Universidad Zaragoza

1542

Master Project

Studies of interactions of glycosyltransferases through molecular dynamics simulations

Author

Alejandro Montesa Gómez

Director

Pedro Merino Filella

Master in Biophysics and Quantitative Biotechnology

BIFI, University of Zaragoza

2024/2025

Index

Abstract

Fringe proteins are GT-A inverting glycosyltransferases that catalyze the addition of GlcNAc (donor) to *O*-fucosylated EGF motifs of Notch receptors (acceptor), modifying its ligand-binding affinity. In this study, molecular dynamics simulations of the mouse homolog, Mfng, in different states of its proposed catalytic cycle, have been performed to obtain deeper insights of its interactions and mechanism. Simulations of UDP and UDP-GlcNAc bound systems revealed how these molecules interact with Fringe's active site and coordinate the Mn^{2+} metal ion, necessary for the protein catalytic activity. Interesting results have been obtained for ternary systems (protein-donor-acceptor), where changes in the secondary structure of a long, flexible loop, which was modelled after crystallization, demonstrate Fringe's activation upon substrate binding, similarly to other GT-A fold glycosyltransferases. These studies establish coherent starting points for more advanced mechanistic studies, for which it is necessary to use QM/MM or metadynamics approaches that will allow the complete elucidation of the reaction mechanism.

Introduction to glycosyltransferases

Carbohydrates play diverse and vital roles in living organisms, ranging from constituting structural components to mediating cell-cell recognition processes. Given the vast amount and variety of carbohydrate-based structures found in nature, a large group of enzymes is necessary not only for their synthesis but also for their degradation and modification. These enzymes are collectively known as carbohydrate-active enzymes (CAZymes) and include glycoside hydrolases (GHs), polysaccharide lyases (PLs), and glycosyltransferases (GTs) (Ardèvol & Rovira, 2015).

Glycosyltransferases (GTs) (EC 2.4.1.x) are enzymes that catalyze the transfer of sugar moieties from activated donor molecules to specific acceptor molecules that can be mono- and oligosaccharides, peptides and proteins, lipids and other molecules, resulting in the synthesis of complex carbohydrates, glycoproteins, glycolipids and other glycoconjugates (Lairson et al., 2008).

Among the vast number of GTs, Leloir glycosyltranferases (65% of all GTs) (Lairson et al., 2008) are a specific family that use nine nucleotide sugars as donors to transfer sugar moieties to acceptor molecules, typically, proteins and oligosaccharides (Figure 1). The donor sugars in GTs are most commonly activated as nucleoside-diphosphate sugars (e.g., UDP-GlcNAc), but monophosphate sugars (e.g., CMP-NeuAc), lipid phosphates or unsubstituted phosphates can also serve as donors.

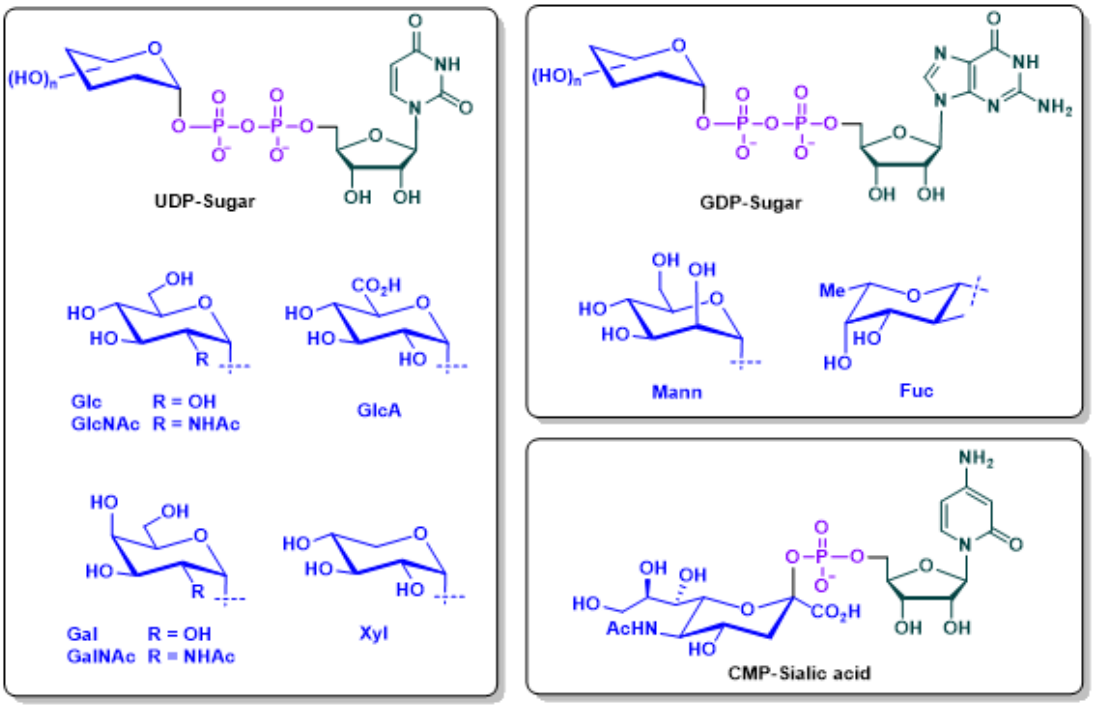


Figure 1. The nine Leloir donors of glycosyltransferases.

Glycosyltransferases can be sub-classified based on various criteria, including substrate class, protein structure, stereochemical preference, and metal dependency for catalytic activity (Mestrom et al., 2019).

Sequence classification

Traditionally, GTs were classified using systems that required demonstrating their function according to the recommendations of the International Union of Biochemistry and Molecular Biology (IUBMB), a task that has become increasingly challenging with the growing volume of genomic sequenced data. Since 1997, thanks to Campbell and colleagues, glycosyltransferases have been classified into families based on similarities in amino acid sequences. This has resulted in a continuously evolving classification that has expanded from the original 27 families described in 1997, to the current 137 families listed in the CAZy database (http://www.cazy.org/).

More than 750.000 glycosyltransferases sequences are known and, although this classification is based on the absence of sequence homology between members of different families, some sequence motifs common to members of distinct families have been identified, like the DxD motif, involved in ion metal binding (discussed below), or motifs related to substrate specificity such as the sialyl motifs (Rini et al., 2022).

Despite some of this families being polyspecific (containing different EC numbers), the tertiary structure of their members remains conserved (just as for GHs), which suggest that slight differences in the three-dimensional structure of GTs may account for their donor/acceptor specificity. Thus, making functional predictions of new glycosyltransferases based on its fold becomes unreliable or inaccurate.

It is worth highlighting that not all GT families are equally populated; as, according to the CAZy database, GT2 and GT4 families account for half of the total number of GTs identified (Lairson et al., 2008).

3D Structure classification

Attending to their three-dimensional structure, GTs are mainly comprised of five different protein folds: GT-A, GT-B, GT-C, GT-D and GT-E (Figure 2); being GT-A and GT-B folds the most common.

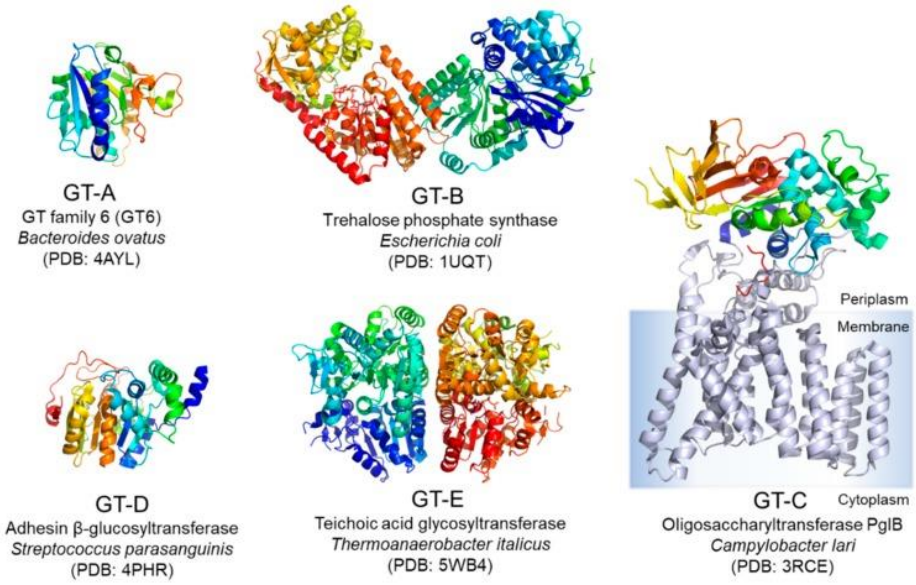


Figure 2. Protein fold of Leloir glycosyltransferases (GT-A, GT-B, GT-D) and non-Leloir glycosyltransferases (GT-C, GT-E). Obtained from (Mestrom et al., 2019).

The GT-A fold consists of two tightly associated β/α/β domains of variable size that form a continuous central β-sheet. Most enzymes in this group exhibit a conserved DXD motif, where the carboxylates coordinate a divalent cation and/or a ribose. Although these motifs are considered defining features of GT-A glycosyltransferases, some GTs lacking this signature do exist (Albesa-Jové et al., 2014). Like the GT-A fold, the GT-B architecture also contains two β/α/β Rossmann-like domains. However, in GT-B enzymes, the domains are less tightly associated and face each other, resulting in the active site being located within the cleft between the two domains. These domains are associated with donor and acceptor substrate-binding sites.

Structural representatives of GT-C folds are hydrophobic integral membrane proteins that use lipid phosphate-linked sugar donors, which is not surprising given their subcellular location (Gloster, 2014).

The GT-D fold, firstly described for DUF1792, consists of three regions: an N-terminal region with two β strands that form sandwich domains containing the metal-binding site (in this case, a DXE motif); a central region with a Rossmann-like fold; and a C-terminal region. This fold is distinct from the classic glycosyltransferase folds, GT-A and GT-B, and does not share any similarities with the previously described GT-C fold (Zhang et al., 2014).

The GT-E fold was first described in the TagA enzyme, which catalyzes the initial step in the biosynthesis of wall teichoic acids in *Staphylococcus aureus* and other pathogenic bacteria. This enzyme adopts a unique α/β tertiary structure, markedly different from previously described glycosyltransferases. TagA lacks the canonical DXD motif found in GT-A glycosyltransferases and differs substantially from GT-B and GT-C enzymes, which adopt multi-domain structures. However, it shows limited homology with the founding

member of the GT-D family, which transfers glucose from UDP-glucose onto protein *O*-linked hexasaccharides. Despite this homology, the number and topology of their secondary structure elements—and thus, their catalytic mechanisms—are different (Kattke et al., 2019).

Stereochemical preference

Reactions that involve the formation of new glycosidic bonds have two possible stereochemical results: the retention or inversion of the anomeric configuration with respect to the transferred sugar. Thus, we can also classify these enzymes as retaining or inverting, depending on the stereochemical outcome of the reaction. Logically, two different outcomes must result from the utilization of different reactions mechanisms.

Most inverting mechanisms occurs via a S_N2 reaction in a single displacement step (Figure 3A), where an hydroxyl group of the acceptor substrate attacks the donor sugar anomeric carbon, displacing the leaving group nucleotide moiety from the opposite face and resulting in the inversion of anomeric configuration of the product. Enzymes that employ this mechanism usually utilize a catalytic base (like Glu, Asp or His) that assist the catalysis by deprotonating the nucleophile. However, some inverting enzymes without a catalytic base near the active site have been identified (e.g., POFUT1 and FUT1). In these cases, a S_N1 reaction mechanism was proposed (Figure 3B), where the β-phosphate oxygen atom of the donor assists in the deprotonation of the attacking hydroxyl group. This reaction may be promoted by a water molecule that provides a proton relay between the hydroxyl and the β-phosphate oxygen (Moremen & Haltiwanger, 2019).

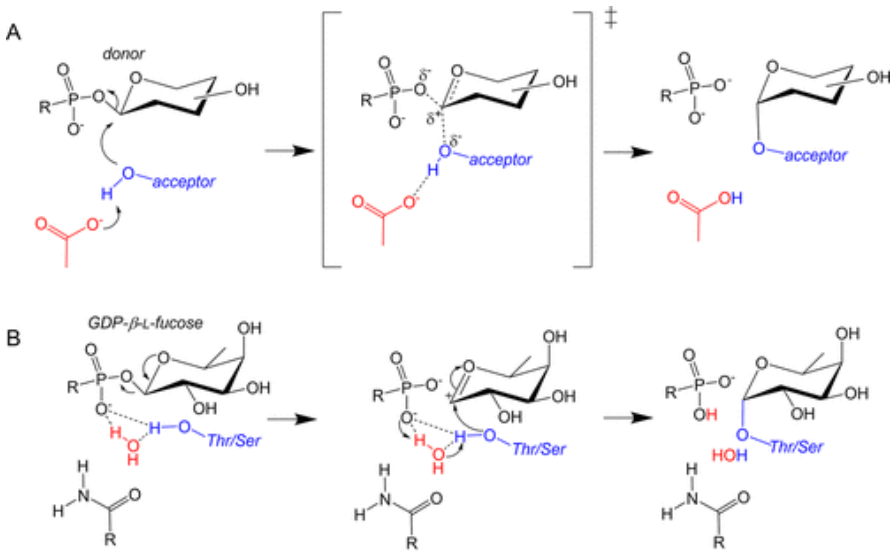


Figure 3. (A) Classical inversion mechanism, in which a carboxylic acid residue acts as a general base (with a general sugar donor). (B) Proposed mechanism of POFUT1 where the transfer of the acceptor proton to the β-phosphate can take place directly or via a water molecule. Adapted from (Piniello et al., 2021).

The retaining mechanism is not as well understood, with two main mechanisms being actually discussed. Early on, inverting GTs were though to proceed via Koshland double-displacement (similar to GHs), where a covalent enzyme intermediate with inverted anomeric configuration is formed after the nucleophilic attack of the donor (Figure 4B). This adduct undergoes a second attack by the acceptor that breaks the glycosyl-enzyme bond, and forms a new glycosidic bond with overall retention of the initial configuration (Forrester et al., 2022).

However, structural evidence for this mechanism remains insufficient. Currently most retaining GTs adhere to an "internal return" S_Ni-like substitution mechanism (Figure 4A), wherein the nucleophilic hydroxyl group attacks the anomeric carbon from the same side that the departure of the leaving group takes place. So, the key point of this mechanism is that both the leaving group and the hydroxyl group need to interact, because the leaving group is acting as a catalytic base, deprotonating the nucleophile (Rao et al., 2024).

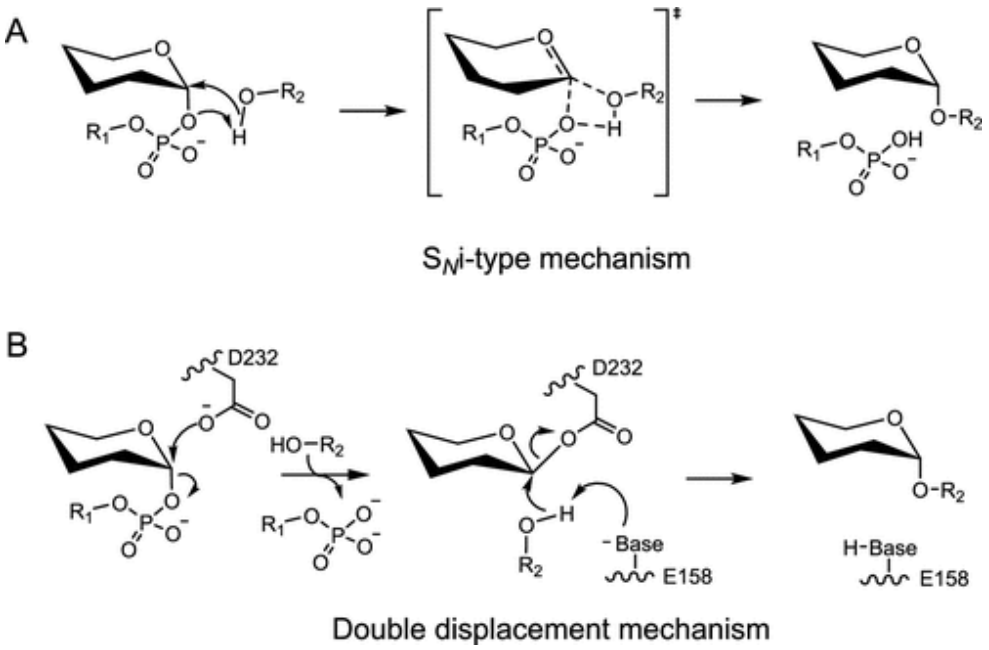


Figure 4. Proposed catalytic mechanism for retaining GTs. (A) S_Ni-like mechanism; (B) double-displacement mechanism. Obtained from (Rao et al., 2024).

Nevertheless, this S_Ni-like substitution mechanism is not aware of discussion; some authors suggest that the reaction occurs in a concerted manner, which means that bond cleavage and formation occurs simultaneously (as depicted in Figure 4, A). Others argue that the "internal return" mechanism involves the formation of a covalent ion-pair intermediate, which is stabilized after the donor group is released but before the acceptor binds.

Fringe

In this work, we have fix our attention in Fringe proteins, Leloir GTs that use UDP-GlcNAc as glycosyl donor. Fringe proteins are inverting Golgi-resident β1,3-N-acetylglucosaminyltransferases that modify Notch receptors by catalyzing the addition of N-acetylglucosamine (GlcNAc) to *O*-fucose glycans attached to specific epidermal growth factor (EGF) motifs of the receptor.

These EGF-like domains (Figure 5) consist of 30–40 amino acids with six conserved cysteines residues that form three internal disulphide bridges. EGF motifs are usually found in the extracellular domains (ECD) of membrane-bound proteins, where they often occur in multiple tandem copies. In fact, 36 tandem copies are found in Notch ECD, and they are responsible for mediating the interactions between the receptor and its ligands, Delta and Serrate (also known as Jagged).

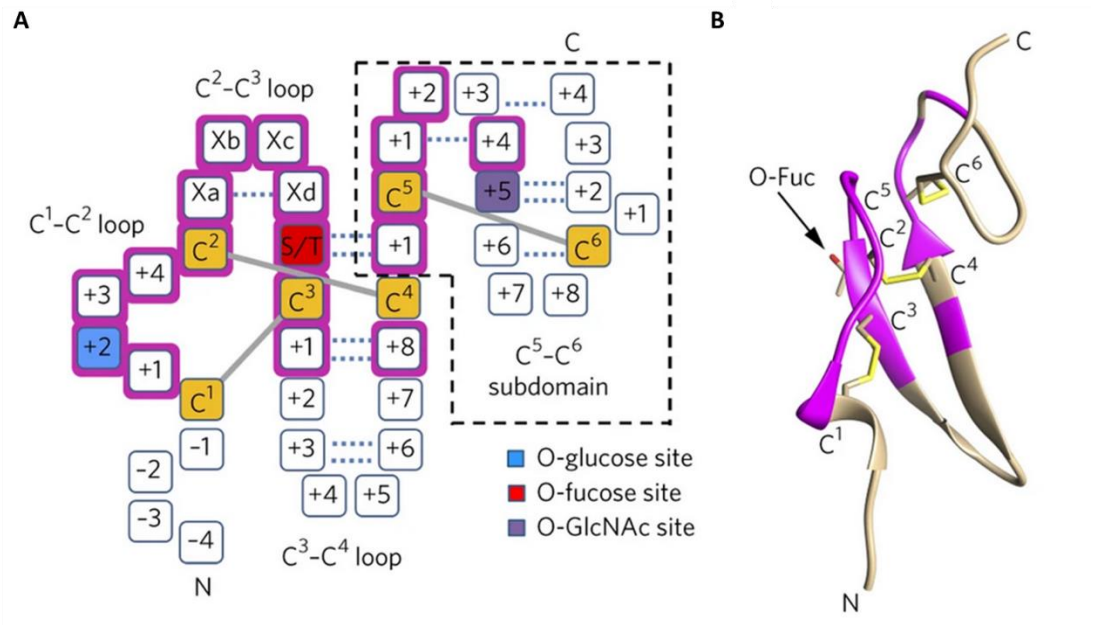


Figure 5. (A) Schematic representation of EGF-like domain. S/T residue is where *O*-fucose is added by POFUT1 protein and elongated with GlcNAc by Fringe. (B) *O*-fucosylated EGF three-dimensional structure. Adapted from (Li et al., 2017).

Notch receptors are single-pass Type I transmembrane proteins composed of a large extracellular domain. After ligand binding, a conformational change is produced in the juxtamembrane domain of Notch, rendering its sensibility to successive cleavages by different metalloproteases. These cleavages release the intracellular domain of the receptor, which is translocated to the nucleus where it forms a transcription activation complex with other transcription factors (Wang et al., 2015).

This signalling pathway, despite being simple, is remarkably pleiotropic in its outcomes. Notch plays a critical role in fundamental processes for the cell, such as cell proliferation, death or development. Thus, aberrant functioning of Notch signalling components is directly linked with human and animal disorders (Zhou et al., 2022).

Fringe enzyme was first described in *Drosophila melanogaster*, where it was identified as a cell-autonomous modulator of Notch activity. In other words, Fringe regulates Notch activity within the same cell, without relying on external factors or signal from other cells. In flies, Fringe enhances Notch activation by one of its ligands, Delta, while inhibiting activation by Serrate (Figure 6). In mammals, three Fringe orthologues have been found: Manic Fringe (Mfng), Lunatic Fringe (Lfng), and Radical Fringe (Rfng). These orthologues also modulate Notch receptor activity, though in a more variable manner, as Rfng can activate signalling from both ligands (Moloney et al., 2000).

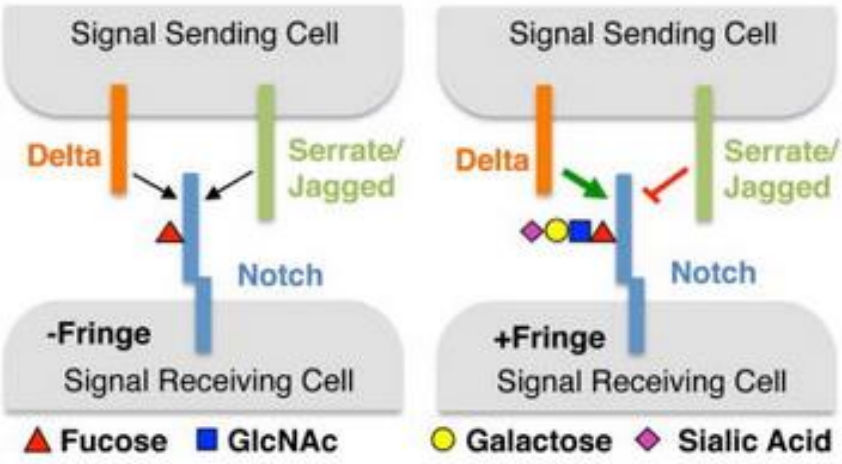


Figure 6. Elongation of O-fucose residues by Fringe modulates Notch activity, resulting in enhanced activation from Delta-family ligands but reduced from Serrate/Jagged-family ligands. Adapted from (Kakuda & Haltiwanger, 2017).

Jinek et al. crystalized the catalytic fragment of mouse Mfng in 2016, and this will be the starting point of the present study. This fragment consist in a seven stranded β-sheet, in which the sixth strand is antiparallel, flanked by α-helices (Figure 7, blue) and a C-terminal domain with three β-strands and a α-helix (Figure 7, pink), which resembles the core structure found in other glycosyltransferases with GT-A fold.

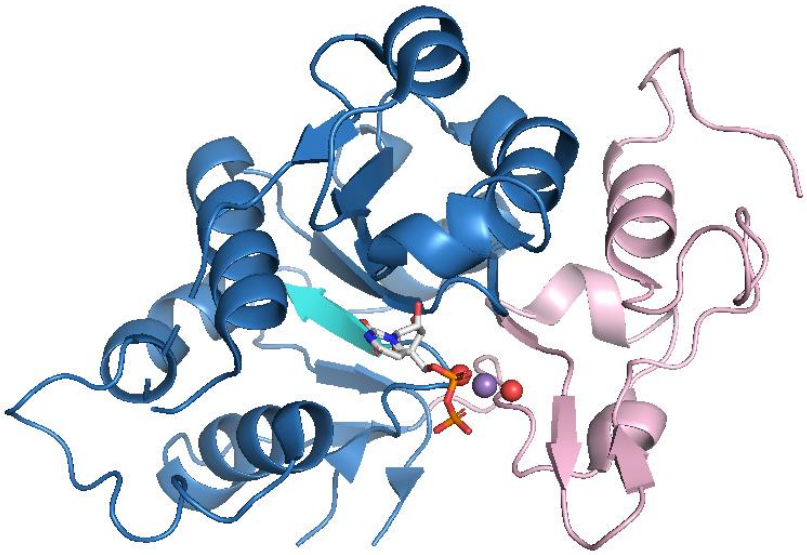


Figure 7. Catalytic fragment of mouse Mfng. The GT-A core is coloured blue, except for the antiparallel strand of the β-sheet which is coloured cyan, and the C-terminal domain pink; following the fashion described in Jinek et al. (2006). UDP is shown as sticks and Mn^{2+} (purple) and water molecules (red) are shown as spheres.

Similarly to other GT-A glycosyltransferases, Mfng presents a DxD motif (in this case, Asp142-Asp143-Asp144) and a Mn^{2+} ion located in the nucleotide sugar-binding site. This Mn^{2+} is coordinated by Asp144, His256, two water molecules, and in turn coordinates the α and β phosphates of UDP, acquiring an octahedral geometry with six atoms coordinating the metal (Figure 8A). On the other hand, Asp143 is in contact with the 2'- and 3'- hydroxyls of the ribose sugar through its carboxylate oxygens and the nitrogen atom of the peptide backbone (Figure 8B).

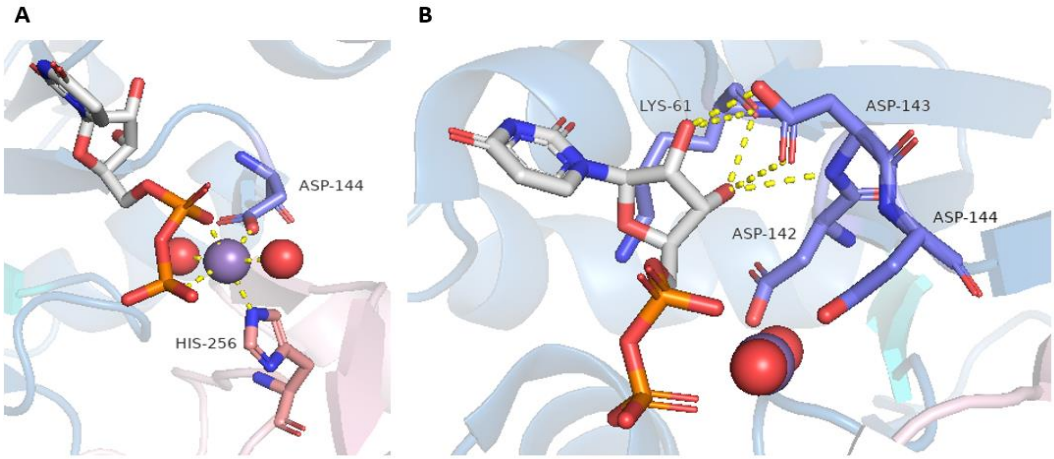
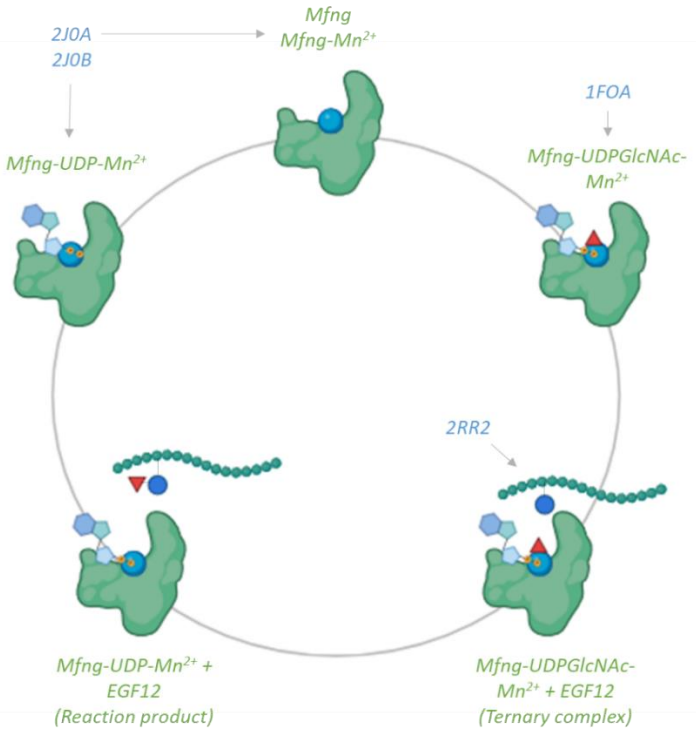


Figure 8. (A) Mn^{2+} coordination sphere, comprised of Asp144 (blue sticks), His256 (pink sticks), two water molecules (red spheres) and UDP. (B) Polar contacts of the 2'- and 3'- hydroxyl groups of the ribose in UDP.

Objectives

The main objective of this work is **to obtain deeper insights into the molecular mechanism of Fringe by analysing the steps before and after the reaction through molecular dynamics (MD) simulations** (Scheme 1). We will use the catalytic fragment of mouse homolog Mfng for the MD studies. Locations of those steps corresponding to the reaction itself, i.e. reactant, transition structure and product, require the use of QM/MM techniques which are beyond the present study and they will be addressed in further studies.



Scheme 1. Catalytic cycle proposed for Fringe. System that are proposed to be constructed and simulated are identified in green, while PDB structures that are going to be used for this purpose are identified in blue.

To assess this main objective some specific objectives have been proposed:

- To construct the different systems to be simulated
- To perform MD simulations of the constructed systems
- To analyse the obtained trajectories
- To extract conclusions from the analysed data

The obtained results could be used to compare the conformational differences at various stages of the catalytic cycle. In the future, by performing metadynamics or QM/MM studies, we could characterize possible transition states that may arise during the reaction mechanism, aiming to a better comprehension of it. This approach could also facilitate the rational design of Fringe inhibitors or mutations with potential applications in the biotechnological industry.

Computational methods

Construction of models

To commence this study, available PDB structures of the Fringe protein (PDB ID: 2J0A, resolution: 1.80 Å; PDB ID: 2J0B, resolution: 2.10 Å) were obtained (Jinek et al., 2006). These structures correspond to the catalytic domain (spanning residues 45 to 321) of the apo-protein and the protein in complex with UDP and Mn^{2+}, respectively. Missing residues from the structures were modelled using ChimeraX software (Pettersen et al., 2021). It is important to take into account that ChimeraX alters the original numbering of the residues in the PDB, starting from 1. Thus, is necessary to calculate the gap between the new structure and the original PDB to follow any original discussion on the paper describing the X-ray. The protonation states of the charged amino acids were assigned by MolProbity (Williams et al., 2018), paying special attention to H256, that has to be protonated only on δ nitrogen (HID) to coordinate the Mn^{2+} ion. Cysteines residues forming disulphide bonds need also to be checked, changing the residue name from CYS to CYX in the corresponding PDB files. In this case C110-C121, C139-C202 and C306-C315 are forming disulphide bonds.

More complex systems were constructed starting from these initial structures. An structure of Fringe in complex UDP-N-acetylglucosamine (UDP-GlcNAc) was obtained after structural superposition of 2J0B structure with the crystal structure of rabbit N-acetylglucosamynyltransferase I in complex with UDP-GlcNAc and Mn^{2+} (PDB ID: 1FOA, resolution: 1.80 Å). Subsequently, a ternary complex structure, consisting of the protein bound to the donor substrate, i.e., UDP-GlcNAc, and the acceptor substrate, i.e., *O*-linked fucose, was obtained after docking the monosaccharide into the expected acceptor site.

The complete substrate structure was after added by superimposition of the ternary complex with the crystal structure of the *O*-fucosylated epidermal growth factor-like repeat 12 of mouse Notch1-receptor (PDB ID: 2RR2). Using GLYCAM-Web tools (http://glycam.org), we were also able to model the N-acetyl-β-D-glucosamine-(1→3)-α-L-fucose disaccharide to generate a structure with the product of the reaction.

Hydrogen atoms, solvation box, and counter ions necessary to neutralize the charge of the systems were added with AmberTools utility tLeaP (Case et al., 2023). Systems were placed in an orthorhombic water box with the protein edge at the closest distance of 12 Å to the box.

MD simulations

Theoretical introduction

Molecular dynamics simulations are computational methods used to study the physical movements of atoms and molecules over time. They rely on classical mechanics, typically using Newton's laws of motion, to predict the behaviour of particles within a

system. By solving the equations of motion for each particle in the system, MD simulations provide a temporal evolution of the system, which is called trajectory. MD simulations are particularly useful in understanding how molecular systems, such as proteins, nucleic acids, or small molecules, behave in various environments like aqueous solutions or within biological membranes (Paquet & Viktor, 2015).

The interactions between particles are defined by force fields, which consist of potential energy functions based on parameters such as bond lengths, angles, van der Waals forces, and electrostatic interactions. The choice of these potential functions is critical in order to obtain accurate simulations.

When simulating macromolecules, such as proteins, it is important to consider that these molecules are not isolated. Therefore, they must be solvated to mimic their natural environment. Solvation can be achieved in two ways: implicit solvation, where the water molecules are replaced by a potential, or explicit solvation, where the macromolecule is surrounded by a box of solvent molecules. The shape of this solvent box can vary (e.g., cubic, spherical), but it is usually designed to minimize the number of molecules required for solvation, reducing computational costs. To avoid or minimize unnatural boundary effects, periodic boundary conditions (PBC**)** are applied. However, certain considerations must be made when using PBC to ensure accurate results (Paquet & Viktor, 2015).

Methodology

MD simulations were performed with Amber22 (Case et al., 2005; Salomon-Ferrer et al., 2013). The force fields used are the following: FF14SB (Maier et al., 2015) for enzyme residues, GLYCAM06 (Kirschner et al., 2008) for sugar ligands, GAFF2 (He et al., 2020) for other organic ligands such as UDP, and TIP3P (Jorgensen et al., 1983) for water solvent. The simulations were carried out through several steps. A two-stage energy minimization was performed for every system: first, energy was minimized keeping the protein and ligands fixed using a restraint of 500 kcal/mol/Å^2, while solvent molecules and ions were allowed to move; then an unrestricted energy minimization for all the system was performed. The system was then gradually heated from 0K to 300K in the NVT ensemble for 100 ps with a weak positional restraint of 25 kcal/mol/Å^2 applied on the protein backbone, allowing the solvent and ions to equilibrate around it. Continuing from the heating phase, the system was equilibrated in two different stages of 100 ps each under the NPT ensemble, using the Berendsen thermostat scheme (Berendsen et al., 1984) to control the temperature and the Berendsen barostat to maintain the target pressure (pressure relaxation time of 2 ps). Finally 1000 ns production runs were performed for every system, in ten cycles of 100 ns each. By default, frames were extracted from the production files every 200 ps. During all the MD simulations bonds containing hydrogen were constrained using the SHAKE method (Kräutler et al., 2001) with an integration step of 2fs. Also to note that an 8 Å cutoff was applied to Lennard-Jonnes and electrostatic interactions.

For every system, various replicas were run. Production runs of these replicas were started using a restart file from the equilibration of their respective original simulation. Thus, to ensure that each replica was independent, but consistent in terms of the MD parameters, allowing reliable comparisons and ensemble analysis of the obtained outputs, random seeds were inputted for each run. As the new production run was not preceded by any minimization or heating, velocities of the system were rescaled using the Langevin thermostat scheme at 300 K.

Trajectory analysis

The obtained trajectories were analysed using PyMOL (Schrödinger & DeLano, 2020) and CPPTRAJ (Roe & Cheatham, 2013). One of the most common types of analysis performed after producing MD simulations is the calculation of the root-mean-squared deviation (RMSD) of the coordinates.

RMSD measures the deviation of the structure coordinates to a reference set of coordinates, being zero if the overlap between both sets is perfect. Thus, RMSD can be defined as:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(r_i^X - r_i^Y)^2}$$

Where *N* is the number of atoms, *X* and *Y* are the target and reference structures, respectively; and *i* is the current atom. RMSD can be calculated for all frames of a simulation to determine the stability of the system itself. A highly dynamical RMSD plot may suggest instability or maybe that the system has not reached equilibrium yet. On the other hand, a relatively balanced RMSD plot may indicate that stability or equilibrium has been reached during the simulation. Therefore, the number of atoms and their order must be the same between the target and the reference structure, for the RMSD to have a meaningful statistical interpretation.

Hydrogen bond analysis was also carried out using CPPTRAJ. Hydrogen bonds are formed between an electronegative atom, the acceptor, which is usually nitrogen, oxygen or fluorine; and a hydrogen atom covalently bonded to another electronegative atom, the donor. Because of the differences in electronegativity, the H atom bears a partial positive charge, while the N, O or F atom bears a partial negative charge.

Hydrogen bonds are a dominant feature in the secondary structure of proteins but also in the formation of complexes between protein and ligands. Thus, analysing hydrogen bonds can provide insights into the stability of the protein or the complex. CPPTRAJ tracks hydrogen bonds over the trajectory using geometric criteria: the distance between the donor and acceptor heavy atoms and optionally the donor-hydrogen-acceptor angle.

Other interactions were analysed using the *nativecontacts* command of CPPTRAJ, thanks to which we were able to track contacts between residues during the simulation. A contact is considered native if it is present in the first frame of the simulation. In contrast, we speak of non-native contacts for those that appear during the simulation and were not present in the initial frame. With this information we were able to compare contacts between protein chains at the start and the end of the simulation.

CPPTRAJ was also utilized to analyse the secondary structure propensities of certain residues using the DSSP method, that assigns a secondary structure type (Table 1) for each residue based on the backbone amide and carbonyl atom positions (Kabsch & Sander, 1983).

Integer	DSSP character	Secondary structure type
0	‘ ’	None
1	‘E’	Parallel β-sheet (Extended)
2	‘B’	Anti-parallel β-sheet (Bridge)
3	‘G’	3-10 helix
4	‘H’	α helix
5	‘I’	π helix
6	‘T’	Turn
7	‘S’	Bend

Table 1. Secondary structure types assigned by DSSP method and their corresponding integer/character. Obtained from https://amberhub.chpc.utah.edu/secstruct/.

Results and Discussion

Apo- and holo-enzyme simulations

The first simulated system comprised the apoenzyme (PDB ID: 2J0A, chain completed with ChimeraX; see Methods), which lacked the Mn^{2+} ion, likely lost during the crystallization process conducted by Jinek et al. (2006). This system it's not biologically relevant, as the absence of the metal is not a feasible state during the catalytic cycle of Fringe. Nevertheless, this simulation was beneficial for grasping the logic processes involved in performing MD simulations and to analyse possible changes in the conformation of the protein. Another MD simulation was carried out for the apoenzyme, this time in the presence of the Mn^{2+} ion, placed by us computationally in the expected place. This system could resemble either the state of the protein before binding the substrates or the state of the protein once the reaction has occurred and the UDP has left the active site.

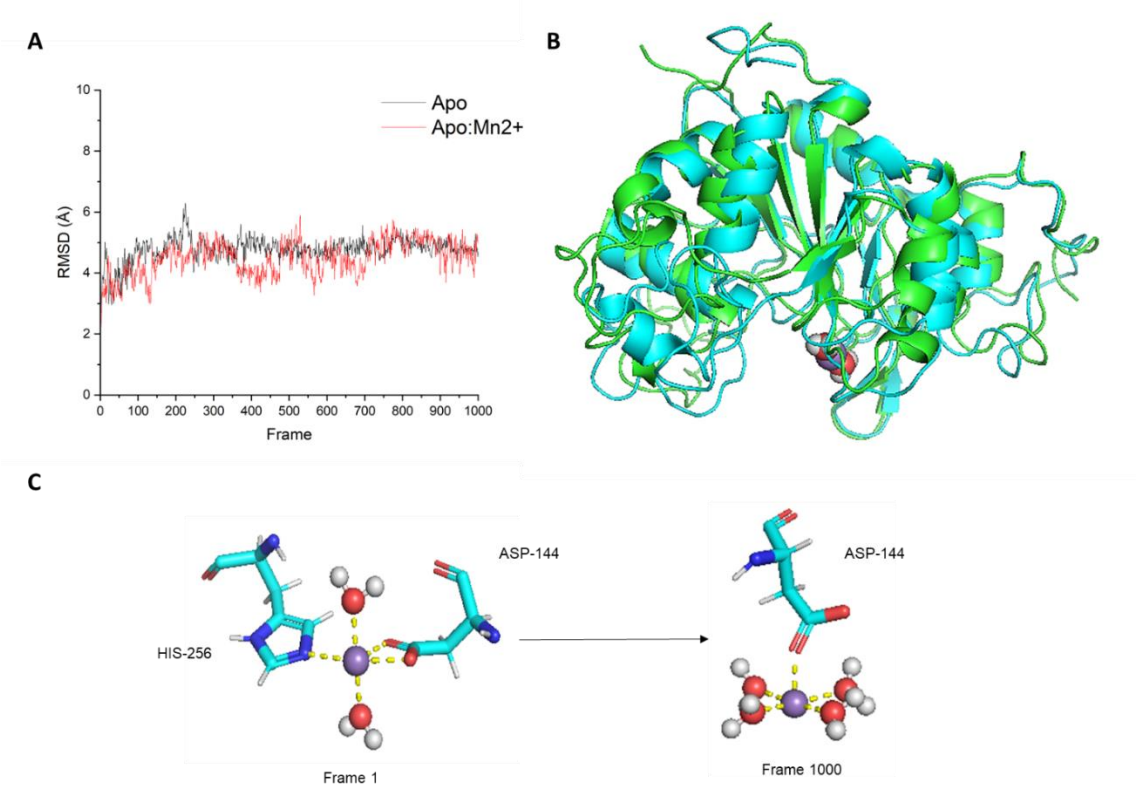


Figure 9. (A) RMSD of each system for all amino acid residues throughout the MD simulations compared to the solvated structure. RMSD for the apoenzyme without the metal ion is shown in black, and for the Mfng-Mn^{2+} complex is shown in red. (B) Structural alignment of stable frames (frame 1000 in each case) from the apoenzyme without Mn^{2+}, coloured in green, and the Mfng-Mn^{2+} complex, in cyan. (C) Comparison of the Mn^{2+} coordination sphere in frame 1 and in frame 1000 of the Mfng-Mn^{2+} simulation.

To evaluate the stability of the dynamics, the RMSD was calculated for each frame with respect to the solvated structure generated after using tleap (Figure 9A). Apart from the disordered loops, no significant structural differences were observed between the apoenzyme with Mn^{2+} (Figure 9B, cyan) and without Mn^{2+} (Figure 9B, green) when stable frames from their respective simulations were aligned (RMSD of 1.294 Å over 200 Cα atoms).

Interestingly, the geometry of the Mn^{2+} coordination sphere changed during the simulation of the Mnfg-Mn^{2+} complex (Figure 9C). Initially, it adopted a bitrigonal pyramidal geometry, where D144, H256 and two water molecules were coordinating the metal. By the end of the simulation, it transitioned to a square pyramidal geometry, where only D144 remained coordinating the Mn^{2+} along four water molecules. Despite these changes, the coordination index, i.e. the number of atoms coordinating the metal, did not change.

Hydrogen bonding between protein residues was studied for both systems by using CPPTRAJ, as implemented in AMBER22 (see Methods). The number of hydrogen bonds per frame had a similar behaviour; in fact, the mean number of hydrogen bonds was the same for both systems, $\bar{x}$ ≈207 (Figure 10A). Thus, the presence of the metal ion does not affect the formation of hydrogen bonds between protein residues. Although many different hydrogen bonds were identified during the simulation, only some of them were present in more than 90% of the frames (Figure 10B).

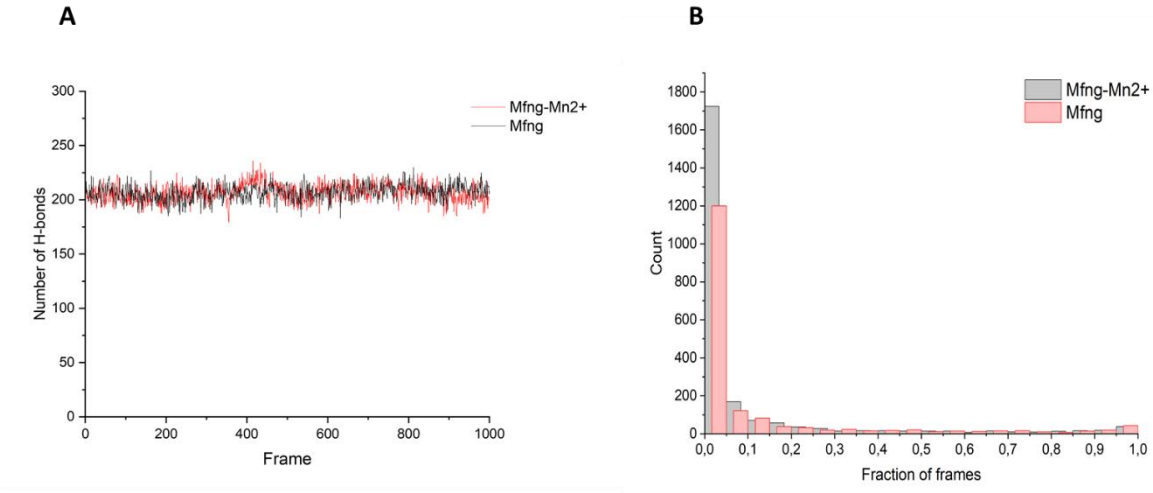


Figure 10. (A) Representation of the number of hydrogen bonds between protein residues per frame for Mfng-Mn^{2+} complex (red) and Mfng without the metal (black). (B) Number of hydrogen bonds per fraction of frames in which they are present, following the same fashion as in panel (A) for both systems.

When simulating the Mfng-UDP-Mn^{2+} complex (PDB ID: 2J0B, chain completed with ChimeraX; see Methods), several approaches were attempted to preserve the Mn^{2+} coordination sphere found in the crystal structure (see Fringe) throughout the entire simulation, but all of them were unsuccessful. Only in a few cases the coordination sphere was maintained for shorts periods of time. Best results were achieved when restraining the entire coordination sphere of Mn^{2+} for 300 ns and then releasing the restraints. The system remained stable during 600 ns and then, H256 was substituted by a water molecule in the coordination sphere (Figure 11B, and 11C). Despite these changes, the octahedral geometry of the coordination sphere was conserved (Figure 11B). The stability of the dynamics was evaluated similarly to the previous systems (Figure 11A).

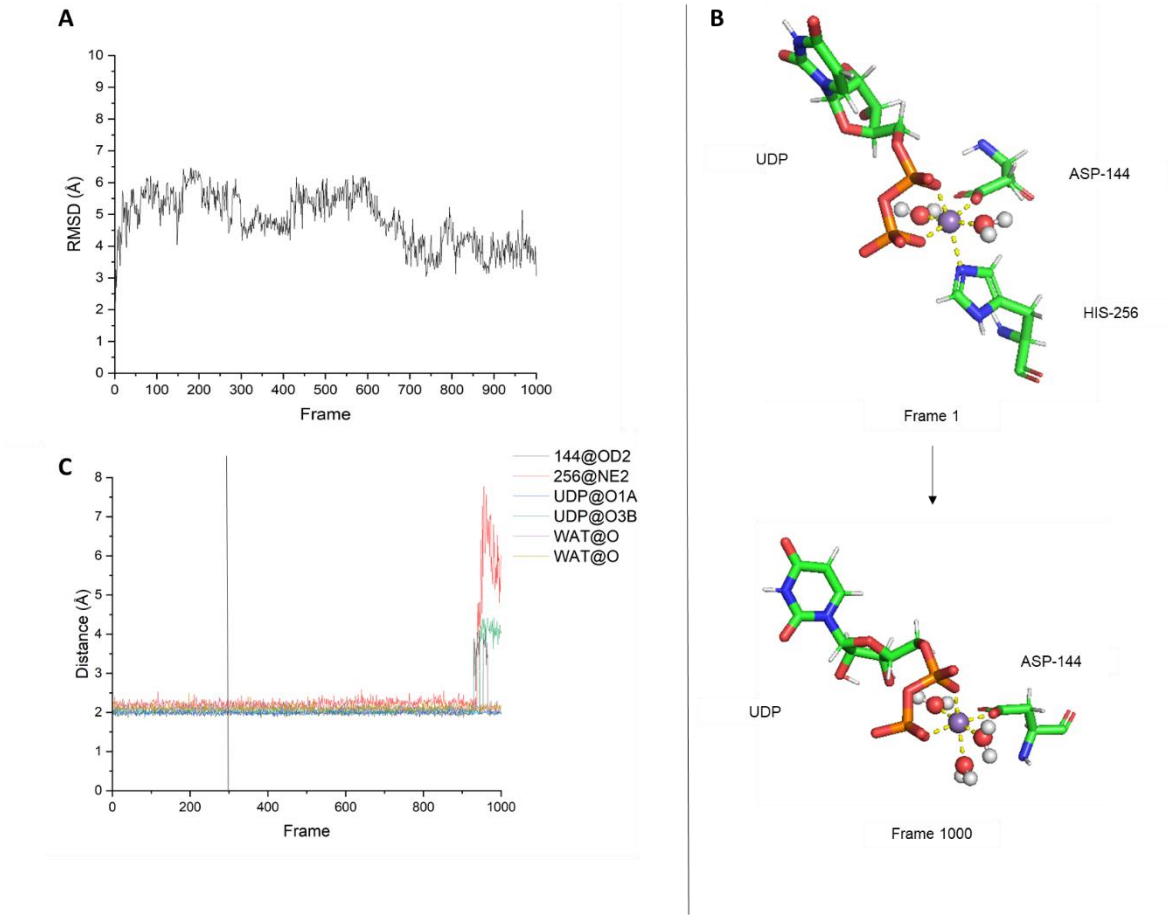


Figure 11. (A) RMSD of the system for all amino acids residues throughout the MD simulation compared to the solvated structure generated by tleap. (B) Comparison of the Mn^{2+} coordination sphere in frame 1 and in frame 1000. (C) Distances of the atoms coordinated to the Mn^{2+} metal, calculated for every frame of the MD simulation. H256 leaves the coordination site during the dynamics (red line), but another water molecule starts coordinating the metal. Vertical line indicates end of restraints.

On the other hand, when simulating the Mfng-UDP-GlcNAc-Mn^{2+} complex, the Mn^{2+} coordination sphere was preserved during the entire simulation time in all the replicas without any restraint (Figure 12B, and Figure 12C). Despite one of the water molecules from the crystal being replaced by another water molecule during the simulation, no significant changes were observed in the metal's coordination sphere (Figure 12B and Figure 12C). This suggests that the presence of the sugar moiety may stabilize interactions between nearby residues and the metal. The stability of the simulation was evaluated in a similar wat to that of previous systems (Figure 12A).

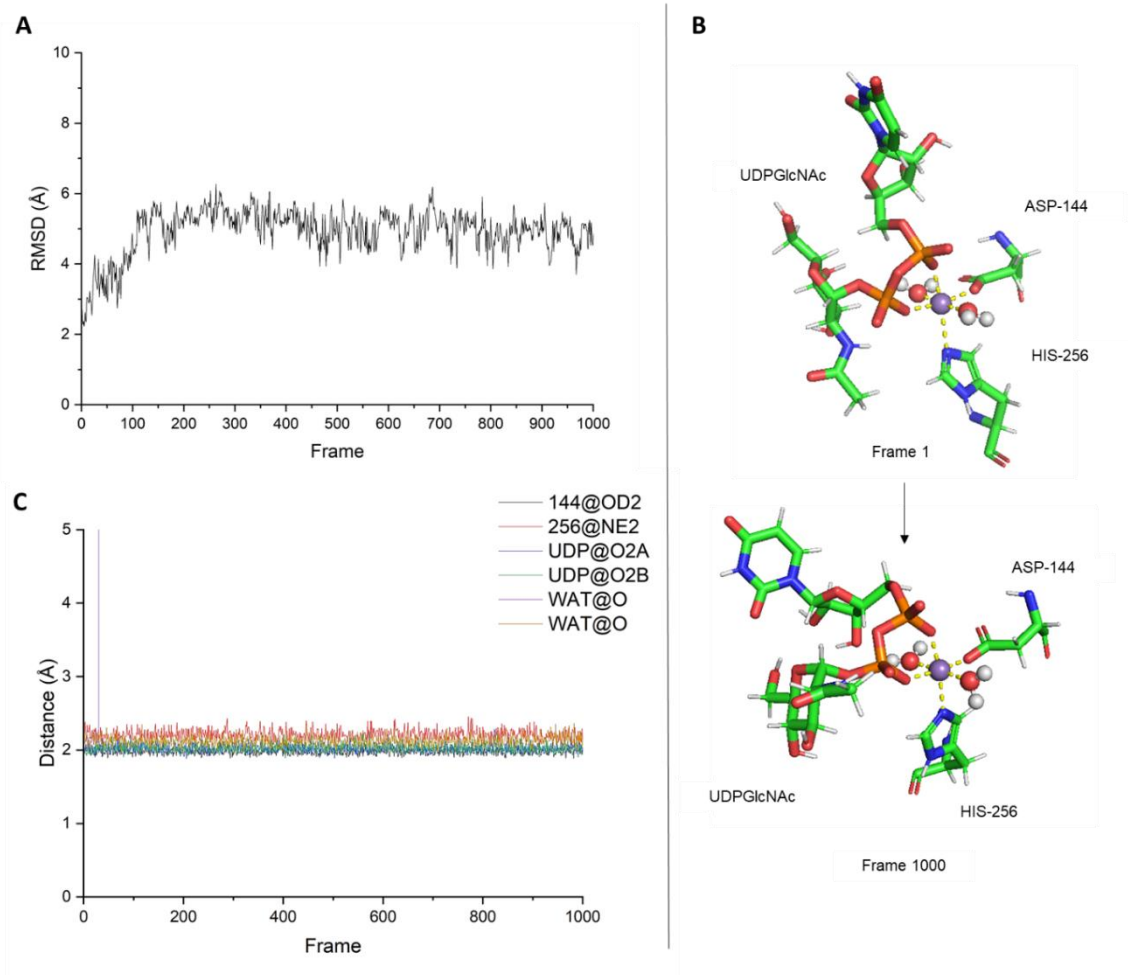


Figure 12. (A) RMSD of the system for all amino acids residues throughout the MD simulation compared to the solvated structure generated by tleap. (B) Comparison of the Mn^{2+} coordination sphere in frame 1 and in frame 1000. (C) Distances of the atoms coordinated to the Mn^{2+} metal, calculated for every frame of the MD simulation. One water molecule leaves the coordination site during the dynamics (vertical purple line), but another water molecule starts coordinating the metal.

No structural changes were identified when aligning a stable frame from the Mfng-UDP-Mn^{2+} simulation (Figure 13A, in purple) with a stable frame from the Mfng-Mn^{2+} simulation (Figure 13A, in cyan). The RMSD for the alignment was 1.376 Å over 204 Cα atoms.

Apart from the long loop modelled by ChimeraX (Figure 13, encircled), where changes in secondary structure elements were visible (though not provoking a conformational change), no significant structural differences were observed when aligning a stable frame from the Mfng-UDP-GlcNAc-Mn^{2+} simulation (Figure 13B, in green) with a stable frame from the Mfng-Mn^{2+} simulation (Figure 13B, in cyan). The RMSD for the alignment was 1.152 Å over 195 Cα atoms.

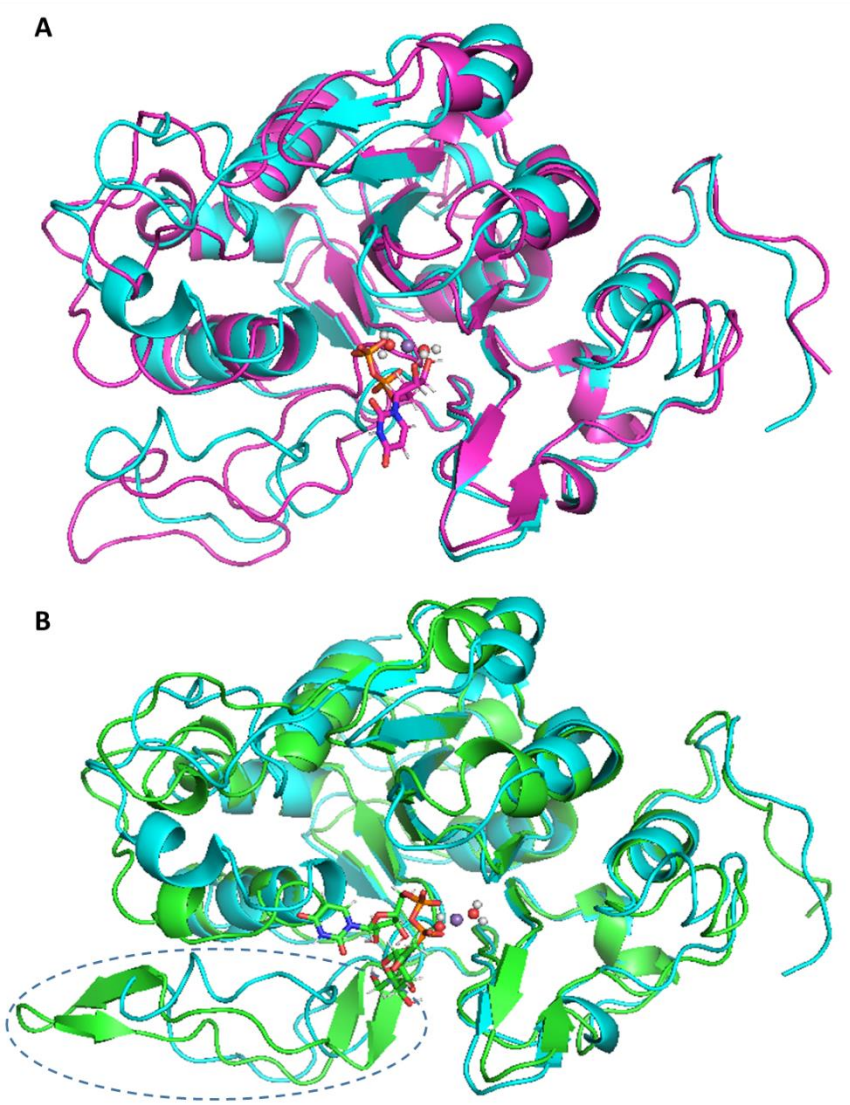


Figure 13. (A) Structural alignment of stable frames (frame 1000 in each case) from Mfng-UDP-Mn^{2+}, coloured in magenta, and Mfng-Mn^{2+}, in cyan (B) Structural alignment of stable frames (frame 1000 in each case) from Mfng-UDP-GlcNAc-Mn^{2+}, coloured in green, and Mfng-Mn^{2+}, in cyan. Equivalent regions that has adopted different secondary structure are encircled.

When analysing hydrogen bonding between protein residues, similar results were obtained for systems with UDP and UDP-GlcNAc compared to the ligand-free systems that have been analysed before, and even when compared to each other (Figure 14A). Notably, the analysis of the hydrogen bonds formed between protein residues and the ligands, UDP and UDP-GlcNAc (Figure 14B, magenta and green, respectively), was also of interest. The mean number of hydrogen bonds between the protein and its respective donor substrate was quite similar for both systems.

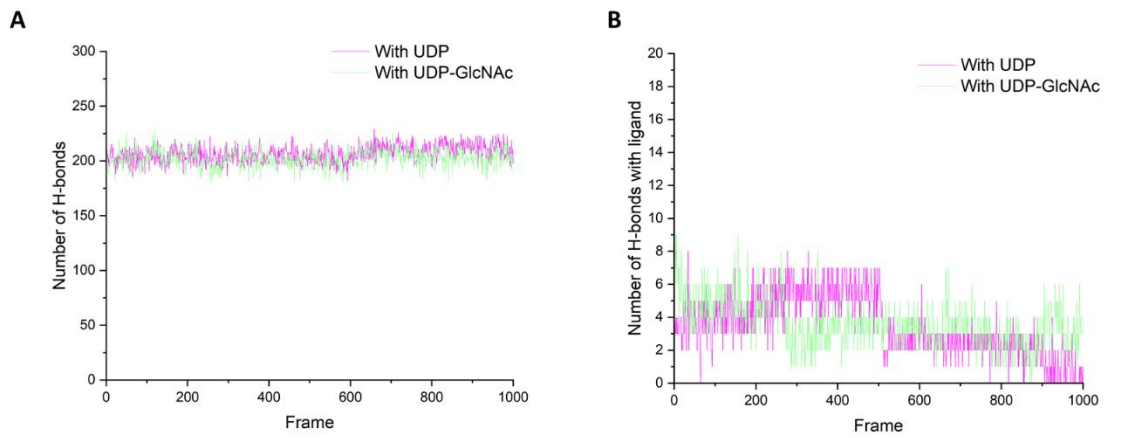


Figure 14. (A) Representation of the number of hydrogen bonds per frame between protein residues for Mfng-UDP-Mn^{2+} complex (purple) and Mfng-UDP-GlcNAc-Mn^{2+} (green). (B) Representation of the number of hydrogen bonds per frame between the cofactor and the protein residues, following the same fashion as in panel A for both systems.

The reason of that similarity could be the location of the diphosphate nucleotide in the active site, positioned and oriented towards the exterior rather than the protein core, limiting the number of residues that are able to interact with it. On the other hand, while the number of interactions between the UDP-GlcNAc and the protein stabilized at 3-4 hydrogen bonds throughout the entire trajectory (Figure 14B, in green; and Figure 15B), the number of hydrogen bonds formed by the UDP decreased, to 1 or even 0 in some frames (Figure 14B, in purple; and Figure 15A). This is not an unexpected result, since the UDP must be released after the reaction occurs, allowing a new molecule of UDP-GlcNAc to enter the active site.

Tables 2 and 3 list hydrogen bonds between the protein and its ligand (UDP and UDP-GlcNAc, respectively) that were present in more than 200 frames (20% of the frames).

Acceptor	Donor	Number of frames	Average Distance (Å)	Average consecutive frames
UDP@O2A	R70@NH2	645	2.96	5
UDP@O2A	R70@NH1	560	2.95	5
UDP@O3’	D143@N	344	3.26	4
UDP@O1A	R70@NH1	305	2.93	3
D143@OD2	UDP@O2’	285	2.70	8
D143@OD2	UDP@O3’	203	2.87	8
UDP@O2’	T63@N	200	3.18	4

Table 2. Hydrogen bonds acceptor and donor atoms between Mfng and UDP. Each atom is specified by the one-letter code for the residue name (e.g., S for Serine), the residue position in the sequence (e.g., 171), and the atom type after the "@" symbol (e.g., O for oxygen). UDP atoms types are assigned based on Figure S1 in the Supplementary Information.

Acceptor	Donor	Number of frames	Average Distance (Å)	Average consecutive frames
UDP-GlcNAc@O1A	R70@NH2	837	2.88	8
UDP-GlcNAc@O1A	R70@NH1	659	2.94	4
UDP-GlcNAc@O2A	R70@NH1	232	3.09	2
D232@OD2	UDP-GlcNAc@O2’	211	2.68	7
UDP-GlcNAc@O4	D231@N	200	3.03	2

Table 3. Hydrogen bonds acceptor and donor atoms between Mfng and UDP-GlcNAc. Each atom is specified by the one-letter code for the residue name (e.g., S for Serine), the residue position in the sequence (e.g., 171), and the atom type after the "@" symbol (e.g., O for oxygen). UDP-GlcNAc atoms types are assigned based on Figure S2 in the Supplementary Information.

For both ligands, most important interactions with the protein were formed between one oxygen atom of the α phosphate and the guanidine group nitrogen atoms of R70. For UDP, interactions with R70 were broken towards the end of the simulation (Figure 16A), while these interactions were preserved for UDP-GlcNAc (Figure 16B).

UDP also formed hydrogen bonds with D143 carboxylate oxygens through the hydroxyl groups of the ribose moiety, as it was observed in the crystal structure (see Fringe). These interactions were also present in the case of UDP-GlcNAc, but appeared in a lower number of frames when compared with UDP.

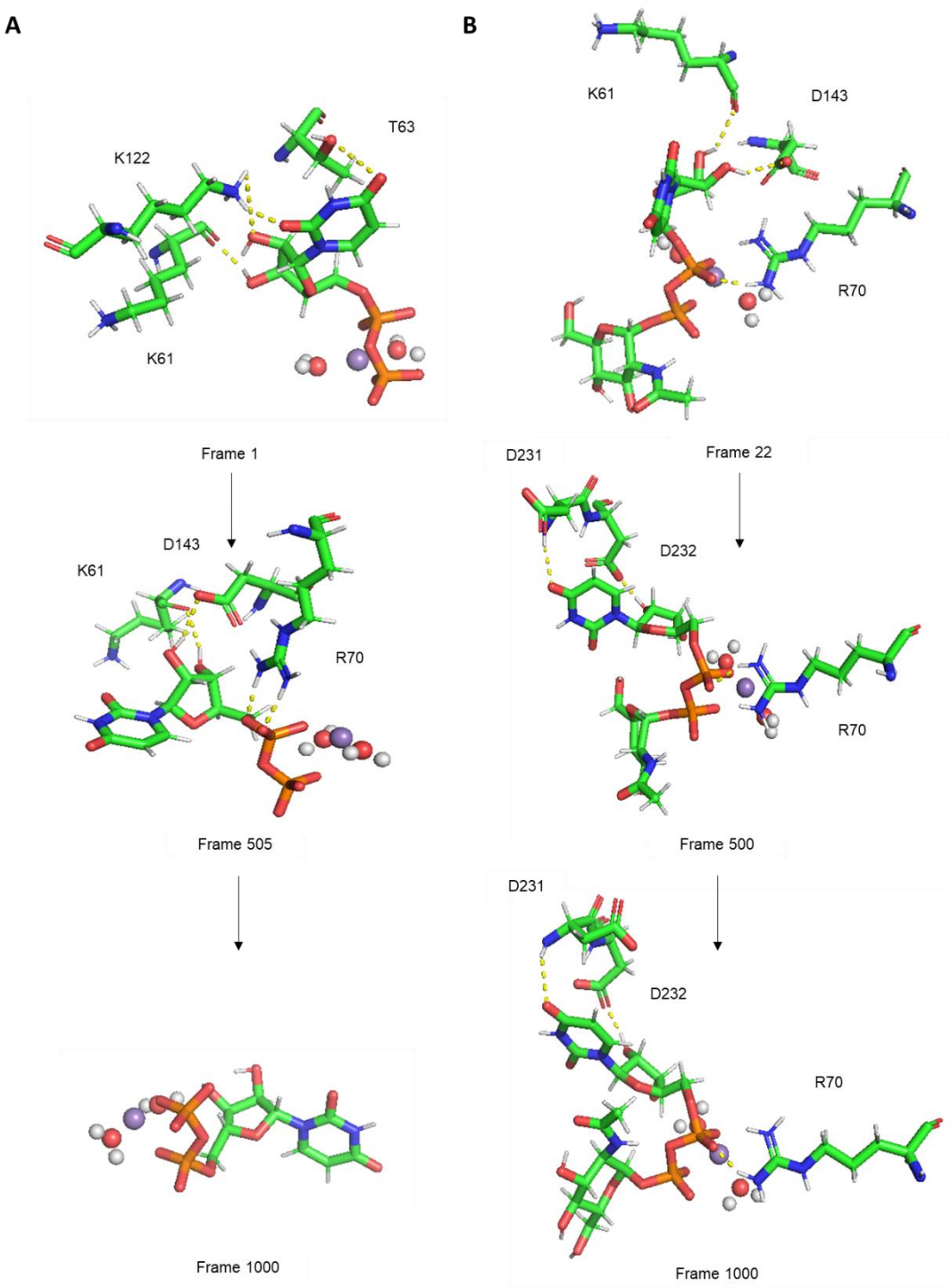


Figure 15. (A) Temporal evolution of the interactions formed between the UDP and the protein. In the last frame, no interactions were present, the UDP was ready to be released from the active site. (B) Temporal evolution of the interactions formed between the UDP-GlcNAc and the protein. The number of hydrogen bonds was almost constant during the dynamics.

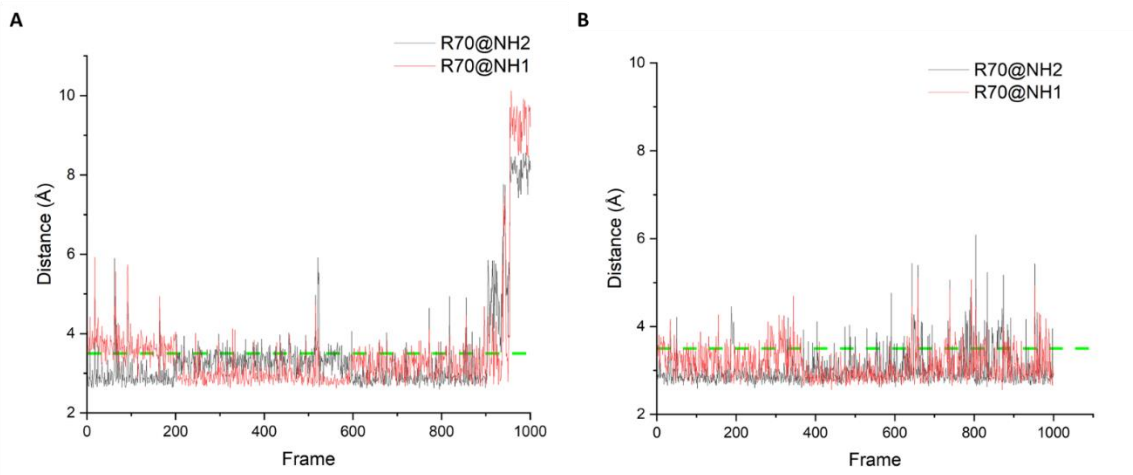


Figure 16. (A) Distance between α phosphate oxygen atom of UDP, and NH2 (black line) and NH1 (red line) atoms of R70. The green-dashed line indicates a cutoff distance of 3.5 Å, above which we have considered that there was no hydrogen bond between involved atoms. (B) Distance between α phosphate oxygen atom of UDP-GlcNAc, and NH2 (black) and NH1 (red) atoms of R70. The same cutoff distance was applied as in panel (A).

Ternary complex simulations

Thanks to the work of Jinek et al. (2006), the location of the acceptor binding site was confirmed through the introduction of point mutations in the pocket of *Drosophila* Fringe, which shares some sequence identity with Mfng (47% sequence identity). These assays confirmed the catalytic function of D232 (D327 in *Drosophila*), but also that many residues that are located near the donor binding site, such as S171 and T196 (S265 and T291 in *Drosophila*), are important for binding the *O*-fucose moiety of the acceptor (Jinek et al., 2006).

Some authors suggest that protein-protein interactions, between Fringe and Notch, are necessary for recognition of the acceptor substrates (Moloney et al., 2000). Thus, we tried to construct a system not only with the *O*-fucose but also with the peptide carrying this monosaccharide to study these interactions and extent our knowledge on how Fringe recognizes Notch EGF motifs to catalyze the addition of GlcNAc to these motifs.

With this data, an *O*-fucose monosaccharide bound to a methyl group, simulating the linkage with a threonine residue, was docked to the Mfng-UDP-GlcNAc structure (Figure 17). This allowed us to generate a complete ternary structure by superimposition of the 2RR2 PDB structure, which corresponds to the crystal structure of the *O*-fucosylated epidermal growth factor-like repeat 12 of mouse Notch1-receptor (see Methods).

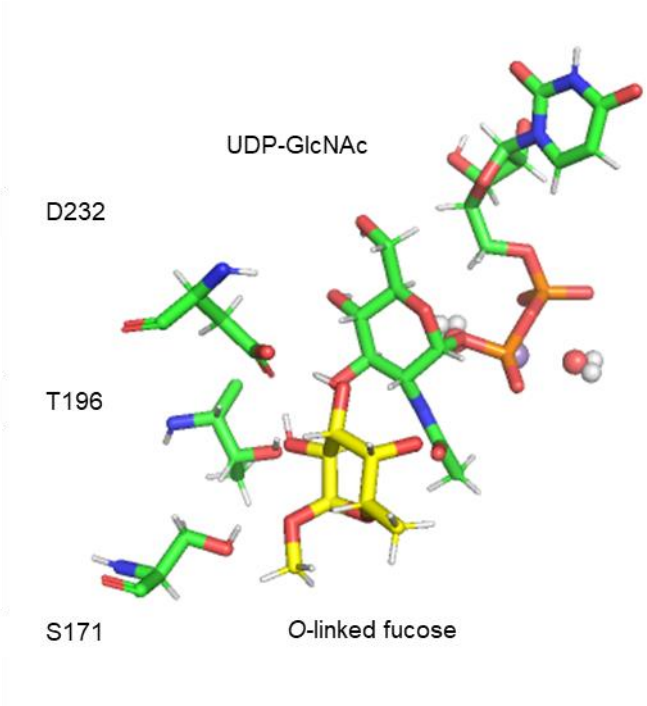


Figure 17. Docked *O*-fucose (yellow sticks) in the suggested acceptor binding pocket of Mfng by Jinek et al. (2006). Residues that are suggested to be important in the acceptor recognition are shown as green sticks.

When analysing the results obtained for this ternary complex simulation, which includes the complete EGF-12 from Notch, some considerations had to be made. Specific restraints were imposed during the production run ensuring that every molecule remained located in its desired position.

To achieve this, the distance between the O3 atom of the fucose and the anomeric carbon of GlcNAc (Figure 18D, in black), the atoms that are going to form the glycosidic bond, was restrained to remain within 3-5 Å during the first 600 ns (Restraint 2). In a similar way, the distance between the O3 atom of the fucose and D232 oxygen atoms (Figure 18D, in red) was also restrained to the same range, ensuring that the hydrogen atom that has to be transferred remained close enough to the catalytic aspartic residue. However, this restraint was only applied during the first 300 ns (Restraint 1).

The stability of the simulation was evaluated as for the previous systems (Figure 18A). Similarly to that observed previously, H256 was substituted in the coordination sphere of the metal by a water molecule. One of the water molecules from the crystal also left the coordination of the Mn^{2+}, being substituted by another water molecule during the simulation. Interestingly, the oxygen atoms from D144 that coordinates with the metal interchange during the simulation (Figure 18B, and Figure 18C). In all moment the metal (Mn^{2+}) kept a octahedral geometry for the coordination sphere (Figure 18B).

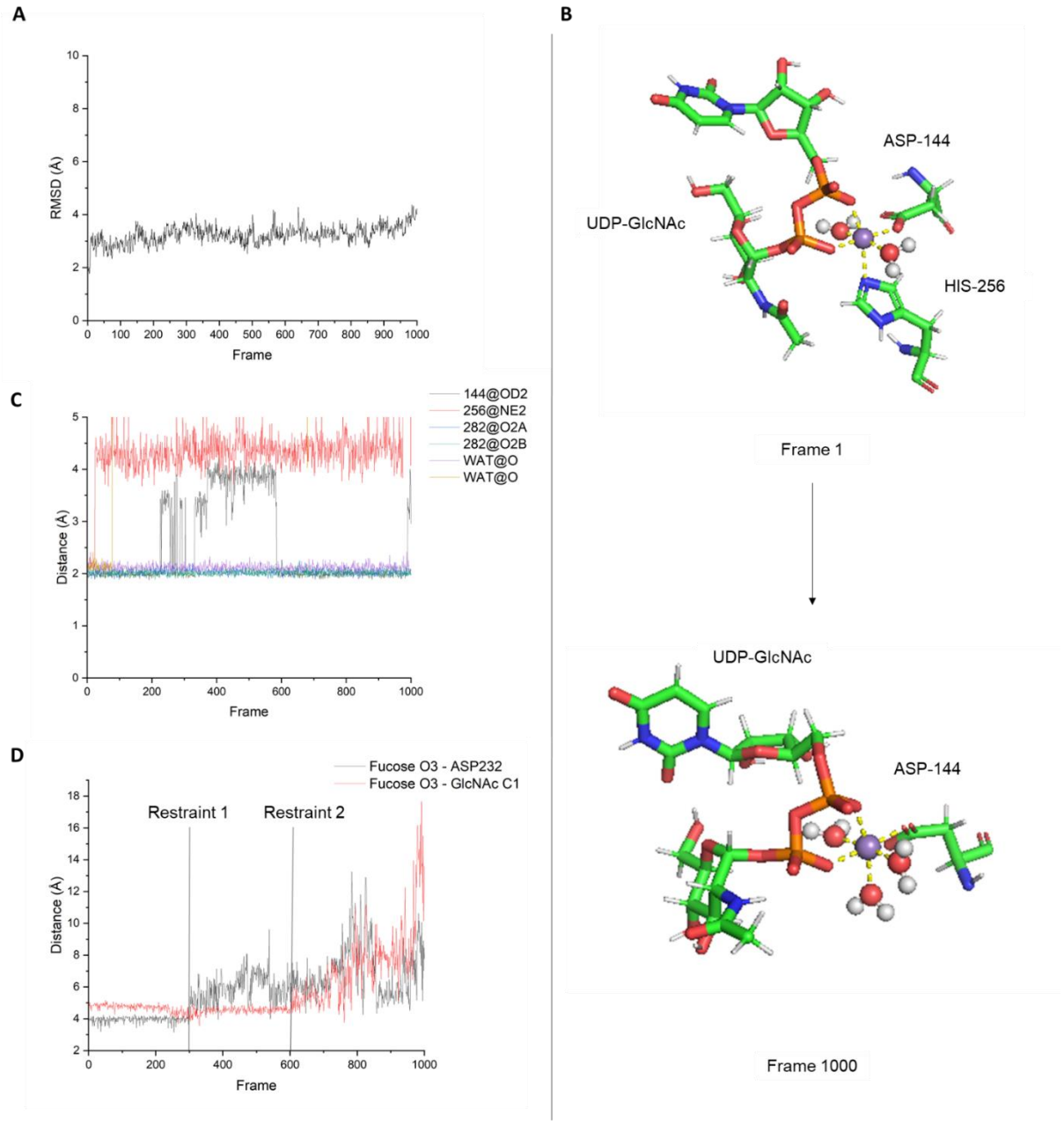


Figure 18. (A) RMSD of the system for all amino acids residues throughout the MD simulation compared to the solvated structure generated by tleap. (B) Comparison of the Mn^{2+} coordination sphere in frame 1 and in frame 1000. (C) Distances of the atoms coordinated to the Mn^{2+} metal, calculated for every frame of the simulation. (D) Distance between fucose O3 atom and D232 OD2 atom (black line) and distance between fucose O3 atom and GlcNAc anomeric carbon (red line). Vertical lines indicate end of restraint.

An interesting structural change was identified in the long loop of the protein (Figure 19, encircled) when aligning a stable frame from the ternary complex simulation (Figure 19, in pink) with a stable frame from the Mfng-Mn^{2+} simulation (Figure 19, in cyan). The RMSD for the alignment was 1.394 Å over 207 Cα atoms. This long loop, that spans from residue 171 to residue 200 (and was not present in the crystal, but modelled with ChimeraX), presented no secondary structure elements in the Mfng-Mn^{2+} complex. However, we have already seen how during the Mfng-UDP-GlcNAc-Mn^{2+} simulation some β-strands appeared in this region (Figure 13, encircled), a structural change not related to the reactivity or activation of the protein, as that system lacked the acceptor substrate. In this case, the presence of an α-helix was observed in the loop 20 ns after the simulation started.

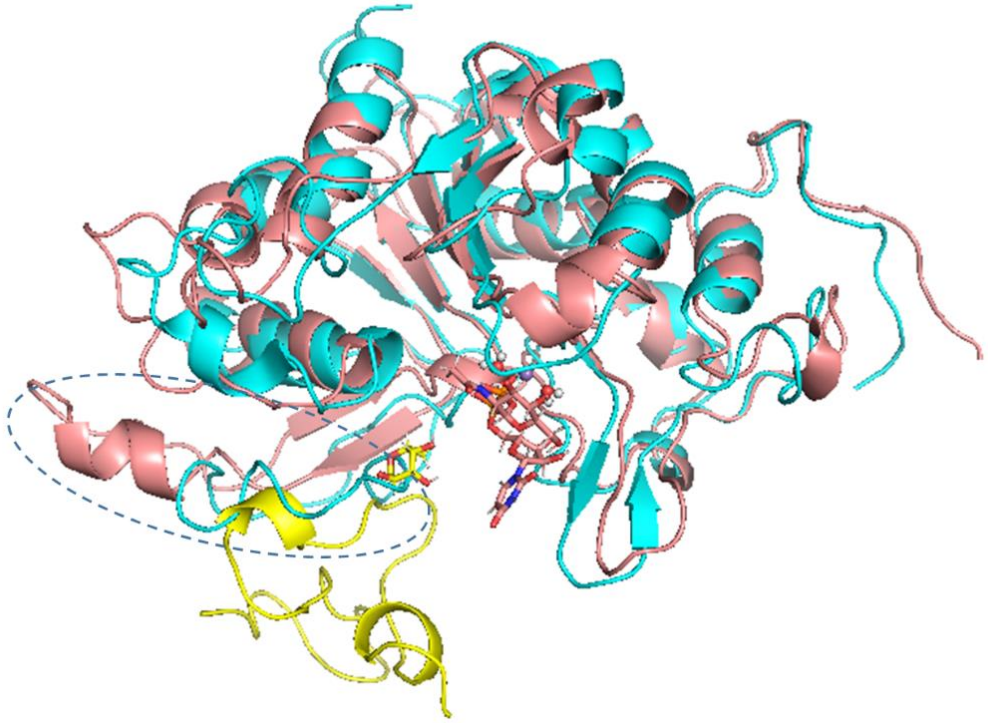


Figure 19. Structural alignment of stable frames (frame 1000 in each case) from the ternary complex, coloured in pink, and Mfng-Mn^{2+}, in cyan**.** Equivalent regions that has adopted different secondary structure are encircled. The EGF-12 bound to the fucose is represented in yellow.

The presence of this α-helix after binding of the acceptor substrate has been described previously in the literature and associated with the activation of GT-A fold proteins for catalysis (Ardèvol & Rovira, 2015). The trend of the involved residues to adopt a helical disposition was evaluated with CPPTRAJ (see Methods), confirming that residues H177 to Q185 had a high tendency to form α-helices (Figure 20).

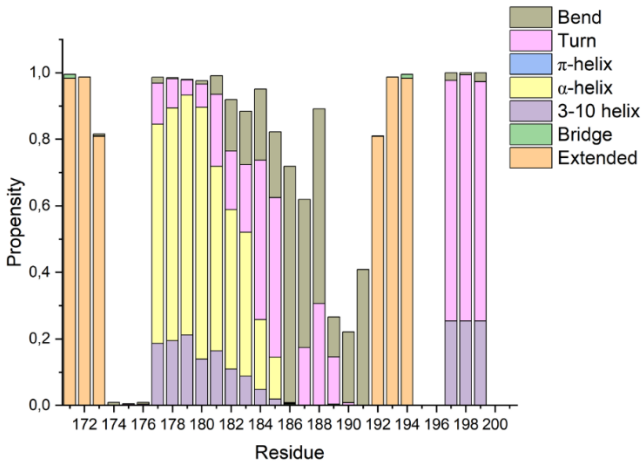


Figure 20. Propensity of residues 171 to 200 of the long N-terminal loop of Mfng to adopt a secondary structure element throughout the simulation.

This statistical analysis of the tendency of each residue to adopt a secondary structure element form was also performed for every other constructed system (Figure 21), to ensure that the presence of this α-helix, which indicates that the protein may be activated, was indeed due to the presence of the acceptor substrate.

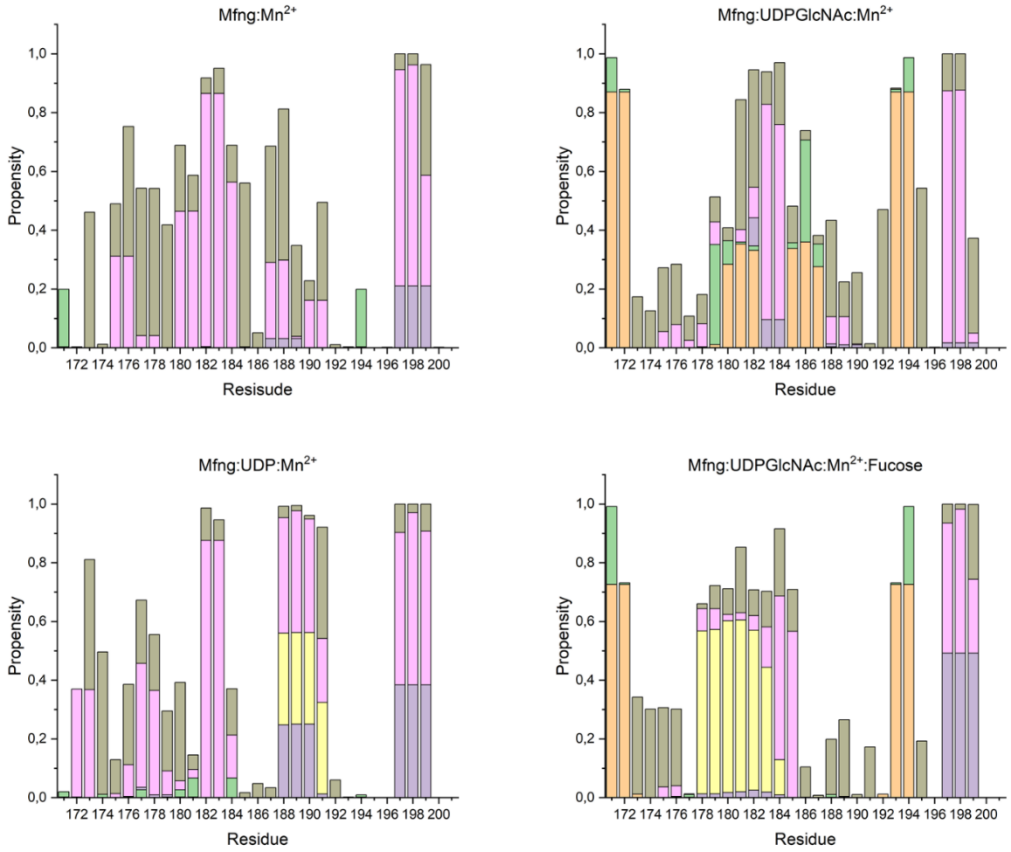


Figure 21. Propensity of residues 171 to 200 of the long N-terminal loop of Mfng to adopt a secondary structure element, calculated for every system, following the same colour fashion as in Figure 20.

This motif did not appear in any other system, except when a simulation of the Mfng-UDP-GlcNAc-Mn^{2+} with the fucose, but not the EGF peptide, was conducted. In this case, residues A178 to K184 showed a high tendency to adopt an α-helix structure (Figure 21, bottom right panel). In the only presence of the fucose, the α-helix appeared later after the simulation started (at 300 ns) when compared to its early appearance in the complete ternary complex simulation (at 20 ns). However, this behaviour could not been reproduced in all the replicas that were run (the formation of the α-helix was induced by the mere presence of the fucose in two of the three produced replicas).

This results suggest that the formation of the helix and, therefore, the activation of Fringe, can be achieved only with the presence of the acceptor monosaccharide, but the recognition of the complete peptide favours the process.

When analysing hydrogen bonding between protein residues (Figure 22A), similar results to other previously analysed systems were obtained. Hydrogen bonds formed between the protein and the UDP-GlcNAc were also analysed (Figure 22B).

Up to frame 600, when the last restraint was removed, the number of hydrogen bonds that both Mfng and EGF12 formed with the cofactor stabilized at 4-5 (Figure 22B). From that point onward, the GlcNAc motif began to move away from the catalytic base and the fucose, and the number of interactions with the cofactor decreased (Figure 22B, and Figure 23). This shift is understandable, considering that the presence of the acceptor, the donor, and the catalytic base represents the intermediate previous state to the reaction – a process that cannot be fully understood by performing MD simulation.

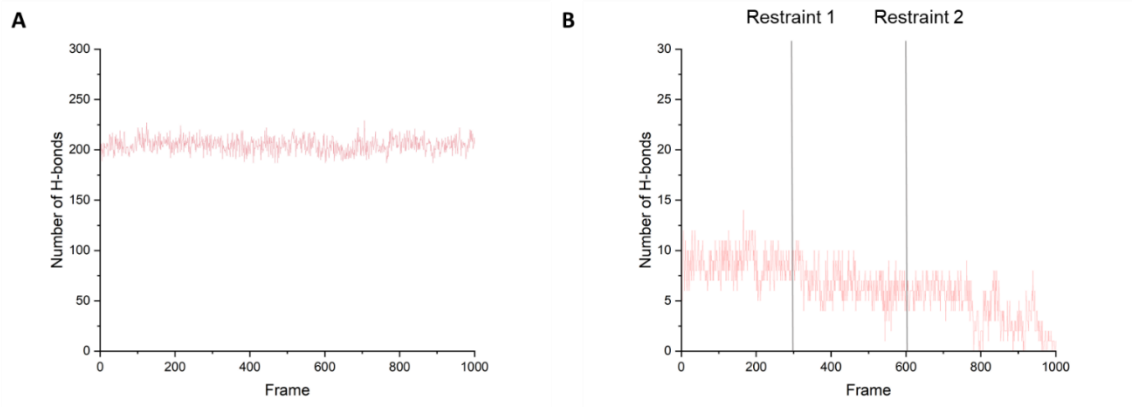


Figure 22. (A) Representation of the number of hydrogen bonds between protein residues per frame. (B) Representation of the number of hydrogen bonds between UDP-GlcNAc-Mfng and UDP-GlcNAc-EGF12 per frame. Vertical lines indicate end of a specific restraint.

In contrast to the Mfng-UDP-GlcNAc-Mn^{2+} simulation, where no interactions between the GlcNAc motif of the cofactor and the protein were observed (Figure 14B), in the ternary complex the main interactions of UDP-GlcNAc and Mfng were formed between residues T196, D232 and D142 and the oxygen atoms of the glucosamine ring of GlcNAc (Figure 23). UDP-GlcNac also interacts with the EGF12 peptide, with the primary interactions formed occurring between the *O*-linked fucose and the acetyl oxygen of the GlcNAc moiety (Figure 23).

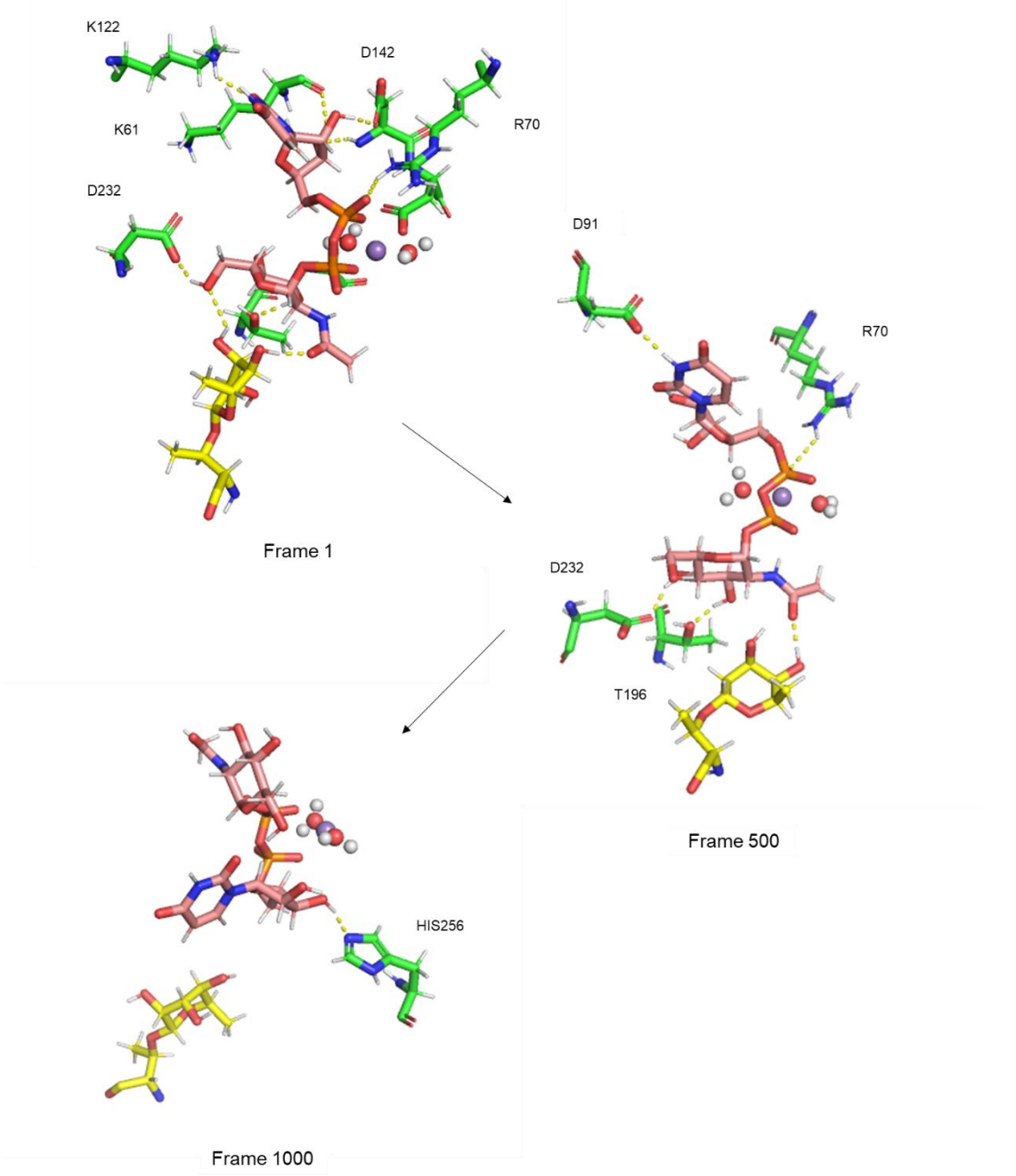


Figure 23. Temporal evolution of the interactions formed between UDP-GlcNAc (pink sticks) and Mfng (green sticks) and EGF12, for which only the *O*-linked fucose is represented (yellow sticks).

Table 4 lists hydrogen bonds between Mfng and UDP-GlcNAc that were present in more than 200 frames (20% of the frames).

Acceptor	Donor	Number of frames	Average Distance (Å)	Average consecutive frames
T196@OG1	UDP-GlcNAc@O3’	577	2.82	21
D232@OD2	UDP-GlcNAc@O4’	362	2.67	12
D142@OD2	UDP-GlcNAc@O4’	341	2.72	23
D232@OD1	UDP-GlcNAc@O6’	285	2.74	3
UDP-GlcNAc@O2A	R70@NH1	282	3.01	4
D91@ OD2	UDP-GlcNAc@N3	253	2.94	3
UDP-GlcNAc@O4’	G198@N	246	3.17	2
D91@OD1	UDP-GlcNAc@N3	210	2.96	3
UDP-GlcNAc@O1A	R70@NH2	200	2.89	10

Table 4. Hydrogen bonds acceptor and donor atoms between Mfng and UDP-GlcNAc. Each atom is specified by the one-letter code for the residue name (e.g., S for Serine), the residue position in the sequence (e.g., 171), and the atom type after the "@" symbol (e.g., O for oxygen). UDP-GlcNAc atoms types are assigned based on Figure S2 in the Supplementary Information.

Table 5 lists hydrogen bonds between EGF12 and UDP-GlcNAc that were present in more than 50 frames (5% of the frames). For the sake of simplicity we numbered the amino acid residues that form part of the EGF12 from 1 to 38, despite being numbered differently in PyMOL (as in the *.pdb* structure the EGF12 is located after the Mfng chain).

Acceptor	Donor	Number of frames	Average Distance (Å)	Average consecutive frames
UDP-GlcNAc@O7’	0FA@O4	631	2.72	7
UDP-GlcNAc@O7’	0FA@O3	291	2.75	4
UDP-GlcNAc@O1B	N3@ND2	54	2.85	3
UDP-GlcNAc@O3’	0FA@O3	50	2.93	1

Table 5. Hydrogen bonds acceptor and donor atoms between EGF12 (in yellow) and UDP-GlcNAc. Each atom is specified by the one-letter code for the residue name (e.g., S for Serine), the residue position in the sequence (e.g., 171), and the atom type after the "@" symbol (e.g., O for oxygen). 0FA refers to the *O*-linked fucose in the EGF12 structure, which atoms types are assigned based on Figure S3 in the Supplementary Information. UDP-GlcNAc atoms types are assigned based on Figure S2 in the Supplementary Information.

Interactions and hydrogen bonds between Mfng and the EGF12 were also analysed. Using *nativecontacts* command from CPPTRAJ, with a cut-off distance of 4 Å, 6 contacts were identified between both protein chains in the first frame of the simulation. However these contacts were not preserved throughout the simulation (Table 6).

Contact Pair (Atom1 - Atom2)	Number of frames	Average Distance (Å)
S171@OG - M28@SD	1	3.68
L172@N - M28@SD	19	3.63
L172@O - M28@CE	2	3.28
P175@CG - I26@CG1	2	3.78
P175@CD - I26@CG1	2	3.84
P175@CD - I26@CG2	1	3.86

Table 6. Contacts between Mfng (Atom1) and EGF12 (Atom2) in the first frame of the simulation. Each atom is specified by the one-letter code for the residue name (e.g., S for Serine), the residue position in the sequence (e.g., 171), and the atom type after the "@" symbol (e.g., O for oxygen).

The number of hydrogen bonds between Mfng and EGF12 stabilized around 12-13, even when restraints between atoms were removed in the production cycles (Figure 24).

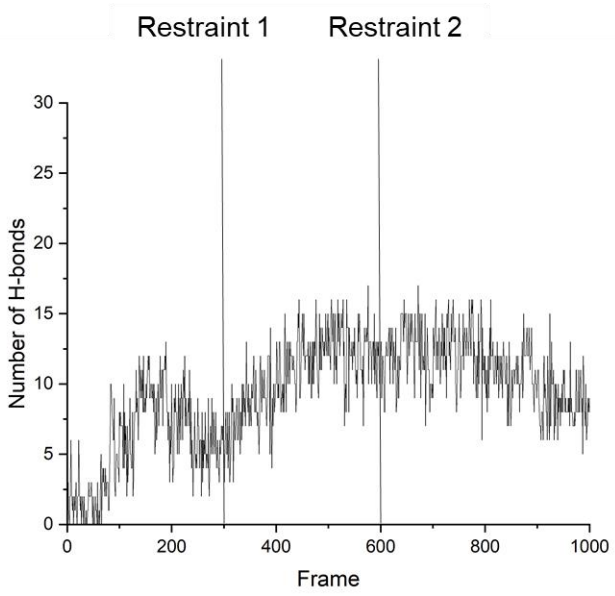


Figure 24. Number of hydrogen bonds per frame between Mfng and EGF12 amino acids chains, not taking into account the hydrogen bonds formed between Mfng and the *O*-linked fucose. Vertical lines indicate end of a specific restraint.

Table 7 list hydrogen bonds between Mnfg and EGF12 peptide chains that appeared in more than 200 frames during the simulation (20% of the frames).

Acceptor	Donor	Number of frames	Average Distance (Å)	Average consecutive frames
D13@O	L172@N	780	3.02	8
N173@OD1	Q19@NE2	717	2.89	6
L172@O	T15@N	662	3.05	4
R174@O	Q19@NE2	636	2.88	7
D13@O	S171@OG	439	2.71	3
I38@OXT	R174@NH1	416	2.83	12
I38@O	R174@NH2	409	2.92	4
E37@OE1	R174@NH2	406	2.92	5
E37@OE2	R174@NH2	399	2.94	4
E37@OE1	R174@NH1	313	2.82	11
E37@OE2	R174@NH1	311	2.82	8
I38@OXT	R174@NH2	288	2.90	3

Table 7. Hydrogen bonds acceptor and donor atoms between Mfng and EGF12. To difference each protein chain and avoid confusions with residue numeration, Mfng atoms are listed in black, while EGF12 atoms are listed in yellow. Each atom is specified by the one-letter code for the residue name (e.g., S for Serine), the residue position in the sequence (e.g., 171), and the atom type after the "@" symbol (e.g., O for oxygen).

Many different hydrogen bonds were identified between Mfng and EGF12 before restraints were removed (Figure 25A). Most of these hydrogen bonds implied residues located on the long N-terminal loop of Mfng (Figure 25, in green). Apart from this loop, R228, located in a α helix near the catalytic base D232, interacts with residues E22, D1 and D18 from EGF12 (Figure 25, in yellow).

Interactions between L171 and EGF12 T15 and D13 atoms were preserved even when restraints were removed in the simulation (Figure 25B), which indicate that these interactions are important in the recognition and stabilization of the peptide.

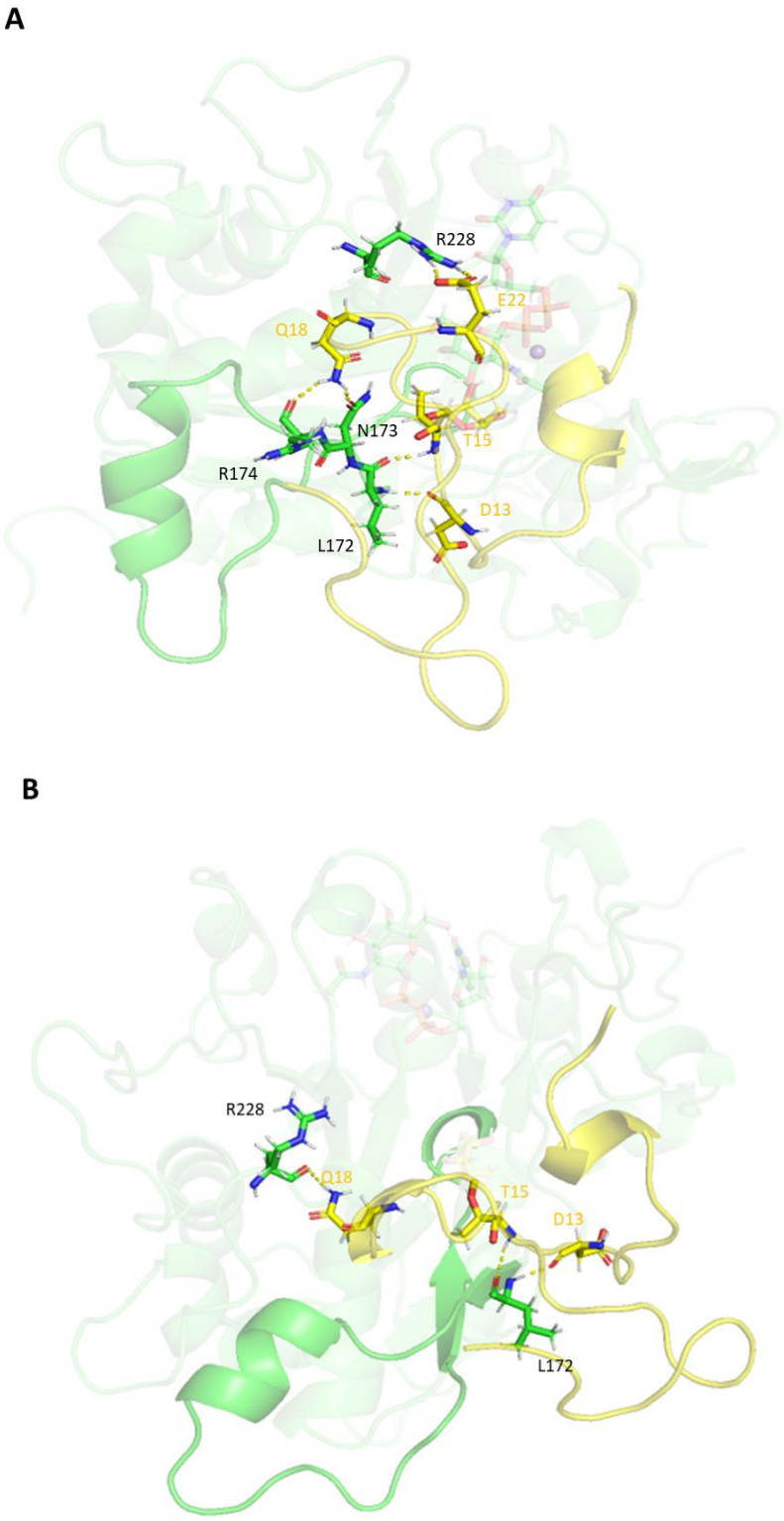


Figure 25. (A) Hydrogen bonds between Mfng N-terminal loop amino acids (in green) and EGF12 peptide chain (in yellow) in frame 1. (B) Hydrogen bonds between Mfng N-terminal loop amino acids (in green) and EGF12 peptide chain (in yellow) in frame 1000.

Finally, interactions between Mfng and the *O*-linked fucose were also investigated, comparing the results obtained from the simulation with the complete EGF12 and the simulation containing only the acceptor monosaccharide.

Tables 8 and 9 list hydrogen bonds between Mfng and the *O*-linked fucose from EGF12 in both simulations (with the complete EGF12 and with the acceptor monosaccharide only, respectively) that were present in more than 50 frames (5% of the frames).

Acceptor	Donor	Number of frames	Average Distance (Å)	Average consecutive frames
S170@OG	0FA@O2	306	2.86	5
0FA@O2	S170@OG	50	2.79	2

Table 8. Hydrogen bonds acceptor and donor atoms between Mfng and *O*-linked fucose from EGF12. Each atom is specified by the one-letter code for the residue name (e.g., S for Serine), the residue position in the sequence (e.g., 171), and the atom type after the "@" symbol (e.g., O for oxygen). 0FA refers to the *O*-linked fucose in the EGF12 structure, which atoms types are assigned based on Figure S3 in the Supplementary Information.

Acceptor	Donor	Number of frames	Average Distance (Å)	Max life time
D232@OD2	0FA@O3	423	2.7	5
0FA@O2	T196@OG1	186	2.83	5

Table 9. Hydrogen bonds acceptor and donor atoms between Mfng and *O*-linked fucose. Each atom is specified by the one-letter code for the residue name (e.g., S for Serine), the residue position in the sequence (e.g., 171), and the atom type after the "@" symbol (e.g., O for oxygen). 0FA refers to the *O*-linked fucose, which atoms types are assigned based on Figure S3 in the Supplementary Information.

In the simulation with the complete EGF12 substrate (Table 8), the main interaction between Mnfg and the fucose was formed with S170, an amino-acid that was predicted to form part of the acceptor binding pocket by Jinek et al. (2006). However, this interaction was found only in 1 frame in the simulation that only presented the fucose (Table 9).

The main interaction between Mfng and the fucose in this later simulation was formed with D232, the catalytic base, but we have to take into account that this distance was restrained for the first 300 ns. Apart from D232, a hydrogen bond between the fucose and T196 (another acceptor binding pocket residue identified by Jinek et al.) was presented for 186 frames.

When simulating the reaction product, Mn^{2+} coordination sphere was maintained for 600 ns. However, after this time, H256 left the coordination of the metal (Figure 26B, and 26C), being substituted by a water molecule that entered the active site, similarly to what occurred in the Mfng-UDP-Mn^{2+} simulation, where H256 was also maintained for 600 ns (in that case after 300 ns of restraints). Notably, the octahedral geometry was conserved despite this changes (Figure 26B). The stability of the dynamics was evaluated as for the previous systems (Figure 26A).

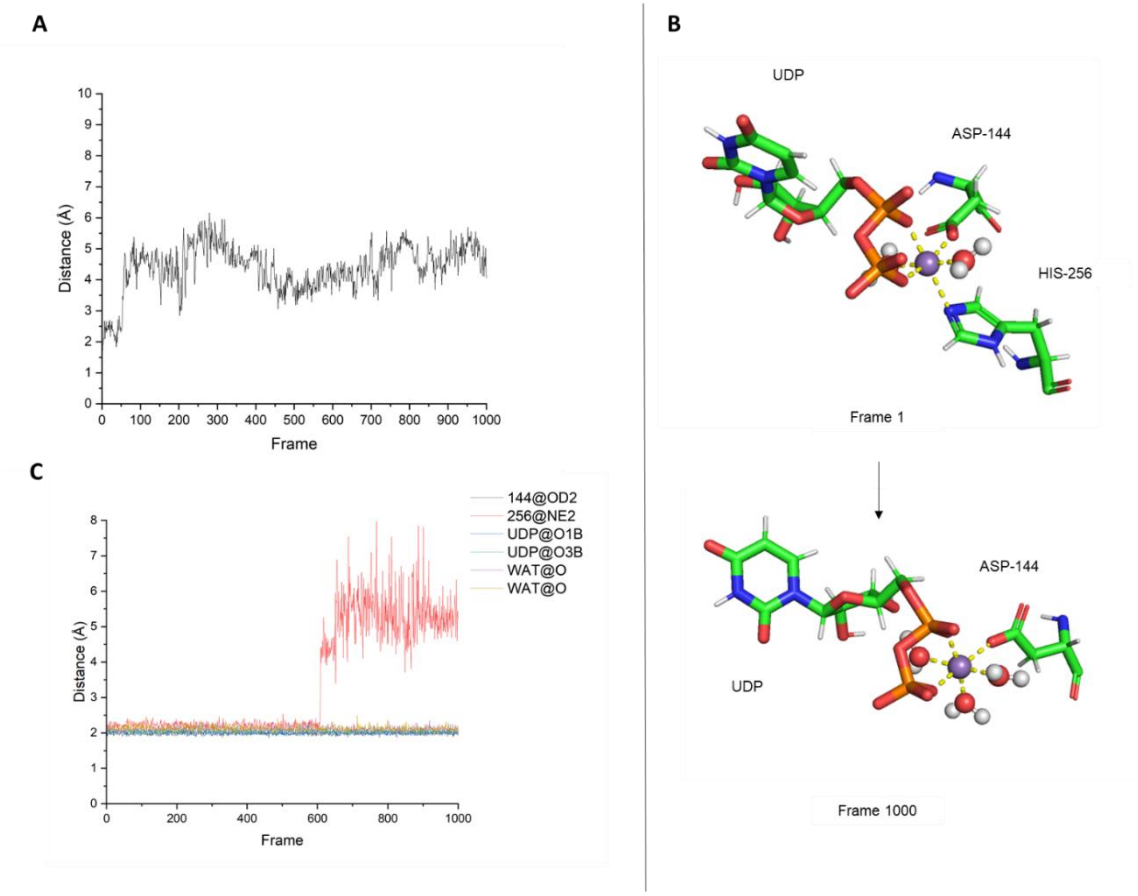


Figure 26. (A) RMSD of the system for all amino acids residues throughout the MD simulation compared to the solvated structure generated by tleap. (B) Comparison of the Mn^{2+} coordination sphere in frame 1 and in frame 1000. (C) Distances of the atoms coordinated to the Mn^{2+} metal, calculated for every frame of the MD simulation. H256 leaves the coordination site during the dynamics (red line), but another water molecule starts coordinating the metal.

Apart from the long N-terminal loop, no significant structural differences were observed when aligning a stable frame from this system (Figure 27B, in green) with a stable frame from the Mfng-Mn^{2+} simulation (Figure 27B, in cyan). The RMSD for the alignment was 1.208 Å over 208 Cα atoms. This result was not surprising since both systems represent states of the protein that are close in time, as after the reaction product is released, the UDP molecule has also to be freed from the active site, leading to the Mfng-Mn^{2+} complex.

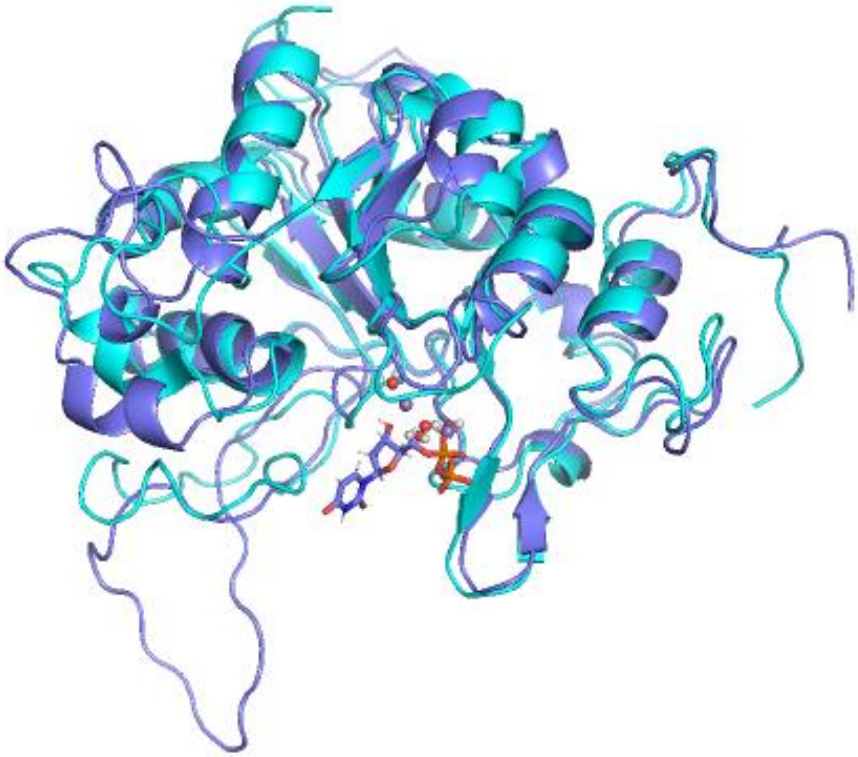


Figure 27. (A) Structural alignment of stable frames (frame 1000 in each case) from the reaction product, coloured in dark blue, and Mfng-Mn^{2+}, in cyan**.**

The initial structure for this system presented the α-helix in the N-terminal loop of Mfng (Figure 28, in green), as for the activated protein (Ardèvol & Rovira, 2015). That secondary structure element was lost during the simulation (Figure 28 encircled), at the same time that glycosylated EGF12 moves away from the acceptor binding pocket. Thus, once the acceptor has been glycosylated and the product is released, the protein adopts the inactive form which causes a low affinity for UDP. Under this situation, UDP-GlcNAc can displace UDP and restart the catalytic cycle.

These obtained results provide a deeper comprehension of Fringe's catalytic activity, from the binding of UDP-GlcNAc to the donor binding pocket, to the release of UDP after the reaction has occurred, that are the pre- and post-reaction steps to the chemical reaction. How the bonds are broken and formed, i.e., the reaction mechanism itself, is still undeciphered, and has to be studied in the future using QM/MM methods and/or metadynamics.

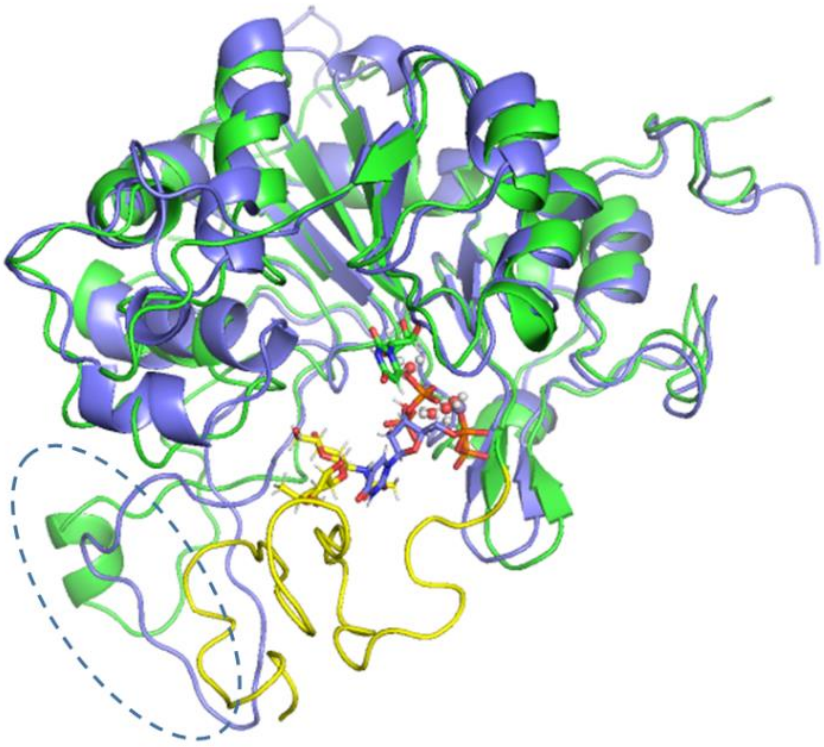


Figure 28. Structural superimposition of frame 1 (green) and frame 1000 (dark blue) of the product simulation. Equivalent regions that has adopted different secondary structure are encircled. The EGF-12 bound to the fucose is represented in yellow, only for frame 1, as it moved far away in frame 1000. UDP for each frame is coloured in the same fashion as the rest of the protein, green for frame 1 and dark blue for frame 1000.

Conclusions

These interactions of the different forms of Manic Fringe glycosyltransferase with its substrates (acceptor and donor) studied through molecular dynamics simulations, showed different dispositions once the reactions has been produced to facilitate the releasing of the products: glycosylated peptide and UDP.

The coordination sphere of the metal atom (Mn^{2+}) is only consistently preserved in the Mfng-UDP-GlcNAc-Mn^{2+} simulation, which may suggest that the presence of the donor sugar may contribute to the stability of the coordination sphere. Indeed, a common observation in other simulations was the departure of H256, which coordinates the metal in the crystal structure obtained by Jinek et al., from the coordination sphere, being always substituted by a water molecule from the dynamics. On the other hand, in all cases the octahedral geometry was maintained, confirming that the Mn^{2+} ion in Fringe remains hexacoordinated throughout the catalytic cycle.

The hydrogen bonds between the protein chain and the UDP and UDP-GlcNAc are limited. When no acceptor substrate is present, main interactions with Mfng, for both UDP and UDP-GlcNAc, are formed with R70. When UDP is going to be released from the active centre, after the reaction has occurred and a new molecule of UDP-GlcNAc is required for catalyzing the reaction, interactions with this residue are broken. In the presence of the acceptor substrate, UDP-GlcNAc also forms important interactions with the peptide chain from EGF12 and the *O*-linked fucose, facilitated by its GlcNAc moiety.

Significant structural changes in Fringe enzymes upon substrate binding have been demonstrated through MD simulations of ternary complexes. These changes are similar to the ones reported for others GT-A fold glycosyltransferases. The appearance of an α-helix in the long N-terminal loop of Fringe, when both UDP-GlcNAc (the donor) and the *O*-fucose (the acceptor) are present, indicates a possible activation-related change. As shown in the results, the presence of the fucose is sufficient to induce this secondary structure change in the loop, but the complete substrate, i.e. the fucosylated EFG motif from Notch receptors, facilitates the process. This change in secondary structure is lost when the reaction product is formed, inducing the inactivation of Fringe, which requires the product to leave the active site and the UDP cofactor to be replaced for a new UDP-GlcNAc molecule.

The work of Jinek et al. (2006), predicting the residues that form part of the acceptor binding site in mouse homolog Mfng using *Drosophila*'s Fringe has been confirmed, as fucose main interactions with Mfng catalytic fragment with residues S170 and T196 has also been demonstrated.

It is believed that Fringe catalyzes the reaction via an S_N2 mechanism, due to the presence of the catalytic base D232 near the acceptor and donor binding pockets, but

the reaction mechanism has yet to be proven. This task will be accomplished in future works by performing QM/MM studies of the reaction centre.

Bibliography

Albesa-Jové, D., Giganti, D., Jackson, M., Alzari, P. M., & Guerin, M. E. (2014). Structure-function relationships of membrane-associated GT-B glycosyltransferases. *Glycobiology*, *24*(2), 108-124. https://doi.org/10.1093/glycob/cwt101

Ardèvol, A., & Rovira, C. (2015). Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from ab Initio Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *Journal of the American Chemical Society*, *137*(24), 7528-7547. https://doi.org/10.1021/jacs.5b01156

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, *81*(8), 3684-3690. https://doi.org/10.1063/1.448118

Case, D. A., Aktulga, H. M., Belfon, K., Cerutti, D. S., Cisneros, G. A., Cruzeiro, V. W. D.,…Merz, K. M., Jr. (2023). AmberTools. *Journal of Chemical Information and Modeling*, *63*(20), 6183-6191. https://doi.org/10.1021/acs.jcim.3c01153

Case, D. A., Cheatham, T. E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr.,…Woods, R. J. (2005). The Amber biomolecular simulation programs. *J Comput Chem*, *26*(16), 1668-1688. https://doi.org/10.1002/jcc.20290

Forrester, T. J. B., Ovchinnikova, O. G., Li, Z., Kitova, E. N., Nothof, J. T., Koizumi, A.,…Kimber, M. S. (2022). The retaining β-Kdo glycosyltransferase WbbB uses a double-displacement mechanism with an intermediate adduct rearrangement step. *Nat Commun*, *13*(1), 6277. https://doi.org/10.1038/s41467-022-33988-1

Gloster, T. M. (2014). Advances in understanding glycosyltransferases from a structural perspective. *Curr Opin Struct Biol*, *28*, 131-141. https://doi.org/10.1016/j.sbi.2014.08.012

He, X., Man, V. H., Yang, W., Lee, T.-S., & Wang, J. (2020). A fast and high-quality charge model for the next generation general AMBER force field. *The Journal of Chemical Physics*, *153*(11), 114502. https://doi.org/10.1063/5.0019056

Jinek, M., Chen, Y.-W., Clausen, H., Cohen, S. M., & Conti, E. (2006). Structural insights into the Notch-modifying glycosyltransferase Fringe. *Nature Structural & Molecular Biology*, *13*(10), 945-946. https://doi.org/10.1038/nsmb1144

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, *79*(2), 926-935. https://doi.org/10.1063/1.445869

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, *22*(12), 2577-2637. https://doi.org/https://doi.org/10.1002/bip.360221211

Kakuda, S., & Haltiwanger, R. S. (2017). Deciphering the Fringe-Mediated Notch Code: Identification of Activating and Inhibiting Sites Allowing Discrimination between Ligands. *Dev Cell*, *40*(2), 193-201. https://doi.org/10.1016/j.devcel.2016.12.013

Kattke, M. D., Gosschalk, J. E., Martinez, O. E., Kumar, G., Gale, R. T., Cascio, D.,…Clubb, R. T. (2019). Structure and mechanism of TagA, a novel membrane-associated glycosyltransferase that produces wall teichoic acids in pathogenic bacteria. *PLoS Pathog*, *15*(4), e1007723. https://doi.org/10.1371/journal.ppat.1007723

Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., & Woods, R. J. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem*, *29*(4), 622-655. https://doi.org/10.1002/jcc.20820

Kräutler, V., van Gunsteren, W. F., & Hünenberger, P. H. (2001). A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *Journal of Computational Chemistry*, *22*(5), 501-508. https://doi.org/https://doi.org/10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V

Lairson, L. L., Henrissat, B., Davies, G. J., & Withers, S. G. (2008). Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem*, *77*, 521-555. https://doi.org/10.1146/annurev.biochem.76.061005.092322

Li, Z., Han, K., Pak, J. E., Satkunarajah, M., Zhou, D., & Rini, J. M. (2017). Recognition of EGF-like domains by the Notch-modifying O-fucosyltransferase POFUT1. *Nature Chemical Biology*, *13*(7), 757-763. https://doi.org/10.1038/nchembio.2381

Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., & Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, *11*(8), 3696-3713. https://doi.org/10.1021/acs.jctc.5b00255

Mestrom, L., Przypis, M., Kowalczykiewicz, D., Pollender, A., Kumpf, A., Marsden, S. R.,…Hagedoorn, P. L. (2019). Leloir Glycosyltransferases in Applied Biocatalysis: A Multidisciplinary Approach. *Int J Mol Sci*, *20*(21). https://doi.org/10.3390/ijms20215263

Moloney, D. J., Panin, V. M., Johnston, S. H., Chen, J., Shao, L., Wilson, R.,…Vogt, T. F. (2000). Fringe is a glycosyltransferase that modifies Notch. *Nature*, *406*(6794), 369-375. https://doi.org/10.1038/35019000

Moremen, K. W., & Haltiwanger, R. S. (2019). Emerging structural insights into glycosyltransferase-mediated synthesis of glycans. *Nature Chemical Biology*, *15*(9), 853-864. https://doi.org/10.1038/s41589-019-0350-2

Paquet, E., & Viktor, H. L. (2015). Molecular dynamics, monte carlo simulations, and langevin dynamics: a computational review. *Biomed Res Int*, *2015*, 183918. https://doi.org/10.1155/2015/183918

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I.,…Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci*, *30*(1), 70-82. https://doi.org/10.1002/pro.3943

Piniello, B., Lira-Navarrete, E., Takeuchi, H., Takeuchi, M., Haltiwanger, R., Hurtado-Guerrero, R., & Rovira, C. (2021). Asparagine Tautomerization in Glycosyltransferase Catalysis. The Molecular Mechanism of Protein O -Fucosyltransferase 1. *ACS Catalysis*, *11*, 9926-9932. https://doi.org/10.1021/acscatal.1c01785

Rao, D., Zhu, L., Liu, W., & Guo, Z. (2024). Molecular Mechanism of Double-Displacement Retaining β-Kdo Glycosyltransferase WbbB. *The Journal of Physical Chemistry B*, *128*(31), 7476-7485. https://doi.org/10.1021/acs.jpcb.4c02073

Rini, J. M., Moremen, K. W., Davis, B. G., & et al. (2022). Glycosyltransferases and Glycan-Processing Enzymes. In A. Varki, R. D. Cummings, & J. D. Esko (Eds.), *Essentials of Glycobiology* (4th ed.). Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/glycobiology.4e.6

Roe, D. R., & Cheatham, T. E., III. (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation*, *9*(7), 3084-3095. https://doi.org/10.1021/ct400341p

Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S., & Walker, R. C. (2013). Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation*, *9*(9), 3878-3888. https://doi.org/10.1021/ct400314y

Schrödinger, L., & DeLano, W. (2020). PyMOL. In. http://www.pymol.org/pymol.

Wang, H., Zang, C., Liu, X. S., & Aster, J. C. (2015). The role of Notch receptors in transcriptional regulation. *J Cell Physiol*, *230*(5), 982-988. https://doi.org/10.1002/jcp.24872

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N.,…Richardson, D. C. (2018). MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci*, *27*(1), 293-315. https://doi.org/10.1002/pro.3330

Zhang, H., Zhu, F., Yang, T., Ding, L., Zhou, M., Li, J.,…Wu, H. (2014). The highly conserved domain of unknown function 1792 has a distinct glycosyltransferase fold. *Nat Commun*, *5*, 4339. https://doi.org/10.1038/ncomms5339

Zhou, B., Lin, W., Long, Y., Yang, Y., Zhang, H., Wu, K., & Chu, Q. (2022). Notch signaling pathway: architecture, disease, and therapeutics. *Signal Transduct Target Ther*, *7*(1), 95. https://doi.org/10.1038/s41392-022-00934-y

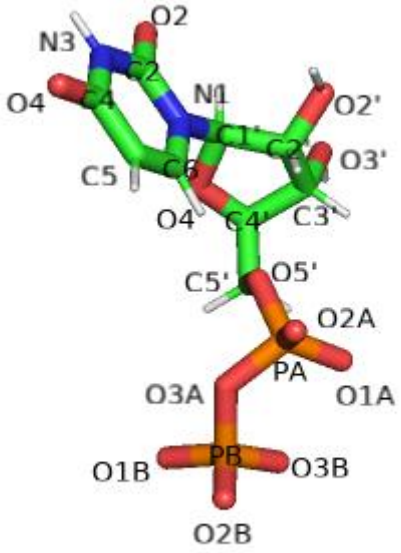


Figure S1. PyMOL Atom types for UDP molecule.

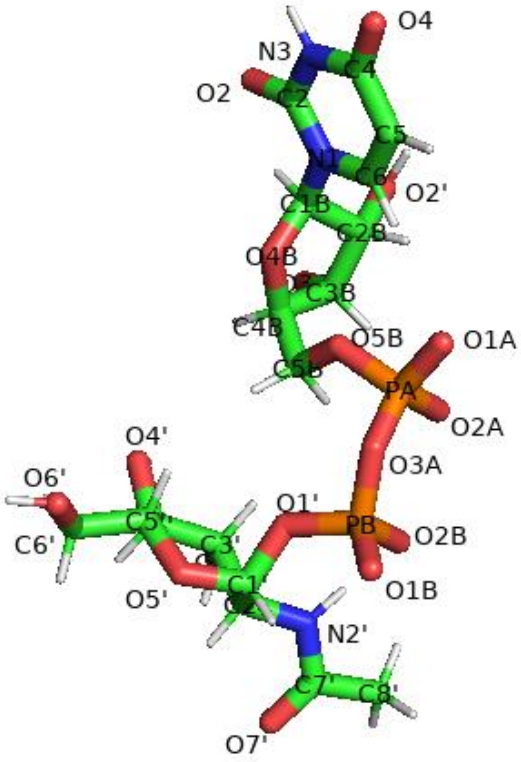


Figure S2. PyMOL atom types for UDP-GlcNAc molecule.

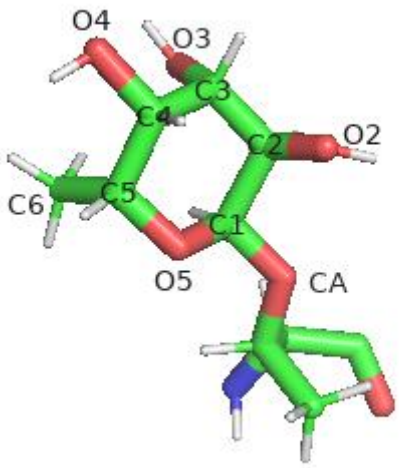


Figure S3. PyMOL atom types for fucose. Fucose is represented bound to T15 residue of EGF12.