ChemPubSoc Europe

DOI: 10.1002/chem.201600467

CHEMISTRY A European Journal Full Paper

Molecular Modeling

Glycomimetics Targeting Glycosyltransferases: Synthetic, Computational and Structural Studies of Less-Polar Conjugates

Mattia Ghirardello,[a] Matilde de las Rivas,[b] Alessandra Lacetera,[c] Ignacio Delso,[a, d] Erandi Lira-Navarrete,[a] Tomás Tejero,[a] Sonsoles Martín-Santamaría,*[c] Ramón Hurtado-Guerrero,*[b, e, f] and Pedro Merino*[a]

Abstract: The Leloir donors are nucleotide sugars essential for a variety of glycosyltransferases (GTs) involved in the transfer of a carbohydrate to an acceptor substrate, typically a protein or an oligosaccharide. A series of less-polar nucleotide sugar analogues derived from uridine have been prepared by replacing one phosphate unit with an alkyl chain. The methodology is based on the radical hydrophosphonylation of alkenes, which allows coupling of allyl glycosyl compounds with a phosphate unit suitable for conjugation to uridine. Two of these compounds, the GalNAc and galactose derivatives, were further tested on a model GT, such as GalNAc-T2 (an important GT widely distributed in human tissues), to probe that both compounds bound in the medium–high micromolar range. The crystal structure of GalNAc-T2 with the galactose derivative traps the enzyme in an inactive form; this suggests that compounds only containing the β-phosphate could be efficient ligands for the enzyme. Computational studies with GalNAc-T2 corroborate these findings and provide further insights into the mechanism of the catalytic cycle of this family of enzymes.

Introduction

Glycosyltransferases (GTs) are key enzymes responsible for the incorporation of carbohydrates into a variety of acceptor biomolecules, including proteins, lipids, oligosaccharides and different metabolites.[1] The resulting glycoconjugates mediate a wide range of functions from structure and storage to signalling and, as a consequence, they are related to important diseases. Therefore, the chemical manipulation of the activity of GTs could lead to the development of useful therapeutic drugs.[2] Thus, considerable synthetic efforts have been directed toward the preparation of efficient GT inhibitors.[3] Transfer of the sugar residue occurs from an anionic nucleotide sugar donor to the acceptor substrate; it can take place with retention or inversion of configuration at the anomeric centre of the sugar residue.[4] In this context, elucidation of the mechanisms used by GTs has been pursued with much interest.[5] Notably, only nine sugar donors are known to be involved in protein glycosylation, which is the most abundant post-translational modification in nature, in mammals.[6] Six of these sugar donors contain the uridine moiety (Figure 1), which is in agreement with the existence of GTs employing uridine diphosphate (UDP) sugars as the most predominant in nature.[7]

Most of designed inhibitors mimic nucleotide phosphate sugars by incorporating anionic groups to emulate binding of the diphosphate bridge.[8] However, such compounds are not capable of permeating into cells due to their high polarity.[9] To overcome this problem, neutral inhibitors have also been prepared,[10] and derivatives in which the phosphate group was replaced by a different apolar group showed enhanced cell internalisation,[10f,g] although some reduction in binding affinity occurred. Vocadlo and co-workers demonstrated that UDP-5SGlcNAc acted as an inhibitor of O-GT,[11] but it should be generated inside the cell from precursor 5SGlcNAc by using the salvage pathway. Nevertheless, most of the designed compounds resulted in poor biological activity because they did

[a] M. Ghirardello, Dr. I. Delso, Dr. E. Lira-Navarrete, Prof. T. Tejero, Prof. P. Merino
Departamento de Síntesis y Estructura de Biomoléculas
Instituto de Síntesis Química y Catálisis Homogénea (ISQCH)
Universidad de Zaragoza, CSIC, 50009 Zaragoza, Aragón (Spain)
E-mail: pmerino@unizar.es

[b] M. d. l. . Rivas, Dr. R. Hurtado-Guerrero
Instituto de Biocomputación y Fisica de Sistemas Complejos (BIFI)
BIFI-IQFR (CSIC) Joint Unit, Universidad de Zaragoza
50009 Zaragoza (Spain)
E-mail: rhurtado@bifi.es

[c] A. Lacetera, Dr. S. Martín-Santamaría
Departamento de Biología Físico-Química
Centro de Investigaciones Biológicas
CIB-CSIC, Ramiro de Maeztu, 9, 28040 Madrid (Spain)
E-mail: smsantamaria@cib.csic.es

[d] Dr. I. Delso
Servicio de Resonancia Magnética Nuclear
Centro de Química y Materiales de Aragón (CEQMA)
Universidad de Zaragoza, CSIC, Campus San Francisco
50009 Zaragoza (Spain)

[e] Dr. R. Hurtado-Guerrero
Fundación ARAID, 50018 Zaragoza (Spain)

[f] Dr. R. Hurtado-Guerrero
Instituto de Investigaciones Sanitarias de Aragón (IIS-A)
Zaragoza 50009 (Spain)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under http://dx.doi.org/10.1002/chem.201600467.

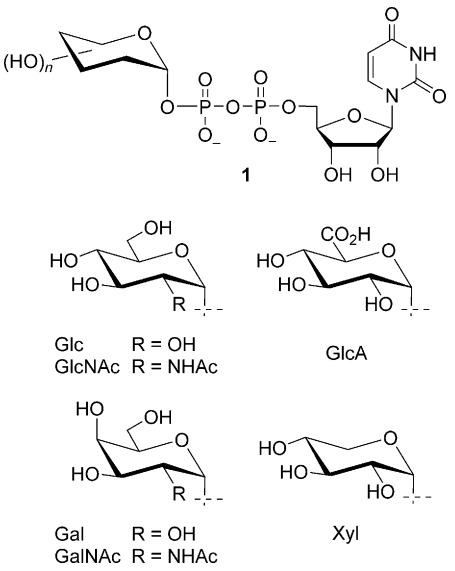


Figure 1. Uridine-based sugar donors.

not incorporate the nucleoside moiety, which diminished recognition and selectivity by the target enzymes.

Herein, we report the design, preparation and binding studies of less-polar uridyl–sugar analogues **6** as suitable GT binders in which the β-phosphate unit has been replaced by an alkyl chain (Figure 2).

Insertion of methylene (**2**)[12] and replacement of the anomeric oxygen by methylene (**3**)[13] or propylene (**4**)[14] groups has already been reported, but compounds **2–4** are still too polar due to the presence of two phosphate units (Figure 2). Compound **3a** was evaluated against three bacterial UDP–galactopyranose mutases and showed only moderate inhibition, although crystallographic studies showed that it was bound to the active site of the enzyme in a novel conformation not observed previously.[13d] As a model enzyme, we have selected GalNAc-T2; a member of the important group of *N*-acetylgalactosaminyl transferases (GalNAcTs E.C. 2.4.1.41) involved in mucin biosynthesis.[15] In particular, a previous computational study with GalNAc-T2 reported by Masgrau and co-workers anticipated that UDP-2′-deoxyGal would not be a good donor substrate, whereas UDP-Gal could be a valid sugar donor in some cases.[16]

Based on these precedents, we designed analogues **6**, in which one phosphate group was replaced by an ethylene group, as suitable GT binders with GalNAc-T2 as a model enzyme. The synthetic strategy is based on coupling between uridine and phosphoalkyl sugars, which are prepared through radical hydrophosphonylation[17] of allyl sugars.

We also demonstrate through tryptophan fluorescence spectroscopy and X-ray studies that two of these compounds bind moderately to the enzyme. In particular, trapping of a galactose moiety containing compound bound to an inactive form of GalNAc-T2 provides hints for improving their affinity through the incorporation of β-phosphate or similar negatively charged groups. Furthermore, in silico studies support the aforementioned findings and that the presence of only one phosphate

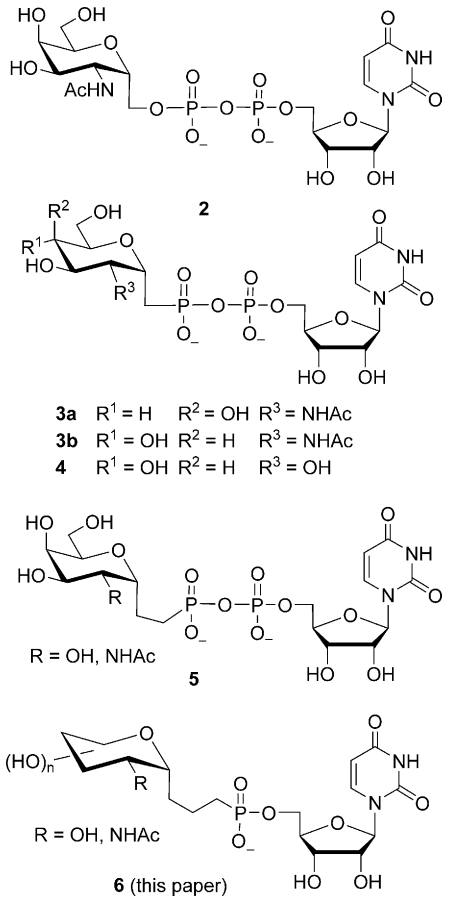


Figure 2. Structures of the C-analogues of uridine dinucleotide sugars discussed herein.

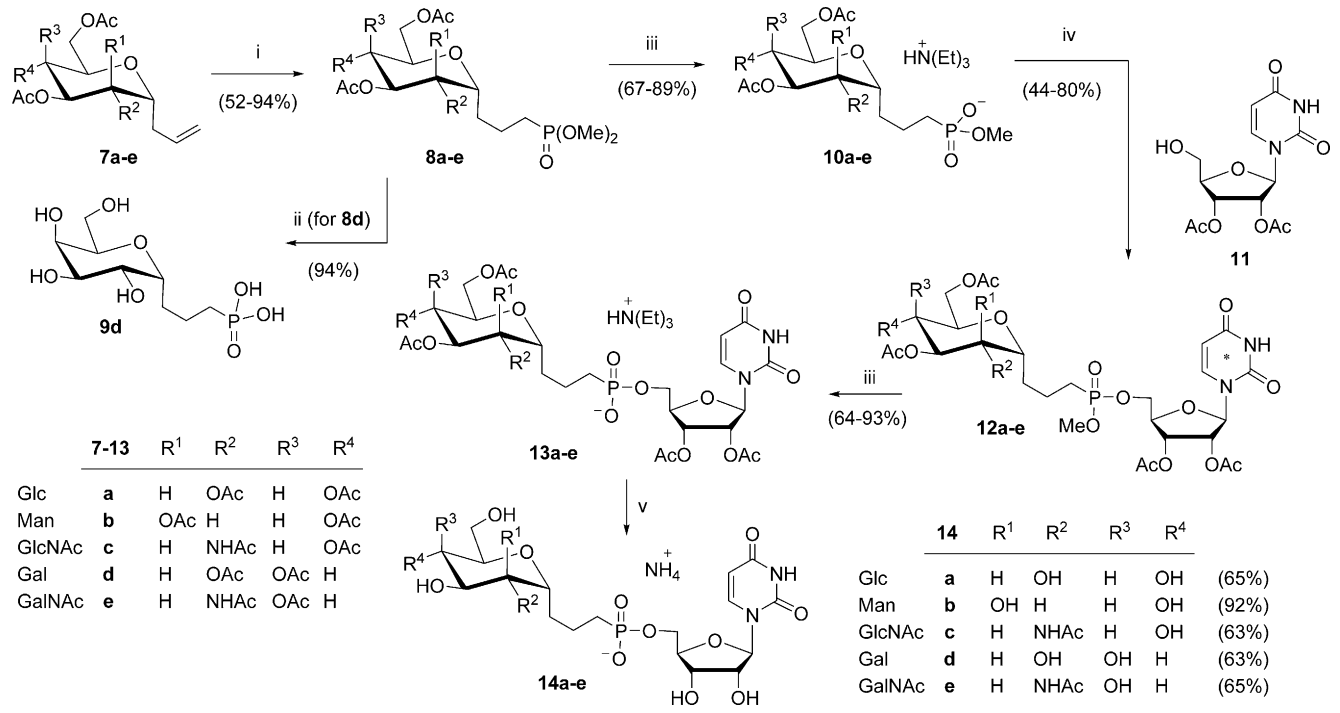
group close to the sugar moiety could be enough for the uridyl–sugar analogue to interact with the protein. These combined studies also provide new evidence of the mechanism of the catalytic cycle of GalNAc-T2.

Results and Discussion

Synthesis of glycomimetics

The starting allyl sugars **7** used herein were prepared from D-glucose for allylsugar **7a**,[18] D-mannose for allylsugar **7b**,[19] D-glucosamine for allylsugar **7c**[20] and D-galactose for allylsugars **7d**[19,21] and **7e** (Scheme 1).[22] Protected diacetyl uridine **11** was prepared from commercially available uridine via the trityl derivative.[23]

Photoinduced free-radical hydrophosphonylation of **7a–e**, following the protocol developed by Dondoni and co-workers,[24] afforded phosphonates **8a–e** with complete regioselectivity and good yields (Scheme 1). Removal of the protecting groups of **8d** under typical conditions[25] led to free phosphonate **9d** in excellent yield. Any attempt to couple compound **9d** with **10** in the presence of *N*,*N*′-dicyclohexylcarbodiimide (DCC)[26] failed, as well as deprotection of both OMe groups. Izumi and co-workers previously reported the condensation of a phosphate monoester with a nucleoside in good yield.[27]



Scheme 1. Synthesis of partially free phosphonates. i) dimethyl phosphite, dieethyl phosphate acetophenone (DPAP), *hν*, 30 min, RT; ii) Me_3SiCl, NaBr, MeCN, 1 h, 40 °C; then 2 h, RT; then 5:1 H_2O/EtOAc, 2 h, RT; then NH_4OH, 2 h, RT; iii) PhSH, Et_3N, dioxane, 4 h, RT; iv) **10** (1.0 equiv), (benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (BOP), DMF, *i*$PrEt_2N$, 4 h, RT; v) 25% $NH_4OH_{(aq)}$, MeOH, 7 h, RT.

Therefore, we decided to prepare monoesters **10a–e** through selective deprotection of the phosphonate group by using thiophenol and triethylamine.[28] Condensation of monoesters **10a–d** and diacetyluridine **11** by using BOP and *N*,*N*-diisopropylethylamine (DIEA) in DMF[29] yielded the fully protected nucleotide sugar analogues **12a–e** (as a mixture of diastereomers due to phosphorous chirality) with good yields. Removal of the methyl group of the phosphonate[28] gave **13a–e**.

Finally, deacetylation of **13a–e** with a 25% aqueous solution of ammonium hydroxide afforded free sugar nucleotide analogues **14a–e** in good yields after purification through a Bio-Gel P-2 column (eluted with NH_4HCO_3 in 4:1 H_2O/MeOH), followed by lyophilisation.

Crystallographic and binding studies

To validate whether our compounds might have a biological effect, we selected the GT GalNAc-T2 as an example. This enzyme is part of a large family of 20 members that transfer a GalNAc moiety from UDP–GalNAc onto Ser or Thr residues of proteins in the presence of manganese.[30] This enzyme predominantly uses UDP–GalNAc as the sugar donor, although, in some cases, it might also use UDP–galactose.[16,31] Prompted by this, we determined whether compounds **14d** and **14e**, that contained a galactose and GalNAc moiety, respectively, might bind to this enzyme. Tryptophan fluorescence spectroscopy studies for compound **14d** gave K_d values of (269±40) and (254±20) μM, and those for compound **14e** gave K_d values of (800±10) and (915±180) μM, in the absence and presence of 2 mM Mn^{2+} (Figure 3), respectively; this suggested that interactions with Mn^{2+} for these ligands might not be crucial for their binding. Contrary to the preference of GalNAc-T2 for UDP–GalNAc as the favourite donor substrate for catalysis, the results also suggest that the compound with the galactose moiety (**14d**) binds about 3.5-fold better than that with the GalNAc moiety (**14e**).

Compound **14d** traps the enzyme in an inactive state. To understand the binding mode of these compounds to GalNAc-T2, we solved the crystal structure of GalNAc-T2 in a complex

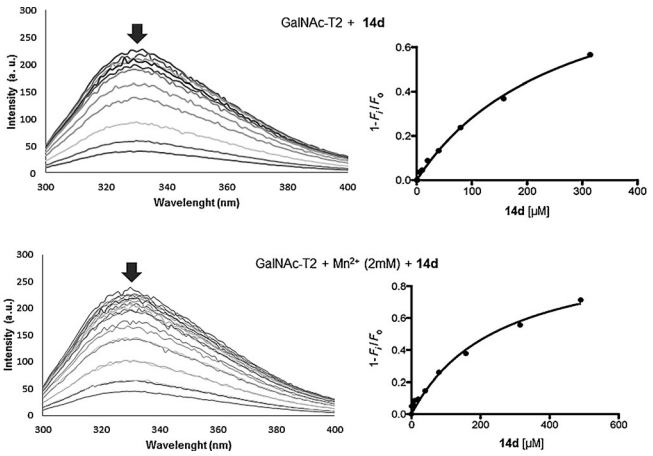


Figure 3. Quenching of intrinsic GalNAc-T2 tryptophan fluorescence measured with increasing concentrations of **14d** in the absence and presence of Mn^{2+}. All data points represent the means ± standard deviations (S.D.) for three measurements. The K_d value for **14d** was determined by fitting fluorescence intensity data against the concentration of **14d**.

with **14d**. Despite numerous attempts with both compounds (**14d** and **14e**) in cocrystallisation and soaking experiments, we only managed to obtain a complex with **14d** through soaking experiments for orthorhombic crystals previously grown with UDP/Mn^{2+}. The structure was solved at 2.07 Å, which allowed us to solve and interpret the density maps (Table S1 in the Supporting Information).

The asymmetric unit, as reported previously, displayed six molecules of GalNAc-T2.[31] Three out of six GalNAc-T2 molecules present in the asymmetric unit clearly showed density maps for the presence of both UDP/Mn^{2+} and **14d** in each active site, whereas the three remaining GalNAc-T2 molecules contained UDP (Figure S1 in the Supporting Information). The crystal structure shows that the typical GT-A fold (catalytic domain) is located in the N-terminal region and the lectin domain is located in the C-terminal region (Figure 4).

Strikingly, GalNAc-T2 in complex with **14d** adopts an inactive-state conformation, as reported previously for other binary complexes containing either glycopeptides or UDP (Figure 4).[15a,31,32] Both active and inactive states are present during the catalytic cycle of this enzyme and are associated with the motion of a flexible loop that can oscillate between closed and open conformations, respectively. In particular, UDP and **14d** are exposed to the solvent due to the open conformation of the flexible loop, whereas UDP-5S-GalNAc is covered by the flexible loop, which adopts a closed conformation, rendering the enzyme in an active state (Figure 4; PDB entry 4D0Z).

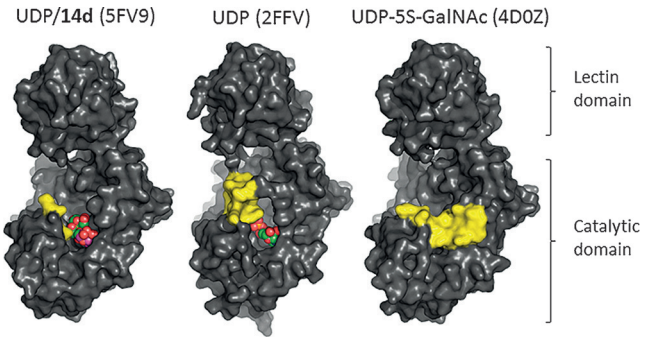


Figure 4. Surface representation of GalNAc-T2 in complex with **14d**/UDP, UDP and UDP-5S-GalNAc. Protein, flexible loop and nucleotides/**14d** are coloured black, dark grey and light grey, respectively. The active and inactive states correspond to closed (4D0Z) and open (5V9 and 2FFV) conformations, respectively.

Compound **14d** induces an inverted uridine conformation and an unusual 4-coordinate Mn^{2+} complex. Closer inspection of the active site reveals that the uridine moiety of both **14d** and UDP adopts an unusual inverted conformation; this was also found earlier in a structure of this enzyme in complex with UDP (Figure 5 and PDB entry 2FFV).[32] This atypical conformation was proposed to represent a final step during the catalytic cycle of GalNAc-T2 in which UDP was ready to exit the enzyme.[31,32] One major difference is apparent between both complexes: the metal is unusually four-coordinate in our crystal structure and five-coordinate in the GalNAc-T2-UDP

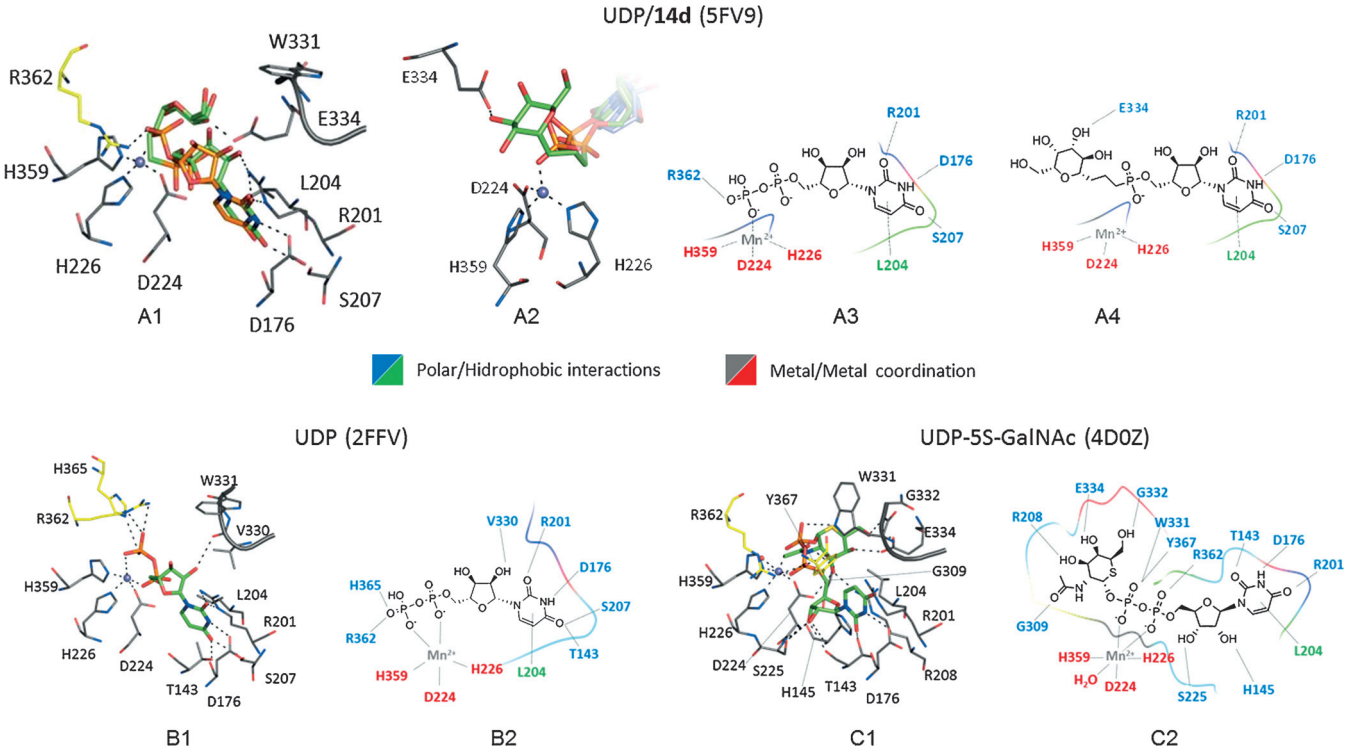


Figure 5. Structural features of the sugar nucleotide binding site for GalNAc-T2 in complex with UDP/**14d**. A1) A close-up view of the active site in complex with UDP-Mn^{2+}/**14d**. A2) Detailed view of metal interactions. Two-dimensional schemes of the interactions between **14d** (A3) and UDP (A4) and the residues of GalNAc-T2. Close-up views of the active site in complex with UDP/Mn^{2+} (B1) and UDP-5S-GalNAc-Mn^{2+} (C1). Two-dimensional schemes of the interactions between the ligands (B2) and residues of GalNAc-T2 (C2).

complex (PDB entry 2FFV; Figure 5). The metal in the former structure is coordinated by the α-phosphate, Asp224, His226 and His359 (Figure 5, A3), whereas in the latter complex the metal is additionally coordinated by the UDP β-phosphate (Figure 5, B1 and B2). However, six-coordinate Mn^{2+}, which requires the previous residues, pyrophosphate and an extra water molecule to form an octahedral complex, is not only the most common coordination in crystal structures of this enzyme, but also the most abundant in proteins in general.[15a,31–34] These results suggest that these UDP–phosphonate scaffolds, with only the α-phosphate, are weaker ligands in terms of binding to GalNAc-T2, but they are still suitable binders for this enzyme.

Previously, different crystal structures obtained from the same type of orthorhombic crystals suggested that UDP might be present in both active and inactive states of the enzyme. On the contrary to our structure, UDP in these crystals always adopted the typical conformation found for UDP-GalNAc, which was required for efficient turnover.[31] Thus, it is tempting to speculate that **14d** has allowed us to trap a conformation in the final step of the catalytic cycle of this enzyme, in which UDP would be ready for immediate departure from the enzyme.

The crystal structure also supports the aforementioned K_d values; this confirms that **14d** does not rely on Mn^{2+} for binding to GalNAc-T2. With regard to direct interactions between the enzyme and nucleotides/**14d** in the different complexes, it is important to note that **14d** shows fewer interactions (Figure 5) than UDP (PDB entry 2FFV) and UDP-5S-GalNAc (PDB entry 4D0Z). It is important to note that Trp331, a critical residue in the catalytic cycle of these enzymes,[32] adopts an "out" (out of the active site) conformation in the complex with **14d**/UDP (Figure 5, A1), which correlates well with the inactive state of GalNAc-T2. On the contrary, Trp331 adopts an "in" (inside the active site) conformation in the complex with UDP-GalNAc that sets up the enzyme in an active state (Figure 5, C1).

Most of the direct interactions between the uridine moiety of **14d** and the enzyme are kept as the UDP from the reported PDB entry 2FFV, whereas a larger number of direct interactions are established between UDP-GalNAc and the enzyme, presumably because the active site is covered by the flexible loop (Figures 4–6). Finally, the phosphonate of **14d** does not interact directly with any residue and the galactose moiety is recognised exclusively by Glu334 (Figure 5, A2); this is likely to be an explanation for the poor binding of this compound to the enzyme. In the docking studies, we found that the Gal moiety adopted several alternative binding poses within the sugar site.

Docking studies

To obtain a 3D picture of the putative binding ability of this family of compounds, docking calculations were performed for analogues **14** in both conformations of GalNAc-T2: closed (active, PDB ID 4D0Z) and open (inactive, PDB ID 2FFV). The calculations were able to predict reasonable docked binding poses for all ligands, although with some slight differences. UDP, UDP-GalNAc and UDP-5S-GalNAc were also docked for comparison purposes to validate the computational protocol. Interestingly, all compounds were predicted to bind in both conformations (open and closed) inside the active site of the enzyme with binding poses similar to those found in the crystallographic structures.

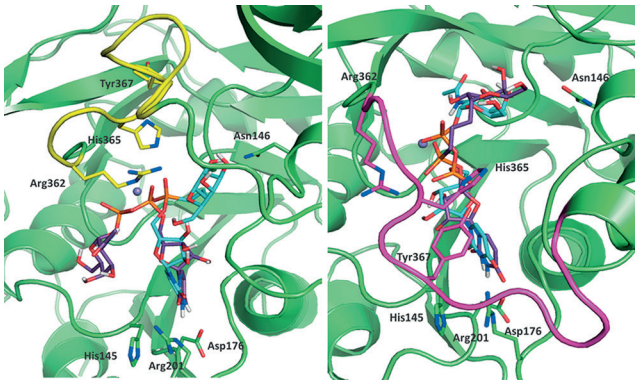


Figure 6. Left: Superposition of docked poses in GalNAc-T2 (PDB-ID 2FFV) of **14d** (cyan) and UDP-GalNAc (purple). The open loop is shown in yellow. Right: Superposition of docked poses in GalNAc-T2 (PDB-ID 4D0Z) of **14d** (cyan) and UDP-GalNAc (purple). The closed loop is shown in magenta.

Predicted binding energies were higher than those of reference diphosphate compounds due to the consequently decreased ionic phosphate–manganese interactions caused by the lack of the β-phosphate group. Nevertheless, docked binding energies were favourable; thus validating the phosphonate scaffold of analogues **14** as a possible platform for the optimisation of GT binders. The best binding energies were always predicted for the active (closed) conformation, which was in agreement with a binding pose with more GalNAc-T2/ligand interactions, although the docking results indicated the ability to bind also to the open conformation, which was preferred in the crystal structure. Both possibilities might coexist in the equilibrium of conformational ensembles. Docking calculations are able to estimate both binding poses, whereas the limitations of X-ray, in this case, have not allowed us to obtain the closed 3D structure. No major differences in the predicted binding energy were found among the phosphonate analogues **14**, which indicated a lack of preferred binding. For compounds **14**, the main contributions to binding are provided by the uridine–enzyme hydrogen bonds, as in the 4D0Z complex, and, to a lesser extent, from ionic phosphonate–Mn^{2+} interactions through one or two of the phosphonate oxygen atoms.

In the case of the active conformation of the GT, an interesting difference was found with regard to binding of the sugar moiety of analogues **14**, which also showed alternative binding poses in which the sugar moiety reached the Ala307 CO group and Asn146 side chain, according to the highly flexible aliphatic chain. For the inactive conformation of GalNAc-T2, the docking calculations placed the synthetic analogues in the equivalent region to that of the active conformation (Figure 6).

The uridine moiety is observed to rotate in a similar way to the crystallographic pose, and the overall conformation of the ligand was found to adopt a more bended conformation, with the sugar moiety placed in front of the uridine plane. For comparison purposes we performed docking calculations for **14d** to our crystallographic structure (PDB 5V9, chain "B ") finding similar results between the crystallographic pose for **14d** and the docked pose for UDP-GalNAc.

We conclude, from the overall docking study, that compounds **14** are putative GalNAc-T2 binders in which the uridine moiety plays a major role in binding, in agreement with the crystallographic binding poses, whereas the phosphonate group and sugar moiety provide additional interactions; there are no differences in the types of sugar.

Targeted molecular dynamics (tMD) simulations

To shed light on the conformational changes that the protein needs to undergo between the inactive and active conformations (opening/closing process), and to analyse the required interactions to host the substrate, targeted molecular dynamics (tMD) simulations were undertaken for the GalNAc-T2/UDP-GalNAc complex. The simulation was started from the complex in the inactive open conformation (from the docked binding pose) towards the complex in the active closed conformation (from the docked binding pose of UDP-GalNAc, very similar to the crystallographic complex with UDP-5S-GalNAc; Figure 7). The analysis of the root-mean-square fluctuation (Figure S2 in the Supporting Information) during the tMD simulation time identified a main fluctuation that corresponded to the flexible loop Arg362–Ser373 with a maximum value for Pro370. This main change corresponds to the folding of the loop over the substrate, sealing it in the active site, and leading to the closed conformation. Additionally to the Asp224, His226 and His359 side chains, the Mn^{2+} ion was found to coordinate to two phosphate oxygen atoms (Figure S3 in the Supporting Information). During the progression of the tMD simulation, one water molecule entered into the coordination sphere, according to a six-coordinate Mn^{2+} model (Figure S4 in the Supporting Information).

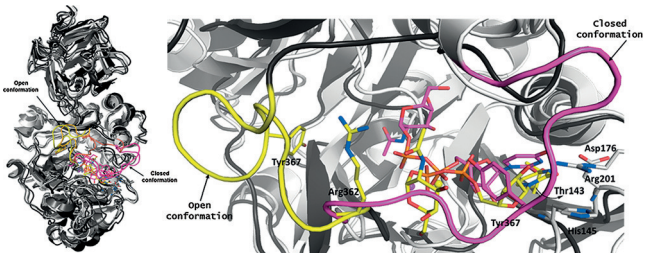


Figure 7. Superposition of five structures of the GalNAc-T2/UDP-GalNAc complex along the tMD simulation of the transition from the open (black, loop in yellow) to the closed (white, loop in magenta) conformation. Intermediate protein structures are shown as a gradient from yellow (starting geometry) to orange (intermediate geometries) to magenta (final target geometry). The ligand UDP-GalNAc is shown in yellow for the open complex and in magenta for the closed complex. Left: Full perspective. Right: Detailed image showing some representative residues.

Several functional groups from GalNAc-T2 establish interactions, especially those involving the uridine moiety. In particular, the NH in position 3 of the uracyl ring establishes a stable hydrogen bond along the simulation with the Asp176 side chain (Figure 7, right, and Figure S5 in the Supporting Information). The CO groups at positions 2 and 4 of the uracyl ring establish alternating hydrogen bonds with Arg201 and Thr143 side chains (Figure 7, right, and Figure S6 in the Supporting Information) during the rotation of the uridine. This rotation allows the torsional change undergone by the diphosphate linker. Also, stacking interactions between the His145 side chain and the uracyl moiety are identified. Most of these interactions are found in the crystallographic structures, in agreement with a specific anchorage for the uracil moiety. On the contrary, the flexibility of the diphosphate linker allows GalNAc to rotate and move during the closing process, which changes the interactions along the simulation to adapt to the new surrounding environment.

For example, in the inactive conformation, GalNAc is involved in two hydrogen bonds: one between the endocyclic GalNAc oxygen and the Arg362 backbone NH, and another one between the N-acetylamido CO group and the Arg362 guanidinium group. After closing, these interactions with Arg362 are lost and new hydrogen bonds are established: one between GalNAc OH-3 and the backbone NH of Gly309, and another one between OH-4 from the GalNAc moiety and the Ala307 CO group. Moreover, new CH–π interactions form between the ribose and the Tyr367 indole ring.

In summary, tMD results may give us an approximation of the binding pose changes that the UDP-GalNAc undergoes during the closing/opening process when bound to the GT. Specific interactions for GalNAc during the closing/opening process may be useful to understand the mechanism of the catalytic cycle of this enzyme, and to help in the design of selective GT binders. In fact, we provide a plausible explanation for the role of the N-acetylamido group in binding, which supports its higher affinity to the enzyme in comparison to UDP-Gal.

Conclusion

GalNAc-T2 has been used as a prototype GT for studying the binding mode of less-polar glycomimetics. Easily accessible sugar nucleotides analogues, in which one phosphate group has been replaced by an apolar alkyl chain, are weaker binders of GalNAc-T2 than the natural substrate UDP-GalNAc, as demonstrated by combined structural and computational studies. However, compound **14d** is capable of trapping the enzyme in an inactive form. Interestingly, docking studies reveal that both compounds can bind to the inactive and active forms of GalNAc-T2, which suggests a more complex scenario. The crystal structure provides additional useful information to design better compounds that might inhibit GalNAc-T2. In particular, it is inferred from the crystal structure that the β-phosphate is required for binding to Mn^{2+}, and thus, replacing the α-phosphate by a propylene group, while keeping the phosphate or

similar groups in the β position, should maintain or improve the binding of these compounds.

Finally, tMD simulations of the complex of GalNAc-T2 with UDP-GalNAc has provided new insights into substrate recognition events involved in the preliminary steps of the catalytic cycle of GalNAc-T2. We have simulated the change from the open, inactive conformation to the closed, active conformation to identify the overall change, with regard to the conformation and interactions with the protein, as experienced by the substrate to finally lead to the active, closed complex. This transition involves the rotation of the uridine and GalNAc moieties and a conformational change of a very dynamic flexible loop. Our studies, combining the synthesis of UDP-GalNAc-based analogues with structural and computational techniques, allowed us a better understanding of the substrate recognition and catalytic cycle of GalNAc-T2.

Experimental Section

General methods

The reaction flasks and other glass equipment were heated in an oven at 130 °C overnight and assembled in a stream of argon. All reactions were monitored by TLC on silica gel 60 F254; the positions of the spots were detected by $\lambda = 254$ nm UV light or by spraying with either a 5% solution of phosphomolybdic acid in ethanol or Mostain solution. Column chromatography was carried out in a Buchi 800 MPLC system or a Combiflash apparatus by using silica gel 60 microns and with solvents that were distilled prior to use. Melting points were uncorrected. Purification by semi-preparative HPLC (column Atlantis® DC18 5 μm, 19×100 mm, flow: 12.5 mL min^{-1}) was carried out in a Waters 515 pump system with photodiode array (PDA) detection. ^{1}H and ^{13}C NMR spectra were recorded on Bruker Avance 400 MHz or AVANCE II 300 MHz instruments in the stated solvent. Chemical shifts are reported in ppm (δ) relative to $CHCl_3$ ($\delta = 7.26$) in $CDCl_3$. NMR assignments were made by using standard 2D experiments. For numbering of NMR data, see Figure 8. Optical rotations were recorded on a JASCO DIP-370 polarimeter. Elemental analyses were performed on a PerkinElmer 240B microanalyzer or with a PerkinElmer 2400 instrument. High-resolution mass spectra were recorded on a QToF spectrometer equipped with an ESI source (microTOF-Q, Bruker Daltonik) by using sodium formate as an external reference. Hydrophosphonylation was carried out in a glass vial (diameter: 1 cm; wall thickness: 0.65 mm), closed with a natural rubber septum, located 2.5 cm away from the household UVA lamp apparatus equipped with four 15 W tubes (1.5×27 cm each).

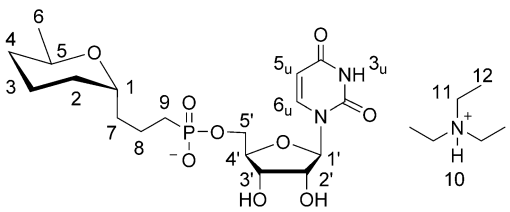


Figure 8. Numbering used for NMR spectroscopy assignments of compounds.

Hydrophosphonylation of allylglycosides 7 a–e: General procedure

DPAP (37.8 mg, 0.148 mmol) was added to a well-stirred solution of allyl glycoside **7** (0.295 mmol) in dimethyl phosphite (2.7 mL, 29.5 mmol). The solution was irradiated at room temperature under stirring for 30 min and then concentrated. The residue was purified by column chromatography on silica gel (EtOAc/MeOH, 1:0 to 95:5 containing 0.2% of Et_3N) to give **8** as an oil.

Dimethyl [(2,3,4,6-tetra-*O*-acetyl-α-D-glucopyranosyl)propyl]-phosphonate (8 a): Yield: 367 mg, 94%; $[\alpha]_D^{27} = +53$ ($c = 1.0$, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$): $\delta = 5.32$ (dd, $J = 9.5$, 8.7 Hz, 1H; H-3), 5.09 (dd, $J = 9.5$, 5.8 Hz, 1H; H-2), 4.99 (dd, $J = 9.6$, 8.7 Hz, 1H; H-4), 4.27 (dd, $J = 12.0$, 5.7 Hz, 1H; H-6), 4.18–4.11 (m, 1H; H-1), 4.09 (dd, $J = 12.0$, 2.6 Hz, 1H; H-6), 3.83 (ddd, $J = 9.6$, 5.7, 2.6 Hz, 1H; H-5), 3.77 (d, $J = 11.0$ Hz, 6H; OCH_3), 2.08 (s, 3H; Ac), 2.04 (s, 3H; Ac), 2.01 (s, 3H; Ac), 2.00 (s, 3H; Ac), 1.94–1.51 ppm (m, 6H; H-7, H-8, H-9); ^{13}C NMR (100 MHz, $CDCl_3$): $\delta = 170.1$, 170.0, 169.9, 169.8, 72.5, 70.5 (2 C), 69.0 (2 C), 62.7, 52.6 (d, $J = 6.3$ Hz), 26.6 (d, $J = 15.5$ Hz), 25.0 (d, $J = 140.5$ Hz), 20.9 (2 C), 20.8 (2 C), 18.7 ppm (d, $J = 5.0$ Hz); ^{31}P NMR (162 MHz, $CDCl_3$): $\delta = 32.0$ ppm; elemental analysis calcd (%) for $C_{19}H_{31}O_{12}P$: C 47.31, H 6.48, P 6.42; found: C 47.56, H 6.30, P 6.24.

Dimethyl [(2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyl)propyl]-phosphonate (8 b): Yield: 254 mg, 42%; $[\alpha]_D^{28} = +7$ ($c = 0.9$, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$): $\delta = 5.22$ (dd, $J = 8.9$, 3.3 Hz, 1H; H-3), 5.17 (dd, $J = 8.9$, 8.5 Hz, 1H; H-4), 5.15 (dd, $J = 3.3$, 3.2 Hz, 1H; H-2), 4.36 (dd, $J = 12.0$, 6.5 Hz, 1H; H-6), 4.10 (dd, $J = 12.0$, 4.0 Hz, 1H; H-6), 3.95 (ddd, $J = 10.6$, 3.8, 3.2 Hz, 1H; H-1), 3.87 (ddd, $J = 8.5$, 6.5, 4.0 Hz, 1H; H-5), 3.74 (d, $J = 9.5$ Hz, 6H; OCH_3), 2.12 (s, 3H; Ac), 2.11 (s, 3H; Ac), 2.06 (s, 3H; Ac), 2.03 (s, 3H; Ac), 1.92–1.60 ppm (m, 6H; H-7, H-8, H-9); ^{13}C NMR (100 MHz, $CDCl_3$): $\delta = 170.7$, 170.5, 170.0, 169.8, 73.8, 70.6 (2 C), 68.8, 66.9, 62.3, 52.2 (d, $J = 9.8$ Hz), 29.2 (d, $J = 15.2$ Hz), 24.0 (d, $J = 140.5$ Hz), 20.8, 20.7, 20.6, 20.5, 18.5 ppm (d, $J = 4.5$ Hz); ^{31}P NMR (162 MHz, $CDCl_3$): $\delta = 31.8$ ppm; elemental analysis calcd (%) for $C_{19}H_{31}O_{12}P$: C 47.31, H 6.48, P 6.42; found: C 47.22, H 6.38, P 6.51.

Dimethyl [(2-acetamido-2-deoxy-3,4,6-tri-*O*-acetyl-α-D-glucopyranosyl)propyl]phosphonate (8 c): Yield: 213 mg, 73%; $[\alpha]_D^{28} = +29$ ($c = 0.9$, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$): $\delta = 6.00$ (d, $J = 8.5$ Hz, 1H; N*H*Ac), 4.97 (dd, $J = 8.5$, 7.5 Hz, 1H; H-3), 4.87 (dd, $J = 7.5$, 7.0 Hz, 1H; H-4), 4.26 (dd, $J = 12.0$, 5.6 Hz, 1H; H-6), 4.19 (dt, $J = 8.5$, 4.7 Hz, 1H; H-2), 4.09–4.01 (m, 2H; H-1, H-6), 3.79 (ddd, $J = 7.0$, 5.6, 3.5 Hz, 1H; H-5), 3.67 (d, $J = 10.6$ Hz, 6H; OCH_3), 2.04 (s, 3H; Ac), 2.01 (s, 3H; Ac), 2.00 (s, 3H; Ac), 1.91 (s, 3H; Ac), 1.79–1.44 ppm (m, 6H; H-7, H-8, H-9); ^{13}C NMR (100 MHz, $CDCl_3$): $\delta = 170.9$, 170.6, 169.9, 169.0, 71.3, 70.2, 70.1, 68.5, 61.7, 52.3 (d, $J = 6.7$ Hz), 52.3 (d, $J = 6.7$ Hz), 50.7, 27.3 (d, $J = 14.3$ Hz), 24.0 (d, $J = 141.5$ Hz), 23.1, 20.8, 20.7, 20.6, 18.5 ppm (d, $J = 4.9$ Hz); ^{31}P NMR (162 MHz, $CDCl_3$): $\delta = 34.2$ ppm; elemental analysis calcd (%) for $C_{19}H_{32}NO_{11}P$: C 47.40, H 6.70, N 2.91, P 6.43; found: C 47.31, H 6.55, N 3.31, P 6.35.

Dimethyl [(2,3,4,6-tetra-*O*-acetyl-α-D-galactopyranosyl)propyl]-phosphonate (8 d): Yield: 561 mg, 78%; $[\alpha]_D^{28} = +60$ ($c = 1.0$, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$): $\delta = 5.36$ (dd, $J = 3.4$, 2.1 Hz, 1H; H-4), 5.21 (dd, $J = 9.5$, 5.1 Hz, 1H; H-2), 5.14 (dd, $J = 9.5$, 3.4 Hz, 1H; H-3), 4.22–4.13 (m, 2H; H-1, H-6), 4.10–3.99 (m, 2H; H-5, H-6), 3.70 (d, $J = 10.7$ Hz, 6H; OCH_3), 2.08 (s, 3H; Ac), 2.04 (s, 3H; Ac), 2.03 (s, 3H; Ac), 1.98 (s, 3H; Ac), 1.81–1.45 ppm (m, 6H; H-7, H-8, H-9); ^{13}C NMR (100 MHz, $CDCl_3$): $\delta = 170.3$, 170.1, 170.0, 169.9, 71.5, 68.4, 68.3, 68.1, 68.7, 61.5, 52.3 (d, $J = 11.1$ Hz), 52.2 (d, $J = 11.1$ Hz), 26.3 (d, $J = 15.0$ Hz), 24.2 (d, $J = 141.0$ Hz), 20.7, 20.7, 20.6, 20.6, 19.1 ppm (d, $J = 4.4$ Hz); ^{31}P NMR (162 MHz, $CDCl_3$): $\delta = 34.0$ ppm;

elemental analysis calcd (%) for $C_{19}H_{31}O_{12}P$: C 47.31, H 6.48, P 6.42; found: C 47.45, H 6.37, P 6.29.

Dimethyl [(2-acetamido-2-deoxy-3,4,6-tri-*O*-acetyl-α-D galactopyranosyl)propyl]phosphonate (8e): Yield: 82.8 mg, 58%; $[\alpha]_D^{28} = 31$ ($c = 1.0$, $CHCl_3$); ^{1}H NMR (400 MHz, $CDCl_3$): $\delta = 6.14$ (d, $J = 8.3$ Hz, 1H; N*H*Ac), 5.30 (t, $J = 3.0$ Hz, 1H; H-4), 5.10 (dd, $J = 9.7$, 3.0 Hz, 1H; H-3), 4.52–4.42 (m, 1H; H-2), 4.29–4.14 (m, 2H; H-1, H-6), 4.07 (dd, $J = 11.6$, 4.9 Hz, 1H; H-6), 3.99 (m, 1H; H-5), 3.72 (d, $J = 10.8$ Hz, 6H; OCH_3), 2.10 (s, 3H; Ac), 1.05 (s, 3H; Ac), 2.03 (s, 3H; Ac), 1.96 (s, 3H; Ac), 1.79–1.44 ppm (m, 6H; H-7, H-8, H-9); ^{13}C NMR (100 MHz, $CDCl_3$): $\delta = 170.7$ (2 C), 170.5, 170.2, 71.8, 68.5, 68.3, 67.0, 61.6, 52.3 (d, $J = 6.7$ Hz), 48.6, 26.7, 24.0 (d, $J = 140.5$ Hz), 23.0, 20.9, 20.7, 20.7, 18.8 ppm (d, $J = 5.2$ Hz); ^{31}P NMR (121 MHz, $CDCl_3$): $\delta = 27.0$ ppm; elemental analysis calcd (%) for $C_{19}H_{32}NO_{11}P$: C 47.40, H 6.70, N 2.91, P 6.43; found: C 47.58, H 6.82, N 3.05, P 6.34.

Deprotection of compounds 13a–e: General procedure

A stirred solution of the corresponding derivative **13** (0.034 mmol) in MeOH (1.7 mL) was treated with a 25% solution of NH_4OH (1.7 mL) and stirred for 7 h at room temperature. The solution was concentrated in vacuo, and the residue was purified on a Bio-Gel P-2 column (1×70 cm), eluted with 100 mmol NH_4HCO_3 in 4:1 H_2O/MeOH, and lyophilised to give pure **14** as a foam.

Ammonium uridin-5′-yl [(α-D-glucopyranosyl)-propyl]phosphonate (14a): Yield: 13.1 mg, 65%; m.p. 153–155 °C; $[\alpha]_D^{27} = +31$ ($c = 0.7$, H_2O); ^{1}H NMR (400 MHz, D_2O, 25 °C): $\delta = 8.00$ (d, $J = 8.1$ Hz, 1H; H-6_u), 6.00 (d, $J = 4.7$ Hz, 1H; H-1′), 5.99 (d, $J = 8.1$ Hz, 1H; H-5_u), 4.39 (dd, $J = 5.0$, 4.7 Hz, 1H; H-2′), 4.36 (t, $J = 5.0$ Hz, 1H; H-3′), 4.32–4.29 (m, 1H; H-4′), 4.18 (ddd, $J = 11.8$, 4.2, 2.5 Hz, 1H; H-5′), 4.10 (ddd, $J = 11.8$, 5.3, 2.9 Hz, 1H; H-5), 4.03 (ddd, $J = 11.4$, 5.9, 2.8 Hz, 1H; H-1), 3.84 (dd, $J = 12.2$, 2.3 Hz, 1H; H-6), 3.75–3.63 (m, 3H; H-6, H-2, H-3), 3.53 (ddd, $J = 9.8$, 5.6, 2.3 Hz, 1H; H-5), 3.37 (dd, $J = 9.8$, 8.9 Hz, 1H; H-4), 1.88–1.50 ppm (m, 6H; H-7, H-8, H-9); ^{13}C NMR (100 MHz, D_2O, 25 °C): $\delta = 166.6$, 152.0, 141.6, 102.4, 88.8, 83.2 (d, $J = 16.2$ Hz), 75.4, 73.9, 73.3, 72.4, 71.3, 70.4, 69.6, 62.6 (d, $J = 5.1$ Hz), 61.1, 25.7 (d, $J = 134.5$ Hz), 25.0 (d, $J = 16.8$ Hz), 19.1 ppm (d, $J = 4.3$ Hz); ^{31}P NMR (121 MHz, D_2O, 25 °C): $\delta = 28.4$ ppm; elemental analysis calcd (%) for $C_{18}H_{32}N_3O_{13}P$: C 40.84, H 6.09, N 7.94, P 5.85; found: C 40.77, H 6.18, N 8.14, P 5.97

Ammonium uridin-5′-yl [(α-D-mannopyranosyl)propyl]phosphonate (14b): Yield: 16.6 mg, 92%; m.p. 134–136 °C; $[\alpha]_D^{28} = +9$ ($c = 0.9$, H_2O); ^{1}H NMR (400 MHz, D_2O, 25 °C): $\delta = 7.96$ (d, $J = 8.1$ Hz, 1H; H-6_u), 5.96 (d, $J = 4.4$ Hz, 1H; H-1′), 5.94 (d, $J = 8.1$ Hz, 1H; H-5_u), 4.35 (dd, $J = 5.0$, 4.4 Hz, 1H; H-2′), 4.32 (t, $J = 5.0$ Hz, 1H; H-3′), 4.28–4.24 (m, 1H; H-4′), 4.14 (ddd, $J = 11.7$, 4.1, 2.4 Hz, 1H; H-5′), 4.06 (ddd, $J = 11.7$, 5.3, 2.9 Hz, 1H; H-5′), 3.95–3.89 (m, 1H; H-1), 3.87 (dd, $J = 3.3$, 1.8 Hz, 1H; H-2), 3.82 (dd, $J = 12.2$, 2.3 Hz, 1H; H-6), 3.81 (dd, $J = 9.4$, 4.3 Hz, 1H; H-3), 3.70 (dd, $J = 12.2$, 6.0 Hz, 1H; H-6), 3.62 (t, $J = 9.4$ Hz, 1H; H-4), 3.50 (ddd, $J = 9.4$, 6.0, 2.3 Hz, 1H; H-5), 1.94–1.46 ppm (m, 6H; H-7, H-8, H-9); ^{13}C NMR (100 MHz, D_2O, 25 °C): $\delta = 166.3$, 151.8, 141.6, 102.3, 88.8, 83.2 (d, $J = 7.9$ Hz), 78.0, 73.9, 73.4, 71.5, 70.8, 69.5, 67.3, 62.6 (d, $J = 5.2$ Hz), 61.3, 28.5 (d, $J = 16.6$ Hz), 25.5 (d, $J = 134.8$ Hz), 19.5 ppm (d, $J = 4.4$ Hz); ^{31}P NMR (121 MHz, D_2O, 25 °C): $\delta = 28.4$ ppm; elemental analysis calcd (%) for $C_{18}H_{32}N_3O_{13}P$: C 40.84, H 6.09, N 7.94, P 5.85; found: C 40.86, H 6.23, N 7.89, P 5.74.

Ammonium uridin-5′-yl [(2-acetamido-2-deoxy-α-D-glucopyranosyl)propyl]phosphonate (14c): Yield: 12.2 mg, 63%; m.p. 161–163 °C; $[\alpha]_D^{25} = +38$ ($c = 1.1$, H_2O); ^{1}H NMR (400 MHz, D_2O, 25 °C): $\delta = 7.85$ (d, $J = 8.1$ Hz, 1H; H-6_u), 5.84 (d, $J = 4.4$ Hz, 1H; H-1′), 5.83 (d, $J = 8.1$ Hz, 1H; H-5_u), 4.24 (dd, $J = 4.9$, 4.4 Hz, 1H; H-2′), 4.20 (t, $J = 4.9$ Hz, 1H; H-3′), 4.17–4.13 (m, 1H; H-4′), 4.06–3.99 (ddd, $J = 11.8$, 4.3, 2.5 Hz, 1H; H-5′), 3.98–3.89 (m, 2H; H-5′, H-1), 3.84 (dd, $J = 10.7$, 5.8 Hz, 1H; H-2), 3.69 (dd, $J = 12.2$, 2.2 Hz, 1H; H-6), 3.64–3.57 (m, 2H; H-3, H-6), 3.43–3.36 (ddd, $J = 9.8$, 5.5, 2.2 Hz, 1H; H-5), 3.29 (dd, $J = 9.8$, 8.8 Hz, 1H; H-4), 1.92 (s, 3H; NAc), 1.79–1.31 ppm (m, 6H; H-7, H-8, H-9); ^{13}C NMR (100 MHz, D_2O, 25 °C): $\delta = 174.3$, 166.3, 151.8, 141.5, 102.3, 88.8, 83.2 (d, $J = 7.9$ Hz), 73.9, 73.3, 72.4, 70.8, 70.7, 69.5, 62.5 (d, $J = 5.2$ Hz), 61.0, 53.4, 25.7 (d, $J = 16.8$ Hz), 25.6 (d, $J = 134.7$ Hz), 21.8, 19.1 ppm (d, $J = 4.3$ Hz); ^{31}P NMR (162 MHz, D_2O, 25 °C): $\delta = 28.3$ ppm; elemental analysis calcd (%) for $C_{20}H_{35}N_4O_{13}P$: C 42.11, H 6.18, N 9.82, P 5.43; found: C 42.28, H 6.23, N 9.71, P 5.65.

Ammonium uridin-5′-yl [(α-D-galactopyranosyl)propyl]phosphonate (14d): Yield: 11.3 mg, 63%; m.p. 152–155 °C; $[\alpha]_D^{29} = +34$ ($c = 0.9$, H_2O); ^{1}H NMR (400 MHz, D_2O, 25 °C): $\delta = 7.93$ (d, $J = 8.1$ Hz, 1H; H-6_u), 5.94 (d, $J = 8.1$ Hz, 1H; H-5_u), 5.94 (d, $J = 4.6$ Hz, 1H; H-1′), 4.27 (t, $J = 4.6$ Hz, 1H; H-2′), 4.23 (t, $J = 4.6$ Hz, 1H; H-3′), 4.20–4.16 (m, 1H; H-4′), 4.05 (ddd, $J = 11.8$, 4.0, 2.5 Hz, 1H; H-5′), 3.97 (ddd, $J = 11.8$, 5.2, 2.8 Hz, 1H; H-5′), 3.93 (dd, $J = 6.1$, 2.8 Hz, 1H; H-1), 3.87 (dd, $J = 9.7$, 5.8 Hz, 1H; H-3), 3.84 (dd, $J = 3.5$, 0.7 Hz, 1H; H-2), 3.68 (dd, $J = 9.9$, 3.5 Hz, 1H; H-4), 3.54–3.63 (m, 3H; H-5, H-6), 1.76–1.35 ppm (m, 6H; H-7, H-8, H-9); ^{13}C NMR (100 MHz, D_2O, 25 °C): $\delta = 166.2$, 151.7, 141.6, 102.2, 88.9, 88.2 (d, $J = 8.0$ Hz), 75.1, 73.9, 71.5, 69.7, 69.4, 69.1, 68.3, 62.3 (d, $J = 5.2$ Hz), 61.1, 25.7 (d, $J = 134.8$ Hz), 24.9 (d, $J = 17.1$ Hz), 19.3 ppm (d, $J = 4.3$ Hz); ^{31}P NMR (162 MHz, D_2O, 25 °C): $\delta = 28.6$ ppm; elemental analysis calcd (%) for $C_{18}H_{32}N_3O_{13}P$: C 40.84, H 6.09, N 7.94, P 5.85; found: C 40.82, H 5.94, N 7.90, P 5.93.

Ammonium uridin-5′-yl [(2-acetamido-2-deoxy-α-D-galactopyranosyl)propyl]phosphonate (14e): Yield: 12.6 mg, 65%; m.p. 159–161 °C; $[\alpha]_D^{26} = +57$ ($c = 1.0$, H_2O); ^{1}H NMR (400 MHz, D_2O, 25 °C): $\delta = 7.88$ (d, $J = 8.1$ Hz, 1H; H-6_u), 5.83 (d, $J = 8.1$ Hz, 1H; H-5_u), 5.82 (d, $J = 5.1$ Hz, 1H; H-1′), 4.25 (t, $J = 5.1$ Hz, 1H; H-2′), 4.20 (t, $J = 5.1$ Hz, 1H; H-3′), 4.17–4.14 (m, 1H; H-4′), 4.09 (dd, $J = 11.0$, 5.9 Hz, 1H; H-2), 4.04 (ddd, $J = 11.7$, 4.0, 2.4 Hz, 1H; H-5′), 4.01–3.92 (m, 2H; H-5′, H-1), 3.83–3.81 (m, 1H; H-4), 3.77 (dd, $J = 11.0$, 3.3 Hz, 1H; H-3), 3.62–3.53 (m, 3H; H-5, H-6), 1.91 (s, 3H; NAc), 1.78–1.32 ppm (m, 6H; H-7, H-8, H-9); ^{13}C NMR (100 MHz, D_2O, 25 °C): $\delta = 174.5$, 166.3, 151.7, 141.6, 102.1, 88.9, 83.1 (d, $J = 7.9$ Hz), 73.9, 73.2, 71.4, 69.3, 68.4, 67.3, 62.3 (d, $J = 5.2$ Hz), 61.2, 49.7, 25.7 (d, $J = 134.9$ Hz), 25.5 (d, $J = 17.3$ Hz), 21.9, 19.2 ppm (d, $J = 4.2$ Hz); ^{31}P NMR (162 MHz, D_2O, 25 °C): $\delta = 28.4$ ppm; elemental analysis calcd (%) for $C_{20}H_{35}N_4O_{13}P$: C 42.11, H 6.18, N 9.82, P 5.43; found: C 42.02, H 6.21, N 9.90, P 5.52.

Expression and purification of *Hs*GalNAc-T2

Homo sapiens GalNAc-T2 (*Hs*GalNAc-T2) was expressed in *Pichia pastoris* and purified as described previously.[15a,32]

Crystallisation

Protein drops were prepared by mixing protein solution (2 μL; the mix contained 7 mg mL^{-1} GalNAc-T2, 5 mM UDP, 5 mM $MnCl_2$ in 25 mM Tris-HCl pH 7.5) and reservoir solution (2 μL; 8–10% polyethylene glycol (PEG) 8000, 8% ethylene glycol, 100 mM Hepes pH 7.0–7.5), as reported previously.[32] Crystals were grown by hanging drop diffusion at 18 °C for 2 days. Then, compound **13d**, as a powder, was soaked onto these crystals for 25 min. The crystals were cryoprotected in the precipitant solution containing an additional 18% ethylene glycol. All crystals were frozen in a nitrogen gas stream cooled to 100 K.

Structure determination and refinement

All data were processed and scaled by using the XDS package[35]and CCP4[36] software; relevant statistics are given in Table S1 in the Supporting Information. The crystal structure was solved by molecular replacement with Phaser[37] and by using PDB entry 2FFV as a template. Initial phases were further improved by cycles of manual model building in Coot[38] and refinement with REFMAC5:[39] The final model was validated with PROCHECK (model statistics are given in Table S1 in the Supporting Information). The asymmetric units of these crystals showed six molecules of GalNAc-T2, forming three independent dimers, as confirmed by the PISA server.[31] Only chain D showed disordered density towards the end of the lectin domain, and thus, we did not include residues with numbering 555 to 569 in the final model. Coordinates and structure factors have been deposited in the Worldwide Protein Data Bank (wwPDB; see Table S1 in the Supporting Information for the pdb codes).

Tryptophan fluorescence spectroscopy

Fluorescence spectroscopy was used to determine the dissociation constant of GalNAc-T2 against the compounds. All experiments were carried out in a Cary Eclipse spectrofluorometer (Varian) at 25 °C with GalNAc-T2 at 1 μM, and concentrations of the compound varying from 1 to 1000 μM in 25 mM Tris, 150 mM NaCl, pH 7.5. Fluorescence emission spectra were recorded in the $\lambda =$ 300–400 nm range with an excitation wavelength of $\lambda = 280$ nm and a slit width of 5 nm. Data analysis was performed in Prism (GraphPad software) by considering a model with a single binding site [see Eq. (1), in which F_0 is the intrinsic fluorescence of the enzyme in the absence of quencher [Q], F_1 is the observed fluorescence at a given quencher concentration, f_a is the fractional degree of fluorescence, and K_d is the dissociation constant].

$$1 - \frac{F_1}{F_0} = \frac{f_a[Q]}{K_d + [Q]} \quad (1)$$

Computational studies

Macromolecule preparation: Two X-ray crystal structures of GalNAc-T2 were selected: one in the closed active conformation (PDB entry 4D0Z, in complex with UDP-5S-GalNAc, EA2 peptide and Mn^{2+}), and the second one in the open inactive conformation (PDB entry 2FFV, in complex with UDP and Mn^{2+}). Chain A was selected in both PDB structures. All water molecules, ethylene glycol molecules and ligands were removed, but Mn^{2+} ions were retained. By using the Maestro package,[40] missing residues were added and modelled, and neutral terminal N and C groups and hydrogen atoms were added. Charges were assigned to all atoms with the OPLS_2005 force field, and the final protein structure was finally minimised with the same force field.

Ligand preparation: The 3D structure for UDP was extracted from PDB 2FFV. The 3D structure for UDP-GalNAc was built from the 3D structure of UDP-5S-GalNAc from PDB 4D0Z. UDP-5S-GalNAc was also prepared for docking as a reference compound. The 3D structures of 16 compounds were built from their SMILES codes. All ligands were subjected to geometry optimisation by using the MMFFs force field implemented in Maestro. This method was chosen by comparing the docking results for reference compounds by applying different force fields.

Docking calculations: Docking was performed by means of the Glide[41] and AutoDock4 programs.[42] In Glide, a cubic grid box of 25^3 Å^3 was generated for each protein, with the centre defined as the centre of mass among Asp176, His145, Arg362 and His226 residues. For all ligands, the torsional angles were allowed to rotate with no restrictions. A standard precision docking was performed for both structures. In the case of AutoDock, different conformers of the starting geometries of the ligands were docked by using the Lamarckian genetic algorithm, by randomly changing the torsion angles and overall orientation of the molecule. A volume for exploration was defined with a grid spacing of 0.375 Å, centred on the centre of mass among Asp176, His145, Arg362 and His226 residues, and a 3D grid of $21\times18\times19$ Å^3 for 4D0Z and $21\times21\times18$ Å^3 for 2FFV. After docking, the 200 solutions were clustered in groups with root-mean-square deviations of less than 2.0 Å. The clusters were ranked by the lowest energy representative of each cluster. Docking results from both Glide and AutoDock methods were analysed and taken together for the discussion.

tMD simulations: The tMD simulation was performed with the AMBER12 software[43] by using the ff10, gaff and Glycam force fields. The Cα atoms of the starting structure (GalNAc-T2 in complex with UDP-GalNAc in the open conformation) were restrained to the corresponding positions of the target structure (GalNAc-T2 in complex with UDP-GalNAc in the closed conformation). Explicit water solvent (TIP3BOX model) was considered by adding the same number of water molecules to both complexes. Also, the same number of counterions (4 chlorine ions) was added. Energy minimisation of all initial structures (closed and open states) was carried out by using the steepest descent method in AMBER12 for 500 steps and the non-bonded cutoff was 9 Å. For the tMD simulation, the SHAKE algorithm was used, the time step was 0.002 ps, and the non-bonded cutoff was 15 Å. The temperature coupling method was used to keep the temperature constant at 300 K and Berendsen temperature coupling scheme was used. No positional restraints were applied. The total simulation time was 5 ns.

Acknowledgements

This work was supported by the Spanish MINECO contracts CTQ2013-44367-C2-1-P (to P.M.), CTQ2013-44367-C2-2-P (to R.H.G.), and CTQ2011-22724 and CTQ2014-57141-R (to S.M.S.). We also acknowledge the Government of Aragón (Spain) (Bioorganic Chemistry group E-10 and Protein Targets group B-89) for financial support. We acknowledge the Institute of Biocomputation and Physics of Complex Systems (BIFI) at the University of Zaragoza (Spain) for computer time at clusters Terminus and Memento. We thank synchrotron radiation sources DLS (Oxford) and, in particular, beamline I04 (experiment number MX8035-26). M.G. thanks the Spanish Ministry of Education (MEC) for a predoctoral grant (FPU program). The European Commission is gratefully acknowledged (Grant GLYCOPHARM FP7-PITN-GA-2012-317297 and BioStruct-X grant agreement no. 283570 and BIOSTRUCTX_5186).

Keywords: carbohydrates · glycomimetics · glycosyltransferases · molecular modeling · proteins

[1] a) J. Voglmeir, S. L. Flitsch, in *Science of Synthesis*, Biocatalysis in Organic Synthesis (Eds.: K. Faber,W.-D. Fessner, N. J. Turner), Thieme, Stuttgart, **2015**, *Vol. 1*, pp. 507 – 542; b) T. M. Gloster, *Curr. Opin. Struct. Biol.* **2014**, *28*, 131 – 141; c) R. Roychoudhury, N. L. B. Pohl, *Curr. Opin. Chem. Biol.*

2010, *14*, 168 – 173; d) L. L. Lairson, B. Henrissat, G. J. Davies, S. G. Withers *Ann. Rev. Biochem.* **2008**, *77*, 521 – 555.
[2] X. Chen, *ACS Chem. Biol.* **2011**, *6*, 14 – 17.
[3] a) S. Wang, S. Vidal, *Carbohydr. Chem.* **2013**, *368*, 78 – 101; b) P. Merino, T. Tejero, I. Delso, R. Hurtado-Guerrero, A. Gomez-SanJuan, D. Sadaba, *Mini-Rev. Med. Chem.* **2012**, *12*, 1455 – 1464; c) T. Kajimoto, M. Node, *Synthesis* **2009**, 3179 – 3210; d) L. J. Whalen, W. A. Greenberg, M. L. Mitchell, C.-H. Wong in *Iminosugars* (Eds.: P. Compain, O. R. Martin), Wiley, Hoboken, **2007**, pp. 153 – 176; e) K.-H. Jung, R. R. Schmidt in *Carbohydrate-Based Drug Discovery, Vol. 2* (Ed.: C.-H. Wong), Wiley-VCH, Weinheim, **2003**, pp. 609 – 659; f) X. Qian, M. M. Palcic, *Carbohydr. Chem. Biol.* **2000**, *3*, 293 – 312; g) L. Tedaldi, G. K. Wagner, *MedChemComm* **2014**, *5*, 1106 – 1125.
[4] D.-M. Liang, J.-H. Liu, H. Wu, B.-B. Wang, H.-J. Zhu, J.-J. Qiao, *Chem. Soc. Rev.* **2015**, *44*, 8350 – 8374.
[5] a) D. Albesa-Jové, M. F. Mendoza, A. Rodrigo-Unzueta, F. Gomollón-Bel, J. Cifuente, S. Urresti, N. Comino, H. Gómez, J. Romero-García, J. M. Lluch, E. Sancho-Vaello, X. Biarnés, A. Planas, P. Merino, L. Masgrau, M. E. Guerin, *Angew. Chem. Int. Ed.* **2015**, *54*, 9898 – 9902; *Angew. Chem.* **2015**, *127*, 10036 – 10040; b) A. Ardèvol, C. Rovira, *J. Am. Chem. Soc.* **2015**, *137*, 7528 – 7547; c) H. Gómez, I. Polyak, W. Thiel, J. M. Lluch, L. Masgrau, *J. Am. Chem. Soc.* **2012**, *134*, 4743 – 4752; d) R. Hurtado-Guerrero, G. J. Davies, *Curr. Opin. Chem. Biol.* **2012**, *16*, 479 – 487; e) I. Tvaroska, *Mini-Rev. Org. Chem.* **2011**, *8*, 263 – 269; f) D. L. Jakeman, *ChemBioChem* **2011**, *12*, 2540 – 2542.
[6] J. Rini, J. Esko, A. Varki in *Essentials of Glicobiology. 2nd ed.* (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, NY, **2009**, Available from: http://www.ncbi.nlm.nih.gov/books/NBK1921/.
[7] V. Malik, W. Black Gary, *Adv. Protein Chem. Struct. Biol.* **2012**, *87*, 87 – 115.
[8] a) K. Descroix, G. K. Wagner, *Org. Biomol. Chem.* **2011**, *9*, 1855 – 1863; b) M. S. M. Timmer, M. V. Chumillas, W. E. Donker-Koopman, J. Alerts, G. A. van der Marel, H. S. Overkleeft, J. H. van Boom, *J. Carbohydr. Chem.* **2005**, *24*, 335 – 351; c) H. C. Dorfmueller, V. S. Borodkin, D. E. Blair, S. Pathak, I. Navratilova, D. M. F. van Aalten, *Amino Acids* **2011**, *40*, 781 – 792; d) F. Stolz, M. Reiner, A. Blume, W. Reutter, R. R. Schmidt, *J. Org. Chem.* **2004**, *69*, 665 – 679.
[9] J. Jiang, M. B. Lazarus, L. Pasquina, P. Sliz, S. Walker, *Nat. Chem. Biol.* **2012**, *8*, 72 – 77.
[10] a) R. Jørgensen, L. L. Grimm, N. Sindhuwinata, T. Peters, M. M. Palcic, *Angew. Chem. Int. Ed.* **2012**, *51*, 4171 – 4175; *Angew. Chem.* **2012**, *124*, 4247 – 4251; b) K. Schaefer, J. Albers, N. Sindhuwinata, T. Peters, B. Meyer, *ChemBioChem* **2012**, *13*, 443 – 450; c) J. C. Errey, S. S. Lee, R. P. Gibson, C. M. Fleites, C. S. Barry, P. M. J. Jung, A. C. O'Sullivan, B. G. Davis, G. J. Davies, *Angew. Chem. Int. Ed.* **2010**, *49*, 1234 – 1237; *Angew. Chem.* **2010**, *122*, 1256 – 1259; d) S. Wang, J. A. Cuesta-Seijo, D. Lafont, M. M. Palcic, S. Vidal, *Chem. Eur. J.* **2013**, *19*, 15346 – 15357; e) S. Wang, J. A. Cuesta-Seijo, A. Striebeck, D. Lafont, M. M. Palcic, S. Vidal, *ChemPlusChem* **2015**, *80*, 1525 – 1532; f) S. Wang, D. L. Shen, D. Lafont, A.-S. Vercoutter-Edouart, M. Mortuaire, Y. Shi, O. Maniti, A. Girard-Egrot, T. Lefebvre, B. M. Pinto, D. Vocadlo, S. Vidal, *MedChemComm* **2014**, *5*, 1172 – 1178; g) J. Jiang, V. Kanabar, B. Padilla, F. Man, S. C. Pitchford, C. P. Page, G. K. Wagner *Chem. Commun* **2016**, *52*, 3955 – 3958.
[11] T. M. Gloster, W. F. Zandberg, J. E. Heinonen, D. L. Shen, L.-H. Deng, D. J. Vocadlo, *Nat. Chem. Biol.* **2011**, *7*, 174 – 181.
[12] A. Schäfer, J. Thiem, *J. Org. Chem.* **2000**, *65*, 24 – 29.
[13] a) R. Chang, T.-T. Vo, N. S. Finney, *Carbohydr. Res.* **2006**, *341*, 1998 – 2004; b) J. Hajduch, G. Nam, E. J. Kim, R. Froehlich, J. A. Hanover, K. L. Kirk, *Carbohydr. Res.* **2008**, *343*, 189 – 195; c) A. J. Clarke, R. Hurtado-Guerrero, S. Pathak, A. W. Schuttelkopf, V. Borodkin, S. M. Shepherd, A. F. M. Ibrahim, D. M. F. van Aalten, *Embo J.* **2008**, *27*, 2780 – 2788; d) S. K. Partha, A. Sadeghi-Khomami, K. Slowski, T. Kotake, N. R. Thomas, D. L. Jakeman, D. A. R. Sanders, *J. Mol. Biol.* **2010**, *403*, 578 – 590; e) S. M. Forget, A. Jee, D. A. Smithen, R. Jagdhane, S. Anjum, S. A. Beaton, D. R. J. Palmer, R. T. Syvitskia, D. L. Jakeman, *Org. Biomol. Chem.* **2015**, *13*, 866 – 875.
[14] S. Vidal, I. Bruyere, A. Malleron, C. Auge, J.-P. Praly, *Bioorg. Med. Chem.* **2006**, *14*, 7293 – 7301.
[15] a) E. Lira-Navarrete, M. d. L. Rivas, I. Companon, M. C. Pallares, Y. Kong, J. Iglesias-Fernandez, G. a. J. L. Bernardes, J. M. Peregrina, C. Rovira, P. Bernado, P. Bruscolini, H. Clausen, A. Lostao, F. Corzana, R. Hurtado-Guerrero, *Nat. Commun.* **2015**, *6*, 6937; b) B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, B. Henrissat, *Nucleic Acids Res.* **2009**, *37*, D233 – D238; c) K. G. T. Hagen, T. A. Fritz, L. A. Tabak, *Glycobiology* **2003**, *13*, 1R – 16R.
[16] H. Gómez, R. Rojas, D. Patel, L. A. Tabak, J. M. Lluch, L. Masgrau, *Org. Biomol. Chem.* **2014**, *12*, 2645 – 2655.
[17] A. Dondoni, A. Marra, *Org. Biomol. Chem.* **2015**, *13*, 2212 – 2215.
[18] a) D. Horton, T. Miyake, *Carbohydr. Res.* **1988**, *184*, 221 – 229; b) J. R. Kramer, T. J. Deming, *J. Am. Chem. Soc.* **2010**, *132*, 15068 – 15071.
[19] R. Y. Tam, S. S. Ferreira, P. Czechura, J. L. Chaytor, R. N. Ben, *J. Am. Chem. Soc.* **2008**, *130*, 17494 – 17501.
[20] a) A. Babič, S. Gobec, C. Gravier-Pelletier, Y. Le Merrer, S. Pečar, *Tetrahedron* **2008**, *64*, 9093 – 9100; b) J. Cui, D. Horton, *Carbohydr. Res.* **1998**, *309*, 319 – 330.
[21] S. Liu, R. N. Ben, *Org. Lett.* **2005**, *7*, 2385 – 2388.
[22] a) V. R. Bouvet, R. N. Ben, *J. Org. Chem.* **2006**, *71*, 3619 – 3622; b) C. H. Röhrig, M. Takhi, R. R. Schmidt, *Synlett* **2001**, 1170 – 1172.
[23] a) M. J. Damha, N. Usman, K. K. Ogilvie, *Can. J. Chem.* **1989**, *67*, 831 – 839; b) P. Barthelemy, C. A. H. Prata, S. F. Filocamo, C. E. Immoos, B. W. Maynor, S. A. N. Hashmi, S. J. Lee, M. W. Grinstaff, *Chem. Commun.* **2005**, 1261 – 1263.
[24] A. Dondoni, S. Staderini, A. Marra, *Eur. J. Org. Chem.* **2013**, 5370 – 5375.
[25] H. Pizova, P. Bobal, *Tetrahedron Lett.* **2015**, *56*, 2014 – 2017.
[26] P. J. Casara, K. C. Jund, A. Clauss, J.-F. Nave, R. D. Snyder, *Bioorg. Med. Chem. Lett.* **1992**, *2*, 145 – 148.
[27] M. Izumi, K. Wada, H. Yuasa, H. Hashimoto, *J. Org. Chem.* **2005**, *70*, 8817 – 8824.
[28] G. W. Daub, E. E. J. Van Tameken, *J. Am. Chem. Soc.* **1977**, *99*, 3526 – 3528.
[29] J.-M. Campagne, J. Coste, P. Jouin, *J. Org. Chem.* **1995**, *60*, 5214 – 5223.
[30] E. P. Bennett, U. Mandel, H. Clausen, T. A. Gerken, T. A. Fritz, L. A. Tabak, *Glycobiology* **2012**, *22*, 736 – 756.
[31] E. Lira-Navarrete, J. Iglesias-Fernandez, W. F. Zandberg, I. Companon, Y. Kong, F. Corzana, B. M. Pinto, H. Clausen, J. M. Peregrina, D. J. Vocadlo, C. Rovira, R. Hurtado-Guerrero, *Angew. Chem. Int. Ed.* **2014**, *53*, 8206 – 8210; *Angew. Chem.* **2014**, *126*, 8345 – 8349.
[32] T. A. Fritz, J. Raman, L. A. Tabak, *J. Biol. Chem.* **2006**, *281*, 8613 – 8619.
[33] C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday, J. M. Thornton, *J. Biol. Inorg. Chem.* **2008**, *13*, 1205 – 1218.
[34] C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday, J. M. Thornton, *J. Biol. Inorg. Chem.* **2008**, *13*, 1205 – 1218.
[35] W. Kabsch, *Acta Crystallogr. Sect. D* **2010**, *66*, 125 – 132.
[36] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, *Acta Crystallogr. Sect. D* **2011**, *67*, 235 – 242.
[37] Collaborative Computational Project, Number 4, *Acta Crystallogr. Sect. D* **1994**, *50*, 760 – 763.
[38] P. Emsley, K. Cowtan, *Acta Crystallogr. Sect. D* **2004**, *60*, 2126 – 2132.
[39] G. N. Murshudov, A. A. P. Skubák, N. S. Lebedev, R. A. Pannu, R. A. Steiner, M. D. Nicholls, F. L. Winn, A. A. Vagin, *Acta Crystallogr. Sect. D* **2011**, *67*, 355 – 367.
[40] Schrödinger Release 2014-4, Maestro, version 10.0, Schrödinger, LLC, New York, NY, **2014**.
[41] Glide, version 6.5, Schrödinger, LLC, New York, NY, **2014**.
[42] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, *19*, 1639 – 1662.
[43] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F.Wong, F. Paesani, J. Vanicek, R. M.Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, *AMBER 12*, **2012**, University of California, San Francisco.

Received: January 31, 2016
Published online on April 13, 2016

