# How and why plants and human N-glycans are different: Insight from molecular dynamics into the "glycoblocks" architecture of complex carbohydrates

Fogarty CA[‡], Harbison AM[‡], Dugdale AR and Fadda E[*]
Department of Chemistry and Hamilton Institute, Maynooth University, Maynooth, Kildare, Ireland

## Abstract

N-glycosylation is one of the most abundant and diverse post-translational modifications of proteins, implicated in protein folding and structural stability, and mediating interactions with receptors and with the environment. All N-glycans share a common core from which linear or branched arms stem from, with functionalization specific to different species and to the cells' health and disease state. This diversity generates a rich collection of structures, all diversely able to trigger molecular cascades and to activate pathways, which also include adverse immunogenic responses. These events are inherently linked to the N-glycans 3D architecture and dynamics, which remains for the large part unresolved and undetected because of intrinsic structural disorder. In this work we use molecular dynamics (MD) simulations to provide insight into N-glycans 3D structure by analysing the effects of a set of very specific modifications found in plants and invertebrate N-glycans, which are immunogenic in humans. We also compare these structural motifs and combine them with mammalian N-glycans motifs to devise strategies for the control of the N-glycan 3D structure through sequence. Our results suggest that the N-glycans architecture can be described in terms of the local spatial environment of groups of monosaccharides. We define these "glycoblocks" as self-contained 3D units, uniquely identified by the nature of the residues they comprise, their linkages and structural/dynamic features. This alternative description of glycans 3D architecture can potentially lead to an easier prediction of sequence-to-structure relationships in complex carbohydrates, with important implications in glycoengineering design.

**Keywords:** N-glycans, xylose, fucose, glycoblocks, molecular dynamics, complex carbohydrates, molecular recognition

## Introduction

[‡] Both authors have contributed equally to the work
[*] To whom correspondence should be addressed to: elisa.fadda@mu.ie

Complex carbohydrates (or glycans) are an essential class of biomolecules, directly implicated in the cell's interactions with its environment, facilitating communication and infection[1, 2]. These processes are often initiated by molecular recognition involving carbohydrate-binding proteins (lectins) or by glycan-glycan interactions[1, 3-5], all events that hinge on specific structural and dynamic features of the glycans. This makes the 3D complementarity of the glycans architecture the key towards the success of these processes and an essential piece of information for us to have in order to understand glycan recognition. Because of their chemical nature, glycans are intrinsically flexible and highly dynamic at room temperature, thus their characterization through experimental structural biology methods is hardly straightforward even in cryogenic environments[6]. As an additive layer of difficulty, glycosylation is only indirectly dependent on the genome, which often results in a micro- (or macro-)heterogeneity of glycan sequences at specific sites[7]. These are very difficult to resolve, requiring high levels of expertise and complex orthogonal approaches[7-10]. Within this complex framework, the contribution of glycoinformatics tools and databases represents an essential resource to advance glycomics[11-15], while molecular simulations fit in very well as complementary and orthogonal techniques to support and advance structural glycobiology research. Indeed, current high performance computing (HPC) technology allows us to study realistic model systems[16, 17] and to reach experimental timescales[18], so that computing can now contribute as one of the leading research methods in structural glycobiology.

One of the most interesting and remarkably challenging areas in glycoscience research that HPC simulations can address is the study of the links between glycans sequence and 3D structure. This direct relationship is a well-recognized and broadly accepted concept in proteins structural biology, according to which the amino acid sequence dictates the functional 3D fold and its stability. However, the same notion is not generally invoked when discussing other biopolymers or complex carbohydrates. In the specific case of glycans, the structural complexity, in terms of the diversity of monosaccharides, the linkages' stereochemistry and the branched scaffolds, makes the already difficult case even more intricate. Nevertheless, the fact that glycoforms follow recurrent sequence patterns, clearly suggests that the glycans 3D structure is also non-random and very likely sequence-dependent. We use computer modelling to gain insight into these relationships and to define a framework to understand how subtle modifications to the glycans sequence can alter their 3D

structure and conformational dynamics, ultimately regulating their recognition[19]. In this work we use molecular dynamics (MD) simulations to analyse the effects of the inclusion of motifs typically found in plants and invertebrates N-glycans and immunogenic in mammals[20-23]. More specifically, we investigate how core α(1-3)-linked fucose (Fuc) and β(1-2)-linked xylose (Xyl) affect the structure and dynamics of plants N-glycoforms[23] and of hybrid constructs with mammalian N-glycoforms[24].

At first glance plants protein N-glycosylation[23] is quite similar to the one of higher species[25], carrying the distinctive trimannose core (Man3), which can be further functionalised with β(1-2) linked GlcNAc residues on the arms. As a trademark feature, shown in **Figure 1**, plants N-glycans can be functionalised with β(1-2)-Xyl linked to the central mannose and with core α(1-3)-Fuc, instead of the α(1-6)-Fuc found in eukaryotic complex N-glycans. Additionally, the arms can be further functionalised with terminal galactose (Gal) in β(1-3) instead of β(1-4)[23], which forces the addition of fucose in the α(1-4) position of the GlcNAc, and results in the occurrence of Lewis A (LeA) instead of Lewis X (LeX) terminal motifs on the arms[23, 26]. In a previous study, we characterized through extensive MD sampling the structure and dynamics of complex biantennary N-glycans commonly found in human IgGs Fc region[24] and indicated clear sequence-to-structure relationships in complex N-glycans, especially in the context the dynamics of the (1-6) arm. More specifically, the outstretched (open) conformation of the (1-6) arm is progressively less populated as the functionalization of the arm grows, i.e. from 85% in Man3, to 52% in (F)A2, (F)A2[3]G1, and (F)A2[3]G1S1, where the (F) indicates the presence or absence of α(1-6) core fucosylation, to 24% in all structures with (1-6) arm terminating with Gal-β(1-4)-GlcNAc or Sia-α(2-6)-Gal-β(1-4)-GlcNAc, irrespective of the functionalization of the (1-3) arm[24]. As a practical implication of these results, positional isomers, such as (F)A2[3]G1 and (F)A2[6]G1, have different conformational propensities, the latter with a much lower population of outstretched (1-6) arm and therefore quite different 3D average structures, which ultimately explains their differential recognition in glycan arrays[27] and the known difficulties in sialylating the (1-6) arm by ST6-Gal1, relatively to the (1-3) arm[28]. Also, the different 3D conformational propensity of the arms in function of sequence can have important consequences in terms of the N-glycans biosynthesis and biodegradation[29]. As an additional interesting point, we found that the folding of the (1-6) arm over the chitobiose region is completely independent of core α(1-6) fucosylation[24].

In this work we discuss how core α(1-3)-Fuc and β(1-2)-Xyl regulate the conformational propensity of the (1-6) arm to push a predominantly outstretched (open) conformation when the arms are functionalized with terminal β(1-3)-Gal. Within this framework, we explored the possibility of integrating these plant-specific motifs in the context of mammalian sequences as an exploratory strategy towards the design of N-glycans with the desired 3D structure. Finally, we discuss these findings within a framework where the different N-glycoforms can be represented as a combination of spatial self-contained units, named "glycoblocks", rather than in terms of monosaccharides and linkages. We find that this approach helps our understanding of N-glycans architecture in terms of equilibrium structures and relative populations and also of how specific modification affect molecular recognition.

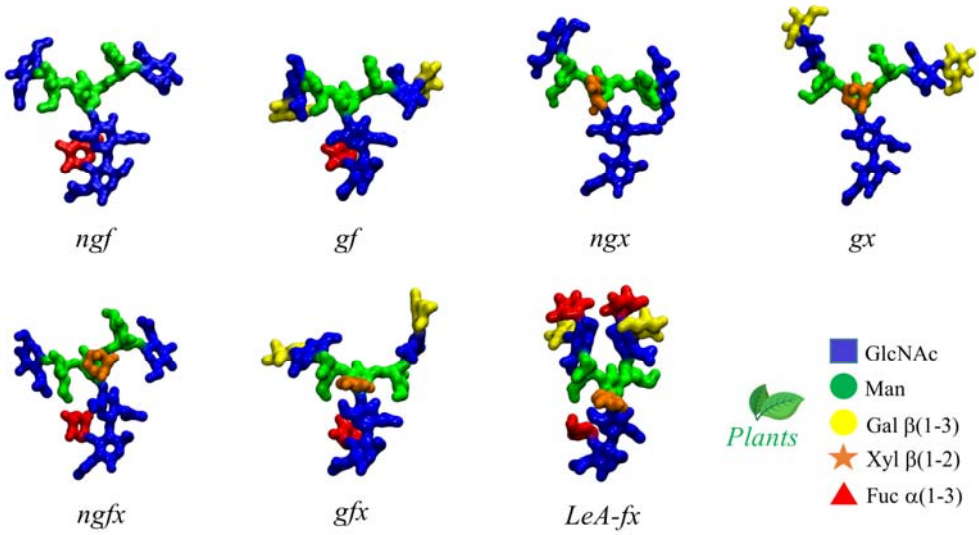



**Figure 1.** Representative structures of the plant N-glycans studied in this work with corresponding nomenclature. The letters *f*, *x*, and *g* indicate the presence of Fuc, Xyl and β(1-3) Gal, respectively, and *ng* the absence of β(1-3) Gal. *LeA* stands for Lewis A antigen. The monosaccharides colouring follows the SFNG nomenclature. The plants N-glycan characteristic linkages are indicated in the legend. Rendering was done with VMD (https://www.ks.uiuc.edu/Research/vmd/).

## Computational Methods

All starting structures were generated with the GLYCAM Carbohydrate Builder (http://www.glycam.org). For each sequence we selected the complete set of torsion angle values obtained by variation of the 1-6 dihedrals, namely the three *gg*, *gt* and *tg* conformations for each 1-6 torsion. The topology file for each structure was obtained using *tleap*[30], with parameters from the GLYCAM06-j1[31] for the carbohydrate atoms and with

TIP3P for water molecules[32]. All calculations were run with the AMBER18 software package[30] on NVIDIA Tesla V100 16GB PCIe (Volta architecture) GPUs installed on the HPC infrastructure *kay* at the Irish Centre for High-End Computing (ICHEC). Separate production steps of 500 ns each were run for each rotamer (starting system) and convergence was assessed based on conformational and clustering analysis. Simulations were extended, if the sampling was not deemed as fully converged. All trajectories were processed using *cpptraj*[30] and visually analysed with the Visual Molecular Dynamics (VMD) software package[33]. Backbone Root Mean Square Deviation (RMSD) and torsion angles values were measured using VMD. A density-based clustering method was used to calculate the populations of occupied conformations for each torsion angle in a trajectory and heat maps for each dihedral were generated with a kernel density estimate (KDE) function. Statistical and clustering analysis was done with the R package and data were plotted with RStudio (www.rstudio.com). Further details on the simulation set-up and running protocol are included as Supplementary Material.

## Results

**Core α(1-3) fucose in plant N-glycans**. One distinctive features of plants N-glycans is the core α(1-3) fucosylation, relative to the core α(1-6)-Fuc found in mammalian N-glycans[23, 24]. To understand the effects on the 3D structure of this modification, we have considered two biantennary systems, one terminating with β(1-2)-GlcNAc on both arms (*ngf*) and the other with terminal β(1-3)-Gal on both arms (*gf*), see **Figure 1**. Core α(1-3)-Fuc occupies a stable position (87.3%), see **Table S.1,** supported by the stacking with the β(1-4)-linked GlcNAc of the chitobiose in a "closed" conformation, which resembles the one observed for LeX[34]. As shown in **Table S.1**, this spatial arrangement imposes a 20° rotation of the GlcNAc-β(1-4)-GlcNAc linkage, relative to the α(1-6) core fucosylated or non-fucosylated chitobiose, where the average psi value is -127.8° (14.8)[24], but doesn't affect the structure of the linkage to the central mannose. The overall dynamics of the N-glycan core remains quite rigid. The slight torsion of the GlcNAc-β(1-4)-GlcNAc linkage imposed by the α(1-3)-Fuc has a dramatic effect on the conformational dynamics of the (1-6) arm, which is found predominantly in an outstretched (66%) conformation, rather than folded over (34%). Interestingly, in the case of α(1-6) core fucosylated N-glycans, and with double fucosylation as discussed later on, the equilibrium of the (1-6) arm was the exact opposite[24]. To note, the folded (1-6) arm

conformation can be over the chitobiose, i.e. 'front fold', see **Figure 2**, or in. a 'back fold' to interact with the α(1-3)-Fuc. As shown in **Table S.1**, the equilibrium of the (1-3) arm is not affected by core α(1-3)-Fuc. Because of the rotated position of the chitobiose, the addition of a terminal β(1-3)-Gal in the *gf* N-glycan pushes the equilibrium towards an outstretched (1-6) arm even further, with the open conformation populated at 72%, see **Table S.2**.

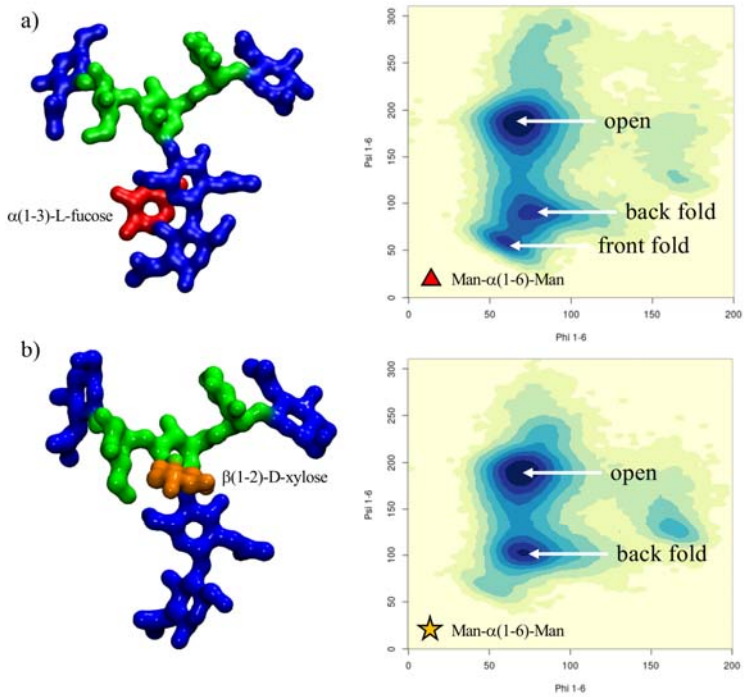



**Figure 2.** A representative structure of the non-galactosylated N-glycan with α(1-3)-linked core fucose is shown in panel a), with on the right-hand side the corresponding heat map showing the conformations accessible to the (1-6) arm in terms of the phi/psi torsion angles. A representative structure of the non-galactosylated N-glycan with β(1-2)-linked xylose is shown in panel b), with on the right-hand side a heat map showing the conformations accessible to the (1-6) arm in terms of the phi/psi torsion angles. The monosaccharides colouring follows the SFNG nomenclature. The structure rendering was done with VMD and the graphical statistical analysis with *RStudio* (www.rstudio.com).

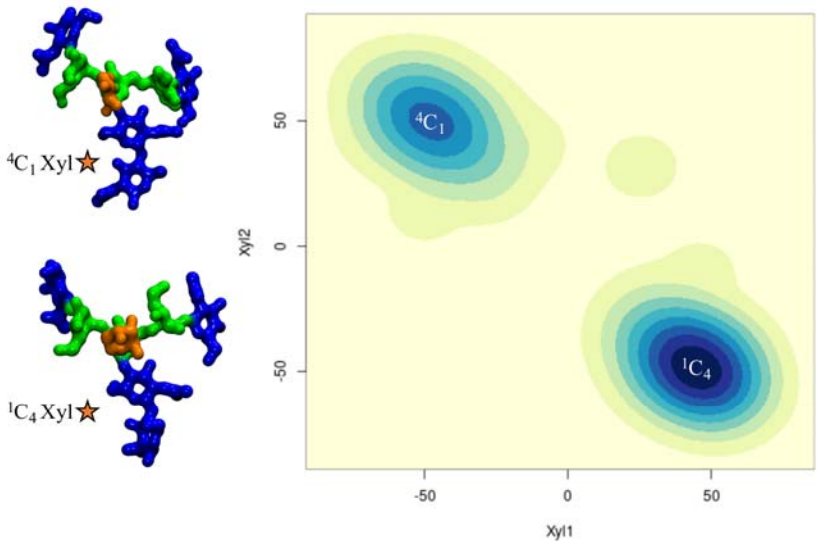



**Figure 3.** β-D-xylose ring pucker analysis over 3 μs of cumulative MD sampling of the *ngx* N-glycan. The two snapshots on the right-hand side are representative *ngx* conformations corresponding to the two different ring puckers. The Xyl1 and Xyl2 axis labels refer to the torsion angles C1C2C3C4 and C2C3C4C5, respectively. The monosaccharides colouring follows the SFNG nomenclature. The structure rendering was done with VMD and the graphical statistical analysis with *RStudio* (www.rstudio.com).

**β(1-2) xylose in plant N-glycans**. Because the β(1-2)-Xyl sits in between the two arms, its presence greatly affects both of their dynamics. Indeed, as a result of steric hindrance the (1-3) arm is much more rigid relative to non-xylylated species, see **Table S.3**, losing its "two conformer" dynamics characteristic of the biantennary mammalian N-glycans[24] and also retained in plant N-glycans with only α(1-3)-Fuc discussed above, see also **Tables S.1** and **S.2**. In regards to the (1-6) arm, as shown in **Figure 2** panel b), the presence of a β(1-2)-Xyl has a very similar effect as the α(1-3)-Fuc, pushing the equilibrium towards an open conformation. To note, in the presence β(1-2)-Xyl, the (1-6) arm cannot fold over the chitobiose core in a 'front fold' either, because of steric hindrance. Also, similarly to the α(1-3) fucosylated glycans, the stability of the open structure is slightly increased when the arm is further functionalized with terminal β(1-3)-Gal, see **Table S.4**. As an additional interesting feature, through the cumulative 3 μs MD sampling, the xylose ring repeatedly inverts its conformation from the all equatorial $^4C_1$ chair, to the $^1C_4$ chair, where all hydroxyl groups are axial, see **Figure 3**. This transition may be energetically facilitated by the hydrogen bonding interaction xylose is able to form when in a $^1C_4$ chair with the α(1-6)-Man, which may compensate for the steric compression, making the $^1C_4$ chair the highest populated conformer at 76% within an N-glycan scaffold. To note, both experimental and *ab-initio* theoretical

studies[35-37] have shown that $^{1}C_{4}$ chair is energetically accessible in isolated β-D-Xyl at room temperature in different dielectric conditions.

**Core α(1-3) fucose and β(1-2) xylose in plant N-glycans.** The presence of both α(1-3)-Fuc and β(1-2)-Xyl brings in the characteristic features highlighted earlier in the analysis of the structures with either α(1-3)-Fuc or β(1-2)-Xyl. Indeed, we see here again the 20° rotation of the chitobiose GlcNAc-β(1-4)-GlcNAc *psi* angle caused by the stacking of the α(1-3)-Fuc to the chitobiose β(1-4)-GlcNAc and the conformational restraints imposed by the β(1-2)-Xyl on the (1-3) arm, see **Table S.5**. We also observed that both α(1-3)-Fuc and β(1-2)-Xyl push the (1-6) arm equilibrium towards an open conformation, which is also the case when both are present in the *ngfx* N-glycan and to an even higher degree, i.e. 87%, in the *gfx* N-glycan, when both arms are functionalized with terminal β(1-3)-Gal, see **Table S.6**. One feature specific to the *ngfx* N-glycan is the higher flexibility of the core Man-β(1-4)-GlcNAc linkage, which allows for the rotation of the trimannose group relative to the chitobiose core. This conformation was accessible, but only populated around 2% when either β(1-2)-Xyl or α(1-3)-Fuc are present, see **Tables S.1** to **S.4**. When both fucose and xylose are present, the population of the rotated trimannose reaches above 20%, see **Table S.5**, which can be considered a synergistic effect as this conformation is stabilized by a hydrogen bonding network involving the core fucose, the GlcNAc on the (1-6) arm and the xylose, as shown in **Figure S.1**. Such folding event has been observed as a stable conformation in two independent simulations. To note, the functionalization of the arms to include terminal β(1-3)-Gal reduces the occurrence of this event down to around 5%, see **Table S.6**.

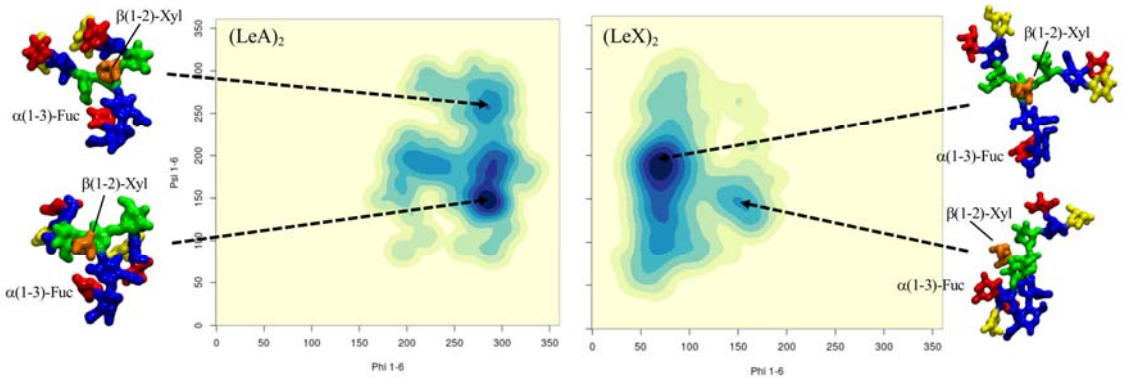



**Figure 4.** Comparison of the different conformational equilibria of the (1-6) arm in a core α(1-3)-Fuc β(1-2)-Xyl A2 N-glycan with terminal LeA and LeX groups on the left- and right-hand side, respectively. Representative structures from 1.5 μs MD sampling of each system are shown to illustrate the conformations corresponding to the different minima. The

monosaccharides colouring follows the SFNG nomenclature. The structure rendering was done with VMD and the graphical statistical analysis with *RStudio* (www.rstudio.com).

**Terminal LeA and LeX motifs in plant N-glycans**. To understand how an increased complexity on the arms would affect the dynamics of the α(1-3) fucosylated and β(1-2) xylylated N-glycans, we considered the functionalization with terminal LeA antigens present in plants N-glycans[26] and with LeX for comparison. As expected[34] the LeA and LeX structures are quite rigid, see **Tables S.7** and **S.15**, and remain in what is known as the "closed" conformation throughout the 1.5 μs cumulative sampling time for each system. One interesting observable is that the branching introduced by functionalizing the terminal GlcNAc residues with α(1-4)-Fuc and β(1-3)-Gal, i.e. LeA, promotes the interaction between the two arms, which is not observed when the arms are linear, neither here for plants N-glycans, nor for mammalian IgG-type complex biantennary N-glycans[24]. The interaction between the arms is promoted by the ability to form complex hydrogen bonding networks, which in this specific case, may also involve the central xylose. As outcomes of the complex interaction the branched arms can establish, the equilibrium of the (1-6) arm is restrained in conformations previously not significantly populated, see **Figure 4** and **Table S.7**, and the GlcNAc-β(1-2)-Man linkage in both arms is remarkably flexible, which is not the case when the arms are linear. Although not natural in plants, to check the corresponding symmetry, we built a core α(1-3)-Fuc and β(1-2)-Xyl N-glycan with terminal LeX on both arms and remarkably, as shown in **Figure 4** and **Table S.15**, the dynamics of the (1-6) arm is completely different. Contrary to the N-glycan with terminal LeA groups, the two arms with LeX are not interacting and the (1-6) arm is predominantly (90%) in an extended (open) conformation, while the closed conformation, which accounts for the remaining 10% is achieved through a rotation around the core Man-β(1-4)-GlcNAc. The lack of interaction between the arms is due to the inability to establish the same stable hydrogen bonding network due to the non-complementary position of the deoxy-C6 of the fucose in LeX relative to LeA.

**Hybrid N-glycans.** To understand how characteristic plant N-glycan motifs can affect the structure of mammalian N-glycoforms, we have designed and analysed the dynamics of a set of hybrid systems. In particular, we were interested in the effect of the addition of β(1-2)-Xyl and α(1-3)-Fuc to (F)A2G2 N-glycans scaffolds in terms of potential alteration of the (1-6) arm dynamics.

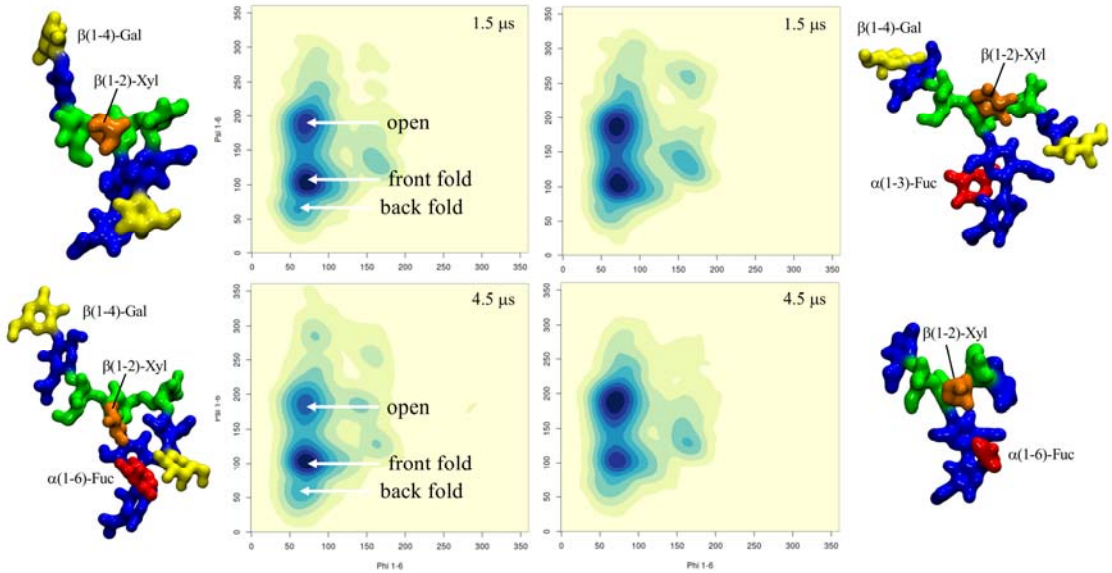



**Figure 5.** Conformational analysis of the (1-6) arm in four hybrid N-glycoforms, β(1-2)-xylylated A2G2 (top-left), β(1-2)-xylylated FA2G2 (bottom-left), β(1-2)-xylylated α(1-3)-core fucosylated A2G2 (top-right) and β(1-2)-xylylated FA2 (bottom-right). The predominant conformations are indicated in the top- and bottom-left heat maps for simplicity. The simulation time relative to each system is indicated in the top right corner of each heat map. The monosaccharides colouring follows the SFNG nomenclature. The structure rendering was done with VMD and the graphical statistical analysis with s*RStudio* (www.rstudio.com).

**β(1-2)-xylylated mammalian N-glycans.** Unlike the case of plants N-glycans, the presence of β(1-2)-Xyl hinders but does not completely prevent the (1-6) arm from folding over when the terminal galactose is β(1-4)-linked, as folding over the chitobiose can be stabilized by stacking, see **Figure 5** and **Table S.8**. The folded conformation with a median *psi* value of 103.5° (± 11.3) is 20° from the average value of 82.9° calculated for the non-xylylated (mammalian) counterpart[24], so slightly distorted, and its population reduced from 74% to 57%. Nevertheless, the closed conformation is still the predominant form, even with β(1-2)-Xyl. The presence of α(1-6)-linked core fucose to create a β(1-2)-xylylated FA2G2 brings in yet another change. As shown in **Figure 5** and **Table S.9**, α(1-6)-Fuc and β(1-2)-Xyl are in an optimal conformation to support the closed (folded) (1-6) arm, by stacking of the terminal galactose by fucose and hydrogen bonding by xylose. Within this context the closed (1-6) arm is the highest populated conformer at 70.0% over 4.5 µs of cumulative sampling of this system. To note that the conformation of the α(1-6)-linked core fucose is the same as the one seen in mammalian N-glycans[24], which on its own we have seen is not enough to affect the (1-6) arm equilibrium, see **Table S.9**. The interaction of the α(1-6)-Fuc with the terminal β(1-4)-Gal is essential to promote the closed conformation of the (1-6) arm as demonstrated

by the results obtained for the xylylated FA2 systems, which recovers a conformational propensity similar to the non-fucosylated, xylylated A2G2, see **Figure 5** and **Tables S.8** and **S.10**.

**α(1-3)-fucosylated mammalian N-glycans**. Because of its orientation tucked "behind" the chitobiose core defined in the context of plants N-glycans earlier, the effect of core α(1-3)-Fuc on the (1-6) arm equilibrium within an A2G2-xylylated scaffold is not as significant as α(1-6)-Fuc. As shown in **Figure 5** and **Table S.11**, this lack of direct effect is demonstrated by the recovery of the same equilibrium as the non-fucosylated A2G2-xylylated system. The dynamics of the chitobiose core is very similar to the one determined for the corresponding plant N-glycan. To analyse the effect of core α(1-3) fucosylation without β(1-2)-Xyl, we have looked at two A2G2 hybrid systems, one with only α(1-3)-linked fucose and one with both core α(1-3)- and α(1-6)-linked fucose, a characteristic "double-fucose" glycosylation motif of insect cells. As shown in **Table S.12** unlike in plants N-glycans, the α(1-3)-Fuc alone does not affect the A2G2 (1-6) arm equilibrium[24], as the folding of the (1-6) arm with terminal β(1-4)-Gal is not obstructed by the rotation of the chitobiose core imposed by the α(1-3)-Fuc position. When both α(1-3)- and α(1-6)-linked fucoses are present the (1-6) arm with terminal β(1-4)-Gal is predominantly folded (closed) at 85%, see **Figure 6** and **Table S.13**, which is higher than in the absence of α(1-3)-Fuc[24]. Indeed, the latter can actively contribute in stabilizing the interaction with the terminal β(1-4)-Gal of the folded (1-6) arm. We also observed interesting events, one representing 10% of 2 μs as indicated by the values of the GlcNAc-β(1-4)-GlcNAc torsion, where the GlcNAc is stacked in between the two fucose residues and another one, contributing to 18% of the simulation time, 14% when the system is also xylylated, in which the GlcNAc ring transitions from $^{4}C_{1}$ to $^{1}C_{4}$ allowing the two fucose to stack, see **Tables S.13** and **S.14** and **Figure S.2**.

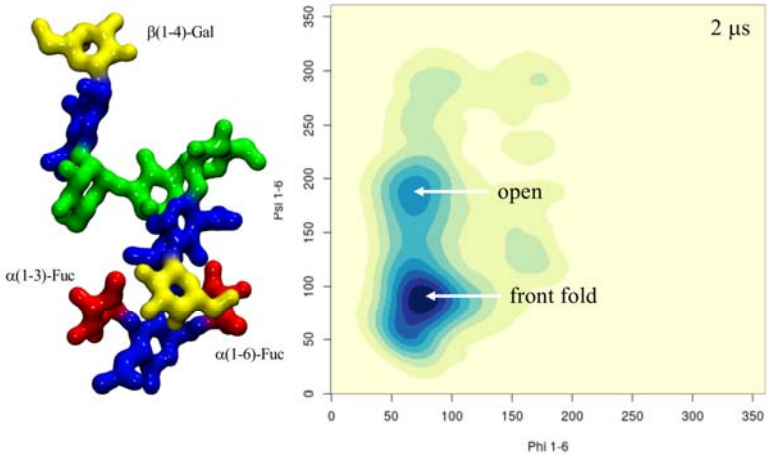


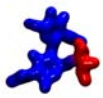
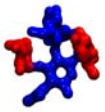
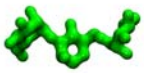
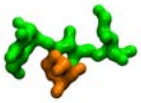
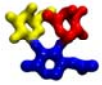

**Figure 6.** Conformational equilibrium of the (1-6) arm in terms of phi/psi torsion angle values for the α(1-3)-fucosylated FA2G2 N-glycoform. The structure with the folded (1-6) arm where the terminal β(1-4)-Gal interacts with both fucose residues is shown on the left-hand side. The monosaccharides colouring follows the SFNG nomenclature. The structure rendering was done with VMD and the graphical statistical analysis with *RStudio* (www.rstudio.com).

## Discussion

Differences and similarities in N-glycans sequences are highly cell-specific as well as important indicators of health and disease states[1, 38]. Exogenous N-glycans motifs can be quite subtle, yet trigger profound differences in terms of molecular recognition[19, 27] and dangerous immunogenic responses[20-22]. In this work we have analysed the effects on the N-glycans structure and dynamics of two motifs in particular, namely β(1-2)-Xyl and core α(1-3)-Fuc, common in plants and invertebrates, but completely absent in mammalian N-glycans. Within the context of plant N-glycans, which have a terminal β(1-3)-Gal, rather than β(1-4)-Gal, both β(1-2)-Xyl and α(1-3)-Fuc contribute independently in promoting an outstretched (open) conformation of the (1-6) arm because of steric hindrance of the xylose and of the rotation forced upon the chitobiose core by the α(1-3)-Fuc. The latter is not an obstruction for the folding of a β(1-4)-Gal terminated (1-6) arm, as we have seen in the hybrid N-glycans constructs. Also, the analysis of the structure and dynamics of the LeA terminating plant N-glycans showed that the specific branching and spatial orientation of the motif allowed for a stable interaction between the arms, which is not observed in complex N-glycans with a linear functionalization of the arms[24]. Notably, the same hydrogen bonding network between the arms cannot be established when the same N-glycan terminates with LeX, because of the non-complementary position of the α(1-3)-Fuc deoxy-$C_6$.

The analysis of all these different complex N-glycoforms clearly shows that every modification, addition or removal of a specific motif, can greatly affect the 3D architecture of the N-glycan, thus its complementarity to a receptor. However, these effects are rather complex to understand or to predict, if we think of the N-glycans 3D structure in terms of sequence of monosaccharides, a view that stems from the way we think about proteins. Our results show that the main effect of all functionalizations is actually local. For example, the core α(1-3)-Fuc forces a rotation of the chitobiose, a degree of freedom very lowly populated otherwise; meanwhile, β(1-2)-Xyl restricts the flexibility of the trimannose core and occupies its centre. Within this framework, the 3D structural and dynamics features of the N-glycoforms can be rationalized by discretizing their architecture in terms 3D units, or "glycoblocks", that group monosaccharides and their linkages within their immediate spatial vicinity, e.g. the core α(1-3)-Fuc and the chitobiose which structure it has modified. A list of the glycoblocks that we have identified with the corresponding descriptors of their 3D features are listed in **Table 1**. The whole N-glycan 3D architecture, in terms of the structures accessible and their conformational propensity, can be then described through the combination of these glycoblocks, together with the knowledge of their dynamic properties and flexibility. Also, consideration of these glycoblocks as spatial units can be useful to understand recognition by lectins and antibodies, which is often affected primarily by the targeted monosaccharide's immediate vicinity and by its accessibility within a specific glycoform. For example, if we consider the 3D structure of the β(1-2)-Xyl Man3 glycoblock *vs.* the Man3 without Xyl, we can understand how the β(1-2)-Xyl position within that unit negates binding to DC-SIGN lectins[19], see **Figure S.3** panels a) and b). Additionally, we can see that the slight rotation on the chitobiose imposed by the core α(1-3)-Fuc does not prevent recognition and binding, see **Figure S.3** panel c).

**Table 1.** List of 3D structural units of monosaccharides (glycoblocks) that regulate the 3D architecture and dynamics of complex biantennary N-glycans from plants and invertebrate sources and hybrid mammalian constructs. The SFNG representation of each glycoblock is indicated in the first column from the left, 3D structures from the highest populated conformers are shown in the second column, rendered with VMD. A brief summary of the conformational features of each glycoblock and the characteristic linkage or its effect on the (1-6) arm conformation are indicates in the last two columns, respectively.

| Glycoblock | Unit 3D structure | Dominant motif descriptor | Characteristic glycosidic linkage |
|---|---|---|---|
| | | Very stable conformation, with a low degree of flexibility around the equilibrium structure | **GlcNAc β(1-4) GlcNAc**<br>phi = -78.7 (11.1)/100 psi = -130.8 (15.7)/99 |
| | | The α(1-3)-Fuc stacking interaction imposes a 20° rotation around the chitobiose glycosidic linkage and further stabilizes the chitobiose structure | **GlcNAc β(1-4) GlcNAc**<br>phi = -72.1 (8.3)/100 psi = -107.1 (7.6)/100 |
| |  | The longer glycosydic linkage provides the α(1-6)-Fuc with more degrees of freedom and does not alter the chitobiose structure | **GlcNAc β(1-4) GlcNAc**<br>phi = -77.8 (11.2)/100 psi = -127.8 (14.8)/100 |
| |  | This motif is structurally similar to the one with the α(1-3)-Fuc, which acts as the dominant restraint. The presence of the α(1-6)-Fuc increases flexibility | **GlcNAc β(1-4) GlcNAc**<br>phi = -73.9 (9.5)/100 psi = -106.4 (14.1)/90 |
| |  | Joint with differently flexible arms. The α(1-6) torsion has two dominant, different conformations accessible that can open and close the (1-6) arm. Populations depend on the attached glycoblock units | **Man α(1-6) Man** (with terminal β(1-4)-Gal)<br>open phi = 76.3 (15.0)/100 psi = -185.1 (22.1)/23<br>closed phi = 76.3 (15.0)/100 psi = 85.3 (15.8)/76 |
| |  | The addition of the β(1-2)-Xyl further restricts the (1-3) arm and hinders the closing forward of the (1-6) arm, which is only favored in (F)A2G2 hybrids | **Man α(1-6) Man** (with terminal β(1-3)-Gal)<br>open phi = 70.5 (10.4)/100 psi = -173.6 (16.3)/70<br>closed phi = 70.5 (10.4)/100 psi = 103.8 (12.7)/26 |
| | | This unit is rather rigid, while the orientation of the terminal Gal in β(1-3/4) is determinant for the interaction with the chitobiose in the closed (1-6) arm. These units in the two arms do not interact | **Gal β(1-3) GlcNAc**<br>phi = -72.9 (11.2)/100 psi = -124.0 (18.2)/96 |
| | | The Lewis A epitope is stable in its closed conformation and its presence greatly affects the (1-6) arm conformation because branching promotes association of the arms | **Man α(1-6) Man** (with terminal LeA and α(1-3)-Fuc)<br>open phi = -70.1 (5.7)/100 psi = -177.1 (16.3)/28<br>closed phi = -70.1 (5.7)/100 psi = 148.2 (7.4)/54 |
| | | This unit is rather rigid, while the orientation of the terminal Gal in β(1-3/4) is determinant for the interaction with the chitobiose in the closed (1-6) arm. These units in the two arms do not interact | **Gal β(1-4) GlcNAc**<br>phi = -76.2 (15.1)/100 psi = -125.7 (15.4)/97 |
| |  | The Lewis X epitope is stable in its closed conformation. Contrary to Lewis A its branching does not favor association of the arms because of the relative positions of the deoxy-$C_6$. | **Man α(1-6) Man** (with terminal LeX, core α(1-3)-Fuc and β(1-2)-Xyl )<br>open phi = 70.7 (9.2)/90 psi = -173.6 (13.4)/90<br>closed phi = 152.4 (12.9)/10 psi = 145.6 (12.3)/10 |

## Conclusions

In this work we used extensive sampling through MD simulations to study the effects on the N-glycan architecture of subtle, yet highly consequential modifications, namely core α(1-3)-Fuc and β(1-2)-Xyl[19]. These are part of standard N-glycoforms found in plants and invertebrates, but immunogenic in humans[21, 22, 26]. Our results show that these modifications can greatly affect the 3D structure of the N-glycan and its structural dynamics, therefore its selective recognition by lectin receptors and antibodies. The atomistic-level of detail information that the MD simulations provide us with, highlights that the effects of different functionalizations, in terms of monosaccharide types and linkages, are primarily local, affecting the immediate spatial vicinity of the monosaccharide within the N-glycan structure. Within this framework, we propose an alternative approach that can help describe and predict the architecture of N-glycans based on the combination of structural 3D units, or glycoblocks. Unlike a description based on monosaccharide sequence and linkages as two separate features, the transition to well-defined and self-contained units, integrating information on both monosaccharides and linkages, can help us rationalize and deconvolute the glycans structural disorder and ultimately understand more clearly the relationships between sequence and structure in complex carbohydrates.

## Acknowledgements

The Irish Centre for High-End Computing (ICHEC) is gratefully acknowledged for generous allocation of computational resources. EF and CAF gratefully acknowledge the Irish Research Council (IRC) for funding through the Government of Ireland Postgraduate Scholarship Programme. EF and AMH gratefully acknowledge the John and Pat Hume Doctoral Scholarship Programme at Maynooth University for funding.

References

1. Varki, A., Biological roles of glycans. *Glycobiology* **2017,** *27* (1), 3-49.
2. Stanley, P.; Taniguchi N; Aebi, M., N-Glycans. In *Essentials of Glycobiology, 3rd Edition*, 3rd ed.; Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press 2015-2017: 2017.
3. Cobb, B. A., The history of IgG glycosylation and where we are now. *Glycobiology* **2020,** *30* (4), 202-213.
4. Ferrara, C.; Grau, S.; Jäger, C.; Sondermann, P.; Brünker, P.; Waldhauer, I.; Hennig, M.; Ruf, A.; Rufer, A. C.; Stihle, M.; Umaña, P.; Benz, J., Unique carbohydrate-carbohydrate interactions are required for high affinity binding between FcgammaRIII and antibodies lacking core fucose. *Proc Natl Acad Sci U S A* **2011,** *108* (31), 12669-74.
5. Day, C. J.; Tran, E. N.; Semchenko, E. A.; Tram, G.; Hartley-Tassell, L. E.; Ng, P. S.; King, R. M.; Ulanovsky, R.; McAtamney, S.; Apicella, M. A.; Tiralongo, J.; Morona, R.; Korolik, V.; Jennings, M. P., Glycan:glycan interactions: High affinity biomolecular interactions that can mediate binding of pathogenic bacteria to host cells. *Proc Natl Acad Sci U S A* **2015,** *112* (52), E7266-75.
6. Nagae, M.; Yamaguchi, Y., Function and 3D structure of the N-glycans on glycoproteins. *Int J Mol Sci* **2012,** *13* (7), 8398-429.
7. Zacchi, L. F.; Schulz, B. L., N-glycoprotein macroheterogeneity: biological implications and proteomic characterization. *Glycoconj J* **2016,** *33* (3), 359-76.
8. De Leoz, M. L. A.; Duewer, D. L.; Fung, A.; Liu, L.; Yau, H. K.; Potter, O.; Staples, G. O.; Furuki, K.; Frenkel, R.; Hu, Y.; Sosic, Z.; Zhang, P.; Altmann, F.; Gru Nwald-Grube, C.; Shao, C.; Zaia, J.; Evers, W.; Pengelley, S.; Suckau, D.; Wiechmann, A.; Resemann, A.; Jabs, W.; Beck, A.; Froehlich, J. W.; Huang, C.; Li, Y.; Liu, Y.; Sun, S.; Wang, Y.; Seo, Y.; An, H. J.; Reichardt, N. C.; Ruiz, J. E.; Archer-Hartmann, S.; Azadi, P.; Bell, L.; Lakos, Z.; An, Y.; Cipollo, J. F.; Pucic-Bakovic, M.; Štambuk, J.; Lauc, G.; Li, X.; Wang, P. G.; Bock, A.; Hennig, R.; Rapp, E.; Creskey, M.; Cyr, T. D.; Nakano, M.; Sugiyama, T.; Leung, P. A.; Link-Lenczowski, P.; Jaworek, J.; Yang, S.; Zhang, H.; Kelly, T.; Klapoetke, S.; Cao, R.; Kim, J. Y.; Lee, H. K.; Lee, J. Y.; Yoo, J. S.; Kim, S. R.; Suh, S. K.; de Haan, N.; Falck, D.; Lageveen-Kammeijer, G. S. M.; Wuhrer, M.; Emery, R. J.; Kozak, R. P.; Liew, L. P.; Royle, L.; Urbanowicz, P. A.; Packer, N. H.; Song, X.; Everest-Dass, A.; Lattová, E.; Cajic, S.; Alagesan, K.; Kolarich, D.; Kasali, T.; Lindo, V.; Chen, Y.; Goswami, K.; Gau, B.; Amunugama, R.; Jones, R.; Stroop, C. J. M.; Kato, K.; Yagi, H.; Kondo, S.; Yuen, C. T.; Harazono, A.; Shi, X.; Magnelli, P. E.; Kasper, B. T.; Mahal, L.; Harvey, D. J.; O'Flaherty, R.; Rudd, P. M.; Saldova, R.; Hecht, E. S.; Muddiman, D. C.; Kang, J.; Bhoskar, P.; Menard, D.; Saati, A.; Merle, C.; Mast, S.; Tep, S.; Truong, J.; Nishikaze, T.; Sekiya, S.; Shafer, A.; Funaoka, S.; Toyoda, M.; de Vreugd, P.; Caron, C.; Pradhan, P.; Tan, N. C.; Mechref, Y.; Patil, S.; Rohrer, J. S.; Chakrabarti, R.; Dadke, D.; Lahori, M.; Zou, C.; Cairo, C.; Reiz, B.; Whittal, R. M.; Lebrilla, C. B.; Wu, L.; Guttman, A.; Szigeti, M.; Kremkow, B. G.; Lee, K. H.; Sihlbom, C.; Adamczyk, B.; Jin, C.; Karlsson, N. G.; Örnros, J.; Larson, G.; Nilsson, J.; Meyer, B.; Wiegandt, A.; Komatsu, E.; Perreault, H.; Bodnar, E. D.; Said, N.; Francois, Y. N.; Leize-Wagner, E.; Maier, S.; Zeck, A.; Heck, A. J. R.; Yang, Y.; Haselberg, R.; Yu, Y. Q.; Alley, W.; Leone, J. W.; Yuan, H.; Stein, S. E., NIST Interlaboratory Study on Glycosylation Analysis of Monoclonal Antibodies: Comparison of Results from Diverse Analytical Methods. *Mol Cell Proteomics* **2020,** *19* (1), 11-30.
9. Mimura, Y.; Katoh, T.; Saldova, R.; O'Flaherty, R.; Izumi, T.; Mimura-Kimura, Y.; Utsunomiya, T.; Mizukami, Y.; Yamamoto, K.; Matsumoto, T.; Rudd, P. M.,

Glycosylation engineering of therapeutic IgG antibodies: challenges for the safety, functionality and efficacy. *Protein Cell* **2018,** *9* (1), 47-62.
10. Thaysen-Andersen, M.; Packer, N. H., Advances in LC-MS/MS-based glycoproteomics: getting closer to system-wide site-specific mapping of the N- and O-glycoproteome. *Biochim Biophys Acta* **2014,** *1844* (9), 1437-52.
11. Alocci, D.; Mariethoz, J.; Gastaldello, A.; Gasteiger, E.; Karlsson, N. G.; Kolarich, D.; Packer, N. H.; Lisacek, F., GlyConnect: Glycoproteomics Goes Visual, Interactive, and Analytical. *J Proteome Res* **2019,** *18* (2), 664-677.
12. Lisacek, F.; Mariethoz, J.; Alocci, D.; Rudd, P. M.; Abrahams, J. L.; Campbell, M. P.; Packer, N. H.; Ståhle, J.; Widmalm, G.; Mullen, E.; Adamczyk, B.; Rojas-Macias, M. A.; Jin, C.; Karlsson, N. G., Databases and Associated Tools for Glycomics and Glycoproteomics. *Methods Mol Biol* **2017,** *1503*, 235-264.
13. Mariethoz, J.; Alocci, D.; Gastaldello, A.; Horlacher, O.; Gasteiger, E.; Rojas-Macias, M.; Karlsson, N.; Packer, N.; Lisacek, F., Glycomics@ExPASy: Bridging the Gap. *Molecular & Cellular Proteomics* **2018,** *17* (11), 2164-2176.
14. Aoki-Kinoshita, K.; Agravat, S.; Aoki, N. P.; Arpinar, S.; Cummings, R. D.; Fujita, A.; Fujita, N.; Hart, G. M.; Haslam, S. M.; Kawasaki, T.; Matsubara, M.; Moreman, K. W.; Okuda, S.; Pierce, M.; Ranzinger, R.; Shikanai, T.; Shinmachi, D.; Solovieva, E.; Suzuki, Y.; Tsuchiya, S.; Yamada, I.; York, W. S.; Zaia, J.; Narimatsu, H., GlyTouCan 1.0--The international glycan structure repository. *Nucleic Acids Res* **2016,** *44* (D1), D1237-42.
15. Rojas-Macias, M. A.; Mariethoz, J.; Andersson, P.; Jin, C.; Venkatakrishnan, V.; Aoki, N. P.; Shinmachi, D.; Ashwood, C.; Madunic, K.; Zhang, T.; Miller, R. L.; Horlacher, O.; Struwe, W. B.; Watanabe, Y.; Okuda, S.; Levander, F.; Kolarich, D.; Rudd, P. M.; Wuhrer, M.; Kettner, C.; Packer, N. H.; Aoki-Kinoshita, K. F.; Lisacek, F.; Karlsson, N. G., Towards a standardized bioinformatics infrastructure for N- and O-glycomics. *Nat Commun* **2019,** *10* (1), 3275.
16. Durrant, J. D.; Kochanek, S. E.; Casalino, L.; Ieong, P. U.; Dommer, A. C.; Amaro, R. E., Mesoscale All-Atom Influenza Virus Simulations Suggest New Substrate Binding Mechanism. *ACS Cent Sci* **2020,** *6* (2), 189-196.
17. Yu, I.; Mori, T.; Ando, T.; Harada, R.; Jung, J.; Sugita, Y.; Feig, M., Biomolecular interactions modulate macromolecular structure and dynamics in atomistic model of a bacterial cytoplasm. *Elife* **2016,** *5*.
18. Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Shaw, D. E., Picosecond to Millisecond Structural Dynamics in Human Ubiquitin. *J Phys Chem B* **2016,** *120* (33), 8313-20.
19. Brzezicka, K.; Echeverria, B.; Serna, S.; van Diepen, A.; Hokke, C. H.; Reichardt, N. C., Synthesis and microarray-assisted binding studies of core xylose and fucose containing N-glycans. *ACS Chem Biol* **2015,** *10* (5), 1290-302.
20. Wilson, I. B.; Harthill, J. E.; Mullin, N. P.; Ashford, D. A.; Altmann, F., Core alpha1,3-fucose is a key part of the epitope recognized by antibodies reacting against plant N-linked oligosaccharides and is present in a wide variety of plant extracts. *Glycobiology* **1998,** *8* (7), 651-61.
21. van Ree, R.; Cabanes-Macheteau, M.; Akkerdaas, J.; Milazzo, J. P.; Loutelier-Bourhis, C.; Rayon, C.; Villalba, M.; Koppelman, S.; Aalberse, R.; Rodriguez, R.; Faye, L.; Lerouge, P., Beta(1,2)-xylose and alpha(1,3)-fucose residues have a strong contribution in IgE binding to plant glycoallergens. *J Biol Chem* **2000,** *275* (15), 11451-8.
22. Bardor, M.; Faveeuw, C.; Fitchette, A. C.; Gilbert, D.; Galas, L.; Trottein, F.; Faye, L.; Lerouge, P., Immunoreactivity in mammals of two typical plant glyco-epitopes, core alpha(1,3)-fucose and core xylose. *Glycobiology* **2003,** *13* (6), 427-34.

23. Strasser, R., Plant protein glycosylation. *Glycobiology* **2016,** *26* (9), 926-939.
24. Harbison, A. M.; Brosnan, L. P.; Fenlon, K.; Fadda, E., Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. *Glycobiology* **2019,** *29* (1), 94-103.
25. Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Darvill, A. G.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Essentials of Glycobiology. 2015.
26. Fitchette-Lainé, A. C.; Gomord, V.; Cabanes, M.; Michalski, J. C.; Saint Macary, M.; Foucher, B.; Cavelier, B.; Hawes, C.; Lerouge, P.; Faye, L., N-glycans harboring the Lewis a epitope are expressed at the surface of plant cells. *Plant J* **1997,** *12* (6), 1411-7.
27. Echeverria, B.; Serna, S.; Achilli, S.; Vivès, C.; Pham, J.; Thépaut, M.; Hokke, C. H.; Fieschi, F.; Reichardt, N. C., Chemoenzymatic Synthesis of N-glycan Positional Isomers and Evidence for Branch Selective Binding by Monoclonal Antibodies and Human C-type Lectin Receptors. *ACS Chem Biol* **2018,** *13* (8), 2269-2279.
28. Barb, A. W.; Brady, E. K.; Prestegard, J. H., Branch-specific sialylation of IgG-Fc glycans by ST6Gal-I. *Biochemistry* **2009,** *48* (41), 9705-7.
29. Moginger, U.; Grunewald, S.; Hennig, R.; Kuo, C.; Schirmeister, F.; Voth, H.; Rapp, E.; Khoo, K.; Seeberger, P.; Simon, J.; Kolarich, D., Alterations of the Human Skin N- and O-Glycome in Basal Cell Carcinoma and Squamous Cell Carcinoma. *Frontiers in Oncology* **2018,** *8*.
30. Case, D.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham III, T.; Cruzeiro, V.; Darden, T.; Duke, R.; Ghoreishi, D.; Gilson, M.; Gohlke, H.; Goetz, A.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R.; Wang, J.; Wei, H.; Wolf, R.; Wu, X.; Xiao, L.; York, D.; Kollman, P. *AMBER 2018*, University of California, San Francisco, 2018.
31. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* **2008,** *29* (4), 622-55.
32. Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M., Comparison of simple potential functions for simulations of liquid water. *Journal of Chemical Physics* **1983,** *79* (2), 926-935.
33. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996,** *14* (1), 33-8, 27-8.
34. Alibay, I.; Burusco, K.; Bruce, N.; Bryce, R., Identification of Rare Lewis Oligosaccharide Conformers in Aqueous Solution Using Enhanced Sampling Molecular Dynamics. *Journal of Physical Chemistry B* **2018,** *122* (9), 2462-2474.
35. Mayes, H. B.; Broadbelt, L. J.; Beckham, G. T., How sugars pucker: electronic structure calculations map the kinetic landscape of five biologically paramount monosaccharides and their implications for enzymatic catalysis. *J Am Chem Soc* **2014,** *136* (3), 1008-22.
36. Rönnols, J.; Manner, S.; Siegbahn, A.; Ellervik, U.; Widmalm, G., Exploration of conformational flexibility and hydrogen bonding of xylosides in different solvents, as a model system for enzyme active site interactions. *Org Biomol Chem* **2013,** *11* (33), 5465-72.
37. Iglesias-Fernández, J.; Raich, L.; Ardèvol, A.; Rovira, C., The complete conformational free energy landscape of β-xylose reveals a two-fold catalytic itinerary for β-xylanases. *Chem Sci* **2015,** *6* (2), 1167-1177.

38. Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J., Glycosylation in health and disease. *Nat Rev Nephrol* **2019,** *15* (6), 346-366.