Bioorganic Chemistry 153 (2024) 107906

Contents lists available at ScienceDirect

Bioorganic Chemistry

journal homepage: www.elsevier.com/locate/bioorg

Multifrequency-STD NMR unveils the first Michaelis complex of an intramolecular *trans*-sialidase from *Ruminococcus gnavus*

Serena Monaco [a,*], Louise E. Tailford [b], Andrew Bell [b], Matthew Wallace [a], Nathalie Juge [b], Jesús Angulo [c,*]

[a] School of Chemistry, Pharmacy & Pharmacology, University of East Anglia, Norwich Research Park, NR4 7TJ Norwich, UK
[b] Quadram Institute Bioscience, Norwich Research Park, NR4 7UQ, UK
[c] Instituto de Investigaciones Químicas (IIQ), Consejo Superior de Investigaciones Científicas and Universidad de Sevilla, Avenida Américo Vespucio, 49, Sevilla 41092, Spain

ARTICLE INFO

Keywords:
Multifrequency STD NMR
Protein-carbohydrate interactions
Intramolecular *trans*-sialidase
Gut microbiota-mucin interactions
DEEP-STD NMR fingerprint
CH-π interactions
Docking calculations

ABSTRACT

*Rg*NanH is an intramolecular *trans*-sialidase expressed by the human gut symbiont *Ruminococcus gnavus*, to utilise intestinal sialylated mucin glycan epitopes. Its catalytic domain, belonging to glycoside hydrolase GH33 family, cleaves off terminal sialic acid residues from mucins, releasing 2,7-anhydro-Neu5Ac which is then used as metabolic substrate by *R. gnavus* to proliferate in the mucosal environment. *Rg*NanH is one of the three intramolecular *trans*-sialidases (IT-sialidases) characterised to date, and the first from a gut commensal organism. Here, saturation transfer difference NMR (STD NMR) in combination with computational techniques (molecular docking and CORCEMA-ST) were used to elucidate the specificity, kinetics and relative affinity of *Rg*NanH for sialoglycans and 2,7-anhydro-Neu5Ac. We propose the first 3D model for the Michaelis complex of an IT-sialidase. This confirms the sialic acid to be the main recognition element for the interaction in the enzymatic cleft and highlights the crucial role of Trp698 to make CH-π stacking with the galactose residue of the substrate 3′-sialyllactose. The same contact is shown not to be possible for 6′-sialyllactose, due to geometrical constrains of the α-2,6 linkage. Indeed 6′-sialyllactose is not a substrate, even though it is shown to bind to *Rg*NanH by STD NMR. These findings corroborate the role of Trp698 for the α-2,3 specificity of IT-sialidases. In this structural study, the use of Differential Epitope Mapping STD NMR (DEEP-STD NMR) approach allowed the validation of the proposed 3D models in solution. These structural approaches are shown to be instrumental in shedding light on the molecular mechanisms underpinning enzymatic reactions in the absence of enzyme-substrate X-ray structures.

1. Introduction

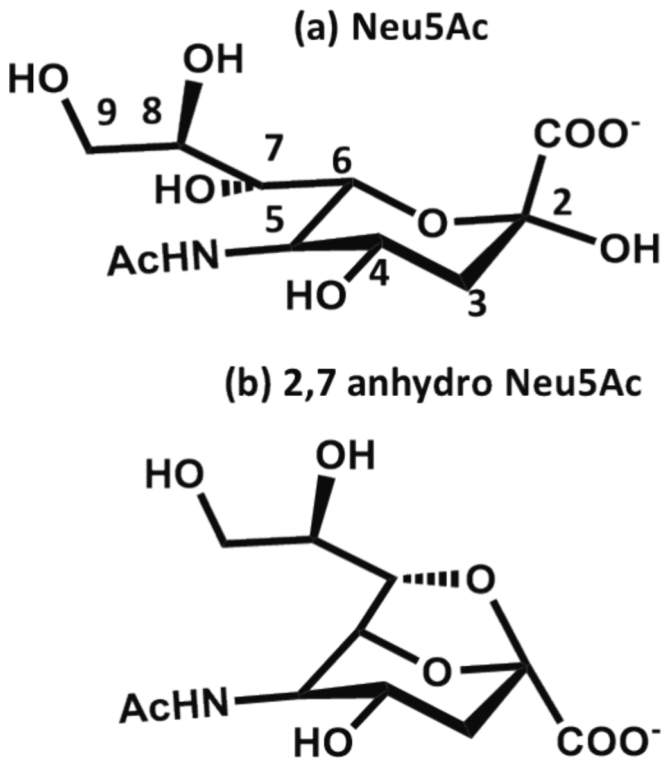
The gut microbiota plays a crucial role in human health, influencing the development of the immune system and pathogen defence, as well as polysaccharide digestion [1]. The microbiota composition varies longitudinally along the gastrointestinal (GI) tract but also transversally from the lumen to the mucosa [2]. The mucus layer covering the gut constitutes the interface between the gut microbiota and the host [3]. The glycan structures present in mucins are complex, consisting of core *O*-glycans elongated by diverse carbohydrate sequences often terminated by α-L-fucose and sialic acid such as *N*-acetylneuraminic acid (Neu5Ac, Scheme 1a), *via* α1–2/3/4 and α2–3/6 linkages, respectively. Due to their terminal location, these sugar residues are key targets for commensal and pathogenic bacteria [4]. The release and utilisation of sialic acid by microbial sialidases confer gut bacteria competitive advantages in the mucus niche [5,6,7].

Based on their substrate specificity and mechanisms of action, bacterial *exo*-sialidases can be divided into three classes: hydrolytic, *trans*-sialidases and intramolecular *trans*-sialidases (IT-sialidases). Hydrolytic sialidases release free sialic acid (Scheme 1a), *trans*-sialidases transfer the cleaved sialic acid to other glycoconjugates, while IT-sialidases release 2,7-anhydro-sialic acid (Scheme 1b) [6]. The possibility to

Abbreviations: STD NMR, Saturation Transfer Difference NMR; DEEP-STD NMR, Differential Epitope Mapping Saturation Transfer Difference NMR; IT-sialidase, intramolecular *trans*-sialidase; Neu5Ac, *N*-acetylneuraminic acid; 3′SL, 3′sialyllactose; 6′SL, 6′sialyllactose; Gal, galactose; Glc, glucose.

* Corresponding authors.
E-mail addresses: s.monaco@uea.ac.uk (S. Monaco), j.angulo@iiq.csic.es (J. Angulo).

https://doi.org/10.1016/j.bioorg.2024.107906
Received 1 August 2024; Received in revised form 17 October 2024; Accepted 20 October 2024
Available online 3 November 2024



Scheme 1. Chemical structure of **a)** Neu5Ac and **b)** 2,7-anhydro-Neu5Ac. The atom nomenclature is given for the carbons of Neu5Ac and it applies to all ligands.

carry out one or more of these enzymatic reactions, together with their capacity to catabolise the released sialic acids, influences the adaptation of the microorganisms to the mucosal environment [8,9].

*Rg*NanH is an IT-sialidase produced by the human gut symbiont *Ruminococcus gnavus*. *R. gnavus* is a prevalent member of the 'normal' gut microbiota in adults [10,11] but it is over-represented in diseases such as inflammatory bowel disease [12,13]. *Rg*NanH has specificity towards α2/3 sialic acid capping the mucin oligosaccharides [6]. The hydrolysis of the terminal sialic acid is followed by intramolecular transglycosylation at the *C*2 position of Neu5Ac, attacked by its own O7-hydroxyl group at the glycerol moiety, thus yielding 2,7-anhydro-Neu5Ac [14]. *Rg*NanH is the third IT-sialidase to be characterised, after NanL from *Macrobella decora* [15] and NanB from *Streptococcus pneumonia* [16], and the first from a gut commensal species [4]. *R. gnavus* ATCC 29149 can grow on 2,7-anhydro-Neu5Ac as a sole carbon source *in vitro* [17], owing to its specific 2,7-anhydro-Neu5Ac transporter and oxido-reductase, that allow the conversion of 2,7-anhydro-Neu5Ac to Neu5Ac [18]. This unique sialic acid metabolism pathway confers *R. gnavus* with a selective advantage over other members of the gut microbiota inhabiting the mucus niche [18a].

*Rg*NanH consists of an active domain belonging to the GH33 family and a carbohydrate binding module (CBM) from the CBM40 family [14,19]. The *R. gnavus* IT-sialidase structure has been solved in complex with 2,7-anhydro-Neu5Ac (PDB ID: 4x4a) [14] but no crystal structure of the Michaelis complex between an IT-sialidase and its substrate is available. In this work, saturation transfer difference NMR (STD NMR) was used to gain molecular insights into the structure of the *Rg*NanH-GH33 catalytic domain bound to sialoglycan substrates, in the form of 3′-sialyllactose (Neu5Acα3Lac or 3′SL) and 6′-sialyllactose (Neu5Acα6Lac or 6′SL).

STD NMR is a robust and versatile technique frequently used for monitoring weak ligand binding to protein receptors (dissociation constant, K_D, ranging from 10^{-8} M to 10^{-3} M). STD NMR relies on selective saturation of the ^{1}H NMR signals of the protein receptor [20]. The generated saturation on specific residues of the receptor is transferred 1) across the whole protein by dipole–dipole relay NOE (intra-molecular saturation transfer via very efficient spin diffusion) and 2) to small molecules binding to the receptor (inter-molecular saturation transfer). Since the highest STD intensities correlate with the closest ligand–protein contacts in the bound state [20a], STD NMR experiments highlight what are the ligand moieties closely contacting the protein surface. Thus, they provide a map of the binding, namely the ligand binding epitope map [21]. Differential epitope mapping STD NMR (DEEP-STD NMR) [22], which relies on a multifrequency STD NMR approach, allows to map which ligand protons are closer to certain type of amino acid side chains, e.g., aliphatic or aromatic patches of the protein surface [23,24,25,26]. In addition, if the protein 3D structure is known, the method allows to gain information on the orientation of the ligand. Comparing the DEEP-STD pattern of a complex of known 3D structure to the DEEP-STD patterns of unknown ligands on the same protein provides information on the binding mode and orientation of the unknown ligands. This approach is termed the *DEEP-STD NMR fingerprinting approach*, and it was previously demonstrated for a set of ligands of cholera toxin unit-B [24].

We hereby exploited single- and multi-frequency STD NMR techniques, in combination with molecular docking and matrix relaxation prediction of STD intensities by CORCEMA-ST [27,28], to obtain experimentally validated 3D models of *Rg*NanH-GH33 bound to 3′SL and 6′SL in solution. While this approach is combined with STD variable temperature and competition experiments to gain information on relative kinetics and binding affinities, this study is not focused on the accurate thermodynamic investigation of the binding processes, but on the structural elucidation of the Michaelis-Menten complex of *Rg*NanH.

2. Results

2.1. Specificity, kinetics and relative binding affinity by STD NMR

The interactions of 3′SL and 6′SL with *Rg*NanH-GH33 D282A catalytic mutant were analysed by STD NMR spectroscopy. Binding was detected for both 3′SL and 6′SL oligosaccharides (Fig. 1). The observed binding to 3′SL was in line with this ligand being the preferential substrate of *Rg*NanH-GH33 [14], as shown by activity assays and supported by *Rg*NanH-GH33 D282A binding to sialoglycan microarray [19]. STD NMR build-up curves at two different frequencies were obtained for 3′SL and 6′SL and, from these, binding epitopes were identified for the two ligands upon saturation at 0.60 ppm. Comparison of the binding epitopes (Fig. 1a,b) showed that the Neu5Ac moiety of both ligands made the closest contact to the protein surface, with a comparable binding epitope pattern for both sialoglycans. The binding epitope also showed that the lactose moiety of 3′SL appeared to have closer contact to the protein surface than the lactose moiety of 6′SL.

To investigate the kinetics and relative affinity of 3′SL and 6′SL to *Rg*NanH, variable temperature and competition STD NMR experiments were performed at 2 s saturation time with protein saturation at 0.60 ppm, monitoring the intensity of the signals in the least crowded spectral regions. Variable temperature experiments (Fig. 2a,b) rely on the fact that the kinetics of exchange increase with temperature and that STD intensities depend heavily on exchange kinetics [27]. Thus, increases in STD signals with increasing temperatures is associated with strong interactions (yet still within the fast exchange conditions required for STD NMR observation). This is due to the relatively slow exchange processes (where STDs are weak due to inefficient accumulation of saturated ligand in the bulk free state) becoming faster and hence producing higher STD intensities upon heating the sample. Such profile was observed for 3′SL (Fig. 2a). In contrast, for 6′SL, decreases in STD signals with increases in temperature were observed (Fig. 2b), reflecting weaker interactions, i.e. a starting fast kinetics becoming faster and hence reducing the STD intensities due to the much shorter residence time in the bound state precluding an efficient transfer of saturation to the

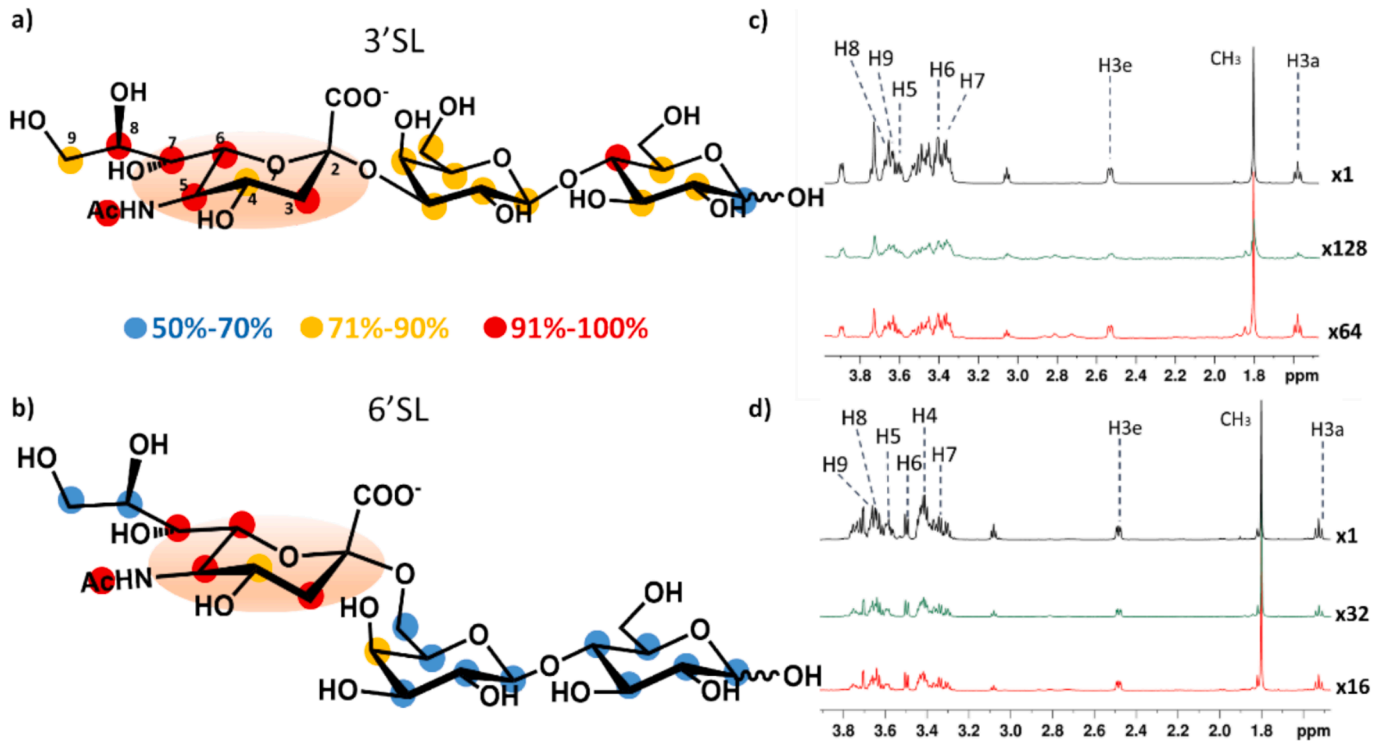


Fig. 1. STD NMR analyses of sialoglycans binding to *Rg*NanH-GH33 D282A. Binding epitope mappings of **a)** 3′SL and **b)** 6′SL from STD NMR build-up curve experiments with irradiation frequency at 0.60 ppm. Relative STD % values were extracted from the spectra for each set of protons of each ligand. For both ligands the methyl group on the acetyl moiety of sialic acid showed the highest STD %, and therefore it was arbitrarily assigned at 100%, while all the other values were normalised relative to this. The pale rose circle highlights Neu5Ac as the main recognition element for both ligands. STD NMR spectra for **c)** 3′SL and **d)** 6′SL and proton assignment. For **c)** and **d)**, reference spectra are shown in black (top); STD NMR spectra with on-resonance irradiation at 6.55 ppm (aromatic side chains) is shown in green (middle); STD NMR spectra with on-resonance irradiation at 0.60 ppm (aliphatic side chains) is shown in red (bottom). Spectra were acquired at 2 s saturation time. The magnification relative to the reference spectrum is given for each difference spectrum. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ligand free state [27].

STD NMR competition is a traditional approach to probe if two or more ligands are binding in the same location on a given protein [29,30], as STD intensity reductions can be associated to the ligands displacing each other from the binding site. Here, we first assessed the competition between 3′SL and 6′SL by performing STD NMR experiments of *Rg*NanH-GH33 D282A in the presence of 3′SL (1 mM), 6′SL (1 mM) and 3′SL + 6′SL (1 mM + 1 mM). In all cases, the STD intensities of 7 well-isolated reporter ligand signals were followed at 288 K (Fig. 2c,d). The STD intensities of the reporter signals of 3′SL were not significantly affected upon addition of 6′SL (Fig. 2c). In contrast, the STD intensities of the protons of 6′SL were considerably reduced upon addition of 3′SL (Fig. 2d). These findings support that both ligands were bound at the same site, with 3′SL having stronger binding affinity.

Additional STD competition experiments (Fig. S1 of the SI) showed that 2,7-anhydro-Neu5Ac could displace both 3′SL and 6′SL, therefore being the strongest binder of the three different ligands. Hence, based on the variable temperature and competition STD NMR experiments, we propose that (i) 3′SL and 6′SL fall into two different kinetics windows: 3′SL binds to *Rg*NanH-GH33 D282A with slow kinetics (higher affinity), while 6′SL binds with faster kinetics; and ii) the ranking of affinities of the three ligands to *Rg*NanH-GH33 D282A is: 2,7-anhydro-Neu5Ac > 3′SL > 6′SL. Importantly, these experiments were performed on the same samples as those used for determining STD NMR binding epitopes, and allowed, to assess the relative affinity and kinetics of the three ligands without the use of additional material.

2.2. Binding mode and ligand orientation in the binding pocket through DEEP-STD NMR

DEEP-STD NMR was carried out to gain information on the binding mode and orientation of 3′SL and 6′SL ligand in the *Rg*NanH-GH33 D282A binding pocket by differential irradiation at 0.60 ppm and 6.55 ppm. These are the same frequencies that had been chosen for the analysis of the complex between *Rg*NanH-GH33 and the product 2,7-anhydro-Neu5Ac in [22]. In that work, these frequencies are shown by protein chemical shift prediction calculations to be centred in the aliphatic and aromatic protein spectral regions, respectively. In particular, the predictions show that by irradiation at 6.55 ppm we are hitting tryptophan and tyrosine residues. Using here the exact same frequencies as in the seminal paper is also useful for comparability with the original work. Indeed, the already available DEEP-STD NMR data for the *Rg*NanH-GH33/2,7-anhydro-Neu5Ac complex [22] (for which the 3D structure has been solved by X-ray crystallography [14]) served as a fingerprint to confirm the binding position and orientation of the sialoglycans in the catalytic cleft of *Rg*NanH-GH33.

In Fig. 3, the DEEP-STD NMR fingerprinting analysis of the two sialoglycans in comparison to the reference 2,7-anhydro-Neu5Ac is reported. Comparison of the DEEP-STD histograms for the Neu5Ac moiety of the thrisaccharides with 2,7-anhydro-Neu5Ac (blue bars) showed a very similar DEEP-STD pattern for the three ligands. This was in agreement with the STD competition experiments, that indicated that all three ligands occupied the same binding pocket. Remarkably, it also showed that the reducing Neu5Ac of the sialoglycans sit in the same orientation as 2,7-anhydro-Neu5Ac, in the *Rg*NanH-GH33 binding pocket.

However, minor differences were observed at the protons of the

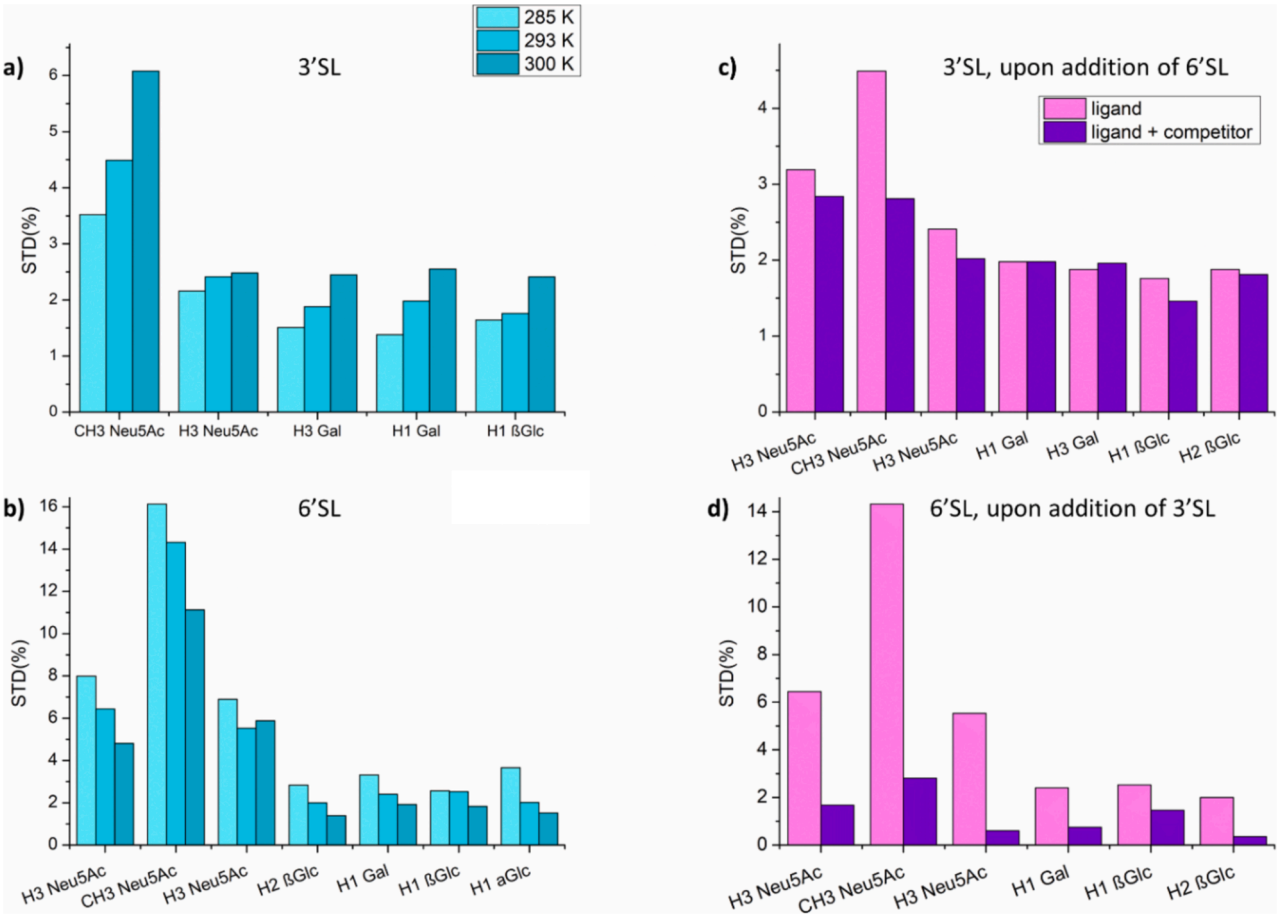


Fig. 2. Variable temperature and competition experiments by STD NMR. Left: Variable temperature STD NMR experiments. STD intensities for **a)** 3'SL and **b)** 6'SL upon binding to *Rg*NanH-GH33 D282A at increasing temperatures. Upon increases in temperature, the STD signals were increased for 3'SL and decreased for 6'SL. This supports that both ligands are located at the two different extremes of the kinetic window favourable for STD NMR observation (3'SL in the low kinetics extreme, and 6'SL in the very fast kinetics extreme). **Right**: STD NMR competition experiments. **c)** 3'SL, upon addition of 6'SL and **d)** 6'SL, upon addition of 3'SL. The samples containing the complexes 3'SL/*Rg*NanH-GH33 D282A and 6'SL/*Rg*NanH-GH33 D282A are compared with the sample containing both ligands 3'SL and 6'SL in the presence of *Rg*NanH-GH33 D282A. 3'SL is able to displace 6'SL, but its binding is not affected by the presence of 6'SL.

polyhydroxy chain (H7 to H9), between the Neu5Ac moiety of the sialoglycans and 2,7-anhydro-Neu5Ac, accounting for the different geometry of the sialic acid rings of 3'SL and 6'SL relative to 2,7-anhydro-Neu5Ac. During hydrolysis, the hydroxyl group at C7 attacks and forms a covalent bond with *C*2, while the lactose is released. The smaller |ΔDEEP-STD| values of protons H8 and H9s of 3'SL, relative to those of H8 and H9s of 2,7-anhydro-Neu5Ac, may be due to the displacement of the polyhydroxy chain further from the aromatic residues resonating at 6.55 ppm (Tyr667, and Trp698), as the reaction takes place.

The DEEP-STD factors for the H8-H9 of 6'SL exhibited a greater difference than those of 2,7-anhydro-Neu5Ac. For 6'SL, the ΔDEEP-STD of H8 is positive and the |ΔDEEP-STD| of H9 was even smaller compared to those of 3'SL and 2,7-anhydro-Neu5Ac. This suggests that the polyhydroxy chain of 6'SL is oriented in a different way in comparison to that of 3'SL (most likely pointing slightly further from the patch of aromatic residues).

In contrast, the DEEP-STD factors of the galactose and glucose moieties differed considerably among 3'SL and 6'SL, indicating a different orientation of their respective lactose moieties in the binding pocket. A homogeneous set of negative ΔDEEP-STD values were observed for the galactose moiety of 3'SL, suggesting that this galactose ring is in the vicinity of one or more aromatic residues. For 6'SL, instead, significantly negative ΔDEEP-STD were only observed for H5 and H6 of galactose and H2, H5 and H6 of glucose. Overall, the analysis of the DEEP-STD NMR fingerprints of the 3'SL and 6'SL compared against 2,7-anyhydro-Neu5Ac confirmed that both ligands bind in the same binding pocket, with analogous orientations of the reducing Neu5Ac but with a different orientation of the lactose moiety. A more detailed analysis of the differences in the DEEP-STD histograms of the lactose moiety is discussed in Sub-sections 2.3.2 and 2.3.3 with reference to the docking results.

2.3. NMR-validated 3D model of the Michaelis complex of the IT-sialidase RgNanH-GH33 by docking calculations and CORCEMA-ST analysis

The crystal structure of the complex of *Rg*NanH-GH33 with 2,7-anhydro-Neu5Ac (PDB ID: 4X4A [14]) was used as a template to generate a 3D model of the structures in solution of *Rg*NanH-GH33 in complex with 3'SL and 6'SL. In the docking calculations, both 3'SL and 6'SL converged to lowest energy solutions showing that the Neu5Ac ring fits deeply into the catalytic cavity (with *glide gscore* and *glide emodel* of *circa* −5.0 kcal/mol and −55 kcal/mol, respectively, comparable for both ligands), as shown in Fig. 4a,b. These results are discussed below in terms of the recognition of the different moieties (i.e. sialic acid, galactose and glucose) of the ligands:

2.3.1. Molecular recognition of the sialic acid moiety

In both models, the Neu5Ac reducing ring was found in a binding orientation comparable to that of 2,7-anhydro-Neu5Ac in the crystal structure (Fig. 4c,d). As deduced from DEEP-STD NMR, the glycerol

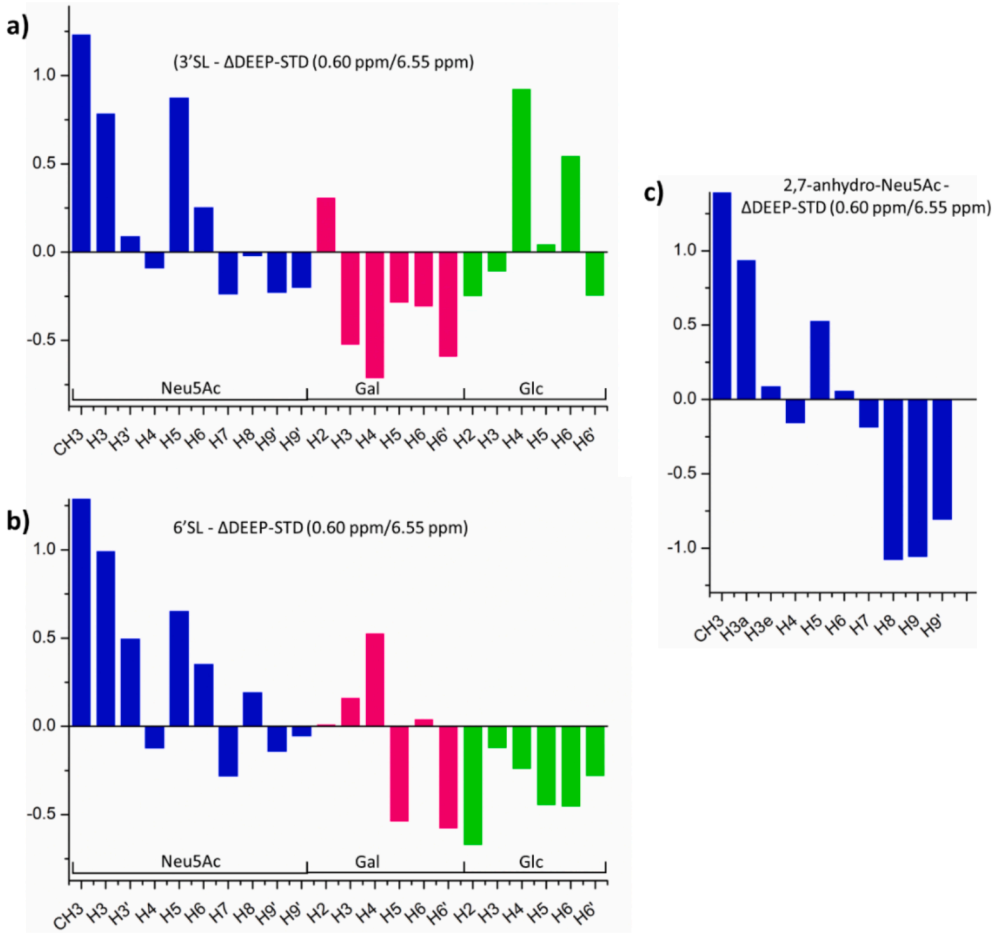


Fig. 3. DEEP-STD NMR fingerprint approach for sialoglycans vs. 2,7-anhydro-Neu5Ac. Differential Epitope Mapping (0.60 ppm/6.55 ppm) ΔDEEP-STD of **a)** 3′SL and **b)** 6′SL, in complex with *Rg*NanH-GH33 D282A, and of **c)** 2,7-anhydro-Neu5Ac in complex with *Rg*NanH-GH33 (histogram in **c**) adapted from data in reference[22]). DEEP-STD factors from protons of the Neu5Ac moiety are shown in blue, those from protons of the galactose moiety are shown in magenta, and those from protons of the glucose moiety are shown in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

moieties of the trisaccharides (purple and cyan arrows in Fig. 4c) were found to be further away from residues Tyr667 and Trp698 (resonances around 6.55 ppm) than the polyhydroxy chain of 2,7-anhydro-Neu5Ac (where the hydroxyl on C7 is covalently bound to *C*2). Protons H8 and H9s of 6′SL (purple histogram in Fig. 4d) appeared to point in the opposite direction relative to the same protons of 3′SL (cyan histogram in Fig. 4d), in agreement with the experimental DEEP-STD fingerprint data.

2.3.2. *Molecular recognition of the galactose moiety*

The main difference between the binding of 3′SL and 6′SL was in the orientation of the lactose disaccharide moiety. The α2/3 glycosidic linkage directs the galactose of 3′SL so that the region encompassing C3, C4 and C5 of the sugar ring can form a very efficient CH-π interaction with the aromatic side chain of Trp698 (Fig. 5a,c). For 6′SL, a distinct bending imposed by its more flexible glycosidic linkage brought the galactose into a lateral orientation with respect to the aromatic side chain of Trp698, exposing only protons H5 and H6 towards the indole, while the rest of the ring protons pointed outwards (Fig. 5b). This is in strong agreement with the DEEP-STD NMR results and the binding epitope mappings, which show weaker contacts for the galactose of 6′SL relative to the galactose of 3′SL (Fig. 1a,b). The lack of CH-π interactions for the α2/6 sugar may also account for the lower affinity compared to the α2/3 substrate.

2.3.3. *Molecular recognition of the glucose moiety*

The glucose residue of 6′SL was in closer proximity to the side chain of Trp698 than the adjacent galactose ring. The glucose *C*2 and C6 were shown to point towards the tryptophan, in strong agreement with the negative DEEP-STD factors observed for H2 and one of the H6s (Fig. 5d). In contrast, the glucose ring of 3′SL pointed towards the opposite side of the binding pocket, accounting for more positive or neutral values of the histogram in Fig. 5c.

To demonstrate that the mutation from aspartate to alanine at position 282 did not affect the binding mode of the sialoglycans into the *Rg*NanH-GH33 binding pocket, a 3D structure of the mutant was generated computationally, from the crystal structure of *Rg*NanH-GH33 [14], mutating the single aspartate 282 residue to an alanine (see Materials and Methods). Docking of 3′SL and 6′SL was repeated in the newly defined grid, and the resulting poses and clustering matched those obtained for *Rg*NanH-GH33 WT (see Supporting Information). Additionally, the DEEP-STD NMR patterns of i) 2,7-anhydro-Neu5Ac/*Rg*NanH-GH33 WT and ii) 2,7-anhydro-Neu5Ac/*Rg*NanH-GH33 D282A were equal within the experimental error (Fig. S2). These results strongly suggest that the D282A mutation does not affect the binding mode of 3′SL and 6′SL in the catalytic cleft.

Finally, to validate the 3D structures of the 3′SL/*Rg*NanH-GH33 and 6′SL/*Rg*NanH-GH33 complexes in solution, we used the full matrix relaxation approach CORCEMA-ST to simulate STD NMR build up curves from the 3D models generated by docking, and compared them

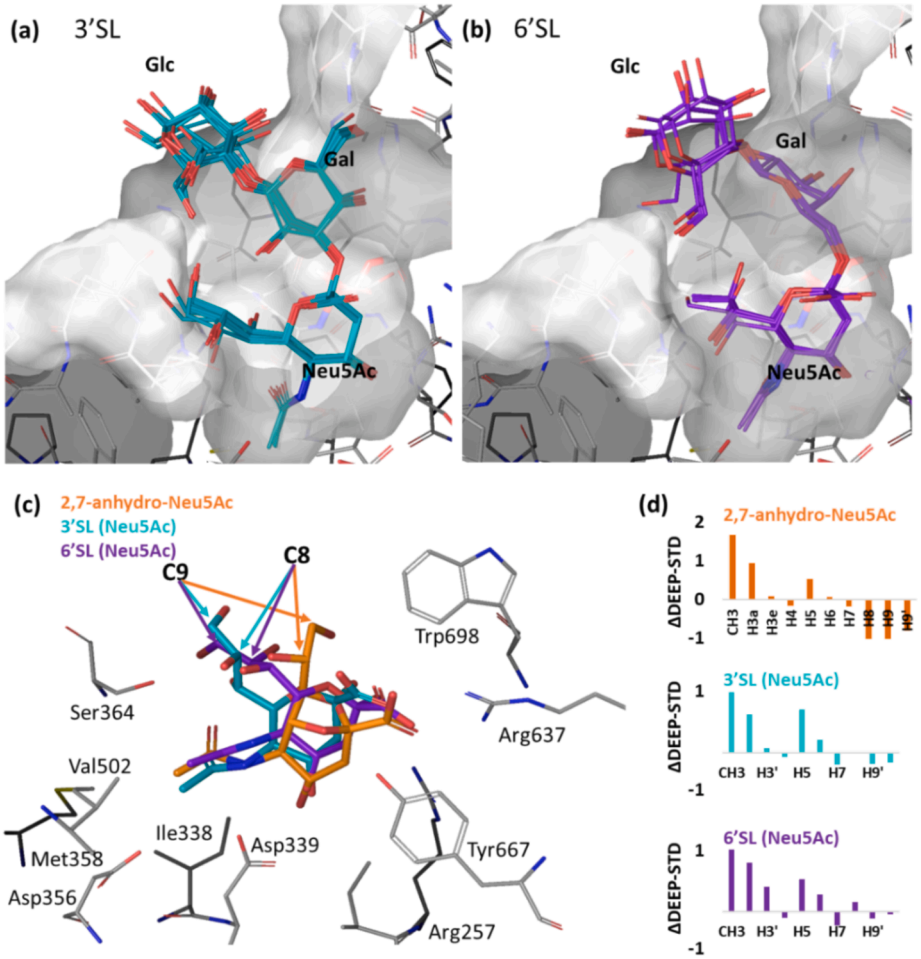


Fig. 4. DEEP-STD NMR data validation of docking poses: Neu5Ac moiety. Superimposition of the lowest energy convergent docking solutions for *Rg*NanH-GH33 in complex with **a)** 3′SL, 10 poses, and **b)** 6′SL, 6 poses. **c)** 2,7-anhydro-Neu5Ac (in orange) from PDB ID: 4X4A, overlapped with the Neu5Ac rings of the best docking solutions for 3′SL (cyan) and 6′SL (purple). Protein residues within 3 Å are shown and labelled. C8 and C9 are pointed by arrows. Protons are omitted for clarity (here and in the following molecular schemes). **d)** ΔDEEP-STD for 2,7-anhydro-Neu5Ac [22] and the Neu5Ac rings of 3′SL and 6′SL, following the same colouring scheme as in **c)**, which is adapted from Fig. 4). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with the experimental STD NMR data at an irradiation frequency of 0.60 ppm [27,28]. The accuracy of the model was validated by calculating an R-factor, called here NOE R factor; NOE R factors lower than 0.3 indicates validation of the 3D models by the experimental STD NMR data. Remarkably, NOE R factors of 0.22 and 0.26 were obtained for the lowest energy poses of 3′SL and 6′SL respectively, validating these as reliable models for the complexes in solution (Figure S3 and S4).

3. Discussion

The formation of the Michaelis complex of the catalytic domain of *Rg*NanH-GH33 with its sialoglycan substrate is the first step in the release of 2,7-anhydro-Neu5Ac in the gut. *Rg*NanH has a reported specificity for α2/3-linked sialic acid, as demonstrated by enzymatic assays using a range of oligosaccharides and by glycan microarrays [14,19].

Here, STD NMR binding epitopes, variable-temperature and competition STD NMR experiments revealed that *Rg*NanH-GH33 D282A binds both 3′SL and 6′SL though with different affinities and kinetics, with 3′SL inhibiting the binding of 6′SL in competition assays. DEEP-STD NMR spectroscopy led to the first NMR-validated 3D model of the Michaelis complex for an IT-sialidase. We showed that 3′SL and 6′SL bound to the active site via the sialic acid ring at the non-reducing end, and mainly differed in the orientation of the lactose disaccharide moiety. For 3′SL, a CH-π interaction with the side chain of Trp698 is formed, while the α2/6 glycosidic linkage of 6′SL directs its galactose ring further away from the tryptophan. Trp698 has been proposed to be one of the key residues involved in the IT-sialidase reaction, providing a hydrophobic environment that would repel water molecules from the catalytic cleft, and promote the O7 nucleophilic hydroxyl group to attack the covalent intermediate of the Neu5Ac bound to the nucleophilic tyrosine, therefore favouring an intramolecular reaction [6]. Trp698 has also been associated with the specificity of the IT-sialidases for α2/3 sialoglycans [31]. Our 3D models showed that the galactose ring is in a favourable position for CH-π interaction with the side chain of Trp698 for 3′SL, but not for 6′SL. This accounts for the lack of enzymatic activity on 6′SL, which is also recognised with much lower affinity. The higher affinity of 3′SL for the catalytic domain of *Rg*NanH can be explained thermodynamically as resulting from: i) favourable enthalpy contribution coming from the CH-π interaction with the side chain of Trp698 (only possible for 3′SL), and ii) unfavourable entropy contribution for 6′SL, due to the flexibility of the additional rotatable bond present at the α2/6 glycosidic bond.

Moreover, since the STD NMR competition experiments showed that 6′SL does not interfere with the interaction of 3′SL with *Rg*NanH-GH33 D282A, the potential for *Rg*NanH-GH33 to recognise and weakly bind α2/6 sialoglycans also present in mucins is not expected to affect the efficiency of hydrolysis of α2/3 sugars under physiological conditions. This is in line with the broad specificity of the carbohydrate binding module CBM40, associated to the GH33 of *Rg*NanH, for a wide range of

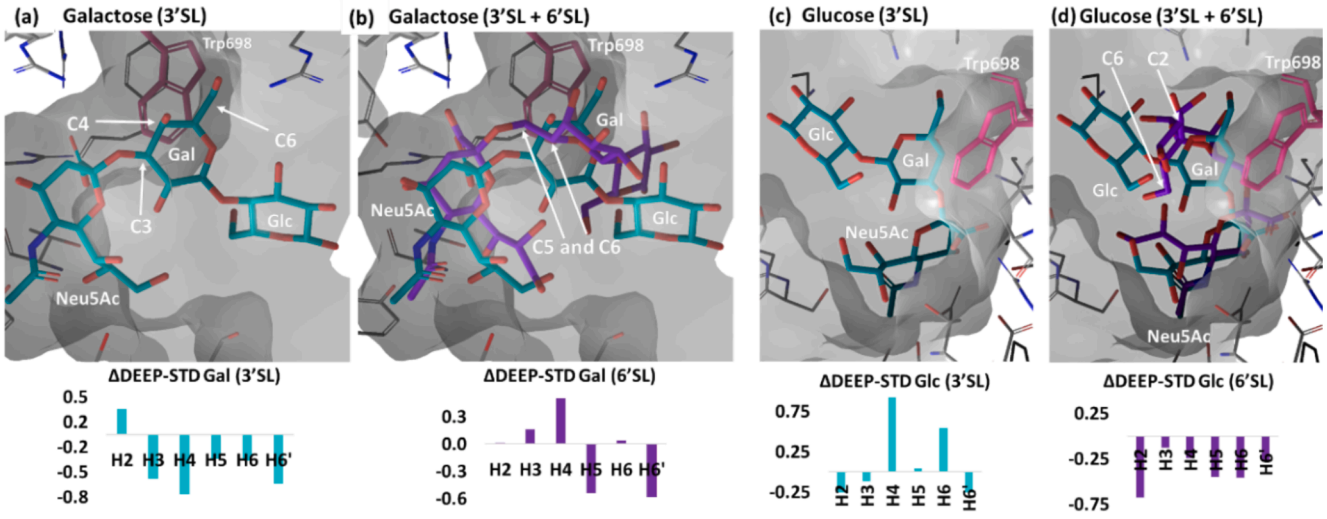


Fig. 5. DEEP-STD NMR data validation of docking poses: lactose moiety. Lowest energy docking solutions for 3′SL **(a, c)** and 6′SL + 3′SL **(b, d)** bound to *Rg*NanH-GH33. **a,b)**: top view of the galactose orientation relative to the side chain of Trp698 (shown in magenta). **c,d)**: side view of Trp698 (shown in magenta) relative to the galactose of 3′SL and the glucose of 6′SL. For the sake of easy comparison, ΔDEEP-STD histograms (0.60 ppm/6.55 ppm) of the galactose and glucose protons of both ligands, respectively, (adapted from Fig. 5) are shown at the bottom. Positions of ligands protons with negative ΔDEEP-STDs are pointed by arrows and labelled to show their close contact to the aromatic residue Trp698. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sialic acids derivatives including both 3′SL and 6′SL (as shown by microarrays, ITC, STD NMR and crystal structures [19]).

4. Conclusions

The use of multifrequency STD NMR in combination with computational techniques (docking and full matrix relaxation calculations with CORCEMA-ST) provided the first 3D structures of the complexes of *Rg*NanH-GH33 with 3′SL and 6′SL ligands. While *Rg*NanH-GH33 has been crystallised with its product 2,7-anhydro-Neu5Ac [14] and its crucial role in sialic acid scavenging to colonise the mucus environment has been demonstrated [18a], the structure of its Michaelis complex remained elusive. Here, we provide NMR-validated high-resolution 3D models for the complexes of *Rg*NanH-GH33 with both the substrate, 3′SL, as well as with 6′SL in solution. Multifrequency DEEP-STD NMR provided atomic level details of the crucial side chain-to-ligand contacts, confirming hypotheses on the residues responsible for the substrate specificity of the IT-sialidase. These results highlight the huge potential of multifrequency STD NMR approaches, such as DEEP-STD NMR, as a powerful tool in the structural biology of enzyme-substrate complexes, to shed light on the architecture and specificity of protein–ligand interactions at the atomic level, when a X-ray complex structure is unavailable.

5. Materials and Methods

5.1. Chemicals and proteins

Neu5Ac, deuterium oxide (99.9 % ^{2}H), sodium chloride and tris-(hydroxymethyl-d_3)-amino-d_2-methane (Tris-d_{11}, 98 % ^{2}H) were obtained from commercial providers. 3′SL and 6′SL were from Carbosynth. Enzymatically synthesised 2,7-anhydro-Neu5Ac and recombinantly produced *Rg*NanH-GH33 D282A were produced as previously described [19].

5.2. NMR measurements and processing

3′SL and 6′SL were assigned on the basis of 1D ^{1}H, 2D ^{1}H,^{1}H-DQF-COSY, ^{1}H,^{1}H-TOCSY, ^{1}H,^{13}C-HSQC and ^{1}H,^{1}H-NOESY experiments ran on samples of the free ligands in unbuffered D_2O, pH 7.0. For STD NMR experiments, all the samples consisted of 1 mM ligand and 50 μM *Rg*NanH-GH33 D282A in a D_2O pH 7.8 buffer solution, adjusted with HCl, containing 10 mM Tris-d_{11} and 100 mM NaCl (ligand: protein ratio 20: 1). To obtain the binding epitope mappings of the ligands, the STD NMR experiments were carried out at different saturation times (0.5, 1, 2, 3, 4 and 5 s) with 512 scans. A sequence that includes 2.5 ms trim pulses and a 3 ms spoil gradient was used (Bruker library name: stddiff.3). Saturation was achieved applying a train of 50 ms Gaussian pulses (0.40 mW) on the f2 channel, at 0.60 ppm and/or 6.55 ppm (on-resonance experiments) and 40 ppm (off-resonance experiments). The protein signal was removed using a 40 ms spinlock (T1ρ) filter. All the experiments were recorded at ^{1}H frequency of 800.23 MHz on a Bruker Avance III spectrometer equipped with 5-mm TXI probe. The STD NMR build-up curves were fitted to a mono-exponential equation, from which the initial slopes were obtained [20b]. The binding epitope was obtained by dividing all the initial slopes by the one of a convenient proton of the sialic acid residue (depending on the ligand), to which an arbitrary value of 100% was assigned. DEEP-STD factors were calculated using the DEEP-STD equation [22]:

$$\Delta\mathrm{DEEP}-\mathrm{STD}_i = \frac{STD_{exp1,i}}{STD_{exp2,i}} - \frac{1}{n}\sum_{i}^{n}\left(\frac{STD_{exp1,i}}{STD_{exp2,i}}\right)$$

5.3. Docking calculations

All molecular modelling was performed with the Glide module within Schrödinger's Maestro software suite, version 11 [32,33,34]. Coordinates for the receptor were obtained from the Protein Data Bank (PDB ID: 4X4A [14]). Where necessary, coordinates for missing atoms were added according to known protein chemistry and sidechain protonation was optimised for neutral pH. A short minimisation was run using the OPLS3 force field, converging heavy atoms to a RMSD of 0.3 Å. The receptor grid was then calculated, centring on the centroid of 2,7-anhydro-Neu5Ac and with a length of 30 Å (WT grid). Re-docking of the 2,7-anhydro-Neu5Ac in the prepped protein was performed and the docking conditions optimised based on the run that provided poses with minimal RMSD to the original XRD pose. The lowest re-docking pose is shown overimposed to the XRD pose in Fig. S5 of the SI. 3D structures of 3′SL and 6′SL were generated using a conformational search,

implementing Monte-Carlo torsional sampling, keeping only unique structures (RMSD > 0.5 Å) and eliminating all structures with an energy 21 kJ mol^{-1} greater than the lowest energy structure. All resulting structures were then minimised using conjugate gradient minimisation. For each ligand, the lowest energy conformations were used to initiate a preliminary round of docking, using the same conditions as for the re-docking run (OPSL3e force field, with standard precision, and sampling ring set to false). The docking consisted of further conformer generation, docking and then minimisation. Ten conformers were generated with 4x enhanced sampling and during docking, the non-bonded term of the potential energy function was softened for non-polar ligand atoms (charge < |0.15|) by applying a scaling factor of 0.8. Finally, minimisation was performed using implicit solvent with a distance dependent dielectric constant of 4. In a second round of docking, the 5 most diverse poses of the first round of docking were exported and re-docked in the same conditions, generating about 50 poses. To determine if the mutation D282A did not affect the binding mode, the D282 of the minimised receptor was mutated to A282 and a new grid was generated (mutant grid) with the same feature as the first. The 3D builder toolbox was used to mutate the residue. A second round of docking was then repeated with the mutant grid and a similar set of poses was generated.

5.4. CORCEMA-ST predictions

For CORCEMA-ST ([27]), the cut-off distance around the protons of the ligand in the binding pocket was 10 Å. The concentrations of ligand and protein used were 2 mM and 50 μM, respectively. k_{on} was set to 1×10^{-8} M^{-1} s^{-1}. The bound ligand correlation time was 50 ns, whereas the free correlation time was 1 ns. The equilibrium constant used was 5000 M^{-1} and 2000 M^{-1} for 3′SL and 6′SL respectively. The non-specific relaxation leakage term (ρ leak) used was 0.4 Hz, to account for the effect of traces of dissolved paramagnetic oxygen, and the irradiation frequency was set to the range 0.1–1.1 ppm to simulate the conditions of irradiation at 0.60 ppm. The NOE R-factor was calculated on the averaged simulated data obtained, with the following equation:

$$NOER = \sqrt{\frac{\sum \left(STD\%_{exp} - STD\%_{calc}\right)^2}{\sum \left(STD\%_{exp}\right)^2}}$$

Author Contributions

All authors have given approval to the final version of the manuscript.

Funding Sources

SM and JA acknowledge support from BBSRC, grant BB/P010660/1. MW and SM thank UKRI for funding via a Future Leaders Fellowship to MW (MR/T044020/1). JA thanks support from Ministerio de Ciencia e Innovación via the grants AEI/10.13039/501100011033/(PID2019-109395 GB-I00 and PID2022-142879NB-I00), co-funded by the European Regional Development Fund (ERDF) "A way of making Europe". LET, AB, NJ gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC), which funded this research through the BBSRC Institute Strategic Programme Gut Microbes and Health BB/R012490/1 and Gut Health and Food Safety BB/J004529/1.

CRediT authorship contribution statement

Serena Monaco: Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Louise E. Tailford:** Resources. **Andrew Bell:** Resources. **Matthew Wallace:** Writing – review & editing. **Nathalie Juge:** Writing – review & editing, Writing – original draft, Validation, Supervision, Funding acquisition, Conceptualization. **Jesus Angulo:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are grateful for the use of the University of East Anglia Faculty of Science NMR facility.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bioorg.2024.107906.

Data availability

Data will be made available on request.

References

[1] I. Sekirov, S.L. Russell, L.C.M. Antunes, B.B. Finlay, Physiol. Rev. 90 (2010) 859–904.
[2] G.P. Donaldson, S.M. Lee, S.K. Mazmanian, Nat. Rev. Microbiol. 14 (2016) 20–32.
[3] M.E. Johansson, J.M.H. Larsson, G.C. Hansson, Proc. Natl. Acad. Sci. 108 (2011) 4659–4665.
[4] A.L. Lewis, W.G. Lewis, Cell. Microbiol. 14 (2012) 1174–1182.
[5] J.P. Ouwerkerk, W.M. de Vos, C. Belzer, Best Pract. Res. Clin. Gastroenterol. 27 (2013) 25–38.
[6] N. Juge, L. Tailford, C.D. Owen, Biochem. Soc. Trans. 44 (2016) 166–175.
[7] L.E. Tailford, E.H. Crost, D. Kavanaugh, N. Juge, Front. Genet. 6 (2015) 81.
[8] A. Bell, N. Juge, Glycobiology 31 (2021) 691–696.
[9] E.R. Vimr, K.A. Kalivoda, E.L. Deszo, S.M. Steenbergen, Microbiol. Mol. Biol. Rev. 68 (2004) 132–153.
[10] J. Qin, R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez and T. Yamada, *nature* **2010**, *464*, 59.
[11] E.H. Crost, L.E. Tailford, G. Le Gall, M. Fons, B. Henrissat, N. Juge, PLoS One 8 (2013) e76341.
[12] C. Manichanh, N. Borruel, F. Casellas, F. Guarner, Nat. Rev. Gastroenterol. Hepatol. 9 (2012) 599–608.
[13] E.H. Crost, E. Coletto, A. Bell, N. Juge, FEMS Microbiol. Rev. 47 (2023) fuad014.
[14] L.E. Tailford, C.D. Owen, J. Walshaw, E.H. Crost, J. Hardy-Goddard, G. Le Gall, W. M. De Vos, G.L. Taylor, N. Juge, Nat. Commun. 6 (2015) 7624.
[15] Y. Luo, S.-C. Li, M.-Y. Chou, Y.-T. Li, M. Luo, Structure 6 (1998) 521–530.
[16] H. Gut, S.J. King, M.A. Walsh, FEBS Lett. 582 (2008) 3348–3352.
[17] E.H. Crost, L.E. Tailford, M. Monestier, D. Swarbreck, B. Henrissat, L.C. Crossman, N. Juge, Gut Microbes 7 (2016) 302–312.
[18] a) A. Bell, J. Brunt, E. Crost, L. Vaux, R. Nepravishta, C.D. Owen, D. Latousakis, A. Xiao, W. Li, X. Chen, Nat. Microbiol. 4 (2019) 2393–2404;
b) A. Bell, E. Severi, M. Lee, S. Monaco, D. Latousakis, J. Angulo, G.H. Thomas, J. H. Naismith, N. Juge, J. Biol. Chem. 295 (2020) 13724–13736;
c) H. Wu, E.H. Crost, C.D. Owen, W. Van Bakel, A. Martínez Gascueña, D. Latousakis, T. Hicks, S. Walpole, P.A. Urbanowicz, D. Ndeh, PLoS Biol. 19 (2021) e3001498.
[19] C.D. Owen, L.E. Tailford, S. Monaco, T. Suligoj, L. Vaux, R. Lallement, Z. Khedri, H. Yu, K. Lecointe, J. Walshaw, S. Tribolo, M. Horrex, A. Bell, X. Chen, G.L. Taylor, A. Varki, J. Angulo, N. Juge, Nat. Commun. 8 (2017) 15.
[20] a) M. Mayer, B. Meyer, Angew. Chem. Int. Ed. 38 (1999) 1784–1788;
b) J. Angulo, P.M. Nieto, Eur. Biophys. J. 40 (2011) 1357–1369.
[21] M. Mayer, B. Meyer, Journal of the American Chemical Society 123 (2001) 6108–6117.
[22] S. Monaco, L.E. Tailford, N. Juge, J. Angulo, Angew. Chem. Int. Ed. 56 (2017) 15289–15293.
[23] J.E. Watt, G.R. Hughes, S. Walpole, S. Monaco, G.R. Stephenson, P.C. Bulman Page, A.M. Hemmings, J. Angulo, A. Chantry, Chem.–A, Eur. J. 24 (2018) 17677–17680.
[24] S. Monaco, S. Walpole, H. Doukani, R. Nepravishta, M. Martínez-Bailén, A. T. Carmona, J. Ramos-Soriano, M. Bergström, I. Robina, J. Angulo, Chem.–A, Eur. J. 26 (2020) 10024–10034.
[25] R. Nepravishta, S. Monaco, M. Distefano, R. Rizzo, P. Cescutti, J. Angulo, Front. Mol. Biosci. 8 (2021) 727980.

[26] K. Malec, S. Monaco, I. Delso, J. Nestorowicz, M. Kozakiewicz-Latała, B. Karolewicz, Y.Z. Khimyak, J. Angulo, K.P. Nartowski, J. Colloid Interface Sci. 638 (2023) 135–148.
[27] V. Jayalakshmi, N.R. Krishna, J. Magn. Reson. 155 (2002) 106–118.
[28] V. Jayalakshmi, T. Biet, T. Peters, N.R. Krishna, J. Am. Chem. Soc. 126 (2004) 8610–8611.
[29] I.K. Leung, M. Demetriades, A.P. Hardy, C. Lejeune, T.J. Smart, A. Szöllössi, A. Kawamura, C.J. Schofield, T.D. Claridge, J. Med. Chem. 56 (2013) 547–555.
[30] Y.S. Wang, D. Liu, D.F. Wyss, Magn. Reson. Chem. 42 (2004) 485–489.
[31] Y. Luo, S.-C. Li, Y.-T. Li, M. Luo, J. Mol. Biol. 285 (1999) 323–332.
[32] R.A. Friesner, R.B. Murphy, M.P. Repasky, L.L. Frye, J.R. Greenwood, T.A. Halgren, P.C. Sanschagrin, D.T. Mainz, J. Med. Chem. 49 (2006) 6177–6196.
[33] T.A. Halgren, R.B. Murphy, R.A. Friesner, H.S. Beard, L.L. Frye, W.T. Pollard, J. L. Banks, J. Med. Chem. 47 (2004) 1750–1759.
[34] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M. P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, J. Med. Chem. 47 (2004) 1739–1749.