Computational characterization of the xanthan gum glycosyltransferase GumK

Davide Luciano,[†] F. Emil Thomasen ,[‡] Kresten Lindorff-Larsen ,[‡] and Gaston Courtade[*,†]

†*Department of Biotechnology and Food Science, NTNU Norwegian University of Science and Technology, Trondheim, Norway*

‡*Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen,DK-2200 Copenhagen N, Denmark*

E-mail: gaston.courtade@ntnu.no

Abstract

The activity of GT-B glycosyltransferases depends on their conformational flexibility and high substrate specificity, but the molecular basis of these features is still not well defined. The GT70 family contains a single well-characterized enzyme, GumK, a glucuronosyltransferase from *Xanthomonas campestris* required for xanthan gum biosynthesis. Here, we applied multiscale molecular simulations and sequence analysis to probe GumK dynamics and substrate specificity. We show that GumK undergoes twisting and bending motions constrained by interdomain contacts and modulated by membrane anchoring. Acceptor-substrate binding within an amphiphilic clamp promotes opening, while donor-substrate binding stabilizes closure, defining a substrate-dependent catalytic cycle. Specificity for glucuronate-UDP is mediated by a conserved electrostatic environment centered on Lys307 and a hydrophobic triad that orients the sugar moiety. On the acceptor side, the binding site selectively accommodates polyisoprenyl carriers up to three isoprene units in length and wraps around the substrate,

constraining the trisaccharide moiety in a catalytically competent conformation. Comparative analysis highlights GumK-specific motifs distinguishing it from homologous GTs. This work provides mechanistic insight into the GT70 family and the dynamic behavior of GT-B enzymes, establishing principles for rational engineering of GumK to modify the monosaccharide composition of xanthan gum.

Introduction

Glycosyltransferases (GTs) are key enzymes involved in the biosynthesis of many biologically relevant molecules. The expression of a specific set of these proteins can profoundly impact an organism's phenotype, influencing cellular processes such as cell signaling, immune response, and structural integrity. GTs glycosylate nucleic acids, proteins, and small molecules and are essential for the biosynthesis of oligosaccharides and polysaccharides, which play crucial roles in metabolism and cellular interactions.[1] Due to their broad biochemical capabilities, GTs have significant industrial and pharmaceutical potential. They can be harnessed as biocatalysts for the synthesis of novel materials and bioactive compounds or targeted by drugs to interfere with pathogenic biosynthetic pathways.[2–4] Despite their importance, the activity of these enzymes is not yet fully characterized. Therefore, research in this field has been particularly active in recent years, and further studies are needed in the future.[3,5–7]

Glycosyltransferases (GTs) exhibit remarkable structural and functional diversity, spanning multiple fold types, including GT-A, GT-B, GT-C, and lysozyme-like architectures.[8,9] This structural variability is reflected in their catalytic mechanisms, which involve the transfer of a monosaccharide unit from an activated donor substrate, typically a nucleotide sugar, to an acceptor molecule. Depending on the reaction mechanism, GTs can follow three distinct catalytic pathways and can be classified as inverting or retaining enzymes, depending on whether they reverse or preserve the anomeric carbon configuration of the donor substrate.[10] The key scientific challenge is to understand and predict how these proteins select their donor and acceptor substrates, as this has a major impact both on their biological

function and their industrial applications.

Out of the 138 glycosyltransferase families listed in the CAZy database,[11] many remain poorly characterized due to difficulties in protein expression and substrate instability. These challenges make it hard to study their biochemical properties, such as substrate specificity and selectivity, which in turn limits our ability to use them in industrial and biotechnological applications. Among these underexplored families, GT70 has a single well-characterized member, GumK. This enzyme plays a key role in the production of xanthan gum, a widely used polysaccharide produced by *Xanthomonas campestris*.

The enzymatic activity of GumK has been extensively studied, and several X-ray crystal structures are available in the Protein Data Bank (PDB).[12] Interest in this enzyme stems from its role in producing xanthan gum, one of the most versatile and widely used polysaccharides in the materials industry.[13–15] Given its broad applications, chemical and biochemical strategies have been explored to improve its functionality.[16] A deeper understanding of the substrate selectivity of GumK could contribute to this effort by tuning its specificity, allowing the incorporation of different sugar units into the polysaccharide. This could lead to novel xanthan derivatives with tailored mechanical and functional properties.

The crystal structure of GumK (PDB: 2HY7) adopts a typical GT-B fold, with both the N- and C-termini located within the same domain. The protein comprises two Rossmann-like domains, traditionally named as the N-domain (residues 13–201 and 362–380) and the C-domain (residues 210–361), connected by a linker (residues 202–209). The C-domain binds glucuronate-UDP (GlcA-UDP) as the donor substrate, while the N-domain interacts with the membrane-anchored acceptor substrate, α-D-Man-(1$\rightarrow$3)-β-D-Glc-(1$\rightarrow$4)-α-D-Glc-1-diphospho-ditrans,polycis-undecaprenol (Man-Cel-UNDPP). The linker provides structural flexibility, allowing domain movements essential for enzymatic function[12] (Figure 1).

Unlike other enzyme folds where the catalytic site is a well-defined pocket, GumK's active site is formed at the cleft between the two domains,[17] making its precise conformation highly dependent on domain motions. GumK follows an inverting catalytic mechanism, in which

the nucleophilic substitution (SN2-like) reaction leads to the transfer of the sugar moiety with an inversion of anomeric configuration. The catalytic residue responsible for this reaction is Asp157, located in the acceptor domain, which activates the acceptor substrate and facilitates the reaction through nucleophilic displacement.[12] The conformational flexibility of the two domains enables two types of motion, twisting and bending, which have been observed in computational studies.[18] These motions are believed to be key for enzymatic activity, facilitating substrate binding and positioning them for catalysis.[17] However, the full conformational landscape of GumK is not known. Identifying the range of possible domain orientations is crucial for defining active conformations, understanding substrate binding and specificity, and determining how these dynamic processes influence reaction kinetics.[6]

Gaining structural and mechanistic insight into the selectivity of the GumK substrate offers two key advantages. First, it can advance the polysaccharide materials industry by enabling targeted enzyme modifications that expand its specificity to non-native substrates, potentially leading to the production of new polysaccharides with tailored properties. Second, it contributes to a broader understanding of how glycosyltransferases achieve selectivity and catalysis.

While experimental studies have provided critical structural data, computational methods now complement them by adding atomistic detail to protein dynamics or enzyme activity, by generating ensembles of conformations compatible with experimental observations. Among the various computational approaches, molecular dynamics simulations can help identify key conformations, reveal hidden molecular interactions, and refine our understanding of how structural flexibility influences function and activity. Ultimately, these insights can guide the design of targeted experiments, enhancing our ability to manipulate and engineer these enzymes for both fundamental research and industrial applications.

Studying GumK computationally is challenging due to the high flexibility of both the protein and its substrates, its interaction with a membrane-embedded acceptor, and the poorly defined catalytic site located in the interdomain cleft. This flexibility creates a com-

plex conformational landscape, complicating the characterization of the structural dynamics that regulate its function and substrate recognition.

To address these challenges, we applied a multi-tiered computational approach, combining normal mode analysis (NMA), unbiased molecular dynamics (UMD) simulations, coarse-gained OPES simulation and Hamiltonian Replica Exchange (HREX) for enhanced sampling. This strategy provided new insights into the various challenges associated with GumK, including its flexibility, substrate interactions, and catalytic site dynamics. Our findings lay the foundation for future experimental validation and potential biotechnological applications.

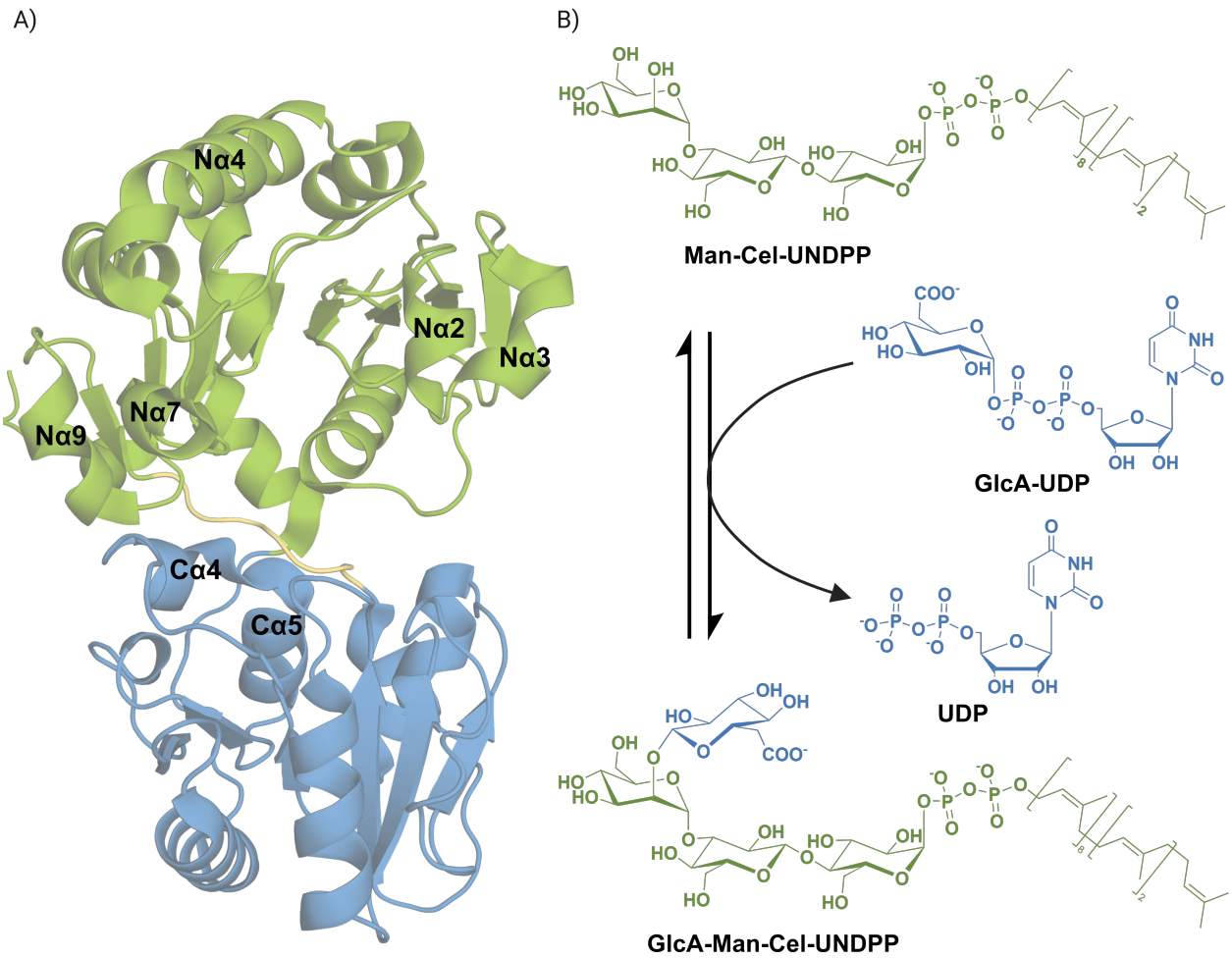

Figure 1: A) Crystal structure of GumK (PDB 2HY7), colored according to domain nomenclature: the N-domain in green, the C-domain in blue, and the linker in yellow. Key regions of the protein, discussed in the main text, are labeled. B) Reaction catalyzed by GumK, with colors corresponding to the domain assignments. The acceptor substrate, Man-Cel-UNDPP, binds to the N-domain, while the donor substrate, glucuronate-UDP (GlcA-UDP), binds to the C-domain. The enzyme catalyzes the formation of a β-1,2-glycosidic bond between glucuronate and mannose in the acceptor substrate, producing GlcA-Man-UNDPP.

Results

The conformational space of GumK is defined by three elementary motions

GumK consists of two distinct domains, with its active site located in the interdomain region. Understanding the dynamics of these domains is crucial for identifying functionally relevant conformations (e.g., closed and open states), pinpointing key residues involved in

interdomain interactions, and characterizing the complexity of its conformational space.

To achieve this, we first performed normal mode analysis (NMA),[19] a coarse-grained approach that models the protein as an elastic network in which α-carbons act as nodes. The resulting normal modes describe intrinsic vibrational movements of the protein, with low-frequency modes that capture large-scale conformational changes and higher-frequency modes that represent localized fluctuations.

GumK crystal structure (PDB 2HY7) is considered a closed conformation,[12] and we use it as the input for NMA. We found that the first two non-zero normal modes correspond to a twisting (NM1) and a bending (NM2) of the two domains (Figure 2A). These two modes exhibit similar frequencies, with the NM1 showing a smaller eigenvalue than NM2, which implies a slower collective vibration. Even though the directionality of motion appears symmetric, the square displacement of the two normal modes is asymmetric (Figure S1), likely due to the uneven distribution of interdomain contacts in the crystal structure.

Further analysis of the first four normal modes revealed a third key motion (NM3), characterized by a second bending of the two domains (Figure 2A). NM2 and NM3 represent perpendicular displacements: NM2 expands the interdomain cleft, increasing the distance between the two domains along their interface, while NM3 tilts one domain relative to the other, causing an asymmetric shift rather than simply widening the cleft. This third mode is the last relevant motion, as the fourth mode has a higher eigenvalue than the first three, and it mainly represents fluctuations localized to the acceptor domain (Figure S1 and S2).

NMA provides insight into global harmonic motions, but it is inherently limited to harmonic deformations around the input structure and does not account for anharmonic conformational transitions or side-chain contributions. To obtain a more comprehensive description of GumK's conformational space, we employed ClustENMD, a method that generates an ensemble of full-atomistic conformations by perturbing the structure along a subset of normal modes and subsequently refining it using the AMBER99SB-ILDN[20] force field.

The final ensemble, projected onto the first three normal modes (Figure 2B and D), reveals

that as conformations shift toward more negative values of NM2, the distribution broadens, with higher RMSD values observed at the ensemble's border. This suggests that when contacts between the two domains are lost, the protein can adopt multiple conformations due to the flexibility of two regions spanning residues 204–221 and 353–361.

Figure 2B further shows that NM1 and NM2 are highly correlated, resulting in a degenerate and less informative space. In contrast, the space defined by NM2 and NM3 reveals two densely populated states: I and II in Figure 2. The first state is centered at zero, suggesting it closely resembles the crystal structure. The second state, located at negative NM2 values and positive NM3 values, corresponds to a rotated donor domain. In this conformation, helix Cα4 (Figure 1A) interacts with residues in the acceptor binding site, aligning with population II in Figure S3B.

NM3 reduces degeneracy in the conformational space, as a consequence of the interdomain interface chemistry (Figure 2C). On the left side, a hydrogen bond network involving Glu192, Ser160, Tyr300, Thr161, Ser186, Asp303, and the catalytic Asp157 surrounds a hydrophobic interaction between Tyr300 and Met189. In contrast, the right side has only a solvent-exposed salt bridge (Arg29–Asp272), resulting in asymmetric domain opening. These interactions are near functional sites: the hydrogen bonds network involves catalytic residue Asp157, likely influenced by acceptor substrate binding, whereas the charged residues on the right side are close to the UDP-binding site. Thus, substrate binding could modulate interdomain dynamics and GumK activity.

The conformations sampled by ClustENMD indicate that GumK comprises two substrate-binding domains—an acceptor domain and a donor domain—connected by two flexible linkers. To avoid confusion with established structural nomenclature, we define the acceptor domain as residues 13–203 and 362–385 (the region binding the acceptor substrate) and the donor domain as residues 222–352 (the region binding the donor substrate), with the linkers spanning residues 204–221 and 353–361 (Figure 2C).

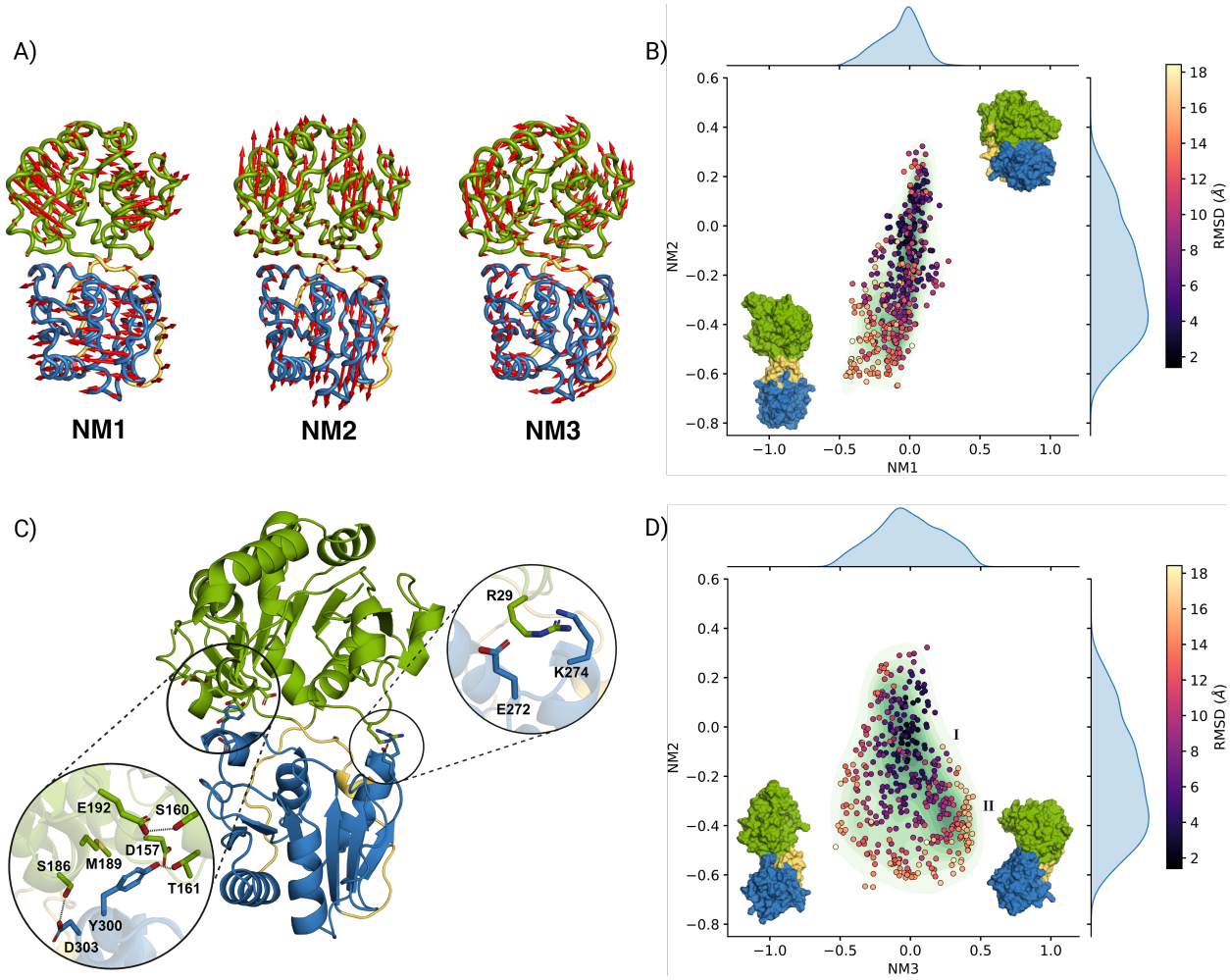

Figure 2: A) First three normal modes from the normal mode analysis of GumK (PDB 2HY7), highlighting its dynamic and functional properties. The N-terminal (green), C-terminal (blue), and linker (yellow) regions reflect distinct motions. Red arrows indicate alpha-carbon displacements: mode 1 involves symmetric twisting, mode 2 symmetric bending along the main axis, and mode 3 bending perpendicular to it. B) ClustENMD ensemble projected onto the first two normal modes. The green contour map represents population density, while dots, colored by RMSD from the crystal structure, reveal functional motions dominated by twisting and bending. C) GumK crystal structure with interdomain interactions: a stabilizing hydrogen bond network and hydrophobic core (Met189, Tyr300) on the left, and a solvent-exposed salt bridge (Arg29–Asp272) on the right. These features influence domain flexibility and function. D) ClustENMD ensemble on the second and third normal modes, colored as in B). Extreme conformations along NM3 show a symmetric distribution with a dense basin (population II) on the positive side.

Charged residues in the acceptor domain mediate interactions between GumK and the membrane

GumK is hypothesized to be a monotopic membrane protein localized on the inner leaflet of the cellular membrane. Based on residue properties in the acceptor domain, Barreras et al. proposed a possible membrane orientation for GumK by analyzing its crystal structure,[12]

but no experimental validation of this interaction has been reported.

To investigate how GumK interacts with the membrane, we performed MARTINI[21] coarse-grained molecular dynamics (MD) simulations using the full-length AlphaFold-predicted structure,[22] which adopts a conformation similar to the crystal structure (PDB: 2HY7), i.e., a closed state. The elastic network in MARTINI constrained the protein to its starting conformation, and OPES metadynamics explore[23] was used to enhance the interaction at the membrane. The closed conformation was chosen as a reference in this experiment. This choice is justified by three arguments. Firstly, an experimental structure is an objective starting conformation. Secondly, using a closed conformation avoids sampling interdomain dynamics, which are not relevant in this experiment since the membrane interaction involves mainly the less soluble acceptor domain. Thirdly, using the same structure for MARTINI and OPM facilitates comparison of results from both methods. A cylindrical-shaped restraining potential was used to restrain the conformations of the protein in the unbound state and favour the sampling of bound modes. (Figure S4). This approach enabled us to model the docking process and assess the stability of different interaction modes.

The free energy surface (FES) (Figure 3A) is expressed as a function of the distance between bead BB of Leu20 which is approximately centered in GumK'acceptor domain and the membrane (d_A) and its tilt angle (τ) (Figure S4). When the d_A is less than 5 nm, two distinct minima appear, corresponding to different orientations. The primary minimum (I in Figure 3A) has τ close to zero, with the acceptor domain interacting with the membrane through helices Nα2 and Nα4 (Figure 1A). This region is rich in arginine residues, while Nα4 contains Trp98, a key residue interacting with the lipid layer, previously hypothesized by Barreras et al. as the primary membrane anchoring site.

OPM web server[24] was used as an alternative method to predict the most likely orientation of GumK relative to the lipid bilayer based on its crystal structure. The conformations in minimum I are closely aligned with the orientation predicted by OPM (Figure 3B). The bound state is enhanced by the N-terminus (red region in the Figure), which can interact

with the membrane.

Analysis of the donor domain's positioning in this state reveals two possible conformations (Figure 3C). In one case, the donor domain interacts with the phospholipid layer, slightly increasing the τ, while in the second case, the donor domain remains free in solution.

The second minimum (II in Figure 3A) of the FES represents an opposite orientation where both termini of the acceptor domain interact with the membrane.

OPES simulations sampled the expected binding mode, as hypothesized by Barreras et al.[12] and supported by the OPM prediction. The second minimum is not compatible with the usual orientation of membrane-associated GTBs[17] but indicates that the two termini tend to bind the membrane, potentially contributing to further stabilization of the protein–membrane complex.

The simulation is not fully converged, as bias potential is still being deposited at the end of the run in the region where the protein binds to the membrane, as predicted by OPM. This indicates that, although this binding mode is sampled, the resulting minimum is not fully explored and the free energy surface can only be used qualitat

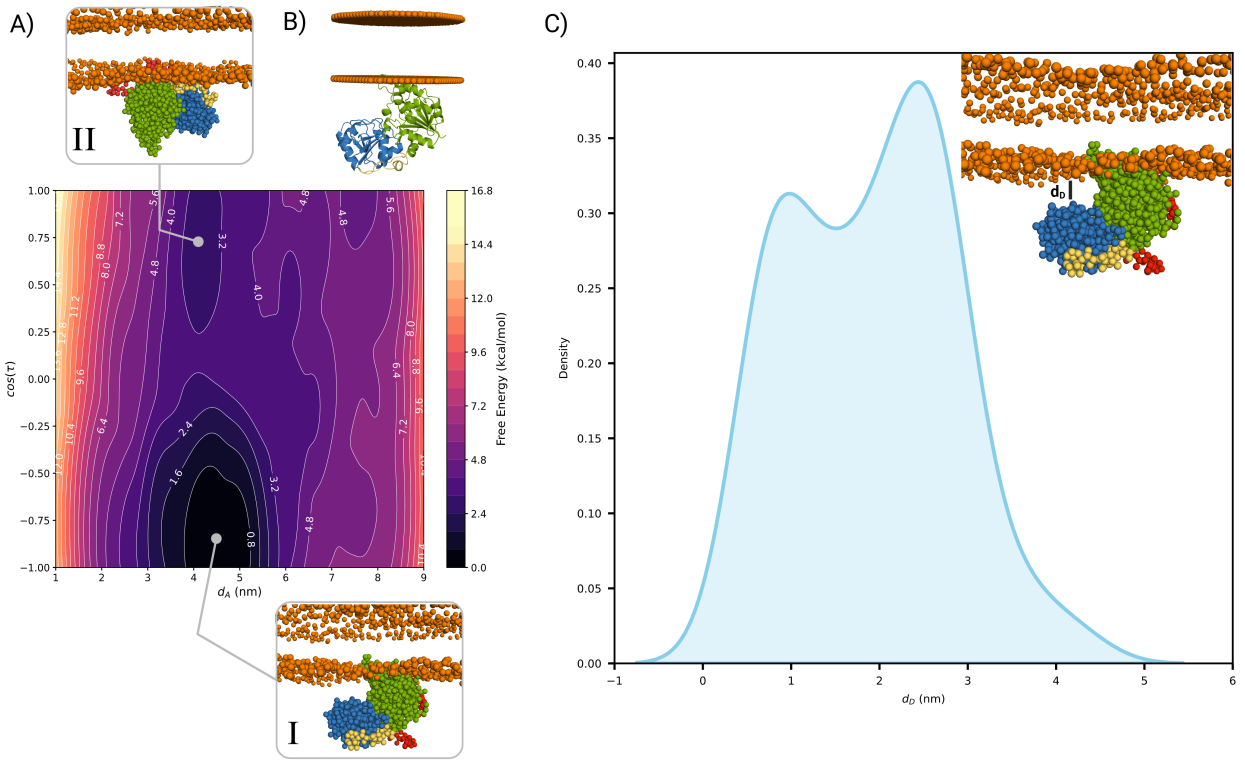

Figure 3: A) Free energy surface (FES) from cylindrical-restrained OPES simulation with the Martini force field, plotted against the two biased collective variables: acceptor domain distance from the membrane (d_A) and tilt cosine (τ). Minima I and II correspond to distinct GumK orientations. Structural snapshots highlight sampled poses, with green for the acceptor domain, blue for the donor domain, red marking GumK termini and orange spheres indicating phospholipid phosphate groups. B) GumK crystal structure (PDB 2HY7) from OPM server aligns with minimum I. C) Probability density of donor domain distance (d_D) from the membrane phosphate plane, measured via Arg259 in Minimu I.

Although ClustENMD effectively explored conformational changes (Figure 2), it lacks membrane context, limiting its ability to fully capture GumK's behavior. Additionally, understanding the kinetics of the opening mechanism and how the protein behaves without a constraining potential bias is crucial. For these reasons, we performed fully-atomistic MD simulations using the OPM-derived orientation in a solvated membrane system.

The system was simulated for 1 µs, and conformational changes were analyzed using essential dynamics,[25] employing principal component analysis (PCA) to construct a covariance matrix based on α-carbon displacements throughout the trajectory. As with normal mode analysis, the resulting principal components (PCs) describe the protein's conformational changes sampled during the simulation. A total of 20 PCs were computed using ProDy.[25] The eigenvalue plot (Figure S5) indicates that the first six PCs capture the majority of vari-

ance, with PC1 alone explaining 67%. The non-Gaussian distributions of these components suggest that they define GumK's essential subspace (Figure S6).[26] PC1 exhibits a bimodal distribution, with one peak near zero corresponding to a conformation similar to the crystal structure (e.i. closed), while a second peak indicates a more open state. The cosine content of the time series of this PC is 43.1% indicating that the transition that leads to the bimodal distribution is a conformational change and not the exploration of a broad basin.[27]

Projection of PC1 onto the first three normal modes (Table 1) reveals that it is a linear combination of NM2 and NM3, representing an opening motion between domains. This motion aligns with the third quadrant of the NM3 vs. NM2 plot. An additional ClustENMD simulation, sampling conformations closer to the crystal structure, confirmed a strong correlation between NM2 and NM3, reinforcing the interpretation that PC1 captures the protein's opening transition (Figure S9).

PC2 follows a more Gaussian distribution, though it retains a high eigenvalue (Figures S5 and S6). Projection onto normal mode space (Table 1) indicates that PC2 combines NM1 and NM3, corresponding to a twisting motion of the two domains.

To asses the convergence level of the simulation, we used the bootstrap normalized block covariance overlap method. The fitting of the inverse of the resulting curve using a triple exponential gives an idea of the time get the convergence of the simulation. The analysis shows that the system is far from convergence and at least 4 us should be used to get a full decorrelation of the covariance overlap matrix (Figure S7).[28] However, the two principal component directions converge faster, as shown in Figure S8. Since we are interested in describing the opening path of the protein and defining a space that can easily discriminate between different conformations, the convergence of PCs directions is more relevant.

By projecting the MD trajectory onto the PC1-PC2 space (Figure 4A), two main populations were identified: one representing the crystal-like closed state and the other one a more open conformation.

The full-atomistic simulations confirmed the observations made in the MARTINI coarse-

Table 1: Projections of the first two principal components computed in the essential dynamics analysis of the GumK-membrane complex unbiased MD simulation.

Vector	NM1	NM2	NM3
PC1	-0.09	+0.13	-0.12
PC2	+0.43	+0.02	-0.34

grained model, showing that over 1 µs, the donor domain also can interact with the membrane. Evidence for this interaction is found in the distribution of distances between Arg259 and the phospholipid headgroup plane (Figure 4D), which is similar to that observed in the MARTINI simulations (Figure 3C). Moreover, additional residues, like Glu236 and Asp234, interact with the membrane via electrostatic attraction to the positively charged phosphatidylethanolamine headgroups.

Further analysis of the acceptor domain revealed key hydrophobic and electrostatic interactions stabilizing GumK at the membrane. Trp98 and Trp109, located on helix N4a, are inserted into the membrane, anchoring the protein (Figure 1A).

During the first 400 ns, a cardiolipin (TYCL2) molecule binds within the cleft between helices Nα4 and Nα2 (Figure 1A and 4C), while a phosphatidylglycerol (DYPG) molecule binds to a groove between helices Nα2 and Nα3.

After 400 ns, DYPG gradually moves from the top of the acceptor domain closer to the bound TYCL2, eventually replacing it. During this process, DYPG is pulled out of the membrane and interacts with Lys60, His90, Arg52, Ser57, and Arg64, located in the Nα4-Nα2 cleft (Figure 1A and Figure 4C), while the leaving TYCL2 interacts with Arg95 and Asn93. The lipid tails of both phospholipids interact with hydrophobic residues in the cleft, particularly Leu56, Val89, and Phe92, which form a groove on the acceptor domain that can accommodate the hydrophobic tails. The phospholipid binding events observed in these simulations suggest the likely location of the acceptor substrate binding site and provide insights into its orientation within the cleft.

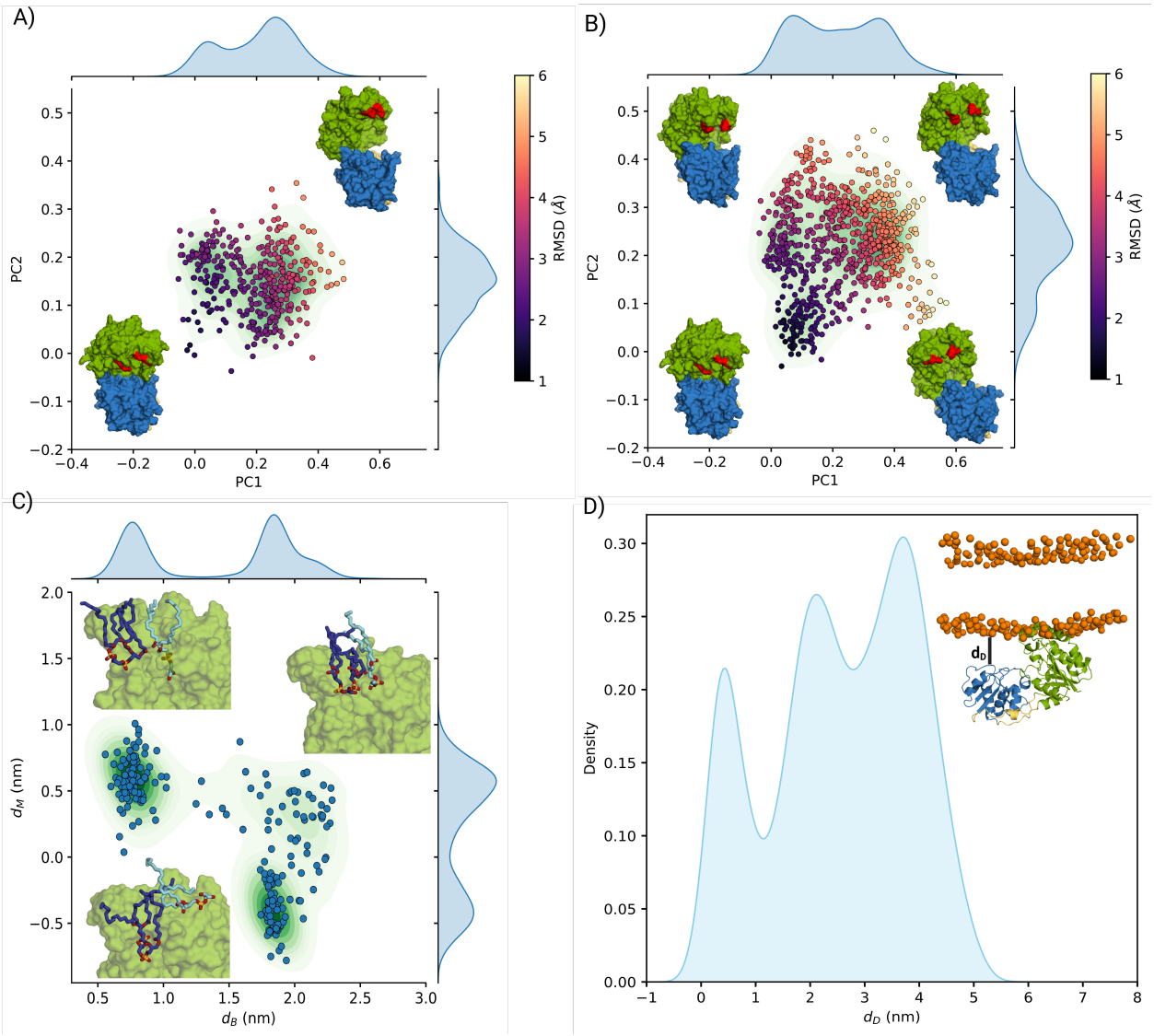

Figure 4: A) Projection of the conformations sampled during the MD on the first two principal components of the essential space. The contour map in green represents the population density of the total ensemble, while some conformations are represented as colored dots based on the RMSD relative to the GumK crystal structure. The two structures highlight representative conformations: a closed conformation on the negative side of the axes and a partially opened conformation on the positive side. The read region is the tail's acceptor substrate binding site used to guide the reading of the picture. B) GumK conformations sampled in the presence of the acceptor substrate, projected onto the same space as panel A. The representative structures show how the protein moves across the four quadrants: open in the first, twisted closed in the second, closed in the third quadrant, and twisted open in the fourth. C) Binding path of phosphatidylglycerol (DYPG) sampled during the simulation. The x-axis represents the distance of DYPG's phosphorus atom from the center of mass of the binding region (d_B), while the y-axis shows the distance of the same atom from the average plane of the membrane's phosphate heads (d_M). The conformations are representative of the binding path. D) Distance distribution of the donor domain from the membrane (d_D) as shown in the structure. The alpha carbon of residue Arg259 is used as a reference for the distance. It shows the donor domain interacts with the membrane.

GumK donor binding site discriminates between its GlcA-UDP substrate and the non-reactive ligand Glc-UDP

Even though Barreras et al. attempted to co-crystallize an inactive mutant of GumK with its donor substrate,[12] GlcA-UDP, the available X-ray crystal structure (PDB: 2Q6V) contains coordinates only for the UDP moiety. This means that the binding pose of the donor substrate's sugar portion remains unknown. To address this, we performed a fully-atomistic simulation using the CHARMM36 force field[29] and the enhanced sampling method Hamiltonian Replica Exchange[30] (HREX) to sample the monosaccharide's conformational space.

To facilitate the exploration of the conformational space, we adopted a three-step approach. First, we simplified the system by simulating only the isolated donor domain instead of the full-length protein to avoid bias from the protein conformation. Second, we employed HREX to enhance sampling efficiency. Third, we constrained the uridine moiety to its crystallographically determined binding conformation, while the sugar portion was confined using a funnel-shaped restraint to prevent unbound conformations. Three independent HREX simulations were run with 19 and 17 replicas, respectively, for GlucA-UDP and Glu-UDP. The results from the first simulation are shown in Figure S10. The other two simulations are used for reproducibility evaluation, and the results are in (Figure S12).

The final replica exchange probability for GlcA-UDP ranged between 33% and 37%, indicating reasonable exchange between neighboring replicas. The replicas move freely in the replica ladder and there is no evidence of serious clustering; however, free diffusion is not observed (Figure S10).

We used the distance (d_k) between the C6 atom of GlcA-UDP and the N^{ζ} of Lys307 as a collective variable (CV) to characterize the binding modes of the sugar moiety. The different populations sampled during the simulation are reported in Figure S10C.

The distance plot revealed two predominant binding modes of GlcA-UDP: one centered at $d_k = 0.3$ nm and another at $d_k = 0.9$ nm. In the first mode, the carboxylate group of GlcA and one of the two phosphate groups interact with Lys307. In the second mode, only

the phosphate groups interact with Lys307, while the carboxylate group remains solvent-exposed.

Although d_k provides valuable insight into the sampling of the substrate, it does not distinguish between different orientations of the sugar ring. To achieve a more detailed characterization of the binding modes, we introduced two additional angular CVs, γ and θ, defined as the angles between the sugar ring's normal vector and the donor domain's principal axes of inertia, which describe the sugar ring's orientation relative to the protein (Figue S10D).

Analysis of the polar space revealed three well-defined populations (see Figure 5). The first population, Cluster I, is centered at $d_k = 0.9$ nm and features the sugar ring oriented with its β-face facing helix Cα7. Cluster II, exhibits an average $d_k = 1.02$ nm, maintaining the α-face orientation toward Cα7 but in a distinct spatial arrangement. Cluster III represents a broad and flexible basin, as the sugar ring is more solvent-exposed when both phosphate groups interact with Lys307, allowing greater conformational variability. The third cluster, Cluster C, has an average $d_k = 0.35$ nm, with the α-face directed towards Cα7.

To evaluate the convergence of these basins, the RMSD of the free energy surface (FES) as a function of the 3 CVs e.i. γ, θ and d_k was plot over an increased number of frames. The reference for the RMSD is the starting FES (Figure S13 A,B). All FESs were computed using all replicas and the binless wham algorithm to estimate the weights for the weighted histogram. This shows that even though the replicas don't show a free diffusion in the replica ladder, the FESs as a function of the target CVs seem to reach a convergence.

We analyzed the sugar-binding region and identified key interactions contributing to substrate recognition and stabilization. Tyr292, Ser305, and Lys307 interact with both the carboxylate and the two phosphate groups across different binding clusters. Additionally, Leu301, Met231, and Leu232 form hydrophobic interactions with each other, effectively excluding the flexible loop between Ala293 and Pro298 from direct contact with the sugar moiety. The distances among these hydrophobic residues vary across different clusters (Fig-

ure S14). This variability suggests they adopt distinct configurations depending on the binding mode of GlcA-UDP, potentially influencing the orientation of the sugar ring.

It has been hypothesized that GumK exhibits high specificity for its native substrate, GlcA-UDP.[31] To further investigate this selectivity, we performed an additional HREX simulation using the same setup but with glucose-UDP (Glc-UDP), which is not a GumK substrate.[18]

For this substrate, the exchange probability was between 34% and 39%, indicating frequent exchange between neighboring replicas. The replicas span the replica ladder, but also in this case, the free diffusion is not observed (Figure S11).

The distance plot for Glc-UDP revealed a single dominant population centered at $d_k =$ 0.8 nm (Figure S11C), corresponding to binding modes where the phosphate groups interact with Lys307. Unlike GlcA-UDP, glucose's C6 does not interact with Lys307, Tyr292, or Ser305. Analysis in polar space (Figure 5B and S11D) indicated greater conformational flexibility for Glc-UDP, highlighted by an additional peak (Cluster II), in which the sugar ring adopts a perpendicular orientation relative to helix Cα7. On average, all visible clusters exhibit a d_k between 0.7 and 0.8 nm.

The same convergence analysis has been performed for Glu-UDP and the behavior is similar; the RMSD tends to flatten over time (Figure S13 C,D).

HREX effectively explored the sugar-diphosphate binding modes when the nucleoside part of the donor domain was kept bound. The donor domain structure appears to impose conformational constraints on the substrate, with the acidic substrate more constrained than the neutral one due to the additional interaction between the carboxylate group and Lys307. Although free diffusion in the replica ladder is not observed, the sampled basins remain reproducible (Figure S12), indicating that the qualitative behavior of the two substrates is consistent. However, lack of free diffusion might indicate incomplete sampling of the polar coordinate subspace used to describe sugar conformations,

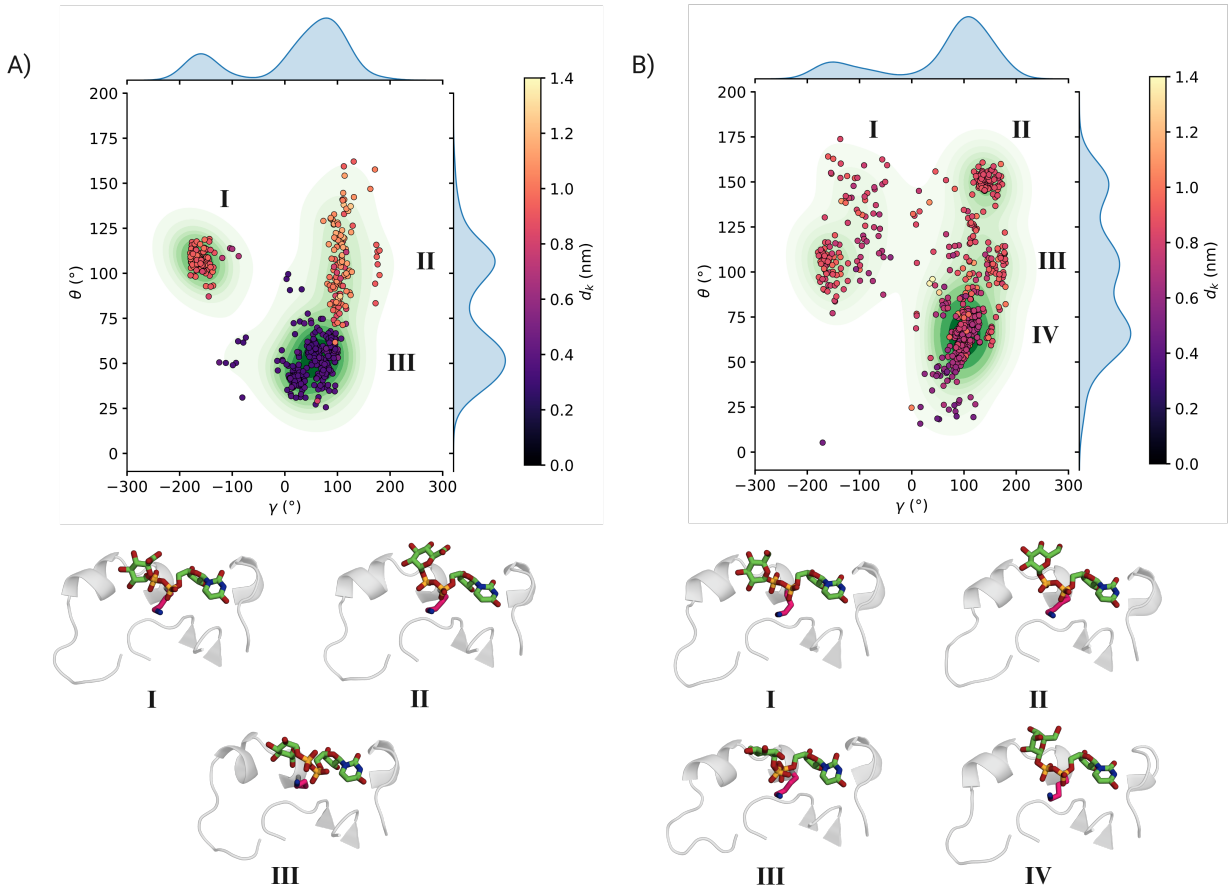

Figure 5: (A) Binding modes of glucuronate-UDP sampled during a constrained Hamiltonian replica exchange simulation of the donor domain. The two axes represent key angles defining the sugar ring orientation relative to the protein (see Supporting Information). The color scale indicates the distance between the sugar C6 and the N^{ζ} atom of residue K307 (pink stick in the structures) and the green contour map is the population density of the sampled conformations. Three distinct clusters, differing in both distance and orientation, are identified. Representative structures for each cluster are shown below. (B) Binding modes of glucose-UDP sampled using the same simulation protocol and plotted in the same space as panel (A). In this case, four clusters are observed, with representative structures displayed below.

Bioinformatic analysis shows conserved residues in the donor domain sugar-binding site

The previous simulations identified key residues in the sugar-binding site—Tyr292, Ser305, Lys307, Met231, Leu232, and Leu301—strategically positioned to orient the sugar ring. If these residues play a functionally critical role in the donor domain, we expect their spatial arrangement to be conserved across different GumK enzymes. To assess this, we performed a structural alignment of all sequences labeled as GumK in UniProt, comparing each protein's reported structure with the crystal structure of GumK.

Similarly, we conducted structural alignments for GumI and GumH (Figure S16), two enzymes involved in xanthan gum side-chain biosynthesis that act on Mannose-GDP (Man-GDP). GumI belongs to the CAZy GT94 family, while GumH is classified within the GT4 family. For these enzymes, we used the AlphaFold models of *Xanthomonas campestris* GumI and GumH as structural references.

One of the most distinctive features of GumK is the presence of a conserved hydrophobic interaction network formed by Leu301, Met231, and Leu232, which plays a critical role in constraining the orientation of the sugar ring (Figure 6). This interaction is closely associated with the presence of the loop spanning Ala293 to Pro298, which, as shown by simulations, contributes to the flexibility of the binding region. While this loop is conserved in GumKs, its amino acid composition is highly variable and does not exhibit a conserved chemical nature. The absence of this loop in GumI and GumH suggests that it is a structural element specific to GumK, potentially contributing to the enzyme's specificity and flexibility in sugar recognition.

Another key difference lies in the distribution of positively charged residues in the binding site. GumK is characterized by the presence of a conserved lysine (Lys307) at the donor-binding site, which interacts with the carboxylate group of GlcA-UDP. Structural alignment revealed that while GumKs consistently maintain a lysine at this position, GumI and GumH typically feature a neutral residue here. However, both GumI and GumH display two positively charged amino acids, the Lys-Arg pair (Figure S16), in a different region of the phosphorylated sugar-binding site, indicating an alternative charge distribution strategy for stabilizing binding of Man-GDP. Furthermore, Tyr300 and Ser304, which contribute to the stabilizing interactions of the hydrogen bond network between domains (see Figure 2 panel C), are structurally conserved across GumKs.

From these observations, the donor monosaccharide-binding region can be divided into four functional segments (Figure 6): (i) the region formed by helices Cα4 and Cα7 (Figure 1A), along with residues Tyr300 and Ser230, which interact with the acceptor domain and

help orient the sugar ring while engaging with the acidic group of GlcA-UDP; (ii) the flexible loop between Ala293 and Pro298, which is unique to GumK, conserved within its family but variable in amino acid composition, and absent in GumI (GT94) and GumH (GT4); (iii) the hydrophobic interaction site formed by Leu301, Met231, and Leu232, a feature exclusive to GumK that stabilizes sugar ring orientation; and (iv) the uridine-binding region.

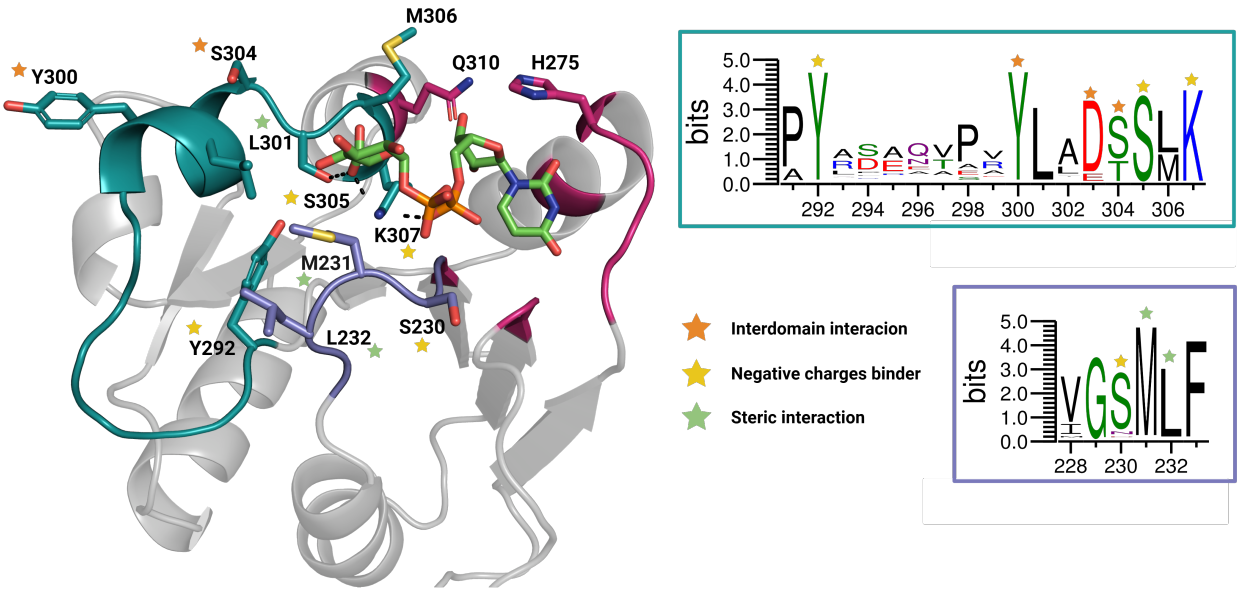

Figure 6: The structure of the GumK donor-binding site is colored based on the three regions analyzed in the structural alignment. The cyan and purple regions correspond to residues interacting with the sugar diphosphate, while the dark pink region represents the uridine-binding site. These colors match the sequence conservation analysis shown on the right. The colored stars indicate key functional residues: orange stars denote residues involved in interdomain interactions with the acceptor domain, yellow stars highlight residues interacting with both the carboxylate group and pyrophosphate of the donor substrate, and green stars mark the hydrophobic bridge (L301, M231, L232), which stabilizes the sugar ring orientation. On the right, sequence conservation is illustrated using WebLogo 3.7.12, where the colors represent the chemical properties of the residues, distinguishing between hydrophobic, polar, positively charged, and negatively charged amino acids.

The acceptor substrate is constrained in a linear conformation by the binding site

GumK specifically binds to the acceptor substrate Man-Cel-UNDPP (Figure 1B), showing a preference for interacting with the pyrophosphate group.[31] The hydroxyl group at C2 of the mannose unit serves as the reactive site, interacting with the catalytic residue Asp157.

To explore the molecular basis of GumK-acceptor interactions, we performed three molecular dynamics (MD) simulations of the GumK-acceptor complex in a membrane environment.

The initial complex was built by positioning the acceptor's tail near the region that interacted with DYPG in the free protein simulation (Figure 4C) and was simulated for 1 µs for each replica for a total of 3µs. The distance between the acceptor's reactive oxygen and the carboxylate carbon of Asp157 (d_acc) was used as a collective variable (CV) to identify the binding event. Among the three replicas, only in one of them the acceptor substrate stays close to the catalytic residue in a reactive conformation for a long period furthermore, only this simulation is used for further considerations (Figure 7A and S18).

During the first 200 ns (region I Figure 7A), the acceptor transiently interacts with Arg29, disrupting the interdomain salt bridge (Arg29-Glu272 Figure 2C) that keeps the protein closed. Despite the trisaccharide's flexibility, the substrate remains near the binding site due to the stable anchoring of its lipidic tail within the cleft between helices Nα2 and Nα4 (Figure 1A).

After 200 ns, the acceptor gets into a more constrained conformation, with d_acc staying below 1 nm for 600 ns (region II in Figure 7A). During this period, the trisaccharide adopts a predominantly linear conformation, as reflected in a narrower α-(1→4) glycosidic bond dihedral distribution compared to the unbound state. Although the α-(1→3) bond has a broader dihedral distribution, it also favors a linear conformation (Figure S17). In this time window, d_acc shows two possible states: one where the reactive hydroxyl of mannose forms a hydrogen bond with Asp157 and another where it remains near the catalytic residue without direct interaction. This observation suggests that the acceptor substrate has conformational freedom. This likely arises from the absence of the donor substrate, which, when present, would impose additional structural constraints on the acceptor. Additionally, the protein undergoes partial opening upon acceptor binding, further increasing the conformational freedom of the acceptor.

Additional contacts help stabilize the substrate, including C6 of β-glucose with Ser22, C4 with Ser132, and α-glucose with Arg52 and Ser22. Furthermore, C6 of α-mannose interacts with His23, whose orientation is stabilized by Glu131 and Ser22. A stable hydrogen bond

network (Glu131, Ser22, Ser132) maintains these residues oriented within the binding site (Figure 7B). However, the bound state exhibits dynamic behavior with short-lived contacts, which implies that it adopts multiple conformations rather than a single stable bining pose.

The lipid tail follows a distinct binding behavior: the first three isoprenyl units remain stably bound, while the rest of the tail is more flexible, as shown by the RMSF of lipidic tail carbons (Figure 7C). This stability arises from the cleft between Nα2 and Nα4 (Figure 1A), which surrounds the tail and enables hydrophobic interactions. The second isoprenyl unit engages Phe92 and Leu56, contributing to substrate retention. Notably, the unsaturated bond of this unit stacks against the aromatic ring of Phe92, further stabilizing the substrate within the binding pocket.

To evaluate how acceptor binding affects GumK dynamics, we projected the trajectory onto the essential space previously defined for GumK (Figure 4B). The analysis revealed that acceptor binding enhances protein opening, with the domains separating more quickly and staying open longer than in the unbound state. This effect is primarily due to interactions with Arg29, which destabilize the interdomain salt bridge and promote domain separation.

Additionally, the acceptor disrupts the interdomain hydrogen bond network, particularly Ser160-Glu192, increasing domain twisting and further facilitating opening. During the simulation, the residues in this network remain close but undergo different contact events. In particular, a new hydrogen bond between Tyr328 and His201 helps to keep the domains together, while Tyr300 shifts under the turn N7a (Figure 1A).

The pyrophosphate group of the lipid tail is the main driver of acceptor binding, clamping tightly between helices Nα4 and Nα2 via electrostatic and hydrophobic interactions. This interaction stabilizes the substrate, keeping the trisaccharide close to the binding site, and inducing conformational changes at the interdomain interface. However, as the reactive hydroxyl approaches Asp157, the binding becomes less stable, suggesting that lipidic tail anchoring plays a dominant role in initial substrate retention.

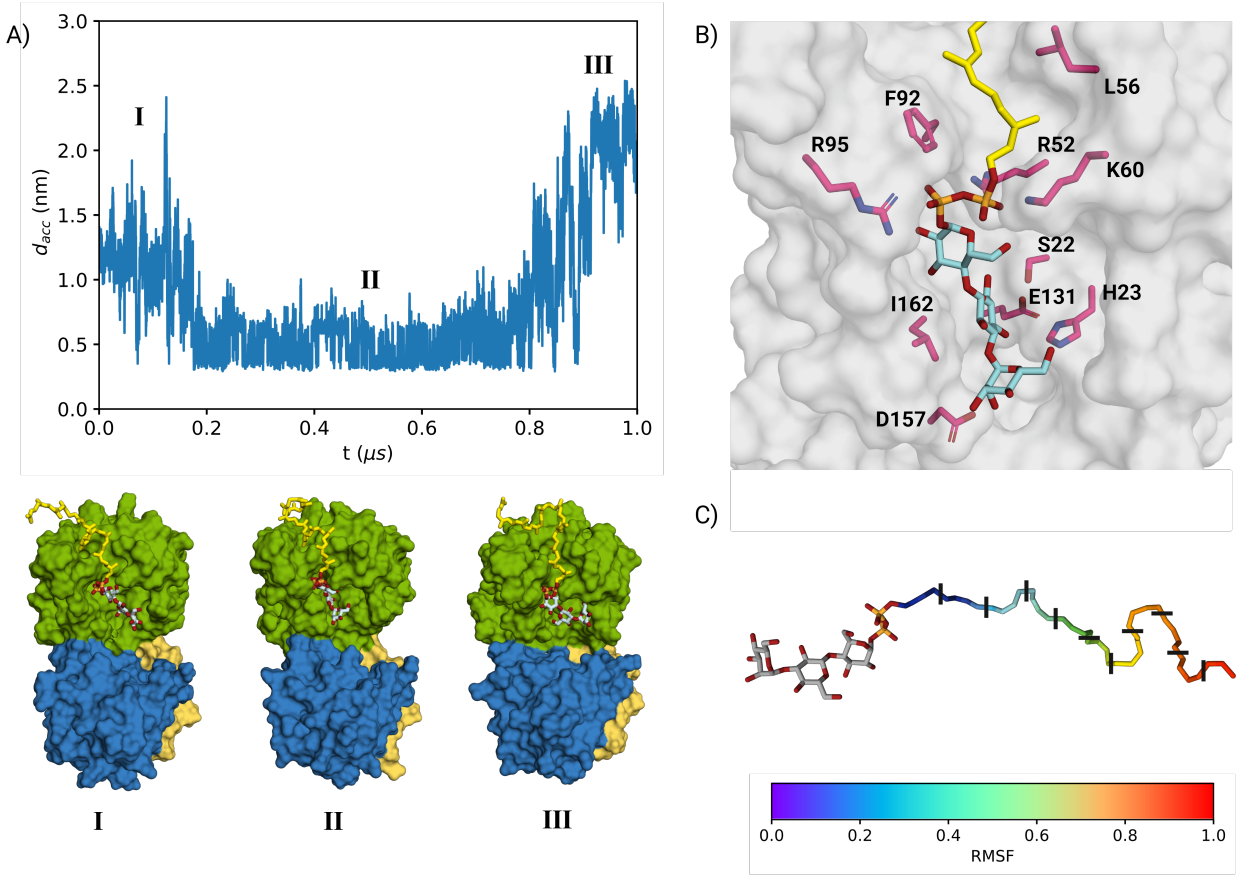

Figure 7: A) Time evolution of the distance between the acceptor substrate and the binding site, measured between the reactive oxygen at C1 of the last sugar unit and the carboxylate carbon of Asp157. The system transitions through three main states, illustrated by the structures at the bottom: I, II, and III. The bound state (II) corresponds to a more open protein conformation, where the acceptor substrate is stabilized in the binding site. States I and III represent unbound conformations, where the substrate interacts with Arg29, promoting protein opening. B) Detailed representation of the bound state, highlighting key residues involved in substrate recognition and binding. C) Normalized RMSF of the carbon atoms in the lipidic tail of the acceptor substrate in the bound state. Black lines mark the isoprenyl units, showing that the first three units remain stably bound within the binding pocket.

Discussion

GT-B enzymes demonstrate remarkable conformational plasticity, with diverse domain orientations driven by distinctive twisting and bending motions. Our computational investigation of GumK has revealed the molecular mechanisms underlying this flexibility through a multiscale approach. This structural dynamism has a fundamental functional role in substrate recognition and catalysis, establishing a direct link between substrate binding, conformational transitions, and enzymatic activity.

The interdomain contacts can tune GumK activity.

Normal mode analysis highlighted three key motions: twisting (NM1) and two bending modes (NM2, NM3) (Figure 2A). However, the topology of the elastic network depends on the initial conformation, as inter-residue contacts rely on a radial cut-off. In the crystal structure (PDB 2HY7), the hydrogen bond network forms a more compact interface than the salt bridge side (Figure 2C). As a result, NM1 is the slowest mode since it disrupts interdomain contacts. Conversely, analyzing open GumK conformations yields different results due to the reduced number of contacts between the domains. Starting from a closed conformation (such as the one observed in the crystal structure) provides a more meaningful set of vectors, as this conformation preserves the potential interactions between the domains. In contrast, when the protein adopts an open conformation, the eigenvalues of the first three modes approach zero (Figure S2B). This indicates that these modes are primarily associated with nearly free rotations and translations of the domains. This occurs because, in the absence of interdomain contacts, the protein behaves approximately as two independent domains that can move freely relative to each other as also shown in ClustENMD results (Figure 2B and C).

ClustENMD provided additional insights into the domain dynamics at an atomistic level. The NM2-NM3 projection (Figure 2) revealed that bending motions make the conformational space broader due to the flexibility of the linker regions (residues 204-221 and 352-361). The bending path is affected by the starting interdomain interaction pattern, as also shown by the small perturbation of GumK in ClustENMD (see supp), when these interactions are fully lost, the two domains get fully free to move. This additional freedom makes the conformational space broader, which in turn makes it less likely for the protein to restore the starting conformation, which in terms of activity it translates in a reduction of the catalytic efficiency e.i the kcat.

This is particularly evident in the sampling of non-active conformations, such as the overtwisted state (Figure S3A) and state II (Figure 2D). The first one is incompatible with

enzymatic activity, as the donor substrate binding site is positioned behind the acceptor substrate binding site. In state II, helix Cα7 occupies the acceptor binding site, preventing potential substrate interaction.

The stable hydrogen bond network (Figure 4C) strongly influences the conformational space. Experimentally disrupting these interactions, such as mutating key residues like Tyr300 could weaken interdomain contacts, enhancing twisting motions, and favoring open conformations. This shift might reduce catalytic efficiency by promoting inactive states.

The membrane affects the interdomain interaction by keeping the domains closer.

In a cellular environment, domain mobility is further constrained by intracellular crowding, which favors compact conformations,[32] and by membrane interactions that contribute to the stabilization of GumK. While future advancements may allow for fully realistic cellular environment simulations, current methods are limited to modeling a protein anchored to a membrane.

Both MARTINI coarse-grained and unbiased all-atom MD simulations showed not only that the acceptor domain binds to the membrane as expected, but also that the donor domain can interact with the bilayer thanks to electrostatic contacts. This additional interaction has previously been described for other membrane GTs.[17] Plant GT-B enzymes like atDGD2 are an extreme example since both domains engage in the membrane anchoring via tryptophan residues.[33] Furthermore, a membrane-dependent enzymatic activity has been observed in PimA, a GT4 glycosyltransferase involved in phosphatidylinositol mannoside biosynthesis in Mycobacteria. Despite being a peripheral enzyme, PimA exhibits increased activity in the presence of acceptor substrates or lipid molecules above the critical micelle concentration.[34] In GumK, donor domain interactions with the membrane might affect interdomain dynamics by keeping the domains near each other. This behavior suggests that a lipidic environment may influence GumK by promoting compact and functionally relevant conformations.

In the MARTINI OPES simulation, the protein is constrained to a closed conformation, indicating that variations in donor domain distance primarily result from minor acceptor domain tilts within the same energy minimum, influenced by membrane interactions. These findings, along with observations on the interactions between the two protein termini (Figure 3A), suggest that the protein can adopt different binding modes and may adapt to different membrane topologies. The simulation indicates that the two termini can interact with the membrane. Since AlphaFold assigns a low confidence score to this region, the structural arrangement of the termini remains uncertain. Although they were modeled as a random coil in the Martini simulation, it remains possible that they adopt an α-helical fold that enables stable binding to the membrane. The specific orientation of GumK at the membrane is strongly influenced by the topology of the interfacial residues, particularly the presence of a Trp98 residue in a specific position that can act as an anchoring point. This orientation can vary among different GT-Bs, likely linked to the intrinsic dynamics of the protein backbone.

During 1 µs of unbiased MD simulations, GumK exhibited various conformational changes (Figure 4), demonstrating its ability to adopt open states even without substrates. ClustENMD gave a qualitative representation of the conformational space, but for a quantitative analysis some enhanced sampling methods should still be used. Protein opening in this fold is often modeled using two collective variables (CVs): an interdomain angle for bending and a dihedral angle for twisting.[35] However, metadynamics on these CVs requires extensive restraints to avoid irrelevant phase space sampling.[36] ClustENMD conformations suggest a similar challenge for GumK. Nevertheless, essential dynamics analysis identified PC1 and PC2 as potential pathways describing domain opening and twisting (Figure 4A,B, Table 1) and they could be a more effective CVs.

The acceptor substrate is clamped by GumK and enhances protein opening

GumK fluctuations in the simulations keep the donor binding site partially buried (Figure 4A), suggesting that spontaneous opening may not be sufficient for donor binding. Conversely, the acceptor binding region remains accessible, as shown by the interactions between the acceptor domain and phospholipids (Figure 4C) which are clamped at the cleft between helices Nα2 and Nα4. Additionally, several annular phospholipids at the protein-membrane interface exchange slowly with the membrane bulk, due to electrostatic interactions with arginines at the membrane binding region. Notably, a phosphatidylglycerol residue involved in phosphatidylcholine replacement, observed during the simulation (Figure 7C), appears bound for a long time to the protein in a groove between helices Nα2 and Nα3. This region may play a key role in the acceptor substrate binding pathway, facilitating its recognition and interaction with the active site.

When the acceptor is placed in a similar position as the bound phosphatidylglycerol, the protein still tends to clamp the polyisoprenyl tail, surrounding the pyrophosphate region. Experimental studies by Barreras et al. showed that GumK remains active on shorter phytanyl tails with only four isoprenyl units, suggesting that substrate specificity is less dependent on tail length.[31] Our simulations align with this observation, showing that while the first three isoprenyl units remain stably bound, the rest of the molecule exhibits greater flexibility (Figure 7C). Additionally, Phe92 interacts with the first isoprenyl unit via π-stacking, reinforcing the importance of unsaturated lipidic chains in substrate binding. These findings indicate that GumK can accommodate shorter oligo-isoprenyl substrates, as long as they retain unsaturation, in a way that resembles the substrate tolerance observed in the GT1 glycosyltransferase MurG from *E. coli*.[37]

The trisaccharidic part of the acceptor substrate promotes enzyme opening by interacting with the salt bridge between Arg29 and Glu272 (Figure 7A). In contrast, phospholipids bound to the Nα2-Nα4 cleft (Figure 4C) do not induce significant opening, as they are too

short to reach the interdomain space.

These findings suggest some independence between the domains. Although phospholipid binding induces local changes at the Nα2-Nα4 cleft, only the acceptor substrate, interacting at the interdomain interface, triggers full opening.

Based on these arguments, we suggest that acceptor binding is the initial step in GumK opening, creating space for donor binding. This implies that state II identified in ClustENMD (Figure 2D), which is more likely when the domains lose contacts, is unfavored in the presence of the acceptor substrate, as it occupies the acceptor cleft, preventing donor domain interactions (Figure S3B). So the presence of the acceptor substrate would change the landscape of the conformational space in Figure 2D by reducing the population II.

This mechanism aligns with findings in GT4 glycosyltransferases, particularly PimA, which transfers mannose from Man-GDP to phosphatidyl-myo-inositol in Mycobacterium.[38] In PimA, the acceptor substrate induces a relaxed conformation, whereas donor binding stabilizes the enzyme in a closed state.[39] Similar substrate-dependent conformational changes occur in other members of the GT4 CAZy family.[40,41]

During the unbiased simulation, only one state, long 600ns over 3μs, can be considered closer to a bound state. The substrate is kept near the catalytic residues thanks to hydrogen bonds to the backbone, stacking interactions with Arg52 and Ile162, and electrostatic/hydrophobic interactions between the lipid tail and the amphiphilic acceptor domain clamp. The binding site constrains an extended conformation for the sugar part, suggesting that volume, regiochemistry, and stereochemistry influence specificity. The tail on the other side, like the UDP motif in the donor substrate, appears to guide and anchor binding.

These structural features suggest that GumK may tolerate various trisaccharides, as long as they fit the binding site. This aligns with experiments showing reduced activity on acetylated acceptor derivatives.[31]

The selectivity for the donor substrate is partially affected by the structure of the donor's binding site

It has been reported that GumK is highly selective for GlcA-UDP;[31] however, the direct experimental evidence of its donor interaction comes from two crystal structures (PDB 2Q6V and 3CV3), where only the UDP moiety is visible due to spontaneous hydrolysis of GlcA-UDP. Structural data show that UDP's β-phosphate interacts with Lys307 and Tyr292, and mutating these residues to Ala abolishes catalytic function.[12] However, the presence of the GlcA moiety could reorient the phosphate group relative to the position in the crystal structures, potentially altering lysine-phosphate electrostatic interactions.

Glycosyltransferases involved in xanthan gum biosynthesis, such as GumI (GT94) and GumH (GT4), utilize Man-GDP instead of GlcA-UDP.[42,43] The structural alignment shows that, unlike GumK, these enzymes lack Lys307 but contain a lysine-arginine pair on the opposite side (Figure 6, S16). In GT4 enzymes this Arg-Lys pair coordinates the diphosphate group and orients the donor substrate.[44,45] GumK lacks this motif, but Arg29 in the acceptor domain may partially compensate by interacting with phosphates or the carboxylate group. Additionally, the position corresponding to Lys307 contains a neutral residue in GumH- and GumI-like enzymes(Figure S16), suggesting that sugar-binding site polarity may influence selectivity.

MD simulations by Salinas et al.[18] did not detect an interaction between GlcA carboxylate and Lys307, although they observed an interaction with Arg29. However, mutating this residue into Ala did not affect enzyme activity.

HREX simulations, with restraints on the UDP moiety and a funnel potential restraint on the sugar diphosphate, revealed distinct GlcA-UDP binding modes (Figure 5). The main clusters show that GlcA-UDP constrains phosphate positioning toward Arg29, potentially stabilizing the interdomain salt bridge and promoting a closed conformation. Similar substrate-induced stabilization of closed states has been observed in other bacterial glycosyltransferases, where the nucleotide region regulates domain interactions and conformational

transitions.[44,45] By contrast, Glc-UDP is more flexible and lacks a direct C6-Lys307 interaction, likely reducing its ability to promote domain closure. These structural distinctions between an active substrate and a structurally similar but non-reactive ligand highlight the key determinants of substrate specificity in GumK.

Our HREX simulations and bioinformatic analysis suggest two hypotheses for GlcA-UDP specificity in GumK. The first proposes that the carboxylate-Lys307 interaction is part of the binding path, enhancing stability and helping to orient the diphosphate toward Arg29 to promote closure. This does not exclude Glc-UDP activity, but implies a higher K_M due to weaker stabilization. The second hypothesis suggests that this interaction is critical for catalysis, orienting the sugar ring and neutralizing a negative charge that could interfere with the nucleophilic attack (acceptor substrate). If correct, Glc-UDP would be a poor substrate or exhibit significantly lower catalytic efficiency (Figure 5). No *in vitro* assays have tested GumK's activity on alternative donor substrates, leaving its specificity unresolved. Experimental validation is needed to determine whether GumK can utilize Glc-UDP or is strictly limited to GlcA-UDP.

Conclusion

In this study, we computationally characterized GumK, a glycosyltransferase involved in xanthan gum biosynthesis. We explored its conformational landscape, substrate interactions, and membrane association. Our findings reveal a complex interplay between structural flexibility and membrane anchoring, which appears to be a shared feature among membrane-associated GTs, potentially representing a generalizable regulatory mechanism in this enzyme class.

We identified that GumK's acceptor substrate binds within a well-defined clamp in the acceptor domain, where it plays a critical role in modulating enzyme conformational dynamics. The acceptor binding enhances domain opening and destabilizes interdomain interactions,

whereas donor substrate binding induces a closing motion, promoting functional cycling. This substrate-driven conformational equilibrium provides a mechanistic framework for understanding how glycosyltransferases regulate catalysis through structural rearrangements. Additionally, our results suggest that GumK selectively accommodates acceptor substrates with lipid carriers limited to approximately three isoprenyl units, which may guide future studies aimed at designing artificial substrates for biochemical assays and crystallization efforts.

At the donor binding site, our simulations indicate that specificity for glucuronate-UDP arises from a conserved electrostatic environment that stabilizes the substrate and facilitates domain closure. This charge-based recognition mechanism differentiates GumK from homologous enzymes such as GumI and GumH, where variations in charge distribution correspond to distinct substrate preferences. These findings provide a broader perspective on the selectivity of GT-B glycosyltransferases, underscoring the role of electrostatic interactions and conformational plasticity in substrate recognition.

A key outcome of this study is the identification of functionally relevant residues that contribute to GumK's catalytic cycle and substrate binding. A detailed summary of these key residues and their proposed functional roles is provided in Table S1, which serves as a reference for future experimental validation and biochemical characterization.

From a computational standpoint, our study highlights the challenges in sampling the full conformational space of GumK due to its intrinsic flexibility and membrane interactions. Although principal component analysis provided meaningful pathways for describing domain motions, future enhanced sampling strategies could further elucidate how substrate binding influences enzyme dynamics. Additionally, the identification of key conformational states in our study suggests potential collective variables (CVs) for free-energy calculations aimed at quantifying GumK's conformational transitions.

Several limitations of the computational approaches still require experimental evaluations. Among these, the complexity of the conformational space of GumK and its substrates

makes the convergence of simulations challenging despite the several approximations and constraints introduced. Furthermore, the force field cannot fully describe the complex chemistry of active carbohydrates, limiting the reliability of a converged population, even if these were attainable.[46]

Our findings not only enhance our understanding of GumK's function and substrate specificity but also offer a framework for studying the broader GT-B enzyme family. The insights gained here provide a foundation for future experimental studies and enzyme engineering efforts, potentially leading to the development of glycosyltransferases with tailored substrate specificities.

Data availability

Data and scripts used for running and analyzing simulations, and for generating the figures are available from https://github.com/gcourtade/papers/tree/master/2025/GumK.

Acknowledgements

G.C. and D.L. were funded by the Novo Nordisk Foundation (grant number NNF18OC0032242).

References

(1) Gabius, H.-J. The sugar code: Why glycans are so important. *164*, 102–111.

(2) Ali, M. Y.; Liaqat, F.; Khazi, M. I.; Sethupathy, S.; Zhu, D. Utilization of glycosyltransferases as a seamless tool for synthesis and modification of the oligosaccharides-A review. *249*, 125916.

(3) Shtaiwi, A.; Khan, S. U.; Khedraoui, M.; Alaraj, M.; Samadi, A.; Chtita, S. A comprehensive computational study to explore promising natural bioactive compounds targeting glycosyltransferase MurG in Escherichia coli for potential drug development. *14*, 7098.

(4) He, B.; Bai, X.; Tan, Y.; Xie, W.; Feng, Y.; Yang, G.-Y. Glycosyltransferases: Mining, engineering and applications in biosynthesis of glycosylated plant natural products. *7*, 602–620.

(5) Yao, W.; Niu, T.; Pan, J.; Yang, X.; Huang, C.; Guan, H.; Yang, L.; Wang, Z.; Wang, R. Discovery and functional characterization of two novel glycosyltransferases associated with the biosynthesis of α-hederin in Dipsacus asperoides. *217*, 109273.

(6) Hassan, B. A.; Milicaj, J.; Ramirez-Mondragon, C. A.; Sham, Y. Y.; Taylor, E. A. Ligand-Induced Conformational and Dynamical Changes in a GT-B Glycosyltransferase: Molecular Dynamics Simulations of Heptosyltransferase I Complexes. *62*, 324–339.

(7) Liu, B.; Zhao, C.; Xiang, Q.; Zhao, N.; Luo, Y.; Bao, R. Structural and biochemical studies of the glycosyltransferase Bs-YjiC from Bacillus subtilis. *166*, 806–817.

(8) Yuan, Y.; Barrett, D.; Zhang, Y.; Kahne, D.; Sliz, P.; Walker, S. Crystal structure of a peptidoglycan glycosyltransferase suggests a model for processive glycan chain synthesis. *104*, 5348–5353.

(9) Mestrom, L.; Przypis, M.; Kowalczykiewicz, D.; Pollender, A.; Kumpf, A.; Marsden, S. R.; Bento, I.; Jarzębski, A. B.; Szymańska, K.; Chruściel, A.; Tischler, D.; Schoevaart, R.; Hanefeld, U.; Hagedoorn, P.-L. Leloir Glycosyltransferases in Applied Biocatalysis: A Multidisciplinary Approach. *20*, 5263.

(10) Ardèvol, A.; Rovira, C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from *ab Initio* Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *137*, 7528–7547.

(11) Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *42*, D490–D495.

(12) Barreras, M.; Salinas, S. R.; Abdian, P. L.; Kampel, M. A.; Ielpi, L. Structure and Mechanism of GumK, a Membrane-associated Glucuronosyltransferase*. *283*, 25027–25035.

(13) Chaturvedi, S.; Kulshrestha, S.; Bhardwaj, K.; Jangir, R. In *Microbial Polymers: Applications and Ecological Perspectives*; Vaishnav, A., Choudhary, D. K., Eds.; Springer, pp 87–107.

(14) Kumar, A.; Rao, K. M.; Han, S. S. Application of xanthan gum as polysaccharide in tissue engineering: A review. *180*, 128–144.

(15) Petri, D. F. S. Xanthan gum: A versatile biopolymer for biomedical and technological applications. *132*.

(16) Patel, J.; Maji, B.; Moorthy, N. S. H. N.; Maiti, S. Xanthan gum derivatives: review of synthesis, properties and diverse applications. *10*, 27103–27136.

(17) Chang, A.; Singh, S.; Phillips, G. N.; Thorson, J. S. Glycosyltransferase structural biology and its role in the design of catalysts for glycosylation. *22*, 800–808.

(18) Salinas, S.; Petruk, A.; Brukman, N.; Bianco, M.; Jacobs, M.; Marti, M.; Ielpi, L. Binding of the substrate UDP-glucuronic acid induces conformational changes in the xanthan gum glucuronosyltransferase. *29*, 197–207.

(19) Atilgan, A.; Durell, S.; Jernigan, R.; Demirel, M.; Keskin, O.; Bahar, I. Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. *80*, 505–515.

(20) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *78*, 1950–1958.

(21) Thomasen, F. E.; Skaalum, T.; Kumar, A.; Srinivasan, S.; Vanni, S.; Lindorff-Larsen, K. Rescaling protein-protein interactions improves Martini 3 for flexible proteins in solution. *15*, 6645.

(22) Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *596*, 583–589.

(23) Invernizzi, M.; Parrinello, M. Rethinking Metadynamics: From Bias Potentials to Probability Distributions. *The Journal of Physical Chemistry Letters* **2020**, *11*, 2731–2736.

(24) Lomize, M. A.; Pogozheva, I. D.; Joo, H.; Mosberg, H. I.; Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *40*, D370–D376.

(25) Palma, J.; Pierdominici-Sottile, G. On the Uses of PCA to Characterise Molecular Dynamics Simulations of Biological Macromolecules: Basics and Tips for an Effective Use. *24*, e202200491.

(26) Lange, O. F.; Grubmüller, H. Can Principal Components Yield a Dimension Reduced Description of Protein Dynamics on Long Time Scales? *110*, 22842–22852.

(27) Hess, B. Convergence of sampling in protein simulations. *65*, 031910.

(28) Romo, T. D.; Grossfield, A. Block Covariance Overlap Method and Convergence in Molecular Dynamics Simulation. *7*, 2464–2472.

(29) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *14*, 71–73.

(30) Bussi, G. Hamiltonian replica exchange in GROMACS: a flexible implementation. *112*, 379–384.

(31) Barreras, M. Functional characterization of GumK, a membrane-associated -glucuronosyltransferase from Xanthomonas campestris required for xanthan polysaccharide synthesis. *14*, 233–241.

(32) Das, N.; Sen, P. Macromolecular crowding: how shape and interaction affect the structure, function, conformational dynamics and relative domain movement of a multidomain protein. *24*, 14242–14256.

(33) Ge, C.; Georgiev, A.; Öhman, A.; Wieslander, A. k.; Kelly, A. A. Tryptophan Residues Promote Membrane Association for a Plant Lipid Glycosyltransferase Involved in Phosphate Stress. *286*, 6669–6684.

(34) Guerin, M. E.; Kordulakova, J.; Schaeffer, F.; Svetlikova, Z.; Buschiazzo, A.; Giganti, D.; Gicquel, B.; Mikusova, K.; Jackson, M.; Alzari, P. M. Molecular Recognition and Interfacial Catalysis by the Essential Phosphatidylinositol Mannosyltransferase PimA from Mycobacteria. *282*, 20705–20714.

(35) Unione, L.; Ortega, G.; Mallagaray, A.; Corzana, F.; Pérez-Castells, J.; Canales, A.; Jiménez-Barbero, J.; Millet, O. Unraveling the Conformational Landscape of Ligand Binding to Glucose/Galactose-Binding Protein by Paramagnetic NMR and MD Simulations. *11*, 2149–2157.

(36) Pang, Z.; Sokolov, M.; Kubař, T.; Elstner, M. Unravelling the mechanism of glucose binding in a protein-based fluorescence probe: molecular dynamics simulation with a tailor-made charge model. *24*, 2441–2453.

(37) Chen, L.; Men, H.; Ha, S.; Ye, X.-Y.; Brunner, L.; Hu, Y.; Walker, S. Intrinsic Lipid Preferences and Kinetic Mechanism of *Escherichia coli* MurG. *41*, 6824–6833.

(38) Boldrin, F. et al. The Phosphatidyl- *myo* -Inositol Mannosyltransferase PimA Is Essential for Mycobacterium tuberculosis Growth *In Vitro* and *In Vivo*. *196*, 3441–3451.

(39) Guerin, M. E.; Schaeffer, F.; Chaffotte, A.; Gest, P.; Giganti, D.; Korduláková, J.; Van Der Woerd, M.; Jackson, M.; Alzari, P. M. Substrate-induced Conformational Changes in the Essential Peripheral Membrane-associated Mannosyltransferase PimA from Mycobacteria. *284*, 21613–21625.

(40) Hassan, B. A.; Milicaj, J.; Tyson, M.; Karki, R.; Sham, Y. Y.; Frantom, P. A.; Taylor, E. A. In Vitro and In Silico Explorations of the Protein Conformational Changes of Corynebacterium glutamicum MshA, a Model Retaining GT-B Glycosyltransferase. *63*, 939–951.

(41) Huang, W.; Song, J.; Sun, T.; He, Y.; Li, X.; Deng, Z.; Long, F. Substrate binding and catalytic mechanism of the Se-glycosyltransferase SenB in the biosynthesis of selenoneine. *15*, 1659.

(42) Salinas, S. R.; Bianco, M. I.; Barreras, M.; Ielpi, L. Expression, purification and biochemical characterization of GumI, a monotopic membrane GDP-mannose:glycolipid 4-β-d-mannosyltransferase from Xanthomonas campestris pv. campestris. *21*, 903–913.

(43) Muniz, J. R. C.; Alves, C. A.; Pieri, C. D.; Beltramini, L. M.; Selistre-de Araújo, H. S.; Vettore, A. L.; Silva, F. R.; Arruda, P.; Garratt, R. C.; Oliva, G.; Souza, D. H. Overexpression, purification, biochemical characterization, and molecular modeling of recombinant GDP-mannosyltransferase (GumH) from Xylella fastidiosa. *315*, 485–492.

(44) Wang, X.-P.; Jiang, Y.-L.; Dai, Y.-N.; Cheng, W.; Chen, Y.; Zhou, C.-Z. Structural and enzymatic analyses of a glucosyltransferase Alr3699/HepE involved in *Anabaena* heterocyst envelop polysaccharide biosynthesis. *26*, 520–531.

(45) Shi, W.-W.; Jiang, Y.-L.; Zhu, F.; Yang, Y.-H.; Shao, Q.-Y.; Yang, H.-B.; Ren, Y.-M.; Wu, H.; Chen, Y.; Zhou, C.-Z. Structure of a Novel O-Linked N-Acetyl-d-glucosamine (O-GlcNAc) Transferase, GtfA, Reveals Insights into the Glycosylation of Pneumococcal Serine-rich Repeat Adhesins. *289*, 20898–20907.

(46) Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate–Protein Modeling. *7*, 3162–3180.

(47) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *1-2*, 19–25.

(48) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *185*, 604–613.

(49) Tiwary, P.; Parrinello, M. A Time-Independent Free Energy Estimator for Metadynamics. *119*, 736–742.

(50) Lee, J. et al. CHARMM-GUI *Membrane Builder* for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *15*, 775–786.

(51) Aktas, M.; Narberhaus, F. Unconventional membrane lipid biosynthesis in *X* *anthomonas campestris*. *17*, 3116–3124.

(52) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *31*, 1695–1697.

(53) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *52*, 7182–7190.

(54) Bussi, G. Hamiltonian replica exchange in GROMACS: a flexible implementation. *112*, 379–384.

(55) Consortium, T. U. UniProt: the universal protein knowledgebase in 2021. *49*, D480–D489.

(56) Ye, Y.; Godzik, A. Flexible structure alignment by chaining aligned fragment pairs allowing twists. *19*, ii246–ii255.

(57) Katoh, K.; Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *30*, 772–780.

(58) Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E. WebLogo: A Sequence Logo Generator: Figure 1. *14*, 1188–1190.