



## Carbohydrate Chemistry

## A Secondary Structural Element in a Wide Range of Fucosylated Glycoepitopes

Thomas Aeschbacher,<sup>[a]</sup> Mirko Zierke,<sup>[b, e]</sup> Martin Smieško,<sup>[b]</sup> Mayeul Collot,<sup>[c, f]</sup>  
Jean-Maurice Mallet,<sup>[c]</sup> Beat Ernst,<sup>[b]</sup> Frédéric H.-T. Allain,<sup>\*[a]</sup> and Mario Schubert<sup>\*[a, d]</sup>

**Abstract:** The increasing understanding of the essential role of carbohydrates in development, and in a wide range of diseases fuels a rapidly growing interest in the basic principles governing carbohydrate-protein interactions. A still heavily debated issue regarding the recognition process is the degree of flexibility or rigidity of oligosaccharides. Combining NMR structure determination based on extensive experimental data with DFT and database searches, we have identified a set of trisaccharide motifs with a similar conformation that is characterized by a non-conventional C–H...O

hydrogen bond. These motifs are present in numerous classes of oligosaccharides, found in everything from bacteria to mammals, including Lewis blood group antigens but also unusual motifs from amphibians and marine invertebrates. The set of trisaccharide motifs can be summarized with the consensus motifs X-β1,4-[Fucα1,3]-Y and X-β1,3-[Fucα1,4]-Y—a secondary structure we name [3,4]F-branch. The wide spectrum of possible modifications of this scaffold points toward a large variety of glycoepitopes, which nature generated using the same underlying architecture.

## Introduction

Carbohydrates are found in great abundance on cell surfaces of all organisms and are crucial for numerous biological processes, including cell–cell adhesion, cellular recognition, and various signaling processes.<sup>[1]</sup> A wide and diverse range of highly specific glycoepitopes function as recognition sites for various receptors—a field receiving increasing attention.<sup>[2]</sup> Branching a linear oligosaccharide with a fucose moiety signifi-

cantly contributes to this diversity, which plays an important role in tissue development, cell-adhesion (such as selectin-mediated leucocyte-endothelial adhesion), fertilization, host-microbe interactions,<sup>[3]</sup> and the human Lewis and ABO blood group antigens.<sup>[4]</sup> Moreover, altered fucosylation has been observed in various cancer cells<sup>[5]</sup> and diseases, including type 1 diabetes mellitus, rheumatoid arthritis, and cystic fibrosis.<sup>[3c]</sup>

Understanding the molecular basis of specificity and affinity of carbohydrate–protein interactions is the key to uncover the glycode and its biological roles. In this context, information on the conformation of the carbohydrate ligands in the unbound and bound form is of particular interest to understand the entropic contribution to binding. The trisaccharide Lewis<sup>x</sup> (Le<sup>x</sup>, Galβ1,4[Fucα1,3]GlcNAcβ) is one of the Lewis blood group antigens that are widely considered to adopt a well-defined structure in solution.<sup>[6]</sup> Le<sup>x</sup> is recognized by DC-SIGN<sup>[7]</sup> and its sialylated version sLe<sup>x</sup> by E-selectin,<sup>[8]</sup> as well as the human zona pellucida.<sup>[9]</sup> Its conformation in solution is virtually identical to the conformations found in the vast majority of lectin complexes, suggesting that the solution conformation is identical to the bioactive conformation. A strong stacking interaction between its Gal and Fuc moiety<sup>[10]</sup> was attributed to van der Waals contacts, steric hindrance, especially by the adjacent GlcNAc moiety, as well as the *exo*-anomeric effect.<sup>[6,11]</sup> To improve the insight into this stabilization, we recently determined the NMR structure of Le<sup>x</sup> attached to a carrier protein. The slower tumbling due to this attachment caused an enhanced nuclear Overhauser effect enabling the measurement of reliable distance restraints, which were suitable for structure determination protocols developed for biological macromolecules. We noticed the presence of a non-conventional C–H...O hydrogen bond that plays an important role in the stabilization

[a] Dr. T. Aeschbacher, Prof. Dr. F. H.-T. Allain, Dr. M. Schubert  
Institute of Molecular Biology and Biophysics, ETH Zürich  
8093 Zurich (Switzerland)  
E-mail: allain@mol.biol.ethz.ch  
mario.schubert@sbg.ac.at

[b] Dr. M. Zierke, Dr. M. Smieško, Prof. Dr. B. Ernst  
Institute of Molecular Pharmacy, University of Basel  
Klingelbergstr. 50, 4056 Basel (Switzerland)

[c] Dr. M. Collot, Prof. Dr. J.-M. Mallet  
Laboratoire des Biomolécules  
Département de Chimie, École normale supérieure  
PSL Research University  
Sorbonne Universités, UPMC Univ. Paris 06, CNRS  
24 rue Lhomond, 75005 Paris (France)

[d] Dr. M. Schubert  
Institute of Molecular Biology, University of Salzburg  
Billrothstr. 11, 5020 Salzburg (Austria)

[e] Dr. M. Zierke  
Department of Chemistry, University of British Columbia  
V6T 1Z1 Vancouver, British Columbia (Canada)

[f] Dr. M. Collot  
UMR 7213 CNRS, Laboratoire de Biophotonique et Pharmacologie  
Faculté de Pharmacie  
74 route du Rhin, CS 60024, 67401 Illkirch (France)

Supporting information and the ORCID numbers for the authors of this article can be found under <https://doi.org/10.1002/chem.201701866>.

of the conformation of  $\text{Le}^x$ .<sup>[12]</sup> The C–H...O hydrogen bond between H5 of fucose and the galactose ring oxygen O5 contributes about  $1.8 \text{ kcal mol}^{-1}$  to the stabilization energy at the Gal–Fuc interface (total  $4.5 \text{ kcal mol}^{-1}$ ), locking the already quite restricted conformational space of that trisaccharide.<sup>[12]</sup> The existence of the C–H...O hydrogen bond in  $\text{Le}^x$  has recently been experimentally confirmed in the tetrasaccharide sialyl Lewis<sup>x</sup> (sLe<sup>x</sup>) by observing a spin–spin (scalar) coupling across this non-conventional hydrogen bond.<sup>[13]</sup> Among the key determinants of a C–H...O hydrogen bond are (i) a distance between H and O that is shorter than the sum of their van der Waals radii,<sup>[14]</sup> (ii) a characteristic downfield shift of the proton resonance<sup>[15]</sup> and (iii) the observation of a (3,-1) bond critical point between H and O in the electron density topological analysis of 3D coordinates.<sup>[15]</sup> The role of C–H...O hydrogen bonds in the stabilization of proteins and nucleic acids has only recently been debated<sup>[14]</sup> and their existence experimentally proven.<sup>[16]</sup> Although generally considered much weaker than a conventional hydrogen bond, electron-withdrawing substitutions, like oxygen in the case of  $\text{Le}^x$  or amide, and carbonyl in the case of amino acids, substantially polarize the C–H bond and thus increase its donor ability. High-level quantum mechanical calculations estimate a stabilization energy of about 30–50% compared to a conventional hydrogen bond for a C $\alpha$ -H...O hydrogen bond in proteins.<sup>[17]</sup>

Here, we raise the question of whether the stabilization of the  $\text{Le}^x$  trisaccharide that includes this typical C–H...O hydrogen bond as a central feature is a unique phenomenon or if similarly stable conformations exist in other oligosaccharides. We therefore conducted an extensive search for C–H...O hydrogen bonds in fucose-containing oligosaccharides and identified in many glycoepitopes a common trisaccharide architecture with a defined three-dimensional conformation that is comparable to secondary structure elements in proteins.

## Results

### The chemical shift as an indicator for C–H...O hydrogen bond confirmed by ab initio calculations

A comparison of the chemical shifts of  $\text{Le}^x$  with the disaccharide Fuc $\alpha$ 1,3GlcNAc (Figure 1a) revealed the expected large differences for C–H groups at the glycosidic linkages (H4, C4, H3 and C3). Interestingly, a very large difference is observed for H5 of Fuc, a proton in close proximity to the free electron pair of Gal O5. Chemical shifts of  $4.7\text{--}4.8 \text{ ppm}$ <sup>[12]</sup> were observed, very much downfield shifted compared to the resonance in isolated fucose ( $4.19 \text{ ppm}$ , BioMagResBank<sup>[18]</sup> entry bmse000036), suggesting that a C–H...O hydrogen bond is involved.<sup>[15]</sup> To get further insights into the correlation between the downfield shift and C–H...O hydrogen bonding, we calculated the dependence of the  $^1\text{H}$  chemical shift on the H...O distance using the model system (*i*Pro-O-Me and Me-O-Me) we used before<sup>[12]</sup> applying DFT calculations (Figure 1b). The shorter the hydrogen bond distance is, the higher the downfield chemical shift contribution. For this model system, the energetic minimum was identified previously at a distance of 2.3–

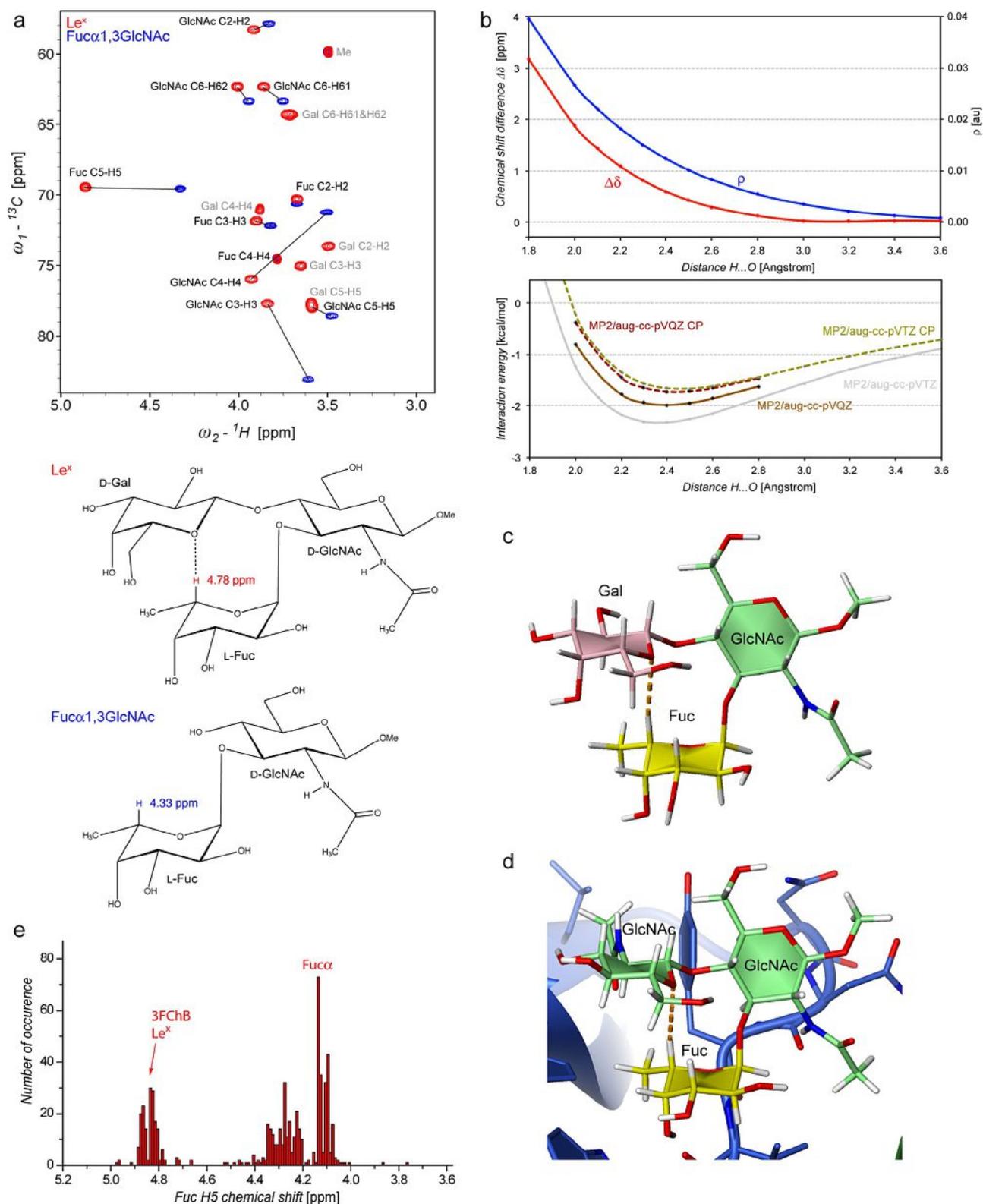
2.5 Å using high level ab initio calculations.<sup>[12]</sup> According to the obtained correlation, this distance range corresponds to a downfield shift  $\Delta\delta$  of 0.4–0.8 ppm, a value in agreement with the experimental shift of  $\Delta\delta \approx 0.45 \text{ ppm}$  observed for  $\text{Le}^x$  in comparison to Fuc $\alpha$ 1,3GlcNAc. Following the computational procedure suggested by Scheiner,<sup>[19]</sup> we were able to show that the main contribution to the downfield chemical shift change is originating from the C–H...O hydrogen bond (Figure S1 and Table S1 in the Supporting Information). We conclude from these calculations that the characteristic H5 Fuc chemical shift can serve as an indicator for the presence of the C–H...O hydrogen bond between the stacking saccharides of Fuc and Gal.

### Similarity between $\text{Le}^x$ and GlcNAc $\beta$ 1,4[Fuc $\alpha$ 1,3]GlcNAc $\beta$

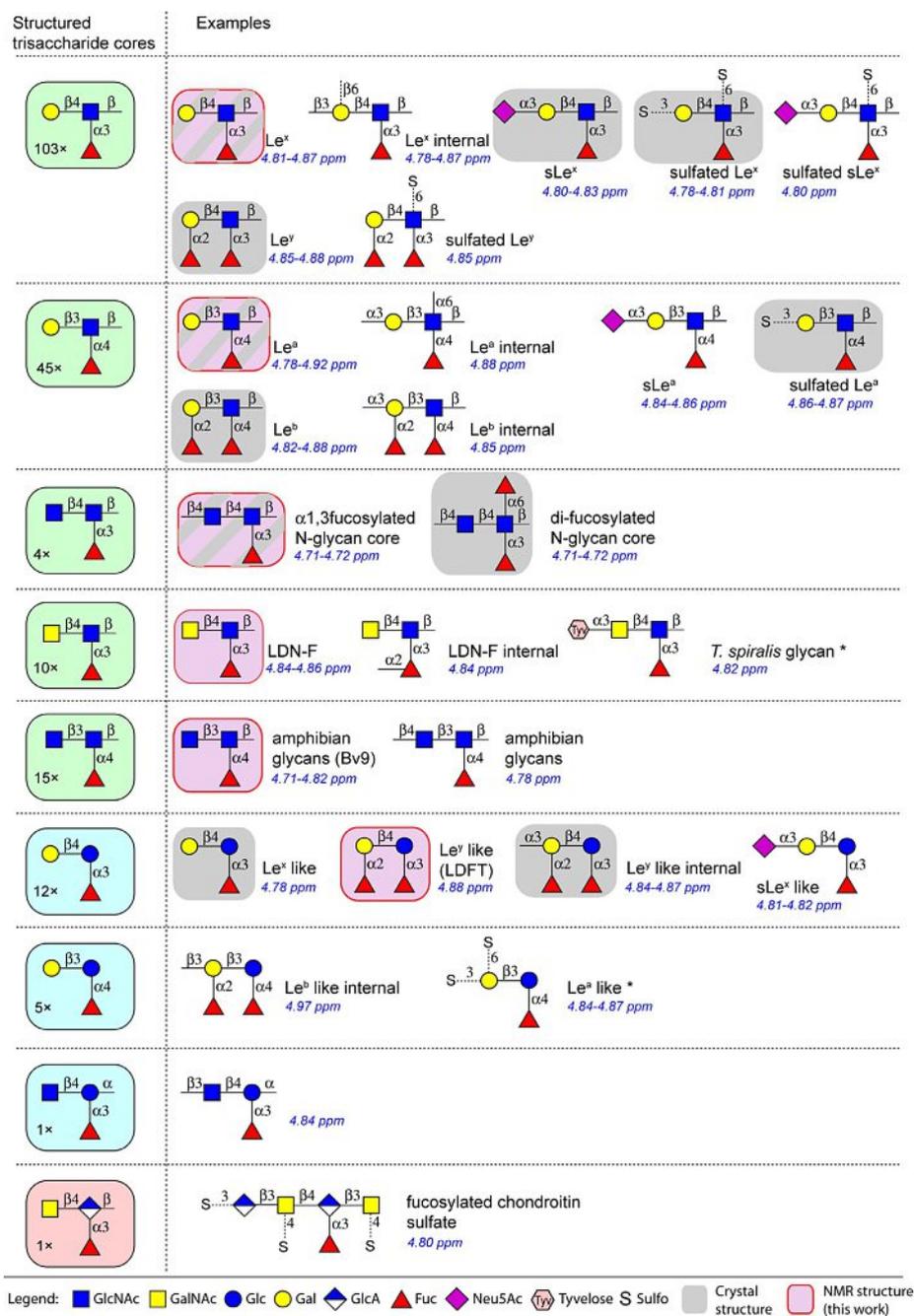
Interestingly, the structure of fucosylated chitobiose (GlcNAc $\beta$ 1,4[Fuc $\alpha$ 1,3]GlcNAc; called 3FChB in the following) in complex with the recently discovered *Coprinopsis cinerea* lectin 2 (CCL2)<sup>[20]</sup> determined by NMR spectroscopy, showed strong similarities to the  $\text{Le}^x$  structure (Figure 1c and d) and most strikingly a similar downfield shift of the Fuc H5 resonance at 5.05 ppm (BioMagResBank entry 17902) was found. A similar core structure and a Fuc H5 chemical shift of 4.75 ppm were observed for 6'-sulfo sialyl Lewis<sup>x</sup> when bound to Siglec-8<sup>[21]</sup> (BioMagResBank entry 25799). This prompted us to use the characteristic Fuc H5 chemical shift as a search criterion for detecting C–H...O hydrogen bonds in similarly stabilized oligosaccharides.

### Statistics of H5 chemical shift of fucose display a discrete cluster

We analyzed the chemical shift distribution of Fuc H5 by searching the NMR database of Glycosciences.de,<sup>[22]</sup> which contains 628 entries for fucose H5 chemical shifts (Figure 1e). Interestingly, about 25% of all observed chemical shifts cluster in a narrow range around 4.84 ppm, which corresponds to the characteristic H5 chemical shifts of  $\text{Le}^x$ , 3FChB and 6'-sulfo sLe<sup>x</sup>. Data of 178 oligosaccharides are found in this cluster (cutoff > 4.6 ppm) that are summarized in Figure 2 (for details, see Table S2 in the Supporting Information). In addition, 21 values from publications were included (marked with + in Table S2). In general, only  $\alpha$ 1,3- and  $\alpha$ 1,4-linked but not  $\alpha$ 1,2- and  $\alpha$ 1,6-linked fucose H5 chemical shifts were found within the cluster > 4.6 ppm. However all entries within that cluster were part of branched oligosaccharides. In the case of  $\alpha$ 1,3-linked Fuc, it was always accompanied with a  $\beta$ 1,4-linked saccharide, typically Gal, GlcNAc or GalNAc, whereas  $\alpha$ 1,4-linked fucose was always accompanied with a  $\beta$ 1,3-linked saccharide, typically Gal or GlcNAc, leading to the following consensus sequences: X- $\beta$ 1,3-[Fuc $\alpha$ 1,4]-Y and X- $\beta$ 1,4-[Fuc $\alpha$ 1,3]-Y in which X stands for the stacking saccharide and Y is a saccharide with equatorial hydroxy groups at positions 3 and 4. Nine differently branched trisaccharides were found as shown on the left of Figure 2 (typical examples on the right). All classes display a chemical shift for H5 of Fuc between 4.7 and 5.0 ppm. This suggests



**Figure 1.** Characteristic chemical shift reports a C–H...O hydrogen bond. (a) Chemical shift differences between Le<sup>x</sup> (red) and Fuca1,3GlcNAc that lacks the stacking Gal (blue) illustrated by an overlay of their <sup>13</sup>C HSQC spectra. Signals of Le<sup>x</sup> are labeled and corresponding signals in the disaccharide are connected by lines to the Le<sup>x</sup> signals (black label). The stacking Gal signals that are missing in the disaccharide are labeled grey. Signals of the methyl group at the reducing end are indicated by “Me”. (b) Dependence of the chemical shift  $\delta$  in a C–H...O hydrogen bond and the electron density at the bond critical point  $\rho(r_c)$  on the H–O distance as calculated with Bader’s atom in molecules (AIM) theory using the model system (iPro-O-Me and Me-O-Me) introduced before.<sup>[12]</sup> For comparison, the energetic minimum for the same model system as a function of the H–O distance<sup>[12]</sup> is shown at the bottom. (c) Representative NMR structure of Le<sup>x</sup>.<sup>[12]</sup> (d) Structure of 3FChB in complex with the lectin CCL2.<sup>[20]</sup> (e) Histogram of all available chemical shifts of H5 of Fuc within the Glycosciences.de database.<sup>[22]</sup> The H5 (Fuc) chemical shifts of Le<sup>x</sup> and of  $\alpha$ -L-fucose are indicated.



**Figure 2.** Structural motifs displaying a characteristic  $^1\text{H}$  chemical shift for Fuc H5. On the left, trisaccharide glycan cores that very likely adopt a defined 3D structure. The number of structures within each category is given as well. Typical examples are shown on the right. Structures indicated with an asterisk were absent in the database search but were added from two publications.<sup>[25,30]</sup> The associated chemical shifts of H5 of Fuc are presented in blue. A light grey background indicates glycans with a three-dimensional structure in the PDB databank. Glycans of which the solution structure was determined by NMR spectroscopy were highlighted in light magenta and framed in red (this work).

that all of those trisaccharide core structures adopt a stabilized 3D structure similar to 3FChB and Le<sup>x</sup>. Only two  $\alpha$ 1,2-fucosylated glycans display a chemical shift close to the cut-off shift of 4.6 ppm, which however, seem to be part of another cluster (Figure 1 e), probably reflecting a different kind of stabilization.

The first two categories contain the widely abundant Lewis-type blood group antigens Le<sup>a</sup>, Le<sup>b</sup>, Le<sup>x</sup>, and Le<sup>y</sup>, as well as substituted versions thereof, including sialyl Le<sup>x</sup> (sLe<sup>x</sup>) and several sulfated Le<sup>x</sup>, Le<sup>y</sup>, and sLe<sup>x</sup> epitopes. Glycans containing 3FChB

are members of the third category that contain  $\alpha$ 1,3-fucosylated *N*-glycan cores typically found in plants and invertebrates.<sup>[23]</sup> Interestingly, several additional motifs that are usually not associated with Lewis antigens were found:  $\alpha$ 1,3-fucosylated LacdiNAc (LDNF) present in helminths<sup>[24]</sup> and sea squirt,<sup>[25]</sup> species-specific amphibian egg jelly coats with an  $\alpha$ 1,4-fucosylation,<sup>[26]</sup> and fucosylated chondroitin sulfate<sup>[27]</sup> also display the downfield-shifted H5 resonance of their Fuc moiety. Moreover, Lewis-type-like antigens in which the reducing end of GlcNAc

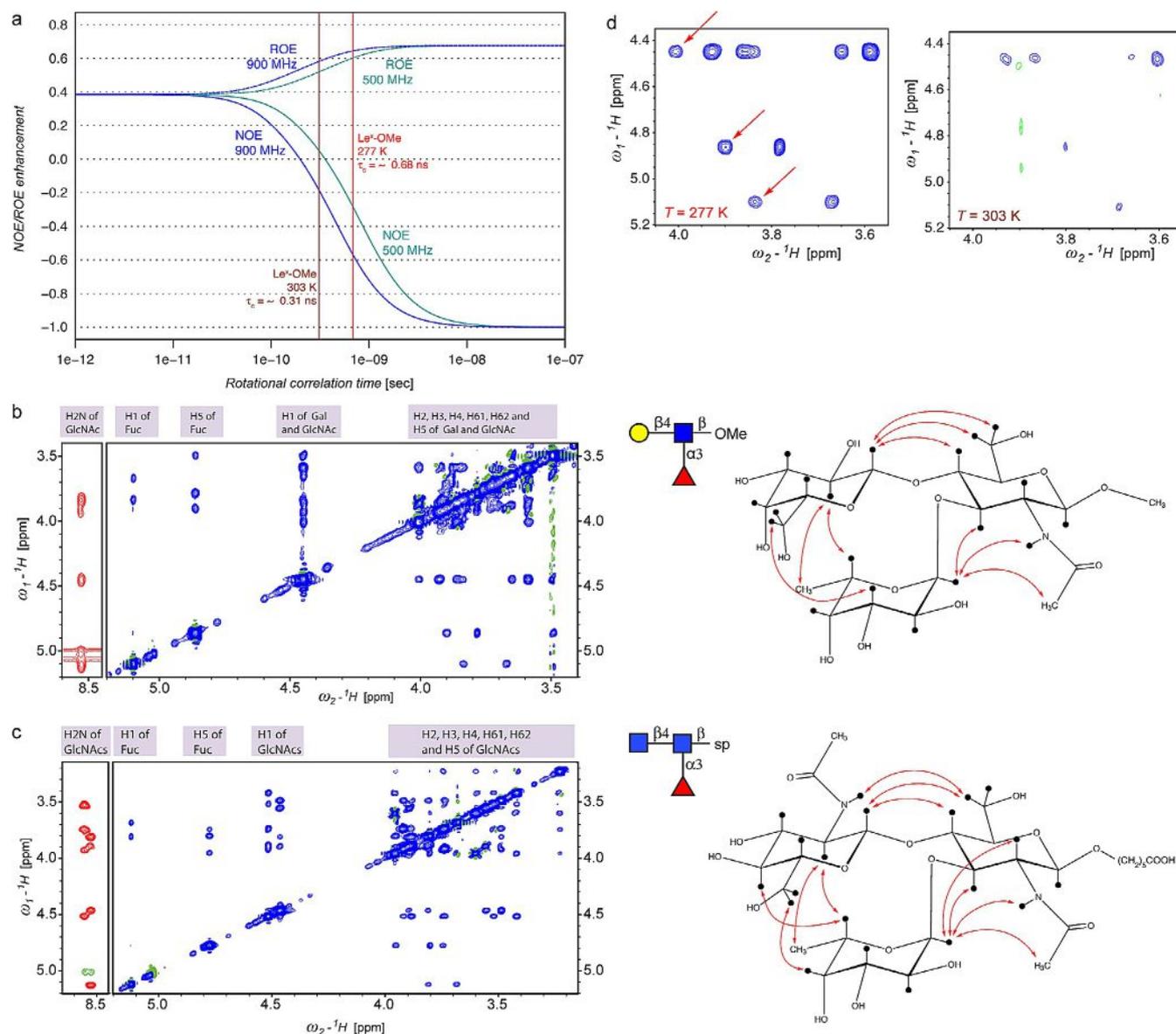
is substituted by a Glc moiety fulfill the search criterion as well. It is noteworthy that  $\alpha$ 1,2-linked Fuc, as found for example in determinants for blood groups A, B and H, do not display the characteristic downfield chemical shift, suggesting that those motifs are not stabilized by a comparable intramolecular C–H...O hydrogen bond.

In summary, the database search revealed nine trisaccharide core structures that all exhibit a H5 chemical shift of Fuc around 4.8 ppm, which is indicative of a C–H...O hydrogen bond suggesting a very similar three-dimensional architecture. To provide more evidence for a common scaffold, we used

NMR structure determination, ab initio calculations, and the analysis of X-ray crystal structures deposited in databases.

### An efficient, generally applicable approach to obtain 3D-structures of carbohydrates in solution by NMR spectroscopy

Three-dimensional structures of carbohydrates in solution are commonly obtained by using molecular dynamics (MD) simulations that typically lack any experimental input and might be biased by the applied force field. A purely experimentally

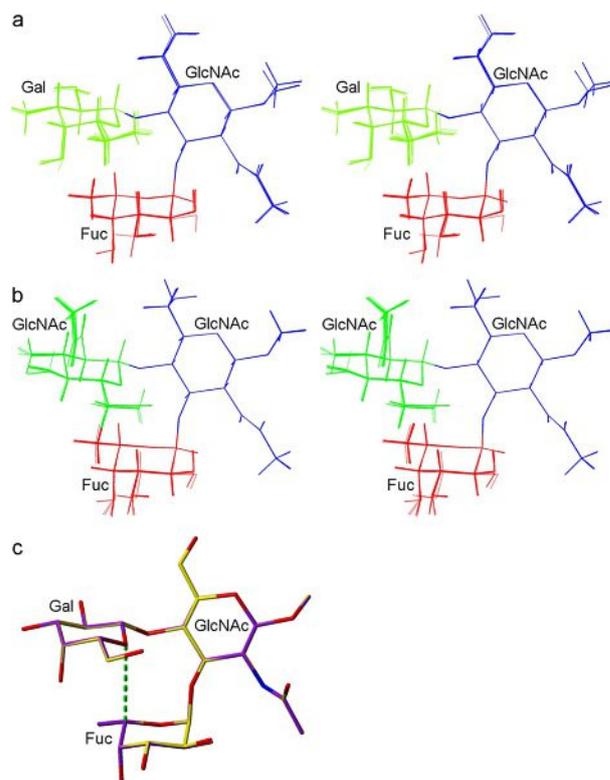


**Figure 3.** Efficient approach for three-dimensional structure determination of oligosaccharides in solution. (a) NOE enhancement factors of a transient NOE experiment as a function of the correlation time  $\tau_c$  indicated for two magnetic fields corresponding to 500 and 900 MHz. Estimated correlation times of Le<sup>x</sup> methyl glycoside (MW: 543.5 g mol<sup>-1</sup>) in D<sub>2</sub>O are indicated by vertical lines. (b) 2D <sup>1</sup>H-<sup>1</sup>H NOESY spectra of methyl Le<sup>x</sup> (3.7 mm) in either H<sub>2</sub>O (red, left) or in D<sub>2</sub>O (blue, right) measured at 900 MHz and 277 K. Chemical shift assignments are indicated on the top for isolated resonances. On the right, schematic presentation of methyl Le<sup>x</sup> with the observed inter-residue NOEs indicated by red arrows. (c) 2D <sup>1</sup>H-<sup>1</sup>H NOESY spectra of 3FChB (GlcNAcβ1,4[Fucα1,3]GlcNAc-O(CH<sub>2</sub>)<sub>5</sub>COOH, 2.8 mm) in either H<sub>2</sub>O (red, left) or in D<sub>2</sub>O (blue, right) recorded at 900 MHz and 277 K. On the right is a schematic presentation of 3FChB showing the obtained inter-residue NOEs by red arrows. (d) Influence of temperature on the quality of NOESY cross-peaks illustrated by two 2D <sup>1</sup>H-<sup>1</sup>H NOESY spectra of 3.7 mm Le<sup>x</sup>-OMe in D<sub>2</sub>O measured at 900 MHz and either 277 or 293 K. Inter-residue NOE cross-peaks are indicated by red arrows.

driven, efficient approach has not been established so far. The most reliable experimental input for the determination of three-dimensional structure in solution is still the nuclear Overhauser effect (NOE). However, for small molecules with masses around 500 Da, the NOE enhancement factors are close to zero, that is, NOEs cannot be observed. This stands in contrast to large biopolymers like proteins for which the enhancement factors reach an optimal value of  $-1$  (Figure 3 a). There are several possibilities to obtain a more favorable NOE enhancement factor: attaching the small molecule to a larger molecule, which we successfully applied recently,<sup>[12]</sup> decreasing the temperature, and applying higher magnetic fields are additional alternatives (Figure 3 a). Whereas the first approach requires the formation of a covalent bond between the small and the large  $^{13}\text{C}/^{15}\text{N}$  labeled molecule, the latter two methods are technically feasible and directly applicable to the small molecule. From the calculated enhancement factors (Figure 3 a), we can deduce that lowering the temperature from 25 to 0 °C increases the estimated tumbling time approximately by a factor of 2. In addition to low temperatures, we used a high magnetic field strength to measure 2D NOESY spectra. With a magnetic field of 21.2 Tesla (900 MHz) together with a temperature just above the freezing point, samples of  $\text{Le}^x$  and  $\alpha$ 1,3-fucosylated chitobiose yielded excellent NOE cross-peaks (Figure 3 b–d). A mixing time of 150 ms was applied which is still in the linear range of NOESY intensities as a function of the mixing time (Figure S2 in the Supporting Information) to avoid spin diffusion effects. The line shapes and the overall quality of the NOESY spectra are much better than the  $^{13}\text{C}$  or  $^{15}\text{N}$  F1-filtered, F2-filtered NOESY spectra measured with our recent approach.<sup>[12]</sup> Exceptionally high numbers of inter-residual NOEs (Figure 3 b and c; Tables S3 and S4), namely 9 for  $\text{Le}^x$  and 11 for fucosylated chitobiose, corresponding to 4.5 and 5.5 NOE distance restraints per glycosidic linkage, respectively, were obtained. These are redundant experimental restraints for defining two angles per glycosidic linkage.

For the experimentally driven structure determination, we first used the NOE-derived distance restraints with a distance geometry approach to avoid any bias from a force field applying CYANA,<sup>[28]</sup> an established software package for proteins and nucleic acids. In a second step, the obtained coordinates were further refined using the GLYCAM force field<sup>[29]</sup> together with the experimental restraints. For both trisaccharides, we obtained a well-defined ensemble (Figure 4 a and b; statistics in Table S5 in the Supporting Information). The structural ensemble of  $\text{Le}^x$  is virtually identical to the one obtained previously when we covalently attached it to a protein and measured at 298 K<sup>[12]</sup> (Figure 4 c). The ensemble of 3FChB corresponds to the structure of the trisaccharide when bound to CCL2.<sup>[20]</sup> This also demonstrates that the lower temperature does not influence the structure significantly, which is also reflected by almost identical chemical shifts when comparing spectra at 277 and 298 K (Figure S3).

Thus, based on collecting a redundant number of experimental distance restraints obtained at ultra-high magnetic fields and low temperature, we established an efficient and reliable method to determine 3D solution structures for small oli-

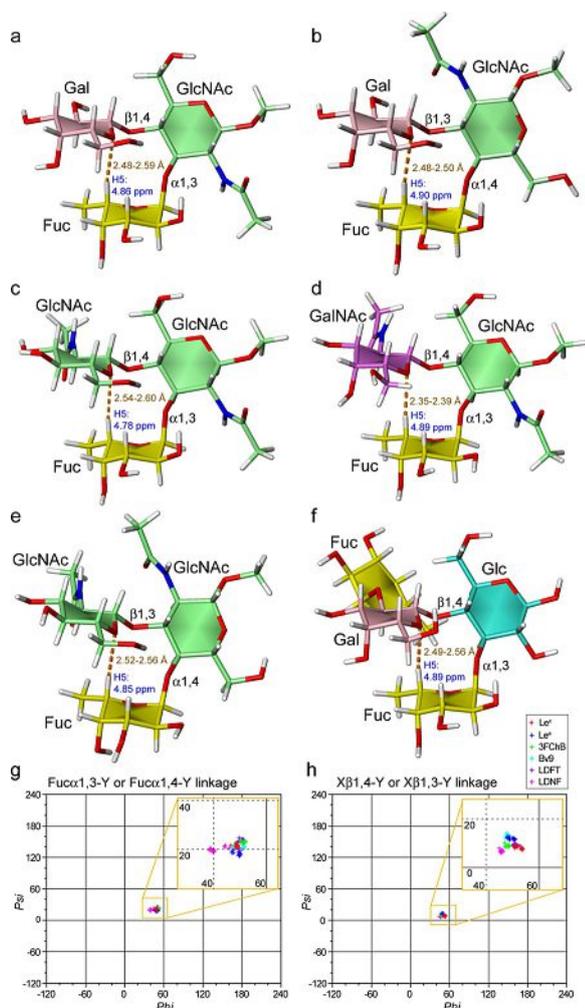


**Figure 4.** Experimental solution structures of  $\text{Le}^x$  methyl glycoside and 3FChB based on NOESY data recorded at 277 K and 900 MHz. (a) Stereo figure of an ensemble of 20 best structures of  $\text{Le}^x$  methyl glycoside. (b) Stereo figure of an ensemble of the 20 best structures of 3FChB. (c) Most representative NMR structure (yellow) superimposed with a representative structure obtained from a  $\text{Le}^x$ -protein conjugate recorded at 293 K (purple). For simplicity only heavy atoms are shown. The C–H...O hydrogen bond is shown as a dashed green line.

gosaccharides. In case of  $\text{Le}^x$ , a virtually identical structure was obtained as in our previous approach, but without the need for a chemical attachment to an isotopically labeled carrier protein. This new approach can be applied directly to every oligosaccharide of interest in aqueous solution.

### 3D solution structures of trisaccharides displaying the characteristic chemical shift

We used the same experimental approach to determine the 3D solution structures of representatives of four additional clusters obtained from the chemical shift statistics of Fuc H5 (Figure 2), in particular the blood group epitope  $\text{Le}^a$ , fucosylated LacdiNAc (LDNF), the amphibian egg glycan Bv9,<sup>[26]</sup> and lactodifucotetraose (LDFT). For the four carbohydrates, 2D NOESY spectra were recorded using temperatures of 275 or 273 K, which resulted in slightly supercooled solutions in  $\text{D}_2\text{O}$ . High quality NOESY spectra were obtained resulting in an exceptional large amount of inter-residual NOE distance restraints (Figure S4 and Tables S6–S9 in the Supporting Information) leading to well-defined structures with a high similarity to 3FChB and  $\text{Le}^x$  (Figure 5 a–f; statistics are found in Table S5). The analysis of the glycosidic linkages (for dihedral-angle plots



**Figure 5.** Comparison of the three-dimensional structures of methyl  $\text{Le}^x$ , methyl  $\text{Le}^a$ , 3FChB, LDNF, Bv9 and LDFT. (a) Representative NMR structure of methyl  $\text{Le}^x$ . The C–H...O hydrogen bond is indicated by an orange dotted line and their C5–O5 distance is given. In addition the chemical shift of H5 of Fuc is indicated in blue. (b–f) Representative NMR structures of methyl  $\text{Le}^a$ , 3FChB, LDNF, Bv9, and LDFT, respectively. (g, h) Comparison of the glycosidic dihedral angles of methyl  $\text{Le}^x$ , methyl  $\text{Le}^a$ , 3FChB, LDNF, Bv9, and LDFT. Left: comparable phi-psi plot of the Fuc-Y linkages. We choose the NMR angle definition ( $\phi$ :  $\text{H}_1\text{-C}_1\text{-O}_1\text{-C}'_X$ ;  $\psi$ :  $\text{C}_1\text{-O}_1\text{-C}'_X\text{-H}'_X$ ) rather than the X-ray dihedral angle definition ( $\phi$ :  $\text{O}_5\text{-C}_1\text{-O}_1\text{-C}'_X$ ;  $\psi$ :  $\text{C}_1\text{-O}_1\text{-C}'_X\text{-C}'_{X-1}$ ), because the latter would lead to different angles between  $\alpha 1,3$ - and  $\alpha 1,4$ -fucosylated glycans, differing by  $\approx 120^\circ$ . Angles were extracted by CARP.<sup>[41]</sup> Right: comparable phi-psi plot of the X-Y linkages in which X stands for the saccharide that stacks with the Fuc and Y stands for saccharide to which X and Fuc are connected.

see Figure 5g and h) exhibited basically identical values for  $\text{Le}^x$ ,  $\text{Le}^a$ , 3FChB, LDNF, Bv9, and LDFT, indicating that they share the same underlying architecture. The three-dimensional arrangements suggest a comparable C–H...O hydrogen bond between H5 of Fuc and O5 of the stacking Gal, GalNAc or GlcNAc (see Figure 5a–f). All structures showed comparable H5...O5 distances in the NMR ensembles (Table 1), which however might be still influenced by the applied force-field.

## Refinement and analysis of 3D solution structures

Because both distance geometry calculations and refinement with Amber do not allow close contacts between non-covalently bound atoms, the C–H...O hydrogen bond might be even shorter in reality. We therefore performed calculations based on first principles: we used DFT<sup>[31]</sup> at the B3LYP/6-31G(d,p) level<sup>[32]</sup> using an implicit water solvation model to minimize the geometry of one representative of each of the six structural ensembles. The resulting structures of  $\text{Le}^x$ , 3FChB,  $\text{Le}^a$ , LDNF, Bv9 and LDFT (Figure S5 in the Supporting Information) are very similar to the initial NMR structures but show shorter H5...O5 distances of 2.3–2.5 Å (Table 1), now significantly shorter than the sum of the van der Waals radii of 2.7 Å. Furthermore, we used the wave functions obtained from the DFT calculations to localize and calculate the electron density at the bond critical points  $\rho(r_c)$  within the C–H...O hydrogen bonds with Bader's atom in molecules (AIM) theory<sup>[33]</sup> (Table S10). The existence of bond critical points with a sufficiently high electron density  $\rho(r_c)$  between 0.010 and 0.014 au is clear evidence for the presence of a C–H...O hydrogen bond in all six structures.

For independently verifying the correctness of our structures, we used the Gauge-Independent Atomic Orbital (GIAO) method<sup>[34]</sup> within Gaussian 09<sup>[35]</sup> to calculate the NMR shielding tensor from the DFT-optimized structures (Table S11 in the Supporting Information). The agreement between the calculated and the experimental chemical shifts is fairly good (RMSD 0.14–0.22 ppm), especially considering the used approximations to mimic the solvent. In particular, the chemical shifts of H5 of Fuc are excellently reproduced with predictions of 4.88–5.10 ppm, corresponding to a downfield shift of 0.55–0.77 ppm compared to the non-stabilized Fuc $\alpha 1,3$ GlcNAc disaccharide. The fact that the distances of 2.3–2.4 Å in the used DFT-optimized structures are typical for a C–H...O hydrogen bond,<sup>[36]</sup> the existence of bond critical points with a sufficiently high electron density and the agreement of the chemical shift predictions with the experimental values, provides direct evidence that the structures are very similar and that a C–H...O hydrogen bond is present in all six trisaccharide structures. Applying the procedure of Scheiner<sup>[19]</sup> further proved the existence of the hydrogen bond and its impact on the chemical shift of Fuc H5 saccharide (Table S1).

## Branched fucosylated oligosaccharides in protein crystal structures

The well-defined NMR structures of  $\text{Le}^x$ ,  $\text{Le}^a$ , 3FChB, LDNF, Bv9, and LDFT prompted us to additionally analyze protein crystal structures that contain representatives of the trisaccharide categories displayed in Figure 2, either as ligand or within a glycan. One hundred crystal structures containing four out of the nine trisaccharide categories are available (marked with a grey background in Figure 2; Tables 1 and S12 in the Supporting Information). Due to its high abundance in N-glycan, cores of plant proteins 3FChB are overrepresented. For comparison, we also included a crystal structure of isolated  $\text{Le}^x$  from the

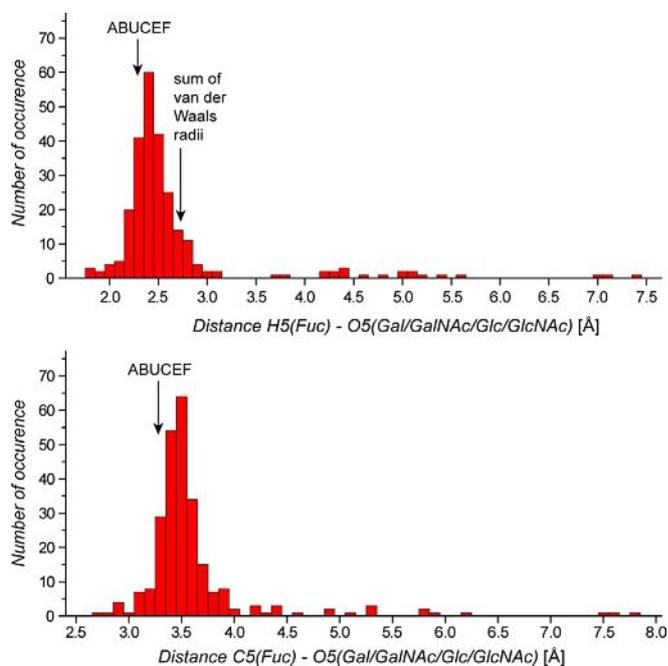
**Table 1.** Distances between H5 of Fuc and O5 of Gal/GlcNAc in the NMR structures, one Cambridge Structural Database (CSD) structure and Protein Data bank (PDB) structures as evidence for C–H...O hydrogen bond. For more details see Table S12 in the Supporting Information.

Method	Glycomotif	Type of structure or PDB/CSD entry identifier	Number of motifs	Distance H5(Fuc)-O5(Gal/GlcNAc) [Å]	Distance C5(Fuc)-O5(Gal/GlcNAc) [Å]
NMR	3FChB	ensemble	20	2.58 ± 0.01 <sup>[a]</sup>	3.62 ± 0.01 <sup>[a]</sup>
		DFT minimized	1	2.35	3.44
	Le <sup>x</sup>	ensemble	20	2.50 ± 0.01 <sup>[a]</sup>	3.56 ± 0.01 <sup>[a]</sup>
		DFT minimized	1	2.32	3.41
	Le <sup>a</sup>	ensemble	20	2.49 ± 0.01 <sup>[a]</sup>	3.55 ± 0.01 <sup>[a]</sup>
		DFT minimized	1	2.33	3.43
	LDNF	ensemble	20	2.36 ± 0.01 <sup>[a]</sup>	3.43 ± 0.02 <sup>[a]</sup>
		DFT minimized	1	2.41	3.50
	Bv9	ensemble	20	2.54 ± 0.02 <sup>[a]</sup>	3.60 ± 0.02 <sup>[a]</sup>
		DFT minimized	1	2.33	3.42 Å
LDFT	ensemble	20	2.53 ± 0.02 <sup>[a]</sup>	3.58 ± 0.02 <sup>[a]</sup>	
	DFT minimized	1	2.47	3.57	
X-ray (CSD)	Le <sup>x</sup>	ABUCEF	2	2.31 and 2.29	3.30 and 3.27
X-ray (PDB)	Le <sup>x</sup>	1FWU, 1G1R, 1G1S, 1G1T, 1KMB, 1SL5, 1SL6, 1UZ8, 2KMB, 2OX9, 2R61, 2RDG, 2Z8L, 3AP9, 3KMB, 3PVD, 3ZNL, 3ZNM, 4CSY, 4DXG, 4KMB, 4P2N, 4RCO, 4RFB, 4UO6, 4UO7, 4USO, 4X0C, 5A70, 5I4D <sup>[b]</sup>	68	2.41 ± 0.06 <sup>[c]</sup>	3.44 ± 0.05 <sup>[c]</sup>
	Le <sup>y</sup>	1CLY, 1CLZ, 1GSL, 1S3K, 2J1T, 3EYV, 3LEG, 3PA2, 3PUN, 4D4U, 4GWI, 4P25, 4RDL, 4WZE, 5ELB, 5ELC, 5ELD, 5ELE <sup>[d]</sup>	46	2.47 ± 0.06 <sup>[c]</sup>	3.50 ± 0.06 <sup>[c]</sup>
	Le <sup>a</sup>	1FWV, 1W8H, 3ASR, 4P3I, 4RMO, 4UT5, 4WZL, 5A6Z <sup>[e]</sup>	23	2.43 ± 0.08 <sup>[c]</sup>	3.46 ± 0.07 <sup>[c]</sup>
	Le <sup>b</sup>	1LED, 3ASS, 3AST, 3LEK, 3SEJ, 4RDK, 4GWJ, 4OPO, 4ZH7, 5F8R, 5F7M, 5F7N, 5F7W, 5F93, 5F9A, 5F9D	34	2.41 ± 0.10 <sup>[c]</sup>	3.43 ± 0.11 <sup>[c]</sup>
	3FChB (N-glycan cores)	1E4M, 1E6Q, 1E6S, 1JU2, 1LK9, 1YM0, 2B9L, 2F9N, 2QQM, 3L9R, 3QW9, 4ARN, 4GWM, 4GWN, 4GZT <sup>[f]</sup>	28 (only selection)	2.40 ± 0.06 <sup>[c]</sup>	3.46 ± 0.05 <sup>[c]</sup>
	Le <sup>x</sup> -like	1W8F	4	2.38 ± 0.08 <sup>[c]</sup>	3.45 ± 0.09 <sup>[c]</sup>
Le <sup>y</sup> -like	2O2L, 3EFX, 5ELF <sup>[g]</sup>	26	2.48 ± 0.06 <sup>[c]</sup>	3.51 ± 0.05 <sup>[c]</sup>	

[a] Average distances and standard deviation of the ensemble consisting of 20 structures. [b] Excluded 3ZW1, 4UNZ, 4UO2, 5AJB, and 5AJC that displayed an elongated Le<sup>x</sup> conformation. [c] Average distances and confidence interval (95%) of all motifs within one category; protons were added to the structures by Maestro (Schrödinger) because the crystal structures lacked protons (details in Table S12). [d] Excluded 2WWMG, 2WWMK, 2 structures of 4D4U that displayed an elongated Le<sup>y</sup> conformation. [e] Excluded 3UET that displayed an elongated Le<sup>a</sup> conformation. [f] Excluded 3UOP that displayed an elongated difucosylated N-glycan core conformation..

Cambridge Structural Database (CSD). In almost all of the 254 extracted oligosaccharide structures from the PDB, the trisaccharide core adopts the same conformation as we found for the six solution structures of Le<sup>x</sup>, Le<sup>a</sup>, 3FChB, LDNF, Bv9, and LDFT. Only in the case of 21 carbohydrate structures from 12 PDB entries do the carbohydrates adopt an obviously different extended conformation, indicated by a large distance between the fucose and the stacking saccharide (Table S12).

In the majority, consisting of 233 oligosaccharides, the distance between C5 of Fuc and O5 of Gal/GalNAc/GlcNAc shows a sharp maximum around 3.5 Å. After adding protons, the measured distances between H5 of Fuc and O5 of Gal/GalNAc/GlcNAc cluster in a narrow range around 2.4 Å (Figure 6, Tables 1 and S12). Even if some of these clustered distances are slightly biased due to repulsive terms between “non-bonded” atoms in the applied force-fields<sup>[12]</sup> the H5–O5 distances are clearly shorter than expected from the sum of the van der Waals radii of 2.7 Å.<sup>[37]</sup> Both sharp distance distributions correspond to typical distances for nearly linear C–H...O hydrogen bonds that were observed in high-resolution neutron diffraction structures of carbohydrate crystals.<sup>[38]</sup> There is no difference in the H5–O5 and C5–O5 distances depending on the kind of trisaccharide (Table 1). In the 21 oligosaccharide structures with extended conformations, the H5–O5 distances



**Figure 6.** Distribution of the H–O and C–O distances between Fuc C5-H5 and its corresponding hydrogen bonded oxygen in crystal structures. Data from 256 oligosaccharides from the PDB and CSD databases are included (Table S12). The sum of the van der Waal radii is indicated.

range from 3.7–7.4 Å and C5–O5 distances between 4.2–7.8 Å (Figure 6). Considering that one of these exceptions is a glycan in complex with an inactivated glycoside hydrolase GH98<sup>[39]</sup> that normally cleaves Galβ1,4GlcNAc linkages, the exact linkage that is distorted in the structure, we speculate that this distortion is required for the enzymatic function. Recently, a detailed investigation of extended Le<sup>x</sup> conformations in silico and bound to β-propeller lectins was reported, which estimated an energetic difference of about 2.5 kcal mol<sup>-1</sup> between the closed and extended conformation.<sup>[40]</sup> Based on MD simulations, it was concluded that the open state exists in 1.4% of the time in solution.

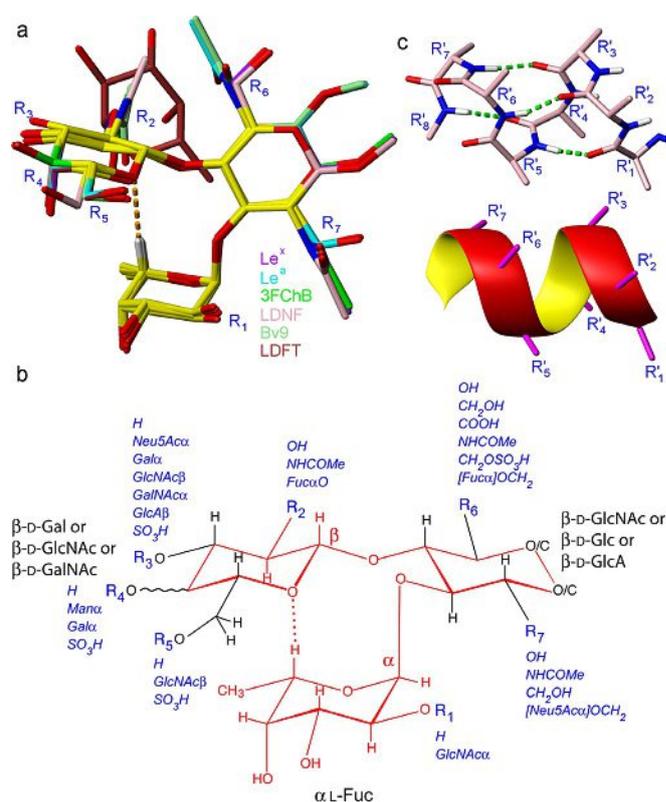
In summary, 92% of oligosaccharide structures from the PDB (Figure 2, indicated by a grey background) confirm that the glycoepitopes of Le<sup>x</sup>, Le<sup>y</sup>, Le<sup>a</sup>, Le<sup>b</sup>, 3FChB, difucosylated *N*-glycan cores, Le<sup>a</sup>-like and a Le<sup>y</sup>-like motifs (Glc instead of GlcNAc) share the same architecture and contain a C–H...O hydrogen bond in their trisaccharide motif when bound or linked to protein.

### A common scaffold and allowed substitutions

The present work suggests the same core architecture of the glycan motifs in Figure 2 consisting of an α-L-Fuc that stacks to a Gal/GlcNAc/GalNAc and a GlcNAc/Glc/GlcA that connects both stacking monosaccharide moieties with either α1,3 and β1,4 or α1,4 and β1,3 linkages (both in equatorial position). A superposition of the six NMR structures of 3FChB, Le<sup>x</sup>, Le<sup>a</sup>, LDNF, Bv9, and LDFT illustrates the shared architecture (Figure 7a). The various substitutions and additional carbohydrate linkages, found in other oligosaccharides with the characteristic Fuc H5 chemical shift, are summarized in Figure 7b. Whereas the L-Fuc moiety is mostly not modified (except at O2 with R<sub>1</sub>), modifications of the Gal/GlcNAc/GalNAc moieties stacked upon Fuc are found on all positions: C2, O3, O4 and O6 (functional groups R<sub>2</sub>, R<sub>3</sub>, R<sub>4</sub>, R<sub>5</sub>). The GlcNAc/Glc/GlcA at the branching site can be connected in two orientations (swapping the equatorial O3 and O4 linkages) and can contain various equatorial substituents at C2 and C6 (R<sub>6</sub> and R<sub>7</sub>). The α-L-Fuc moiety is the integral part of the conserved secondary structure since α-configuration is a prerequisite for the proper orientation of the C–H...O hydrogen bond. L-Saccharides are rare among the common glycans and only α-L-rhamnose exhibiting deviating chirality from α-L-Fuc at C2 and C4 might fulfill the role of the α-L-saccharide.

### Discussion

The NMR structures of α1,3- and α1,4-fucose-branched oligosaccharides determined here together with existing protein/carbohydrate crystal structures (Figure 2) show that all adopt a defined conformation and share the same spatial architecture. Common to those structures is a characteristic chemical shift of H5 of L-Fuc that is indicative of a C–H...O hydrogen bond between H5 of Fuc and the ring oxygen of a neighboring Gal/GlcNAc/GalNAc moiety. Six NMR solution structures, two crystal structures of an isolated carbohydrate and additionally 233 oli-



**Figure 7.** A common scaffold—a secondary structure element in glycans. (a) Superposition of a representative NMR structure of Le<sup>x</sup> (purple), Le<sup>a</sup> (cyan), 3FChB (green), LDNF (pink), Bv9 (pale green), and LDFT (brown). The shared architecture is colored in yellow and the C–H...O hydrogen bond is illustrated by a dotted, orange line. Protons are omitted for clarity except Fuc H5 that is colored in grey. (b) Consensus structure of branched trisaccharides stabilized by a C–H...O hydrogen bond. Observed modifications are given in blue. (c) Schematic presentation of a short α-helix including the N–H...O hydrogen bonds as green dashed lines (top) and as ribbon model (bottom). The amino acid side chains are indicated as R'<sub>1</sub>–R'<sub>7</sub>.

gosaccharide structures found in the PDB database support a common scaffold of X-β1,4-[Fucα1,3]-Y or X-β1,3-[Fucα1,4]-Y glycoepitopes (Figure 2) that contains a C–H...O hydrogen bond as a central feature. We name the structural element “[3,4]F-branch” standing for a glycan branching point with either α1,3- or α1,4-Fuc. The consensus structure illustrated in Figure 7b provides a conceptual basis for understanding a large part of glycoepitopes by their shared architecture that is similar to secondary structure elements in proteins. For comparison, an α-helix as a typical secondary structure of proteins is depicted in Figure 7c. The exemplary heptapeptide contains 12 rotatable backbone angles, indicating the fragility related to large degree of freedom. Four conventional N–H...O hydrogen bonds (≈5–6 kcal mol<sup>-1</sup> each) are required to stabilize this α-helix. In contrast, the [3,4]F-branched oligosaccharides (Figure 7b) are already quite restricted in their conformation due to three pyranose rings with a favored pucker and only four torsion angles of the two glycosidic linkages that are rotatable within a certain range avoiding steric hindrance. A C–H...O hydrogen bond between Fuc H5 and O5 of the stacking pyranose would be able to stabilize the conformation even with a much smaller energetic contribution in comparison to an α-

helix. In the case of  $\text{Le}^x$ , we concluded earlier from *ab initio* calculations a stabilization energy between Gal and Fuc of  $\approx 4.5 \text{ kcal mol}^{-1}$  consisting of stacking interactions ( $\approx 2.7 \text{ kcal mol}^{-1}$ ) and a C–H $\cdots$ O hydrogen bond ( $\approx 1.8 \text{ kcal mol}^{-1}$ )<sup>[12]</sup> that restrains the four glycosidic torsion angles, resulting in a comparable stabilization energy per rotatable bond as in an  $\alpha$ -helix ( $1.1 \text{ kcal mol}^{-1}$  versus  $1.6\text{--}2.0 \text{ kcal mol}^{-1}$ ). Steric hindrance at the glycosidic linkages or sometimes between bulky substitutions, and further stabilization by the *exo*-anomeric effect might play an additional role. There could be even an additional contribution from a second C–H $\cdots$ O hydrogen bond between Fuc H5 and GlcNAc O4, which displays a H–O distance of 2.7–2.9 Å in our refined structures and 2.5–2.7 Å in the ABUCEF crystal structure (no force field applied) as suggested but not experimentally confirmed by Battistel et al.<sup>[13]</sup> This scenario would be called a 3-centered hydrogen bond because the very same Fuc H5 is involved. Such shared hydrogen bonds have been reported for example in collagen previously.<sup>[42]</sup> However, the observed H5–O4 distances are in the range of the sum of the van der Waals radii of about 2.6 Å at the observed C–H–O angle of  $145^\circ$ .<sup>[38]</sup> The close distances between H5 and the two oxygens have been already noticed in  $\text{Le}^a$  in 1980 by Lemieux et al.,<sup>[6]</sup> but at this time non-conventional hydrogen bonds were not widely known or considered.

In analogy to an  $\alpha$ -helix in which the variety of amino acid side chains ( $R'_1\text{--}R'_7$ ) provide a specific surface for interactions, the recognition of the glycoepitopes (Figure 7b) is mediated by their exposed functional groups ( $R_1\text{--}R_7$ ). Small modifications of the substitutions change the glycoepitope that can lead to markedly different recognition events, for example a 6'-sulfo-modification makes  $\text{sLe}^x$  invisible for L-selectin<sup>[43]</sup> but generates a glycoepitope that can be recognized by Siglec-8.<sup>[44]</sup>

Carbohydrates with a pre-defined, biologically active solution structure are perfect ligands because they do not need to change their conformation upon binding and thus have the advantage of a lower entropic penalty and a faster on-rate ( $k_{\text{ON}}$ ) upon binding to their target lectin. An example is 3FChB, which is bound by the lectin CCL2 with a moderately high affinity ( $K_D = 1 \mu\text{M}$ ) and adopts the same conformation in solution and in the complex with the lectin.<sup>[20]</sup> The advantage of a pre-organized conformation is best illustrated by  $\text{sLe}^x$  analogues that were developed as E-selectin antagonists.<sup>[10b,45]</sup> In the case of a replacement of the GlcNAc moiety by either cyclohexane-1,2-diol or flexible ethylenglycol, a more than 100-fold higher inhibitory concentration  $\text{IC}_{50}$  is observed for the flexible analogue compared to the cyclohexane derivative. In the case of the more rigid cyclohexane-1,2-diols, a Fuc H5 chemical shift of 4.6–4.77 ppm indicates the presence of a C–H $\cdots$ O hydrogen bond, whereas for the flexible version a chemical shift of only 4.12 ppm is observed, indicating the lack of such stabilization.

Our studies do not exclude a small degree of dynamics like a fast transition to a very low population of an extended conformation as suggested by Topin et al.<sup>[40]</sup> Such an exchange with low populated higher energy states have been experimentally observed in structured biomolecules like proteins<sup>[46]</sup> and DNA,<sup>[47]</sup> and their presence is probably more common

than expected. However, the observed downfield chemical shifts resulting from the C–H $\cdots$ O hydrogen bond are time averaged and indicate with values very close to the ones obtained by quantum mechanics that in the overwhelming majority of the time, the conformation is in the closed state. The existence of such low abundant dynamics in  $\text{Le}^a$  is supported by the small temperature-dependence of the downfield chemical shift of Fuc H5 ( $\Delta\delta/\Delta T = -2.2 \text{ ppb K}^{-1}$ ).<sup>[48]</sup> In the rare case that a protein is able to stabilize an open, extended conformation, the chemical shift of Fuc H5 will not be downfield shifted anymore as our GIAO calculations of previously reported open conformations suggest (Figure S6 in the Supporting Information).

Among the glycans adopting the presented scaffold are physiologically as well as pathophysiologically highly important glycoepitopes:  $\text{sLe}^x$  and 6-sulfo  $\text{sLe}^x$ , natural ligands of selectins that are involved in leucocyte rolling and homing,<sup>[49]</sup>  $\text{sLe}^x$  in the zona pellucida is crucial for human fertilization,<sup>[9]</sup> 6'-sulfo  $\text{sLe}^x$  is specifically recognized by Siglec-8,<sup>[21,44]</sup> an immunosuppressive co-receptor present on eosinophiles, whereas 6-sulfo  $\text{sLe}^x$  is a ligand for Siglec-9.<sup>[50]</sup> In addition, LDNF is a central part of an allergenic sea squirt pentasaccharide<sup>[25]</sup> and is present in glycans of parasitic helminths,<sup>[24]</sup> and  $\alpha$ 1,3-fucosylated *N*-glycan cores, major allergic determinants of pollen and insect venoms,<sup>[23]</sup> are found as well. Surprisingly, fucosylated chondroitin sulfate found in crab<sup>[27]</sup> and sea cucumber<sup>[51]</sup> that was earlier promoted as selectin antagonist,<sup>[51]</sup> is also among the stabilized glycoepitopes.

## Conclusion

The discovery of a secondary structure element that includes a C–H $\cdots$ O hydrogen bond and thus stabilizes carbohydrate conformations adds a new dimension to the glycode<sup>[52]</sup> and we speculate that more such stabilizing interactions among glycoepitopes remain to be uncovered. Critical for our present investigation were carbohydrate databases that included searchable chemical shifts and structural data,<sup>[22,53]</sup> and we are convinced that such databases will reveal other exciting insights into carbohydrate structures in the future.

## Experimental Section

**Carbohydrate samples:**  $\text{Le}^x$  methyl glycoside,  $\text{Le}^a$  methyl glycoside, Fuc $\alpha$ 1,3GlcNAc methyl glycoside and LDFT were purchased from Carbosynth (UK). The synthesis of LDNF and of 3FChB is described in the Supporting Information. Bv9 was a generous gift of the laboratory of Yann Guerardel (Lille). Carbohydrates were dissolved in either  $\text{D}_2\text{O}$  or 94%  $\text{H}_2\text{O}/6\%$   $\text{D}_2\text{O}$  with concentrations between 2.8 and 3.7 mM.

**NMR spectroscopy:** All spectra were acquired on Bruker Avance III 500 MHz, 600 MHz and 900 MHz spectrometers equipped with cryogenic triple-resonance probes. Spectra were processed in Topspin 2.1 (Bruker, Germany) and analyzed in Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). Resonance assignment was achieved with 2D  $^1\text{H}\text{--}^{13}\text{C}$  HSQC, 2D  $^1\text{H}\text{--}^{13}\text{C}$  HMBC and 2D  $^1\text{H}\text{--}^1\text{H}$  TOCSY spectra with mixing times ranging from 13–120 ms. 2D  $^1\text{H}\text{--}^1\text{H}$  NOESY experiments were recorded with mixing times of 150 ms. The temperature was calibrated with

MeOD according to Findeisen et al.<sup>[54]</sup> NMR tubes that withstand freezing aqueous solutions (5 mm TA, Armar, Switzerland) were used for all measurements. Samples dissolved in D<sub>2</sub>O did not freeze even at 273 K. All spectra are referenced to 2,2-dimethyl-2-silapentanesulfonic acid (DSS). <sup>13</sup>C chemical shifts were indirectly referenced using a scaling factor  $\Xi$  of 0.251449530 according to Markley et al.<sup>[55]</sup> Assigned <sup>1</sup>H-<sup>13</sup>C HSQC spectra of all investigated oligosaccharides are shown in Figure S7 in the Supporting Information.

**NMR structure calculations:** NOE cross-peaks extracted from 2D <sup>1</sup>H-<sup>1</sup>H NOESY spectra were quantified and converted into proton-proton distances that were used in structure calculations as upper-limit restraints (Tables S3, S4, S6-S9). More precisely, signal to noise (S/N) ratios of the NOE cross peaks were extracted using Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). The S/N ratios were correlated to distances assuming a  $r^{-6}$  dependence. The S/N ratios of cross peaks originating from degenerate CH<sub>2</sub> or CH<sub>3</sub> groups were divided by 2 or 3, respectively. Distances were calibrated with H3-H5 correlations, normally nicely isolated in the NOESY spectra that correspond to a distance of 2.65 Å in a typical chair conformation. Initial calibration attempts with the fixed H61-H62 distance of 1.77 Å gave similar results, but due to the proximity of those signals to the diagonal, their S/N ratios could not be extracted reliably for all carbohydrates. Initial coordinates were obtained by the Biomolecule Builder from the GLYCAM website (<http://glycam.ccruc.uga.edu/AMBER/index.html>) that were used to generate carbohydrate parameters in the CYANA library file. Preliminary structures were calculated by CYANA<sup>[28]</sup> starting from 300 structures with randomized conformation using upper limit constraints calculated from the S/N ratios and expanded by a 0.2 Å tolerance as explained in Tables S3, S4, S6-S9 in the Supporting Information and the associated footnotes. The 30 best structures (lowest target function) were further refined with Amber 9.0<sup>[56]</sup> using the Glycam06 force field<sup>[29]</sup> together with an implicit solvent model. 20 structures with the lowest distance violations were used for the final structural ensemble.

**Databank search for specific glycan chemical shifts:** The chemical shift search of GLYCOSCIENCES.de<sup>[22]</sup> was used to search for fucose H5 chemical shifts between 4.5 and 5.5 ppm. The glycan structures and chemical shifts of the 178 results were extracted and transferred into a file. For glycans containing multiple fucose residues, the chemical shifts were unambiguously assigned to each of them. The list contained one line per chemical shift with the associated glycomotif that was highlighted in case of a larger glycan. Finally the entries were ordered and categorized after their glycomotif. Data from five publications<sup>[26,27,30,57]</sup> were added to support categories with only few data. The full list of results is found in Table S2 in the Supporting Information. To estimate the percentage of glycans containing a C-H...O hydrogen bond, all entries containing a H5 (Fuc) chemical shift were searched, revealing 628 entries. The 178 entries that displayed a H5 (Fuc) chemical shift of 4.7–5.0 ppm present then about 25%. A similar search in structural data is described below.

**Databank search for glycan coordinates:** A substructure search within GLYCOSCIENCES.de<sup>[22]</sup> was used to search protein structures for the individual glycomotifs. A broader search for X-β1,4-[Fucα1,3]-Y and X-β1,3-[Fucα1,4]-Y did not reveal any additional results. However, a search using the GlycomeDB<sup>[53]</sup> and the protein database (PDB) directly revealed additional PDB entries. The obtained protein crystal structures containing either glycans or carbohydrate ligands were used to extract the distances between Fuc C5 and O5 of the stacking monosaccharide using XtalView.<sup>[58]</sup> To all structures, protons were added with Maestro (Schrödinger) and

the distances measured between Fuc H5 and O5 of the stacking monosaccharide in order to detect C-H...O hydrogen bonds. The results are summarized in Table S12. The percentage of fucose-containing glycans that are stabilized by a C-H...O hydrogen bond was estimated from 4271 entries containing an α-L-fucose, 232 entries containing X-β1,3-[Fucα1,4]-Y and 910 entries containing X-β1,4-[Fucα1,3]-Y as found in GLYCOSCIENCES.de.

**Accession codes:** The atomic coordinates, chemical shifts, and restraints used for the structure calculations were deposited with the help of Protein Data Bank Japan (PDBj) and PDBj-BMRB in the Biological Magnetic Resonance Bank (BMRB) with the accession codes 21031, 21032, 21034, 21053, and 21054. The LDFT structure and restraints were deposited at the Protein Data Bank (PDB) with accession number 2MK1 (PDB accepts oligosaccharides with four or more residues) and the chemical shifts at the BMRB with accession number 19748.

## Computational Methods

**Estimation of the correlation time  $\tau_c$  and the nuclear Overhauser enhancement:** The correlation time was estimated according to the Stokes' law and an estimation of the hydrodynamic radius from the molecular weight according to Cavanagh et al.,<sup>[59]</sup> assuming an average specific volume of the carbohydrate of 0.65 cm<sup>3</sup>g<sup>-1</sup> and a hydration layer of 1.6 Å. Viscosity values reported for D<sub>2</sub>O<sup>[60]</sup> were used. The nuclear Overhauser enhancement was calculated for a transient NOE experiment in dependence of the magnetic field according to Neuhaus and Williamson.<sup>[61]</sup>

**Geometry optimization and chemical shift calculation:** The geometry of one representative of each NMR structure was optimized using DFT<sup>[31]</sup> at the B3LYP/6-31G(d,p) level of theory.<sup>[32]</sup> Solvent effects were accounted for applying the CPCM model (implicit solvent).<sup>[62]</sup> NMR shielding tensors were obtained using the gauge-independent atomic orbital (GIAO) method.<sup>[34]</sup> Protons were added to the open Le<sup>x</sup> structures from the PDB by Maestro (Schrödinger). Absolute chemical shift values were referenced to tetramethylsilane (fully optimized in solvent using the same level of theory). Vibrational mode analysis was performed at the optimized geometry to confirm the stability of the obtained minimum. No imaginary frequencies were found. All ab initio geometry optimizations and chemical shift calculations were performed using Gaussian 09.<sup>[35]</sup>

**Bond critical point of C-H...O hydrogen bonds:** The molecular wave functions obtained at the B3LYP/6-31G(d,p) level were exported to the program AIMAll,<sup>[63]</sup> which was used to localize and characterize the bond critical points on the basis of the quantum theory of atoms in molecules (AIM).<sup>[33]</sup>

## Acknowledgements

We are grateful for the generous gift of the Bv9 oligosaccharide from GlycoBase, Emmanuel Maes, and Yann Guerardel ([glycobase.univ-lille1.fr](http://glycobase.univ-lille1.fr)). We thank Gerhard Wider (ETH Zürich) for his suggestion to measure at low temperature, Markus Blatter (ETH Zürich) for his help with Amber, Oliver Schwardt (University of Basel) for measuring high resolution mass spectra, Thomas Lütteke (University of Giessen) and Martin Frank (Biognos AB) for fruitful discussions and keeping the Glycosciences.de database running, Robert J. Woods (University of Georgia and NUI Galway) for providing the Glycam website, Markus Künzler for his comments on the manuscript, and Daniel Roe (University of Utah) and Garib Murshudov (MRC in Cambridge)

for discussing the handling of C–H...O hydrogen bond in force fields. Authors would like to thank Steve Scheiner (Utah State University) for his kind help with disentangling the contribution of the C–H...O bond to the shielding using QM calculations. We like to acknowledge Markus Aebi, Robert Konrat, Yaakov Levi, Andreas Trabesinger and Vikram Panse for helpful discussions and suggestions, María del Carmen Fernández-Alonso for suggesting AIM, and Hans-Heinrich Limbach and José Elguero for sharing their enthusiasm for hydrogen bonds. This work was supported by the Swiss National Science Foundation (SNF sinergia grant CRSII3\_127333).

## Conflict of interest

The authors declare no conflict of interest.

**Keywords:** carbohydrates · hydrogen bond · NMR spectroscopy · secondary structure · solution conformation

- [1] a) J. D. Marth, P. K. Grewal, *Nat. Rev. Immunol.* **2008**, *8*, 874–887; b) M. E. Taylor, K. Drickamer, *Curr. Opin. Cell Biol.* **2007**, *19*, 572–577; c) Y. Y. Zhao, M. Takahashi, J. G. Gu, E. Miyoshi, A. Matsumoto, S. Kitazume, N. Taniguchi, *Cancer Sci.* **2008**, *99*, 1304–1310.
- [2] R. D. Cummings, *Mol. Biosyst.* **2009**, *5*, 1087–1104.
- [3] a) E. Staudacher, F. Altmann, I. B. Wilson, L. Marz, *Biochim. Biophys. Acta Gen. Subj.* **1999**, *1473*, 216–236; b) D. J. Becker, J. B. Lowe, *Glycobiology* **2003**, *13*, 41R–53R; c) B. Ma, J. L. Simala-Grant, D. E. Taylor, *Glycobiology* **2006**, *16*, 158R–184R.
- [4] a) M. R. Combs, *Immunohematology* **2009**, *25*, 112–118; b) P. Stanley, R. D. Cummings in *Structures Common to Different Glycans* (Eds.: A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart, M. E. Etzler), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, **2009**, pp. 175–198.
- [5] J. Heimbürg-Molinario, M. Lum, G. Vijay, M. Jain, A. Almogren, K. Rittenhouse-Olson, *Vaccine* **2011**, *29*, 8802–8826.
- [6] R. U. Lemieux, K. Bock, L. T. J. Delbaere, S. Koto, V. S. Rao, *Can. J. Chem.* **1980**, *58*, 631–653.
- [7] K. Pederson, D. A. Mitchell, J. H. Prestegard, *Biochemistry* **2014**, *53*, 5700–5709.
- [8] B. J. Graves, R. L. Crowther, C. Chandran, J. M. Rumberger, S. Li, K. S. Huang, D. H. Presky, P. C. Familletti, B. A. Wolitzky, D. K. Burns, *Nature* **1994**, *367*, 532–538.
- [9] P. C. Pang, P. C. Chiu, C. L. Lee, L. Y. Chang, M. Panico, H. R. Morris, S. M. Haslam, K. H. Khoo, G. F. Clark, W. S. Yeung, A. Dell, *Science* **2011**, *333*, 1761–1764.
- [10] a) K. E. Miller, C. Mukhopadhyay, P. Gagas, C. A. Bush, *Biochemistry* **1992**, *31*, 6703–6709; b) D. Schwizer, J. T. Patton, B. Cutting, M. Smiesko, B. Wagner, A. Kato, C. Weckerle, F. P. C. Binder, S. Rabbani, O. Schwardt, J. L. Magnani, B. Ernst, *Chem. Eur. J.* **2012**, *18*, 1342–1351; c) A. Imberty, S. Perez, *Chem. Rev.* **2000**, *100*, 4567–4588; d) S. Pérez, N. Mouhous-Riou, N. E. Nifant'ev, Y. E. Tsvetkov, B. Bachet, A. Imberty, *Glycobiology* **1996**, *6*, 537–542; e) H. F. Azurmendi, M. Martin-Pastor, C. A. Bush, *Biopolymers* **2002**, *63*, 89–98; f) Y. Ichikawa, Y. C. Lin, D. P. Dumas, G. J. Shen, E. Garciajunceda, M. A. Williams, R. Bayer, C. Ketcham, L. E. Walker, J. C. Paulson, C. H. Wong, *J. Am. Chem. Soc.* **1992**, *114*, 9283–9298.
- [11] R. U. Lemieux, A. A. Pavia, J. C. Martin, K. A. Watanabe, *Can. J. Chem.* **1969**, *47*, 4427.
- [12] M. Zierke, M. Smiesko, S. Rabbani, T. Aeschbacher, B. Cutting, F. H. Allain, M. Schubert, B. Ernst, *J. Am. Chem. Soc.* **2013**, *135*, 13464–13472.
- [13] M. D. Battistel, H. F. Azurmendi, M. Frank, D. I. Freedberg, *J. Am. Chem. Soc.* **2015**, *137*, 13444–13447.
- [14] S. Horowitz, R. C. Trievel, *J. Biol. Chem.* **2012**, *287*, 41576–41582.
- [15] E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci, D. J. Nesbitt, *Pure Appl. Chem.* **2011**, *83*, 1637–1641.
- [16] F. Cordier, M. Barfield, S. Grzesiek, *J. Am. Chem. Soc.* **2003**, *125*, 15750–15751.
- [17] a) R. Vargas, J. Garza, D. A. Dixon, B. P. Hay, *J. Am. Chem. Soc.* **2000**, *122*, 4750–4755; b) S. Scheiner, T. Kar, Y. Gu, *J. Biol. Chem.* **2001**, *276*, 9832–9837.
- [18] E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. Kent Wenger, H. Yao, J. L. Markley, *Nucleic Acids Res.* **2007**, *36*, D402–408.
- [19] a) S. Scheiner, *Molecules* **2016**, *21*, 1426; b) S. Scheiner, *ChemPhysChem* **2016**, *17*, 2263–2271.
- [20] M. Schubert, S. Bleuler-Martinez, A. Butschi, M. A. Walti, P. Egloff, K. Stutz, S. Yan, I. B. Wilson, M. O. Hengartner, M. Aebi, F. H. Allain, M. Kunzler, *PLoS Pathog.* **2012**, *8*, e1002706.
- [21] J. M. Pröpster, F. Yang, S. Rabbani, B. Ernst, F. H. Allain, M. Schubert, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E4170–4179.
- [22] T. Lutteke, A. Böhne-Lang, A. Loss, T. Goetz, M. Frank, C. W. von der Lieth, *Glycobiology* **2006**, *16*, 71R–81R.
- [23] K. Hoffmann-Sommergruber, K. Paschinger, I. B. Wilson, *Biochem. Soc. Trans.* **2011**, *39*, 360–364.
- [24] a) M. Wuhrer, C. A. Koeleman, J. M. Fitzpatrick, K. F. Hoffmann, A. M. Deelder, C. H. Hokke, *Glycobiology* **2006**, *16*, 991–1006; b) I. van Die, V. Gomord, F. N. Kooyman, T. K. van den Berg, R. D. Cummings, L. Verdelde, *FEBS Lett.* **1999**, *463*, 189–193; c) C. Aranzamendi, B. Tefsen, M. Jansen, L. Chiumiento, F. Bruschi, T. Kortbeek, D. F. Smith, R. D. Cummings, E. Pinelli, I. Van Die, *Exp. Parasitol.* **2011**, *129*, 221–226.
- [25] Y. Kato, M. Ohta, T. Munakata, M. Fujiwara, N. Fujii, S. Shigeta, F. Matsuura, *Magn. Reson. Chem.* **2001**, *39*, 259–266.
- [26] A. Coppin, D. Florea, E. Maes, D. Cogalniceanu, G. Strecker, *Biochimie* **2003**, *85*, 53–64.
- [27] H. Kitagawa, Y. Tanaka, S. Yamada, N. Seno, S. M. Haslam, H. R. Morris, A. Dell, K. Sugahara, *Biochemistry* **1997**, *36*, 3998–4008.
- [28] P. Gunter in *Protein NMR Techniques, Methods in Molecular Biology, Vol. 278*, 2nd ed. (Ed.: A. K. Downing), Humana Press, New York, **2004**, pp. 353–378.
- [29] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. Gonzalez-Outeirino, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, *29*, 622–655.
- [30] J. W. Kurutz, L. L. Kiessling, *Glycobiology* **1997**, *7*, 337–347.
- [31] P. Hohenberg, W. Kohn, *Phys. Rev. B* **1964**, *136*, B864–B871.
- [32] A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- [33] R. F. W. Bader, *Chem. Rev.* **1991**, *91*, 893–928.
- [34] K. Wolinski, J. F. Hinton, P. Pulay, *J. Am. Chem. Soc.* **1990**, *112*, 8251–8260.
- [35] *Gaussian 09, Revision A.02*, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford, CT, **2016**.
- [36] R. K. Castellano, *Curr. Org. Chem.* **2004**, *8*, 845–865.
- [37] A. Bondi, *J. Phys. Chem.* **1964**, *68*, 441–451.
- [38] T. Steiner, W. Saenger, *J. Am. Chem. Soc.* **1992**, *114*, 10146–10154.
- [39] M. A. Higgins, G. E. Whitworth, N. El Warry, M. Randriantoa, E. Samain, R. D. Burke, D. J. Vocadlo, A. B. Boraston, *J. Biol. Chem.* **2009**, *284*, 26161–26173.
- [40] J. Topin, M. Lelimosin, J. Arnaud, A. Audfray, S. Perez, A. Varrot, A. Imberty, *ACS Chem. Biol.* **2016**, *11*, 2011–2020.
- [41] T. Lutteke, M. Frank, C. W. von der Lieth, *Nucleic Acids Res.* **2004**, *33*, D242–D246.
- [42] J. Bella, H. M. Berman, *J. Mol. Biol.* **1996**, *264*, 734–742.

- [43] J. B. Lowe, *Immunol. Rev.* **2002**, *186*, 19–36.
- [44] B. S. Bochner, R. A. Alvarez, P. Mehta, N. V. Bovin, O. Blixt, J. R. White, R. L. Schnaar, *J. Biol. Chem.* **2005**, *280*, 4307–4312.
- [45] a) G. Thoma, J. L. Magnani, J. T. Patton, B. Ernst, W. Jahnke, *Angew. Chem. Int. Ed.* **2001**, *40*, 1941–1945; *Angew. Chem.* **2001**, *113*, 1995–1999; b) H. C. Kolb, B. Ernst, *Chem. Eur. J.* **1997**, *3*, 1571–1578.
- [46] A. J. Baldwin, L. E. Kay, *Nat. Chem. Biol.* **2009**, *5*, 808–814.
- [47] I. J. Kimsey, K. Petzold, B. Sathyamoorthy, Z. W. Stein, H. M. Al-Hashimi, *Nature* **2015**, *519*, 315–320.
- [48] E. Säwén, F. Hinterholzinger, C. Landersjö, G. Widmalm, *Org. Biomol. Chem.* **2012**, *10*, 4577–4585.
- [49] a) M. Sperandio, *FEBS J.* **2006**, *273*, 4377–4389; b) R. P. McEver, *Curr. Opin. Cell Biol.* **2002**, *14*, 581–586.
- [50] P. R. Crocker, J. C. Paulson, A. Varki, *Nat. Rev. Immunol.* **2007**, *7*, 255–266.
- [51] L. Borsig, L. Wang, M. C. Cavalcante, L. Cardilo-Reis, P. L. Ferreira, P. A. Mourao, J. D. Esko, M. S. Pavao, *J. Biol. Chem.* **2007**, *282*, 14984–14991.
- [52] K. T. Pilobello, L. K. Mahal, *Curr. Opin. Chem. Biol.* **2007**, *11*, 300–305.
- [53] R. Ranzinger, S. Herget, C. W. von der Lieth, M. Frank, *Nucleic Acids Res.* **2011**, *39*, D373–376.
- [54] M. Findeisen, T. Brand, S. Berger, *Magn. Reson. Chem.* **2007**, *45*, 175–178.
- [55] J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright, K. Wuthrich, *J. Biomol. NMR* **1998**, *12*, 1–23.
- [56] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- [57] a) J. Zhang, A. Otter, D. R. Bundle, *Bioorg. Med. Chem.* **1996**, *4*, 1989–2001; b) Y. Ishizuka, T. Nemoto, M. Fujiwara, K. Fujita, H. Nakanishi, *J. Carbohydr. Chem.* **1999**, *18*, 523–533.
- [58] D. E. McRee, *J. Struct. Biol.* **1999**, *125*, 156–165.
- [59] J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, M. Rance, N. J. Skelton, *Protein NMR Spectroscopy: Principles and Practice*, Academic Press Inc., San Diego, **2006**, p.
- [60] C. H. Cho, J. Urquidi, S. Singh, G. W. Robinson, *J. Phys. Chem. B* **1999**, *103*, 1991–1994.
- [61] D. Neuhaus, M. P. Williamson, *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, Wiley-VCH, Weinheim, **2000**, p.
- [62] V. Barone, M. Cossi, *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
- [63] T. A. Keith in *AIMAll (Version 13.05.06)*, Todd A. Keith, TK Christmill Software, Overland Park, KS, **2013**; aim.tkgristmill.com.

---

Manuscript received: April 26, 2017

Accepted manuscript online: June 27, 2017

Version of record online: August 3, 2017

---