalian glycans.[7] We currently rely on our understanding of the glycan biosynthesis to make predictions regarding the stereochemistry of glycoconjugates with the assumptions that (i) all biosynthetic enzymes are known and (ii) that they have strict substrate specificities. These assumptions expose the opportunity for structural misinterpretation, potentially hindering the identification of novel glycan structures with unexplored functions.

The biosynthesis of glycans is mediated by the sequential action of a series of glycoenzymes which are considered to act with a high degree of donor-, acceptor- and linkage specificity.[8] However, the specificity of these enzymes has not been extensively investigated due to a lack of high-throughput analytical techniques. Here we describe the use of peptide microarray technology to investigate glycoenzyme specificity. During these studies, we have discovered a highly unexpected activity of the common human UDP-GalNAc: peptide α-*N*-acetylgalactosaminyltransferase 2 (pp-α-GanT2) which is able to mediate the synthesis of epimeric glycopeptides containing both GlcNAc and GalNAc residues *in vitro* (Fig 1). These results evoked our interest in developing an analytical technique with the necessary sensitivity and resolution to distinguish epimeric glycoconjugates which may be present in biological samples. We now describe the use of travelling wave ion mobility-mass spectrometry (TWIM-MS) for the separation of diastereomeric glycoconjugates (shown in Figure 1) containing multiple asymmetric centres which differ only in their configuration at a single position. Additionally, we demonstrate that the GlcNAc and GalNAc oxonium ions generated following collision-induced dissociation (CID) of the different glycopeptides are also distinguishable by ion mobility. These results highlight

the potential utility of this technique for screening native glycopeptides for the different isobaric monosaccharide building blocks in a manner that is independent of peptide sequence. More generally, the distinction of epimeric glycan product ions generated following CID suggests the potential application of IM-MS in the development of a carbohydrate sequencing strategy, additionally permitting elucidation of stereochemical information. To explore this possibility, a series of isobaric disaccharides were analysed by IM-MS and CID. The results presented demonstrate that the nature of both the reducing and non-reducing glycans can be determined using this approach, enabling us to move towards a general glycan sequencing strategy. Combined, the data presented clearly demonstrates that IM-MS provides a platform with the necessary resolution and sensitivity to overcome the current limitation of stereoisomer differentiation associated with glycomic and glycoproteomic analysis.

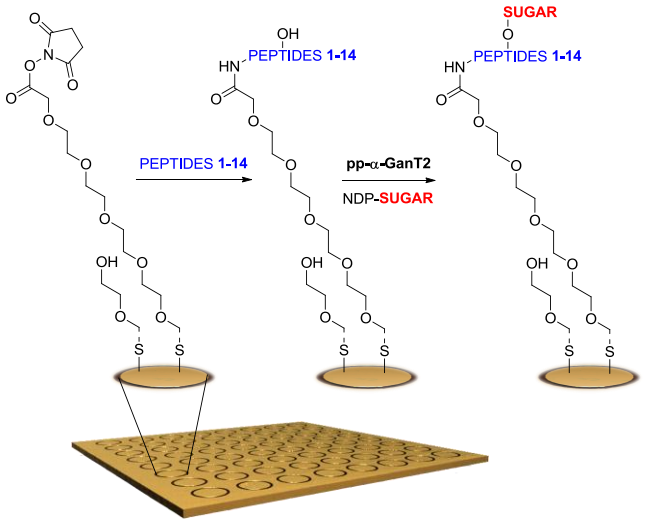
Results and Discussions

Promiscuity of the glycosyltransferase pp-α-GanT2

pp-α-GanT2 is one of the most ubiquitous members of a family of approximately twenty human pp-α-GanTs. These enzymes mediate the first step in the biosynthesis of mucin-type *O*-glycans by catalysing the transfer of an α-linked GalNAc residue from the activated sugar UDP-GalNAc to the hydroxyl groups of serine and threonine residues in proteins (Fig. 1, top). As a result of their pivotal role in mucin biosynthesis, there has been considerable interest in elucidating the specificity of individual pp-α-GanT2 isoforms towards defined acceptor peptide sequences.[9] However, the sugar donor promiscuity of these enzymes remains largely unexplored and it is generally assumed that these glycosyltransferases display a high degree of specificity towards UDP-GalNAc. We have recently reported the use of peptide and carbohydrate microarrays to probe glycoenzyme specificity in a high-throughput manner.[10][11] Acceptor substrates of interest are covalently linked to a self-assembled monolayer (SAM) of alkanethiols on a gold platform. This technology offers the distinct advantage that the modification of immobilized substrates can be conveniently monitored using MALDI-ToF MS, providing a highly sensitive and label free method of analysis.[12] We now describe the use of this technology to investigate the sugar donor promiscuity of recombinant human pp-α-GanT2, which was expressed in *Pichia pastoris* (see supplementary information 2.2). Since reactions can be performed on a minute scale in a high-throughput manner, we were able to rapidly screen a set of 15 naturally occurring and synthetic sugar donors[13][14] against a panel of peptide sequences **1**-**14** immobilized on the array platform (Fig. 2). Full structures of the sugar donors screened are given in supplementary information Fig. S1. Peptides **1**-**13** represent fragments of naturally occurring proteins and contain threonine or serine residues which are reported to be glycosylated *in vivo,* whilst peptide **14** is a non-natural sequence which has previously been shown to be an optimised sequence for pp-α-GanT2.[9]

The results of the microarray studies are summarised in Fig. 2 and indicate that pp-α-GanT2 displays a surprising degree of promiscuity with regards to its sugar donor substrate. As expected, the natural donor substrate UDP-GalNAc was shown to be the best substrate and the enzyme was successful in transferring a GalNAc residue to thirteen of the fourteen immobilised acceptor peptides (**1**-**2**, **4**-**14**). Multiple glycosylation of a number of these peptides was observed by MALDI-ToF MS. UDP-Gal was also shown to be a substrate for human pp-α-GanT2, with a galactose residue being transferred to five of the immobilised peptides. The ability of pp-α-GanT2 to accept UDP-Gal as a donor substrate has been reported previously.[15] Further modifications at the 2-position of the glycan donor have also been shown to be tolerated by the enzyme. Qasba *et al.* have demonstrated that pp-α-GanT2 is able to transfer modified galactose units presenting an azido or a ketone functional group as a chemical handle for further bioconjugation.[16] Importantly, these structures all retain the *galacto*-configuration of the enzyme's natural substrate, UDP-GalNAc. More interestingly, results in Fig. 2 show that wild-type human pp-α-GanT2 is able to transfer a GlcNAc residue from UDP-GlcNAc to seven of the fourteen immobilised acceptor peptides. The ability of pp-α-GanT2 to utilise UDP-GlcNAc, a naturally occurring sugar donor with the alternative C-4 configuration to the natural substrate, is highly unexpected. To our

knowledge, the resulting α-GlcNAc modification has been observed in unicellular eukaryotes,[17][18] but not identified in humans thus far.



		UDP-GalNAc	UDP-Gal	UDP-GalA	UDP-Glc	UDP-GlcNAc	UDP-GlcA	UDP-Xyl	UDP-Ara	UDP-Man	GDP-Man	GDP-Fuc	CMP-NANA	UDP-GlcNAcA	UDP-GlcdiNAc	UDP-GlcdiNAcA
1	GAPGSTAPPA (Muc1)	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-
2	GAHGVYSAPD (Muc1)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	GAPDTRPAPG (Muc1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	GPITTTTVT (Muc2)	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-
5	GVTPTPTPTG (Muc2)	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-
6	GTQTPTTTPI (Muc2)	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-
7	GPISTTTMVT (Intest. Muc)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	GPTGPTTTPI (Intest. Muc)	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-
9	YSPTSPS (RNA II)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	YPGGSTPVSSANMM (CKII)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	RQIHATPTPVTAIG (αDG)	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-
12	PPDAASAAPLR (EPO)	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
13	PPDAATAAPLR (EPO S->T)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	GAPGPTPGPAG (optim. subst.)	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-

Figure 2 | The application of peptide microarrays to investigate the sugar donor promiscuity of pp-α-GanT2. (*top*) Peptides are immobilised on a self-assembled monolayer (SAM) on a gold platform through the *N*-terminus of the peptide by formation of an amide bond. The immobilised peptides are then incubated with pp-α-GanT2 and the relevant NDP-sugar donor and the reaction is monitored by MALDI-ToF MS. (*bottom*) The table shows the peptide sequences **1**-**14** immobilised with their parent protein in parentheses. The activated sugar donors screened against **1**-**14** are listed along the top of the table (for full structures see supplementary information). Positive activity as detected by MS is denoted with +.

Having identified unexpected catalytic activity of pp-α-GanT2 towards UDP-GlcNAc on the microarray platform, our attention turned to validating this result in solution phase. Reactions were carried out with both UDP-GalNAc and UDP-GlcNAc in combination with a panel of biologically significant peptide sequences (see supplementary information 6 for full details). The enzymatic modification of the peptides was analysed using MALDI-ToF MS. The results proved to be consistent with those obtained during the microarray studies. For unambiguous characterisation of the glycopeptide products, the glycosylation of peptide **14** was carried out on a preparative scale using both UDP-GlcNAc and UDP-GalNAc as the donor substrate (Fig. 3). After isolation and

purification, the structures of the resulting glycopeptides **15** and **16** were then unambiguously assigned using detailed NMR studies. Such characterisation was important to exclude the potential activity of a contaminating epimerase, which is well known to convert UDP-GlcNAc to UDP-GalNAc.[19] The anomeric resonances of the monosaccharides in the glycopeptides were assigned using multiplicity-edited $^{1}H,^{13}C$-HSQC spectra, in which a single cross-peak corresponding to a hexopyranosyl residue was observed at δ_H/δ_C 4.98/99.40 and 5.00/99.55 in **15** and **16**, respectively. The ^{1}H and ^{13}C chemical shifts of the sugar residues were assigned using $^{1}H,^{1}H$-TOCSY and multiplicity-edited $^{1}H,^{13}C$-HSQC experiments (see supplementary information table S1) and compared to those of the corresponding monosaccharides.[20][21] From the analysis of $^{1}H,^{1}H$-TOCSY spectra (Fig. 3a and 3b) and ^{1}H and ^{13}C chemical shifts, the sugar residue in **15** has the *gluco*-configuration whereas the sugar residue in **16** has the *galacto*-configuration. Note that whereas all the protons can be traced from H1 of GlcNAc, only those up to H4 can be identified from H1 of GalNAc, due to the small coupling constant between H4 and H5, which is typical for a hexopyranose residue with the *galacto*-configuration. The residues are α-linked since the $^{1}J_{C1,H1}$ coupling constants are 174 and 172 Hz and the $^{3}J_{H1,H2}$ coupling constants 3.6 and 3.7 Hz in **15** and **16** respectively. In addition, only one intra-residue NOE correlation from the anomeric proton to H2 was observed in the $^{1}H,^{1}H$-NOESY spectra of **15** and **16** (Fig. 3c and 3e respectively), which is consistent with the α-configuration of the anomeric centers. The substitution of a threonine residue (T) with the corresponding sugar residue was confirmed by long-range NOE correlations from the respective anomeric proton to the resonances at 4.29 and 1.36 ppm (H3 and H4 in T) in **15** (Fig. 3c and 3d) and to 4.29 and 1.36 ppm (H3 and H4 in T) in **16** (Fig. 3e and 3f). Corresponding nOe signals were previously observed for oligosaccharides in which α-D-Man*p* residues were attached to threonine in glycopeptides related to the α-DG glycoprotein.[22] The spin systems of the threonine residues in **15** and **16** were assigned using $^{1}H,^{1}H$-TOCSY experiments, and correlated to the respective ^{13}C resonances using the multiplicity-edited $^{1}H,^{13}C$-HSQC (supplementary information table S1). Thus, **15** contains an α-D-Glc*p*NAc group and **16** contains an α-D-Gal*p*NAc group, both of which are linked to a threonine residue in their glycopeptide.

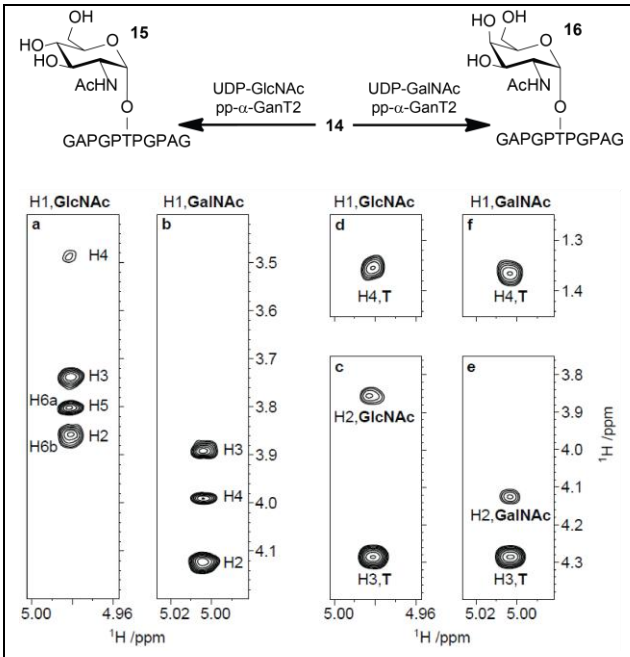


Figure 3 | Synthesis and characterisation of epimeric glycopeptides 15 and 16 – (*top*) Treatment of peptide **14** with ppGanT2 and either UDP-GlcNAc or UDP-GalNAc results in the synthesis of glycopeptides **15** and **16** respectively. (*bottom left*) Selected regions from the $^{1}H,^{1}H$-TOCSY spectra (τ_{mix} 100 ms) of **15** (a) and **16** (b) showing the spin systems originating from the anomeric protons, and (*bottom right*) selected regions from the $^{1}H,^{1}H$-NOESY spectra (τ_{mix} 300 ms) of **15** (c and d) and **16** (e and f) showing, inter alia, inter-residue correlations from the anomeric protons to H3 and H4 in the corresponding threonine residues.

The ability of human pp-α-GanT2 to transfer an α-linked GlcNAc residue from UDP-GlcNAc to biologically significant peptide sequences raises a number of interesting points. Firstly, this observation provides evidence regarding the potential evolutionary origin of the ppGanT family of glycosyltransferases. The evolutionary pressure to increase the structural diversity of glycans which are able to serve as discrete ligands for various recognition events is well documented.[23] In unicellular eukaryotes, mucin-type *O*-glycans are initiated by an α-linked GlcNAc moiety in place of the α-linked GalNAc structure which is observed in mammals. Based on detailed bioinformatics and mechanistic studies it has been proposed that mammalian ppGanTs have evolved from the early UDP-GlcNAc: peptide α-*N*-acetylglucosaminyltransferases (ppGnTs) responsible for the attachment of the GlcNAc residue in these organisms.[24] The discovery of residual UDP-GlcNAc activity in human pp-α-GanT2 lends strong evidence in support of this theory. The fact that closely related UDP-activated sugars with the *gluco*-configuration were not substrates for pp-α-GanT2 (Fig. 2), demonstrates that this enzyme is not simply displaying general promiscuity.

Secondly, the discovery of substrate promiscuity associated with an important human glycosyltransferase suggests that glycan microheterogeneity may occur *in vivo* as a result of glycoenzyme promiscuity, which has not been considered previously. In this instance, the ability of a human pp-α-GanT2 to transfer a GlcNAc residue to biologically relevant peptide sequences *in vitro* raises the possibility that α-GlcNAc could be a natural post-translational modification (PTM) in humans. The kinetic parameters for UDP-GalNAc and UDP-GlcNAc with pp-α-GanT2 in combination with acceptor peptide **14** were determined (UDPGalNAc: K_m = 18 μM, K_{cat} = 20 min^{-1}, UDPGlcNAc: K_m = 273 μM, K_{cat} = 1.3 h^{-1}). Although the K_m for UDP-GlcNAc is an order of magnitude greater than that for UDP-GalNAc, the value lies in the range of other mammalian glycosyltransferases that utilise this substrate[25][26] (see supplementary information Fig. S6). However, the low K_{cat} value observed for UDP-GlcNAc suggests that it is unlikely to be a viable substrate for pp-α-GanT2 *in vivo*. Additionally, incubation of peptide **14** with mixtures of UDP-GlcNAc and UDPGalNAc (up to 5:1 UDP-GlcNAc:UDPGalNAc) failed to result in competitive transfer of the GlcNAc moiety to the peptide *in vitro*. Studies are currently ongoing to determine whether alternative isoforms of the ppGanT family display similar or enhanced activity towards UDP-GlcNAc. Initial investigations suggest that pp-α-GanT1 is unable to utilise UDPGlcNAc as a substrate. A detailed phylogentic analysis (supplementary information Fig. S7) of the catalytic domains of available pp-α-GanTs and pp-α-GnTs shows that human pp-α-GanT11 is the most closely related isoform to the pp-α-GnTs from lower eukaryotes, suggesting that this may be a good candidate to display enhanced activity towards UDP-GlcNAc.

Differentiation of α-GlcNAc and α-GalNAc modified peptides by Ion Mobility-Mass Spectrometry (IM-MS)

The studies described above highlight the advantages of MS as a sensitive analytical technique which meets the significant demands associated with glycomic analysis. However, the results presented also illustrate the limited stereochemical information provided by this technique, and detailed NMR studies using relatively large amounts of purified material were required to characterize the stereochemistry of the glycan transferred to the peptide sequences. The development of new analytical methods which provide more precise stereochemical information on glycoconjugates whilst retaining the high sensitivity associated with MS would greatly facilitate the identification of novel glycosylation patterns occurring *in vivo*. IM-MS is an emerging analytical strategy now finding utility in a number of fields including glycomics and glycoproteomics.[27][28][29][30] This technology has the potential to greatly enhance the analysis of complex glycopeptide mixtures due to its ability to separate ions of identical *m/z* in the gas phase by virtue of changes in their collisional cross section (CCS). Creese *et al.* recently reported the separation of glycopeptides with identical peptide sequences but with different sites of glycosylation using high field asymmetric wave ion mobility spectrometry (FAIMS)[27] and Li *et al.* were able to separate a series of methyl glycosides by both travelling wave ion mobility spectrometry (TWIMS) and drift tube ion mobility spectrometry (DTIMS).[28] During the course of preparation of this manuscript, the same authors described the development of an instrument capable of performing tandem IM-

MS (a hybrid IMS-TWIMS instrument), which they used to analyse select isomeric carbohydrates.[29] Encouraged by these studies, we investigated the potential application of IM-MS for the distinction and characterization of epimeric glycoconjugates, which presents a considerably greater challenge and has not been described previously.

TWIMS-MS (14 repeats) was initially performed on epimeric glycopeptides **15** and **16**, containing α-linked GlcNAc and GalNAc motifs respectively. The mobilograms associated with the $[M+H]^+$ peak at *m/z* 1081.54 for these glycopeptides are shown in Fig. 4 and drift time data is shown in table S2 (supplementary information). Interestingly, the α-GalNAc peptide **16** has a significantly longer drift time than the α-GlcNAc peptide **15** ($p = 2 \times 10^{-12}$), indicating that the α-GalNAc induces a more open/extended conformation in these glycopeptide structures. The broader arrival time distribution further suggests the possibility of multiple extended conformations which cannot be resolved under these conditions. CID of each of the glycopeptides preferentially resulted in neutral loss of the sugar moiety and formation of the unmodified peptide ($[\Delta M+H]^+$, *m/z* 878.51). The mobilograms and drift time data associated with the deglycosylated peptides are shown in Fig. 4a and table S2a (supplementary information) respectively, and demonstrate that the drift times of the unmodified peptide ion following neutral loss of the sugar is identical in both cases ($p = 0.63$), as would be expected. This difference in conformation of the modified peptides is empirically represented by a box-whisker plot (Table S2b supplementary information) of the difference in drift times between glycosylated and deglycosylated peptides. The significance of these differences in drift times was assessed by means of independent T-tests ($p<3 \times 10^{-14}$). The drift times of the corresponding β-linked GlcNAc glycopeptide **17**, before and after CID, were also determined. β-linked GlcNAc is a naturally occurring human PTM mediated by UDP-GlcNAc: peptide β-*N*-acetylglucos-aminyltransferase (OGT).[31] The results clearly demonstrate that isobaric stereoisomers of glycopeptides containing either α-GalNAc or β-GlcNAc linked to threonine, both commonly occurring PTMs in humans, can be separated by IM-MS. No significant difference was observed in the drift times of the singly charged β-GlcNAc glycopeptide **17** compared to the analogous α-linked GlcNAc glycopeptide **15** under the conditions employed ($p = 0.67$). However, a significant difference in drift times were observed between the corresponding doubly charged species (Fig S2a supplementary information).

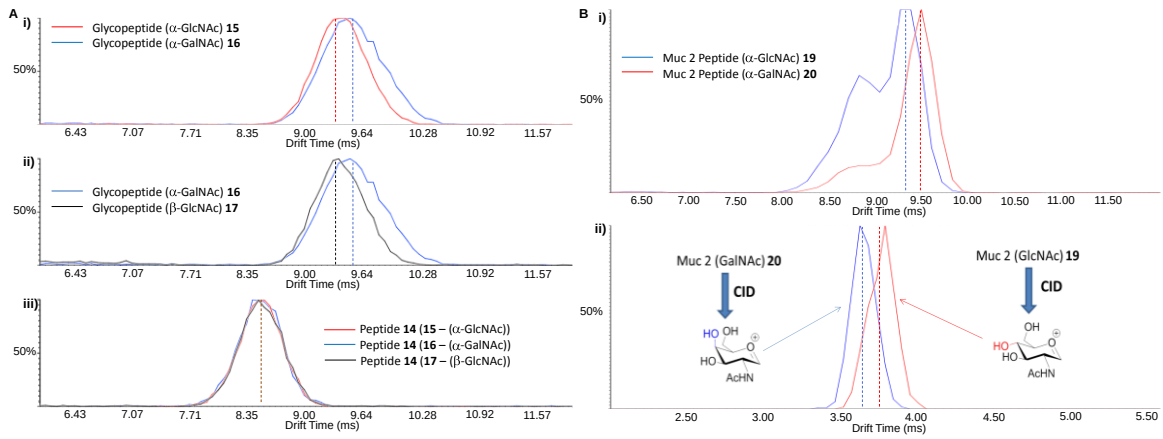


Figure 4 | Mobilograms showing the discrimination of epimeric glycopeptides 15-17 and 19-20 and the distinction of HexNAc oxonium ions generated following CID. (A,i) Mobilogram showing that glycopeptide **16** (α-GalNAc) (blue) can be distinguished from the isobaric glycopeptide **17** (β-GlcNAc) (red). (A,iii) CID of glycopeptides **15**, **16** and **17** (α-GalNAc/α-GlcNAc/β-GlcNAc) prior to ion mobility separation results in neutral loss of the glycan moiety leaving the unmodified peptide that cannot be differentiated by IM-MS (as would be expected). (B,i) Mobilograms associated with both the glycosylated Muc 2 (PTTTPITTTTTVTPTPTPTGTQT) species **19** and **20** demonstrates separation of multiple glycoforms which can be partially resolved under the conditions employed. The more abundant conformations of the epimeric glycopeptides can still be differentiated by IM-MS by differences in their drift time. (B, ii) CID of glycopeptides **19** and **20** prior to ion mobility separation, results in the formation of epimeric oxonium ions that are distinguishable by TWIMS-MS. Vertical lines represent 1/2fwhh.

Having demonstrated the utility of IM-MS to distinguish epimeric glycopeptides derived from the optimized peptide sequence **14**, our attention turned to the separation of glycopeptides based on a structurally more complex natural peptide (PTTTPITTTTTVTPTPTPTGTQT **18**), that is a tandem repeat unit of the Muc 2 protein (Fig. 4b). IM-MS was performed on glycopeptides derived from the Muc 2 sequence containing both a single α-GlcNAc and a single α-GalNAc modification. The mobilograms associated with the $[M+2H]^{2+}$ ion for these two glycopeptides are depicted in Fig. 4b. The presence of multiple threonine residues in this sequence presents the added complication that glycosylation occurs on a number of sites, leading to the formation of a heterogenous mixture of glycoforms (see supplementary information Fig. S10-11 for details regarding sites of glycosylation). These glycoforms could be partially resolved under the ion mobility conditions employed, as evidenced by the presence of two peaks in the mobilograms associated with each of the glycopeptides **19** and **20**. In contrast with the optimal glycosylated peptides, the α-GalNAc glycopeptide **20** has a significantly shorter drift time than that of the α-GlcNAc species **19** ($p = 1.2 \times 10^{-7}$), indicating that the differences in observed drift times are not simply due to differences in conformation of the GlcNAc and GalNAc glycans themselves, but rather differential glycosylation-induced changes in peptide structure. This observation of glycan-induced changes in glycopeptide structure is further supported by the mobility analysis of this peptide comprising sequential addition of multiple α-GalNAc moieties, as the differences in drift times after each addition are not uniform (Table S4 supplementary information).

In order to utilise IM-MS as a generally applicable tool to enable the distinction of epimeric glycans linked to peptides, it is necessary to develop a method for the separation of stereoisomeric glycoconjugates that is independent of peptide sequence. This was achieved by combining CID with subsequent ion mobility separation. The $[Muc\ 2 + GlcNAc + 2H]^{2+}$ and $[Muc\ 2 + GalNAc + 2H]^{2+}$ ions at *m/z* 1260.13 were isolated and subjected to CID resulting in the formation of epimeric GlcNAc and GalNAc oxonium ions which were distinguishable by TWIMS. The GlcNAc oxonium species was shown to have a significantly longer drift time than the corresponding GalNAc species ($n=3$, $p = 2.9 \times 10^{-6}$) (Fig. 4b). Since the structure of these product ions will be specific to the nature of the hexose, rather than the precursor ion from which it is derived, this combination of CID and IM-MS provides a broadly applicable strategy for the distinction and characterisation of epimeric glycans linked to peptides, and could potentially be developed as a screening strategy for biologically-derived glycopeptides. More generally, the observation that epimeric glycan CID product ions can be distinguished by IM-MS highlights the potential application of this technique for sequencing more complex oligosaccharides presented on cell surfaces.

Towards carbohydrate sequencing using IM-MS

The characterisation of oligosaccharides from biological samples typically involves the initial release of the glycans from the peptide or protein (using beta elimination for O-glycans or glycosidase treatment for N-glycans) followed by multistage tandem MS (MS^n). Analysis of the resulting fragment ions provides detailed structural information including regiochemistry of the glycosidic linkages and branching patterns. The major limitation of this technique lies in its inability to provide adequate (detailed) stereochemical information that is required for a comprehensive carbohydrate sequencing strategy. IM-MS has great potential to overcome this limitation due to its ability to distinguish epimeric structures.

To explore the application of IM-MS in carbohydrate sequencing, TWIMS-MS was initially performed on the isobaric hexoses **21**-**23** (Glc, Gal and Man) and HexNAcs **24**-**26** (GlcNAc, GalNAc and ManNAc) (n=3). Only sodiated ions of these monosaccharides were observed under the conditions employed. The drift times and collision cross sectional areas of **21**-**26** (as determined by comparison with a standard with known cross-sectional area)[32] are given in Table 5. The data demonstrates that the three HexNAcs **24**-**26** are all distinguishable using IM-MS. Interestingly, the mobilogram associated with GlcNAc consists of a minor peak followed by a major peak at a longer drift time, suggesting the presence of multiple conformations, potentially arising due to differential sites of sodiation associated with this structure in the gas phase. The drift time of Glc

21 is significantly longer than that of both Man ($p = 7.7 \times 10^{-5}$) and Gal ($p = 7.1 \times 10^{-5}$). Although a small difference in drift time was observed between Gal **22** and Man **23**, this difference was not significant ($p = 0.1$) under the conditions employed. This result is unsurprising since both structures are similar in that they possess a single axial hydroxyl substituent.

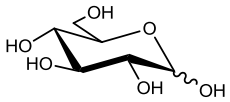
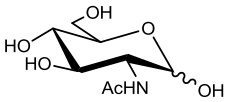
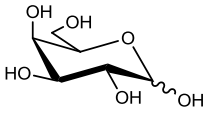
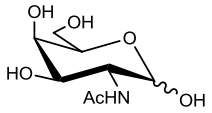
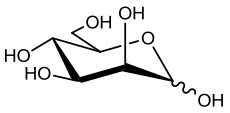
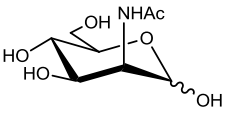
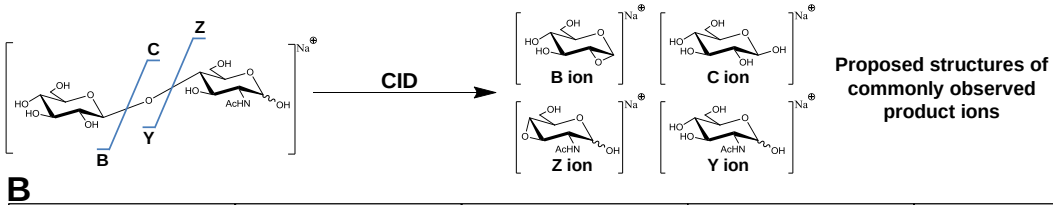
Hex	Structure	Drift-time (ms) (and CCS (Å²))	HexNAc	Structure	Drift-time (ms) (and CCS (Å²))
Glc (**21**)		3.84 (78)	GlcNAc (**24**)		4.58 (90)
Gal (**22**)		3.68 (75)	GalNAc (**25**)		4.25 (85)
Man (**23**)		3.70 (76)	ManNAc (**26**)		4.23 (84)

Table 1 | Drift times and collision cross section (CCS) values of naturally occurring, isobaric monosaccharide building blocks: hexoses (Glc, Gal, and Man) and HexNAcs (GlcNAc, GalNAc and ManNAc). A significant difference in drift time was observed for all monosaccharides with the exception of Gal and Man. Although a small difference in CCS was determined for these two structures, this difference was not significant under the conditions employed and further optimization would be required to potentially enhance the separation of these two structures. The mobilogram associated with GlcNAc consisted of two peaks, suggesting the presence of multiple conformations or sites of sodiation. In this instance, the drift-time and CCS measurements are reported for the most intense species. All drift-times and CCS values were determined for the $[M+Na]^+$ species. CCS values were calculated following calibration of the drift times in the TWIMS device using polyglycine peptides with known CCS values.[32]

A series of isobaric beta-1,4 linked disaccharides **27**-**31**, each containing one hexose and one HexNAc unit was subsequently analysed using the same strategy. Disaccharides **28**-**31** are commercially available and disaccharide **27** was synthesised from *N*-acetyl glucosamine in nine linear steps (see supplementary section 17). TWIMS analysis demonstrated that all intact disaccharides **27**-**31** were distinguishable under the conditions employed, with the exception of structures **27** and **29** which displayed identical drift times. Interestingly, separation of these structures was achieved under optimised conditions on the first generation Synapt HDMS instrument (Fig S14 and Tables S6 and S7 supplementary), suggesting that further optimisation of separation conditions on the G2 may enable discrimination of these species, albeit under slightly modified conditions. In an attempt to derive glycan sequence information using a combination of IM and MS data, each of the disaccharides **27**-**31** was then subjected to CID prior to TWIMS. The commonly observed product resulting from collisional dissociation of a disaccharide into its monosaccharide components are B and C ions (charge retained on the non-reducing glycan) and Y and Z ions (charge retained on the reducing glycan) (Fig 5a), observation of which allows glycan sequencing to be performed from either the reducing and/or non-reducing termini. The mobility data and CCS values of the CID product ions are presented in Fig 5b. All the product ions derived from the three hexoses and the two HexNAcs that we expected to separate are distinguishable (see below) and follow the same trend in drift times reported for the monosaccharides (Table 1). Significantly, the C-ions derived from disaccharides **28** and **29/30** (with Man and Gal at the reducing terminus) are distinguishable, in contrast with the identical drift times observed for the native monosaccharides **22** and **23**. The finding that the Y-ions derived from carbohydrates presenting GlcNAc or GalNAc at the reducing terminus can be discriminated by IM-MS, suggests that this strategy can be used to explore the possibility that α-linked GlcNAc is a naturally occurring modification present in mucin-type glycans, possibly arising due to ppGanT promiscuity (as discussed above).

The Y-ions derived from glycans **27-29** (*ie* with the same reducing sugar) are identical, as are the C-ions from the glycans with the same non-reducing sugar, as would be expected. Also as expected, the C-ion derived from disaccharides **29 or 30** has an identical CCS as the Y-ion derived from disaccharides **31**. The formation of a Z-ion arises as a result of loss of the C-4 hydroxyl group (Fig 5a), and these fragment ions derived from disaccharides **27**-**30** (with either GlcNAc or GalNAc residues at the reducing terminus) also have identical drift times (as expected). Only disaccharide **31,** which has a GalNAc residue at the non-reducing end, yielded a B-ion with sufficient intensity to enable mobility data to be obtained. The drift time of this B-ion is consistent with previously reported enhancement of B-ion formation at non-reducing HexNAc (cf hexose) glycosidic bond cleavage.[33] The mobilograms associated with the B-ion derived from disaccharide **31**, and of the Y-ions derived from disaccharides **27-29** (GlcNAc at the reducing terminus) consist of two peaks, again indicative of multiple conformations (potentially due to different sites of sodiation). The mobilograms associated with the Y-ions consist of a minor peak followed by a major peak at a longer drift time, in agreement with the mobilogram associated with the native monosaccharide (Fig X supplementary info). Combined, the results presented demonstrate that monosaccharide product ions corresponding to the most abundant isobaric glycan building blocks (Glc/Man/Gal and GlcNAc/GalNAc) are all distinguishable by IM-MS. Additionally, the drift times (and calculated CCS values) of these species are independent of the starting disaccharide structures from which they are derived. These observations highlight the potential of IM-MS to contribute to the development of a broadly applicable carbohydrate sequencing strategy.

A



B

Glcβ1-4GlcNAc (27) Drift-time (ms) (and CCS (Å²))	Manβ1-4GlcNAc (28) Drift-time (ms) (and CCS (Å²))	Galβ1-4GlcNAc (29) Drift-time (ms) (and CCS (Å²))	Galβ1-4GalNAc (30) Drift-time (ms) (and CCS (Å²))	GalNAcβ1-4Gal (31) Drift-time (ms) (and CCS (Å²))
6.84 (ND)	6.91 (ND)	6.84 (ND)	6.62 (ND)	6.46 (ND)
CID Product Ions	**CID Product Ions**	**CID Product Ions**	**CID Product Ions**	**CID Product Ions**
Y ion	**Y ion**	**Y ion**	**Y ion**	**Y ion**
4.64 (90)	4.64 (90)	4.64 (90)	4.32 (85)	4.76 (76)
C ion	**C ion**	**C ion**	**C ion**	
3.89 (78)	3.71 (75)	3.75 (76)	3.75 (76)	Low Intensity
Z ion	**Z ion**	**Z ion**	**Z ion**	
4.13 (82)	4.13 (82)	4.13 (82)	4.13 (82)	Low Intensity
Low Intensity	Low Intensity	Low Intensity	Low Intensity	**B ion**
				3.99 (80)

Figure 5 |Drift times of isobaric disaccharides 27-31 and the monosaccharide product ions generated following CID. (A) CID of a disaccharide commonly results in the formation of B-, C-, Y- and Z-ions which arise as a result of a single bond cleavage.[37] The combination of CID and TWIMS allows the drift-times and CCS values of these product ions to be determined. (B) Structures of the β-1,4 linked disaccharides **27**-**31** and representative structures of the monosaccharide product ions (highlighted in red) formed as a result of CID of the respective disaccharide precursor ion. Drift-times for the $[M+Na]^+$ species and calculated CCS values are shown below each of the structures. For cases where two conformations (potential sites of sodiation) were observed, the drift-time and CCS measurements are reported for the most prominent species. CCS values were determined following calibration of the TWIMS device using polyglycine peptides with known CCS values.[32] CCS values for the disaccharides were not determined (ND) as the drift times values lie outside the range of this calibration.

Conclusions

IM-MS is an emerging analytical technique which combines the benefits traditionally associated with mass spectrometry such as high sensitivity and mass accuracy, with the ability to distinguish ions with identical masses such as regio- and stereo-isomers.[34] The development of instrumentation with significantly enhanced sensitivity and mobility resolution will undoubtedly further improve the degree of separation of these species and it is predicted that this powerful technology will find broad application in a variety of research areas.

We have demonstrated the application of IM-MS to distinguish stereoisomers of glycans and glycoconjugates which represents a major limitation currently associated with glycomic and glycoproteomic analysis. This technique has the necessary resolution to distinguish epimeric glycoconjugates which contain multiple asymmetric centres and crucially is able to distinguish epimeric glycan product ions formed following CID of glycopeptide or oligosaccharide structures.. The methodology described is expected to facilitate the identification of novel glycosylation patterns occurring *in vivo*. It is expected that the sequencing of complex carbohydrates will be made possible by the incorporation of an ion trapping device prior to the mobility cell, allowing multi-stage tandem MS to be performed prior to ion mobility. This methodology may ultimately enable the development of a comprehensive carbohydrate sequencing strategy, a goal which remains a significant challenge at the forefront of glycomic analysis.

Methods

Travelling wave ion mobility-mass spectrometry (TWIMS-MS) separation of glycopeptides

Glycopeptides **15**, **16** and **17** were diluted to 1 pmol/μL and Muc 2 glycopeptides **19**, **20**, (2xGalNAc) and (3xGalNAc) to 2 pmol/μL with 1:1 acetonitrile:water 0.1% (v/v) formic acid prior to analysis. Free monosaccharides and disaccharides were diluted to 5 pmol/μL in 1:1 methanol:water. Samples were infused into a Synapt HDMS instrument (See supplementary) or Synapt G2 HDMS (Waters, UK) using borosilicate emitters (Proxeon, Odense, Denmark). For the Synapt G2 HDMS instrument, the capillary, cone voltage and source temperature were typically set to 1.1-1.5 kV, 35 V and 80^{o}C respectively. The IM travelling wave speed was set to 700 m/s and the wave height set at its maximum 40 V for the analysis of glycopeptides and the disaccharides. For the analysis of glycan product ions (post CID), the travelling wave velocity was increased to 1200 m/s. The nitrogen drift gas flow was set at 90 mL/min for all experiments. CID of glycopeptides or disaccharides was induced in the trap cell using an argon collision gas at collision energy (CE) of 45 and 35 kV respectively and prior quadrupole mass selection. Data were collected between 50-2500 m/z for glycopeptides **15**-**17 and 19-20** and between 50 and 1200 for monosaccharides **21**-**26** and disaccharides **27**-**31**with a product ion tolerance of ±100 ppm. The significance of analyte separation was determined using independent T-tests. The mass measurements were calibrated using product ions following CID of Glu-fibrinopeptide (trap CE 35 kV) by infusion (500 fmol/μL; 1:1 acetonitrile:water, 0.1% (v/v) formic acid) at a flow rate of 0.1 μL/min. Drift times were calibrated to calculate collision cross sectional area for glycopeptides by infusion of a tryptic digest of QCal-IM, according to the procedure described previously.[35] Product ion and monosaccharide collision cross sections were calibrated against a series of $[Gly_n+H]+$ peptides (n=3-6), with previously verified collision

cross sections.[32]Mass spectra were processed using MassLynx V4.1 and mobilograms using DriftScope v2.1 (both Waters, UK).

Acknowledgements

This research was supported by grants from the EPSRC, the BBSRC, Swiss Scheme Foundation (MA), the Royal Society (Wolfson Award to SLF), the Knut and Alice Wallenberg foundation, the Swedish Research Council and the European commission (FP7). Work at JIC was supported by BBSRC Institute Strategic Programme Grant BB/J004561/1 and the John Innes Foundation.

References

[1] Hart, G. W. & Copeland, R. J. Glycomics hits the big time. *Cell* **143,** 672-676 (2010).

[2] Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **13,** 448-62 (2012).

[3] Hanisch, F-G. Top-down sequencing of *O*-glycoproteins by in-source decay matrix-assisted laser desorption ionisation mass spectrometry for glycosylation site analysis. *Anal. Chem.* **83**, 4829-4837 (2011).

[4] Kolarich, D., Lepenies, B. & Seeberger, P. H. Glycomics, glycoproteomics and the immune system. *Curr. Opin. Chem. Biol.* **16**, 214-220 (2012).

[5] Antonopoulos, A., North, S. J., Haslam, S. M. & Dell, A. Glycosylation of mouse and human immune cells: insights emerging from *N*-glycome analyses. *Biochem. Soc. Trans*. **39,** 1334-1340 (2011).

[6] Devakumar, A., Mechref, Y., Kang, P., Novotny, M. V. & Reilly J. P. Identification of isomeric *N*-glycan structures by mass spectrometry with 157 nm laser-induced photofragmentation. *J. Am. Soc. Mass Spectrom*. **19,** 1027–1040 (2008).

[7] Adibekian, A. *et al*. Comparative bioinformatics analysis of the mammalian and bacterial glycomes. *Chem*. *Sci*. **2**, 337-344 (2011).

[8] Rini, J. M., Esko, J. D. & Varki, A. *Essentials of glycobiolgy 2nd ed*. Ch. 5 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2008).

[9] Gerken, T. A. *et al*. Emerging paradigms for the initiation of mucin-type protein *O*-glycosylation by the polypeptide GalNAc transferase family of glycosyltransferases. *J. Biol. Chem.* **286,** 14493-14507 (2011).

[10] Laurent, N. *et al*. SPOT synthesis of peptide arrays on self-assembled monolayers and their evaluation as enzyme substrates. *ChemBioChem.* **9,** 2592-2596 (2008).

[11] Laurent, N. *et al*. Enzymatic glycosylation of peptide arrays on gold surfaces. *ChemBioChem.* **9,** 883-887 (2008).

[12] Errey, J. C., Mukhopadhyay, B., Kartha, R. & Field, R. A. Flexible enzymatic and chemo-enzymatic approaches to a broad range of uridine-diphospho-sugars. *Chem. Commun*., 2796-2707 (2004).

[13] Rejzek, M., et al. Chemical synthesis of UDP-Glc-2,3-diNAcA, a key intermediate in cell surface polysaccharide biosynthesis in the human respiratory pathogens *B. pertussis* and *P. aeruginosa*. *Org. Biomol. Chem.* **7**, 1203-1210 (2009).

[14] Su, J. & Mrksich, M. Using mass spectrometry to characterize self-assembled monolayers presenting peptides, proteins, and carbohydrates. *Angew. Chem. Int. Ed.* **41,** 4715-4718 (2002).

[15] Wandall, H. H. *et al*. Substrate specificities of three members of the human UDP-*N*-acetyl-alpha-D-galactosamine:polypeptide *N*-acetylgalactosaminyltransferase family, GalNAc-T1, -T2, and -T3. *J. Biol. Chem.* **272,** 23503-23514 (1997).

[16] Ramakrishnan, B., Boeggeman, E. & Qasba P. K. Novel method for *in vitro O*-glycosylation of proteins: application for bioconjugation. *Bioconjug. Chem.* **18,** 1912-1918 (2007).

[17] Jung, E., Gooley, A. A., Packer N. H., Karuso P. & Williams, K. L. Rules for the addition of *O*-linked *N*-acetylglucosamine to secreted proteins in Dictyostelium discoideum - *in vivo* studies on glycosylation of mucin Muc 1 and Muc 2 repeats. *Eur. J. Biochem.* **253,** 517-524 (1998).

[18] Previato, J. O. *et al*. Biosynthesis of O-*N*-acetylglucosamine-linked glycans in Trypanosoma cruzi. *J. Biol. Chem.* **273,** 14982-14988 (1998).

[19] Thoden, J. B., Wohlers, T. M., Fridovich-Keil, J. L. & Holden, H. M. Human UDP-galactose 4-epimerase. Accommodation of UDP-*N*-acetylglucosamine within the active site. *J. Biol. Chem.* **276,** 15131-15136 (2001).

[20] Jansson, P. -E., Kenne, L. & Widmalm, G. Computer-assisted structural analysis of polysaccharides with an extended version of casper using ^{1}H- and ^{13}C-N.M.R. data. *Carbohydr. Res.* **188,** 169-191 (1989).

[21] Lundborg, M. & Widmalm, G. Structural analysis of glycans by NMR chemical shift prediction. *Anal. Chem.* **83,** 1514-1517 (2011).

[22] Šardzík, R. *et al*. Chemoenzymatic synthesis of *O*-mannosylpeptides in solution and on solid phase. *J. Am. Chem. Soc.* **134,** 4521-4524 (2012).

[23] Gagneux, P. & Varki, A. Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology* **9,** 747-755 (1999).

[24] Ercan, A. & West, C. M. Kinetic analysis of a golgi UDP-GlcNAc:polypeptide-Thr/Ser *N*-acetyl-α-glucosaminyltransferase from *Dictyostelium*. *Glycobiology* **15,** 489-500 (2005).

[25] Taniguchi, N., Honke, K. & Fukuda, M. *Handbook of glycosyltransferases and related genes* Ch. 9-14 (Springer-Verlag, Tokyo, 2003).

[26] Lau, K. S. *et al*. Complex *N*-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation. *Cell* **129,** 123-134 (2007).

[27] Creese, A. J. & Cooper, H. J. Separation and identification of isomeric glycopeptides by high field asymmetric waveform ion mobility spectrometry. *Anal. Chem.* **84,** 2597-2601 (2012).

[28] Li, H. *et al.* Resolving structural isomers of monosaccharide methyl glycosides using drift tube and travelling wave ion mobility mass spectrometry. *Anal. Chem.* **84,** 3231-3239 (2012).

[29] Li,H., Bendiak, B., Siems, W. F., Gang, D. R. & Hill, H. H. Carbohydrate structure characterization by tandem ion mass spectroscopy (IMMS)^{2}. *Anal. Chem.* **85,** 2760-2769 (2013).

[30] Pagel, K. & Harvey, D. J. Ion mobility mass spectrometry of complex carbohydrates - Collision cross sections of sodiated N-linked glycans. *Anal. Chem.* DOI: 10.1021/ac400403d

[31] Hurtado-Guerrero, R., Dorfmueller, H. C. & van Aalten, D. M. Molecular mechanisms of *O*-GlcNAcylation. *Curr. Opin. Struct. Biol.* **18,** 551-557 (2008).

[32] Wyttenbach, T., Bushnell, J. E. & Bowers, M. T. Salt Bridge Structures in the Absence of Solvent? The Case for the Oligoglycines. *J. Am. Chem. Soc.* **120,** 5098-5103 (1998).

[33] Zaia, J. Mass spectrometry of oligosaccharides. *Mass Spectrom. Rev.* **23,** 161-227 (2004).

[34] Borsdorf, H. & Eiceman, G. A. Ion mobility spectrometry: principles and applications. *Appl. Spectrosc. Rev.* **41,** 323-375 (2006).

[35] Chawner, R. *et al.* QconCAT standard for calibration of ion mobility-mass spectrometry systems. *J. Proteome Res.* **11**, 5564−5572 (2012).

[36] Muller, R., Jenny, A. & Stanley, P. The EGF repeat-specific O-GlcNAc-transferase Eogt interacts with Notch signaling and pyrimidine metabolism pathways in *Drosophila*. *PLoS ONE* **8,** e62835 (2013)

[37] Domon, B. & Costello, C. E. Structure elucidation of glycosphingolipids and gangliosides using high-performance tandem mass spectrometry. *Biochemistry* **27,** 1534-1543 (1988).