Article

Conserved Conformational Hierarchy across Functionally Divergent Glycosyltransferases of the GT-B Structural Superfamily as Determined from Microsecond Molecular Dynamics

Carlos A. Ramirez-Mondragon [1], Megin E. Nguyen [1], Jozafina Milicaj [2], Bakar A. Hassan [2], Frank J. Tucci [2], Ramaiah Muthyala [3], Jiali Gao [1,4,5], Erika A. Taylor [2,*] and Yuk Y. Sham [1,6,*]

[1] Bioinformatics and Computational Biology Program, University of Minnesota, Minneapolis, MN 55455, USA; rami0255@umn.edu (C.A.R.-M.); nguy1892@umn.edu (M.E.N.); gao@jialigao.org (J.G.)

[2] Department of Chemistry, Wesleyan University, Middletown, CT 06459, USA; jmilicaj@wesleyan.edu (J.M.); bhassan@wesleyan.edu (B.A.H.); Ftucci@wesleyan.edu (F.J.T.)

[3] Department of Experimental and Clinical Pharmacology, College Pharmacy, University of Minnesota, Minneapolis, MN 55455, USA; muthy003@umn.edu

[4] Department of Chemistry, University of Minnesota, Minneapolis, Minneapolis, MN 55455, USA

[5] Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen 518055, China

[6] Department of Integrative Biology and Physiology, Medical School, University of Minnesota, Minneapolis, MN 55455, USA

* Correspondence: eataylor@wesleyan.edu (E.A.T.); shamx002@umn.edu (Y.Y.S.);
  Tel.: +1-(860)-685-2739 (E.A.T.); +1-(612)-625-6255 (Y.Y.S.); Fax: +1-(860)-685-2211 (E.A.T.);
  +1-(612)-625-5149 (Y.Y.S.)

Abstract: It has long been understood that some proteins undergo conformational transitions en route to the Michaelis Complex to allow chemistry. Examination of crystal structures of glycosyltransferase enzymes in the GT-B structural class reveals that the presence of ligand in the active site triggers an open-to-closed conformation transition, necessary for their catalytic functions. Herein, we describe microsecond molecular dynamics simulations of two distantly related glycosyltransferases that are part of the GT-B structural superfamily, HepI and GtfA. Simulations were performed using the open and closed conformations of these unbound proteins, respectively, and we sought to identify the major dynamical modes and communication networks that interconnect the open and closed structures. We provide the first reported evidence within the scope of our simulation parameters that the interconversion between open and closed conformations is a hierarchical multistep process which can be a conserved feature of enzymes of the same structural superfamily. Each of these motions involves of a collection of smaller molecular reorientations distributed across both domains, highlighting the complexities of protein dynamic involved in the interconversion process. Additionally, dynamic cross-correlation analysis was employed to explore the potential effect of distal residues on the catalytic efficiency of HepI. Multiple distal nonionizable residues of the C-terminal domain exhibit motions anticorrelated to positively charged residues in the active site in the *N*-terminal domain involved in substrate binding. Mutations of these residues resulted in a reduction in negatively correlated motions and an altered enzymatic efficiency that is dominated by lower K_m values with k_{cat} effectively unchanged. The findings suggest that residues with opposing conformational motions involved in the opening and closing of the bidomain HepI protein can allosterically alter the population and conformation of the "closed" state, essential to the formation of the Michaelis complex. The stabilization effects of these mutations likely equally influence the energetics of both the ground state and the transition state of the catalytic reaction, leading to the unaltered k_{cat}. Our study provides new insights into the role of conformational dynamics in glycosyltransferase's function and new modality to modulate enzymatic efficiency.

Keywords: GT; glycosyltransferase; HepI; heptosyltransferase I; heptose; L-*glycero*-D-*manno*-heptose; ADP-heptose or ADPH; ADP-L-*glycero*-D-*manno*-heptose; LPS; lipopolysaccharide; ODLA; O-deacylated *E. coli* Kdo_2-lipid A; CD; circular dichroism; MD; molecular dynamics; PDB; Protein Data Bank; PCA; principal component analysis; DCC; dynamic cross-correlation

1. Introduction

Glycosylation is a highly regulated, ubiquitous biochemical process catalyzed by glycosyltransferase (GT, E.C. 2.4.x.y) enzymes. The implications of glycosylation in cellular processes are broad, as it modulates the structure, stability, and hence function of the target molecule. Although GTs constitute approximately 1 to 2% of the genomes that have been sequenced from species across living kingdoms [1], the details of the molecular mechanism of many GTs remain poorly understood. Significant research endeavors have identified and characterized key residues and regions linked to the enzymatic cycle in a diverse range of GTs [2–9]. These investigations provided a wealth of data, making it possible to develop an atomistic description of the molecular mechanism of GTs. An understanding of the functional dynamics of this important class of enzymes is paramount for inhibitor discovery, chimeric protein design, enzymatic control and regulation, among other studies that offer potentially beneficial clinical and industrial applications.

GTs mediate the transfer of a single sugar (monosaccharide) from an activated sugar complex (sugar donor) onto a specific acceptor substrate [10]. The identities of the substrates (both the sugar donor and acceptor) are GT-specific, with the pool of available sugar donors encompassing a myriad of monosaccharides and activation moiety conjugates (with the latter including nucleotide-monophosphate, nucleotide-diphosphate, lipid-phosphate, or unsubstituted-phosphate compounds) [10]. The sugar acceptor substrates range from single biomolecules to macromolecular complexes of varying compositions (e.g., monosaccharides, oligosaccharides, organic molecules, lipids, polypeptides, DNA) [10]. The resulting sugar conjugate, in most reported cases, serves as the preferred substrate for a subsequent GT in a series of glycosylation reactions that lead to the final multiglycosylated product [3–5,11].

The largest subset of identified GTs in the Carbohydrate-active enzyme (CAZy) database corresponds to the bidomain Leloir-GTs (monosaccharide-nucleotide dependent) [12]. The GTs in the CAZy database have been classified into 111 different enzyme families based upon their primary sequences, with individual domains of Leloir-GTs named according to their relative primary sequence arrangement (*N* and *C* domains), while their combined spatial configurations determine the structural superfamily (GT-A, GT-B, among others) [13]. Regardless of the structural superfamily classification, Leloir-GT structures depict the nucleotide binding Rossman fold (βαβ) motif as a conserved feature that comprises the core of all resolved domains. In GT-A proteins, two Rossman domains are aligned to create a seven-stranded β-sheet at the core of the proteins which undergoes modest structural rearrangements over the course of the transfer reaction, mostly the result of loop region movements [14,15]. GT-B glycosyltransferases such as Heptosyltransferase I (HepI) and TDP-*epi*-vancosaminyltransferase (GtfA) consist of topologically identical Rossman-like domains joined by an extended linker (loop-α-loop) region, also known as the spine region, that are positioned in a stacked configuration (Figure 1). This arrangement generates a central inter-domain cavity that serves as the binding site for both ligands and contains the enzyme's catalytic core. The binding position of each ligand in the GT-B cavity is domain-specific, as the *N* and *C* domains demonstrate selective binding affinity towards the acceptor substrate and sugar donor substrate, respectively, and the two domains undergo conformational transitions to bring the two domains proximal prior to catalysis [3–5]. The location of the ligands in the binding pocket further divides the cavity into catalytic (proximal to active site residues) and non-catalytic regions. GTs are further classified mechanistically as either configuration-retaining or configuration-inverting (Figure 1) [10]. The classification is given by comparing the stereochemistry of the newly formed glycosidic linkage of the product to that of the starting sugar donor substrate.
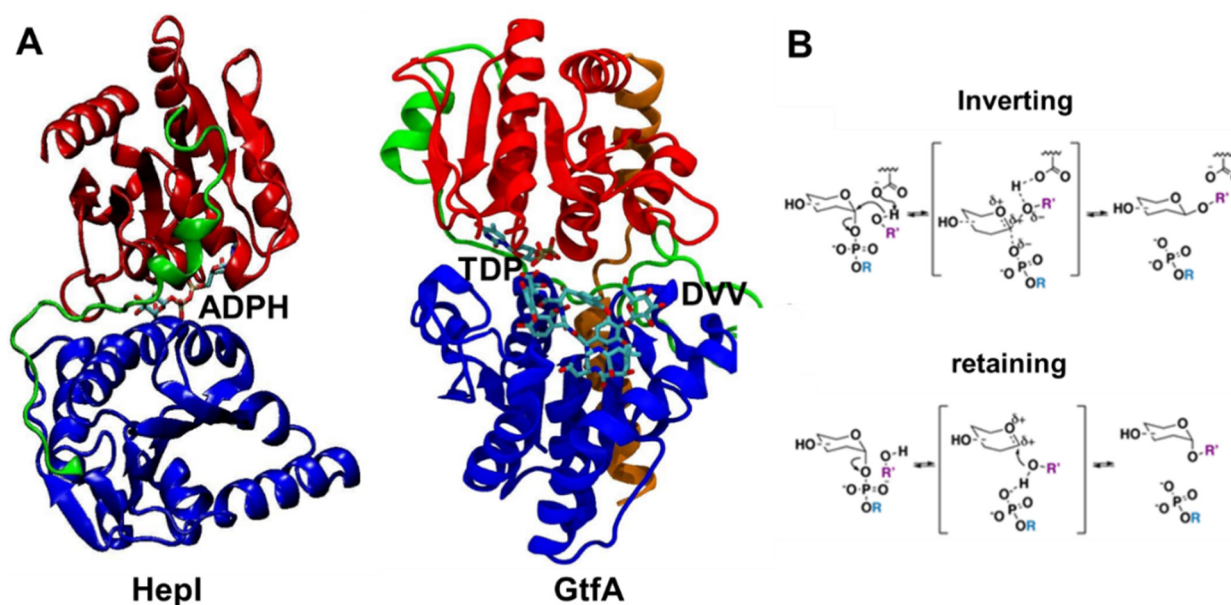
Figure 1. Crystal structures of ligand-bound glycosyltransferases: (**A**) HepI (PDB: 2H1H) and GtfA (PDB: 1PN3) colored for N domain (blue), *C* domain (red), linker (green), and spine (orange, only in GtfA) regions. HepI and GtfA cofactors, ADPH and TDP, respectively, and GtfA substrate DVV, are shown as ball and stick representation, which are located between *N* and *C* domains. Ligands are included to highlight the location of the binding catalytic site. (**B**) Inverting and retaining mechanisms by glycosyltransferases. Our simulations were performed without any substrates for the unbound HepI and GtfA.

As observed experimentally, the transition from a catalytically inactive (open) to active (closed ternary) state in GT-Bs is induced by substrate binding and is proposed to proceed via domain motions including hinge-bending centered on residues of the linker and/or the extended C-terminal spine region (if present), as well as twists that align the two substrates [4,5,11,16]. These motions are accompanied by different degrees of peripheral α-helix repositioning and loop restructuring with no significant alteration to the underlying β-sheet of the Rossman fold of each domain [4,5]. The end result of this dynamic process, described here as domain flexibility, positions the domains closer to one another, emphasizing pre-existing or creating new intra- and inter-domain contacts while bridging the distance between the enzymatic and ligand reactive centers [4,5]. Crystallographic examples of domain flexibility in GT-Bs include Gram-positive bacterial enzymes *N*-acetyl-glucosamine transferase (MshA, retaining, EC: 2.4.1.250, CAZy family GT4), which mediates the first step of mycothiol biosynthesis [4], and TDP-epi-vancosaminyltransferase (GtfA, inverting, EC: 2.4.1.311, CAZy family GT1), involved in biosynthesis of the natural glycopeptide antibiotic chloroeremomycin [5]. Further evidence of conformational changes in GT-Bs upon substrate binding comes from intrinsic tryptophan fluorescence (ITF) experiments of the Gram-negative bacterial enzyme Heptosyltransferase I (HepI, inverting, EC: 2.4.99.B6, CAZy family GT9) [17,18]. HepI catalyzes the addition of the first heptose monosaccharide to the nascent polysaccharide core of the outer membrane lipopolysaccharide (LPS) [19–22]. Binding of the acceptor substrate analog (O-deacylated *E. coli* Kdo_2-lipid A, ODLA) to HepI resulted in a spectral blue-shift of the enzyme's fluorescence profile. Such a shift is indicative of associated structural changes that desolvate one or more tryptophan residues. The extent and degree of the HepI motion, however, cannot be explicitly described due to the nature of the ensemble averaged fluorescence experiments that were performed [17,18].

Some molecular dynamics simulations have been performed to date on GTs; however, these simulations have yet to elucidate the conformational transitions necessary for GT-B proteins to undergo the open-to-closed structural transition. The majority of the simulations to date were performed using protein structures built from homology models as their structural starting point due to the lack of available crystal structures for the desired proteins. Several GT-As, including GT MG517 from *Mycoplasma genitalium* [23], and

LARGE1 from humans [24], together with two GT-Bs, GnT-V from humans [25] and HepIII from *Klebsiella pneumonia* [26], have been modeled and short molecular dynamics simulations have been performed to refine the structure and to corroborate or provide new insight for substrate/protein interactions. For those GTs that have crystal structures available, short (<300 ns) all atom or longer (10 µs) course grain molecular dynamic simulations have been performed. These simulations have largely focused upon substrate binding interactions (i.e., LgtC (GT-A), Gpgs (GT-A), and GumK (GT-B)) and in some cases, interactions of the protein with membranes (i.e., alMGS (GT-B) and PglH (GT-B) [2,27–30]. While multiple enzymes of both the GT-A and GT-B structural classes have been simulated, detailed analysis of the trajectories have only been performed in the GT-A structural scaffolds. Substrate binding and catalysis in GT-A, unlike GT-B, only involves rearrangement of loop regions, as the protein domains are continuous and largely unchanging over the course of the reaction.

We hypothesized that the evolutionary conserved ternary structure, chemical mechanism, and prerequisite domain flexibility required for function across inverting GT-Bs would result in conformational dynamics also being a conserved feature of these familial enzymes, despite low levels of sequence conservation (<30% sequence identity) between enzymes of different CAZy families. To test this, we characterize the type, degree, and order of the bound to unbound transition observed domain flexibility motions in HepI and GtfA as reported by microsecond molecular dynamics (MD) trajectories generated on the Anton1 supercomputer [31]. The advent of Anton1 has allowed for the atomistic examination of the dynamic behavior and associated properties of large biomolecular systems in explicit solvent at timescales that range from microseconds to sub-milliseconds [32–35]. The selection of the GT-B's GtfA and HepI for MD study pertains to: (1) the availability of high resolution bound-form structures with no unresolved backbone or C_α atom coordinates; (2) a shared inversion catalytic mechanism; (3) pairwise sequence identity of 26.7%; and (4) a common catalytic core, discovered during this study, which has been evolutionary conserved in both sequence and three-dimensional (3D) space despite the divergent function of each enzyme. Our recent experimental and computational study has shown nearby positively charged residues within the enzyme active site play an essential role in its conformational transitions and the binding of its negatively charged substrates [36]. To further explore how distal residues may play a similar role, distal nonionizable residues were specifically identified from our simulations situated in the C-terminal domain with motions anticorrelated to positively charged active site residues in the *N*-terminal domain and mutagenesis was carried out to examine their effect on catalytic efficiency.

2. Results

2.1. Differential Dynamic Flexibilities in Open and Closed Configurations in HepI and GtfA

The overall stability of tertiary structures of both GT-B systems was assessed by calculating the enzymes' global Radius of Gyration (C_αRGYR), C_α Root Mean Square Deviation (C_αRMSD) and C_α Root Mean Square Fluctuation (C_αRMSF). The computed C_αRMSD for the binary complex structure of HepI shows an oscillatory behavior that is roughly 1 µs time scale, whereas short-time fluctuation is significant in a range between 1.5 and 3.5 Å (Figure S1). We attribute this long-time scale process to the overall domain–domain bending motions in the open configuration of the full protein. Of the structural ensembles represented by the full HepI trajectory, 11.5% of the frames report enzyme structures with C_αRMSD values >3.0 Å, while 49.4% are between 2.0 and 3.0 Å. The C_αRMSD for ternary complex GtfA, which adopts a closed configuration with respect to the two structural domains, shows an initial increase, which is leveled to an average value of about 2.7 Å after 750 ns. We note that while it takes a relatively longer time for the ternary complex GtfA to reach equilibrium, it is interesting to point out that the long-time oscillating motion has been quenched in the closed configuration. A locally weighted scatterplot smoothing (lowess) curve fit to the C_αRMSD data shows the increase to be biphasic: an initial ~0.75 µs long phase of rapid, continuous rise in the C_αRMSD followed

by a final phase represented by steady fluctuations between 2.5 and 3.0 Å. Of the entire trajectory, 1.35% of the structures of GtfA report C_αRMSD values >3.0 Å, while 78.15% have values spanning from 2.0 to 3.0 Å.

The C_αRGYR data of HepI and GtfA (Figure 2A,D) demonstrate the change in compactness of each protein to highlight the opening of the closed structure GtfA structure as compared to the opened HepI. The plot continually exhibits different features for the binary open configuration and ternary closed form, respectively.
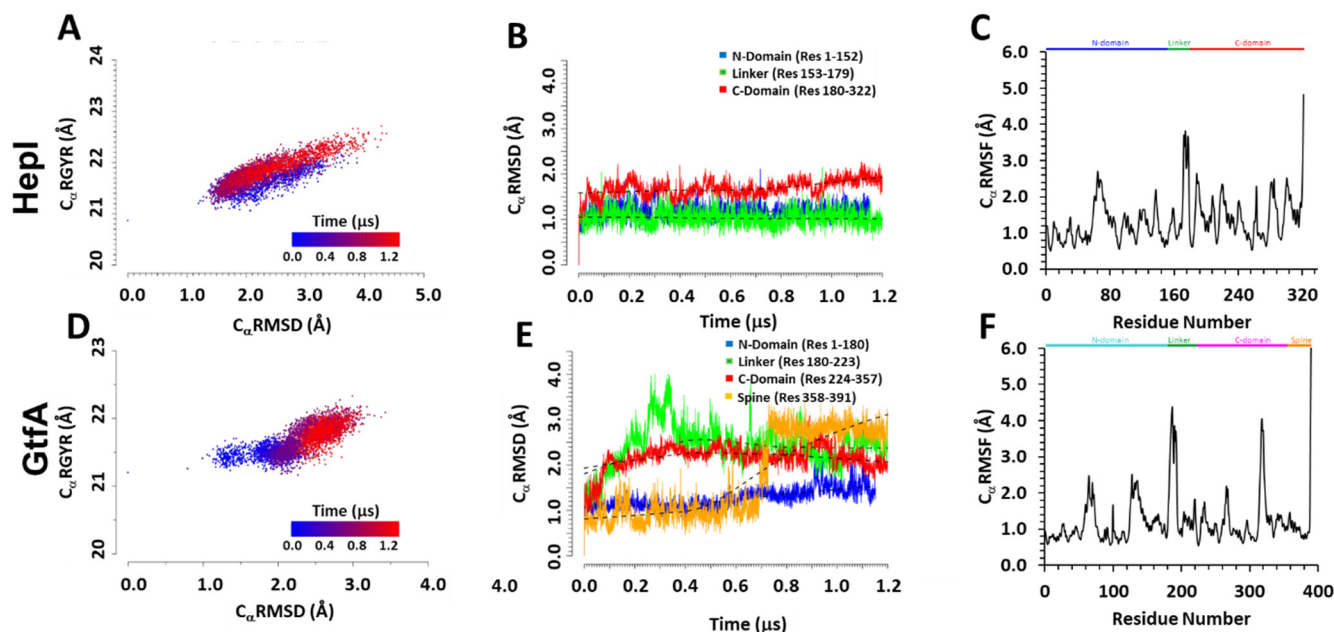

Figure 2. Computed radius of gyration (C_αRMSD), root-mean-square deviation (C_αRMSD), and root-mean-square fluctuation (C_αRMSF) of alpha-carbon atoms in HepI (**A**–**C**) and GtfA (**D**–**F**). (**A,D**) The correlation between C_α RGYR and C_αRMSD, highlighting the change in conformation population over the course of MD simulation from that determined at 0.0 (blue) to 1.2 μs (red); (**B,E**) the variations of C_αRMSD divided into different glycosyltransferase domains over the course of the entire simulation. Black dotted curves represent lowess curves for each domain. (**C,F**) The C_αRMSF plot with secondary structures, α-helix and β-sheet regions, highlighted in blue and yellow, respectively.

The correlation between computed radius of gyration and C_aRMSD for HepI is simulation time-independent, with the same trend in spread and variation, which is consistent with the above assignment of an overall structural oscillation. The positive correlation of expanding radius of gyration with increased C_aRMSD values is also consistent with the overall domain–domain breathing/bending motions. On the other hand, for the GtfA ternary complex, the radius of gyration is relatively independent with increasing C_aRMSD in GtfA fluctuations before 0.5 μs, but as the system reached equilibrium in the later stages of simulation, a positive correlation between the two properties was observed as it should be. Importantly, the fluctuations of C_αRGYR and C_αRMSD in HepI cover broader ranges and reach greater values in both quantities than those in GtfA, indicating that the latter closed configuration is relatively more compact than that of the open configuration of HepI. The data suggest that changes in the compactness of the two enzymes are connected to the underlying global structural changes as measured by the C_αRMSD. To further recapitulate the transition of the starting "closed" structure of GtfA to the "open" form, we evaluated the inter-domain distance between the *N*-and C-terminal domains and determined their probability density distribution over the course of the simulation (Supplemental Figure S6A,B). The observed bimodal distribution of GtfA showed a compact "closed" form with inter-domain distance of 9 Å and an enlarged "open" form with an inter-domain distance of 24 Å that coincides with the natural distribution of the starting "open" form of HepI.

2.2. Distinguishing Characters in Dynamic Fluctuations from Different Secondary Structures

To explore whether the data generated from the present simulations are representative of structural instability or physical microsecond-dynamic motions (e.g., loop restructuring, domain repositioning), the structural stability and dynamic rearrangements of the *N*, *L*inker, *C*, and *S*pine (GtfA only) domains were examined. To this end, we have determined (1) the C_αRMSD of each domain individually (per-domain: *N*, *L*, *C*, and *S* C_αRMSD) (Figure 2B,E), and (2) the C_αRMSD of the specific secondary structural elements in each domain ([Domain]$_{\alpha/\beta}$-ssC_αRMSD) (Figure S2). The data for these localized C_αRMSD analyses were calculated with a common reference to the corresponding crystal structure coordinates that are aligned by the protein regions of interest. The resulting data report exclusively on the changes to the structure of the chosen protein segment(s). As such, the per-domain C_αRMSD analysis functions as a macroscopic evaluation of the domain's total structure, while the per secondary structure type C_αRMSD serves as a microscopic assessment that focuses on changes to the underlying architecture of the domains.

The average HepI C_αRMSD and ssRMSD for each domain fall under 2.0 Å and display a near-linear trend, with both the *N* and *C* domain β-sheet cores consistently having lower value, less fluctuating ssRMSD data than the enveloping α-helices. Unlike the HepI *N* and *L* domains, the HepI *C* domain is in a higher RMSD value range and steadily increases after ~0.6–0.7 μs (Figure 2B). The pattern observed in the C_αRMSD beyond ~0.6 μs is both homologous to that of the C_β-ssRMSD and matches the rise in the C_α-ssRMSD during the same time period (Figure S2). The HepI ssRMSD of the β-sheet cores of *N* and *C* domains have lower averages and smaller fluctuations than the enveloping α-helices. The *N* and *L* domain RMSD are congruent and fit linearly to a lowess curve. On the other hand, with the exception of the *N* domain, increased variance is observed in the per-domain C_α-RMSD GtfA. As in HepI, the C_α-ssRMSD values all fall below 2.0 Å. The GtfA *N*-terminal domain C_α-RMSD has a more constant, near-linear trend that begins to increase at ~0.6 μs. This sudden increase concurs with the start of the inward displacement of the *C*-terminal end of third *N*-terminal β-strand (*N*-β3; res: 50–52) towards the more central *N*-β2 (res: 30–35), as reported in the clearly defined jump at ~0.6 μs in the N_β-ssRMSD. The start of a slight increase in the N_α-ssRMSD was also observed at ~0.4 μs. This corresponds to the outward motion of residues 134–149, which comprise the *N*-terminal segment of *N*-α5 (res: 134–159).

The dynamic fluctuations of the spine domain of GtfA show that an initially fluctuating section in the computed *S*-C_αRMSD initially is shifted to a tighter, slowly decreasing segment after approximately 0.7 μs (Figure 2E). Examination of the structures saved near the transition time clearly portrays the unwinding of the protein's last three C-terminal residues (res: 389–391 on *S*-αII); no other significant alterations were observed in the rest of the spine's backbone structure. GtfA *C* and *L* C_αRMSD result in a near equivalent logarithmic dataset, peaking at ~0.2 μs and ~0.3 μs, respectively, followed by a period of steady decline. The C_α-, C_β-, or l_α-ssC_αRMSD data do not detail any significant structural changes to account for the shape of the *C* and *L* C_αRMSD data (Figure 2B,E). With the exception of the global C_αRMSD, none of the domain or ss-RMSD analyses consider loop (lp) or 3_{10}-helix regions. From the C_α Root Mean Square Fluctuation (C_αRMSF), the largest time-averaged atomic fluctuations (C_αRMSF > 2.0 Å) in both enzymes primarily correspond to loop, 3_{10}, and adjacent N/C-terminal secondary structure residues (Figure 2C,F). In HepI, these areas are: 63–70 (*N*-α3 C-term), 65–67 (*N*-lp5), 68–70 (*N*-α4 N-term) 136–137 (*N*-lp10), 188–191 (*C*-lp1 C-cap), 189–191 (*C*-3_{10}1), 218–221 (*C*-lp2 C-cap), 219–221 (*C*-α2), 263 (*C*-α4 N-cap), 280–285 (*C*-lp7), 300–303 (*C*-3_{10}3), and 321–322 (*C*-α5 C-term). In GtfA, the residues are: 64–65 and 70 (*N*-lp2), 127–128 (*N*-3_{10}4), 183–193 (*L*-lp1), 266 and 268 (*C*-lp3), 315–322 (*C*-lp7), and 389–391 (*C*-α7 C-term). The locations of these high C_αRMSF regions on the enzymes' structures generally correspond to the side boundaries of the binding cavity on the *N* and *C* domains. The equivalence is solely spatial and does not always correlate to similar structural elements. A larger proportion of the high RMSF areas are positioned on the catalytic side of the cavity, primarily encompassing regions in the vicinity of the volume occupied (HepI) or predicted to be occupied (GtfA) by the sugar

moiety of the cofactor. β-sheets in both systems consistently have the lowest C_αRMSF values (<1.0 Å) and are data trough points. The α-helices demonstrate more fluctuation, with the high C_αRMSF points corresponding to the helical N/C-terminal end residues.

2.3. Conformation Distribution Due to Low-Frequency Large Amplitude Motions

Principal component analyses (PCAs) of the trajectories reveal that large structural fluctuations of both systems are not limited to loop and 3_{10}-helix movements (Figure 3).
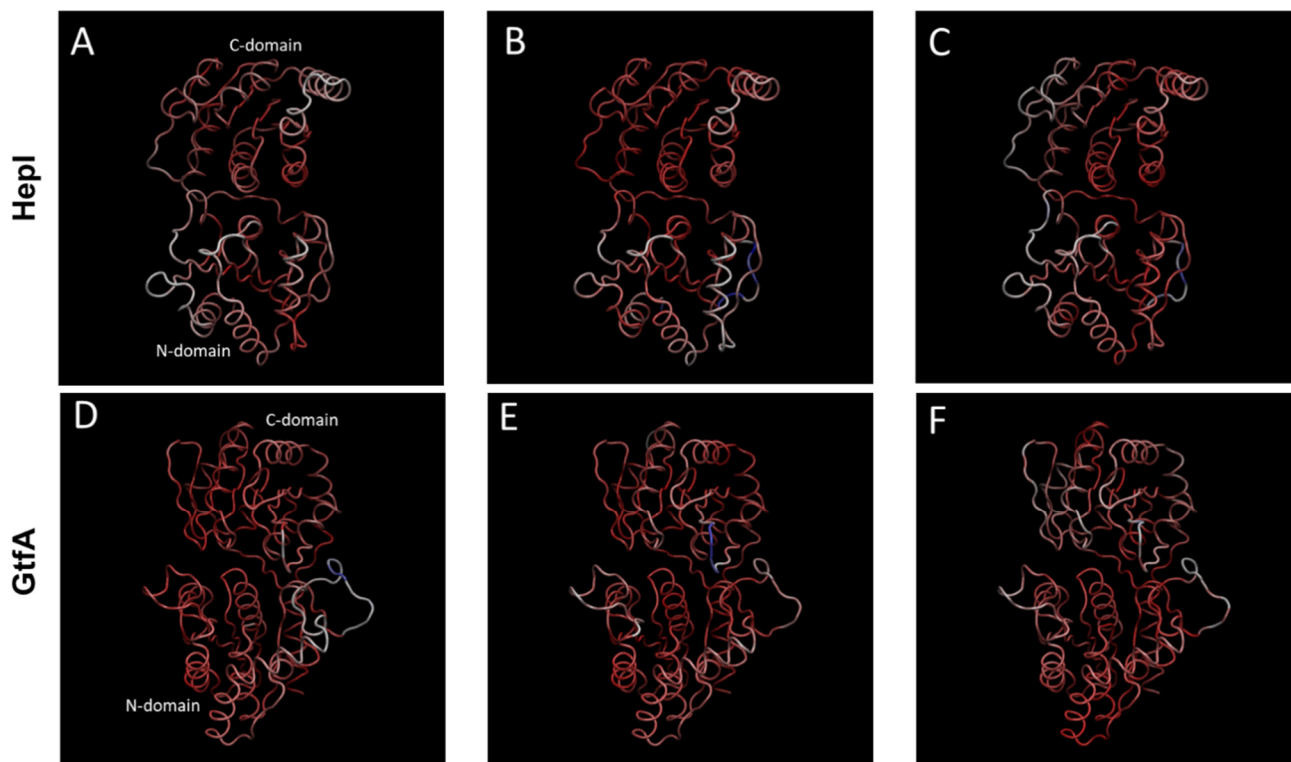

Figure 3. The first three principal components of quasiharmonic motions sampled by the 1.2 μs molecular dynamics trajectories for HepI (**A**–**C**) and for GtfA (**D**–**F**) (see Supplementary Figure S5 for animation). HepI and GtfA domains, C-domain, N-domain, and linker region are represented by white, blue, and green ribbons, respectively. These three large amplitude components account for 60.1 and 50.5% of dynamic motions, respectively, for HepI (top) and GtfA (bottom).

We found that the values of 55.9 and 50.5%, respectively, of the overall structural variance from the HepI and GtfA simulations, are associated with the first three principal components (Figure S3). To better describe the underlying structural motions featured by each PC, an animated trajectory that interpolates between the extremes points of each principal component through an even number of time steps was created (Figure S4). The extreme points represent a single trajectory frame whose protein backbone arrangement (C_α coordinates) is projected onto a PC, reaching the farthest stretch displacement. The motions in the top three motions for both enzyme systems correspond to domain–domain movements hinged about the linker/spine region with slight variances between each principal component.

PC1 of HepI (PC1$_H$, 35.3% structural variance), pertains to quasiharmonic motion, centered about the linker region, around residues 168–175 of the second loop of the linker region (*L*-lp2), with an antiparallel domain twist as the hinge moves. PC2$_H$ (11.1% structural variance) corresponds to an antiparallel rotation of the *C* and *N* domains, hinged around linker residues 164–166 and 168–174 of *L*-lp2, identified from nonstructured regions of C_αRMSF. PC3$_H$ (9.5% structural variance), involves hinge motion with no twisting of domains, anchored about linker (164–165 [*L*-lp1], 166–169 [*L*-α1]) and *C* domain (180–184 [*C*-β1], 262 [*C*-β4], 263 and 266–267 [*C*-α6], 282–284 and 288–291 [*C*-lp7]) residues.

To observe the interplay of the dominant principal component motions of HepI over the trajectory we analyzed the top three principal components by time-clusters plotted in 3D space with relative PC coordinates defined as (±PC1, ±PC2, ±PC3) (Figure 4).

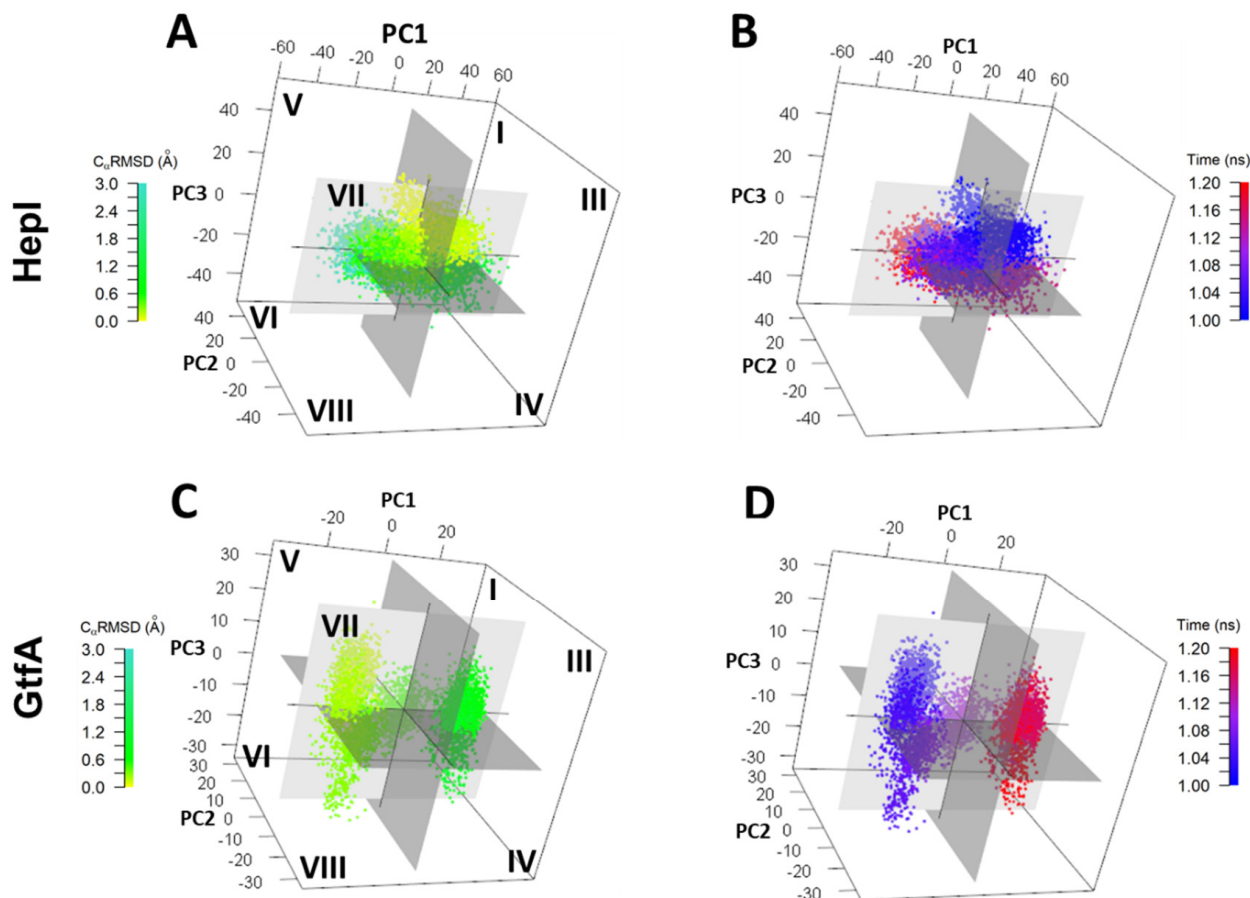**Figure 4.** Conformation distribution of the top three principal components (3D-PCA) plotted against (**B**,**D**) C_αRMSD (green to teal) and (**A**,**C**) time from 0.0 μs (blue) through 1.2 μs (red) for HepI (top) and Gtf A (bottom).

For HepI, although the global C_αRMSD is periodic, the dominant principal components are not necessarily harmonic. The conformational dynamics of HepI system begins at octant III and ends in octant VIII. HepI crosses between negative and positive PC1 space six times chronologically over 1.2 microseconds, whereas it only crosses zero eigenvalue in PC2 and 3 once, indicating the motion of PC1 occurs in a faster timescale than the motions of PC2 and 3, respectively. In 3D space, HepI occupies each octant I-VIII for 8.3, 14.0, 11.4, 13.6, 11.4, 12.3, 13.2, and 16.1% of the trajectory, respectively (Table S1). There are, however, single points or segments of the trajectory that deviate from the above described progression and are found in an octant other than that of the time-cluster it would be affiliated with. Overall, these low-frequency large amplitude motions are consistent with the distribution of radius of gyration and C_αRMSD shown in Figure 2A,E.

In PC1 of GtfA (PC1$_G$, 33.5% structural variance), the motion is akin to the movement of the *N* and *C* domains observed in PC3$_H$ (Figure S5C,D) with the loop moving perpendicularly from the N domain. The rotation is anchored on specific linker (177–179 [*N*-β6], 180–201 [*L*-lp1]) and spine residues (369–370 [*S*-3$_{10}$1], 373–374 and 379–380 [*S*-α2]). PC2 of GtfA shows a rotational displacement of the *C* and N domains, next to the linker (181–194, 196–197, and 201–202 of *L*-lp1) and spine residues (372–376 and 378–379 [*S*-α2]) (Figure S5E). The movement in PC2$_G$ is similar to the motion represented by PC2$_H$. PC3$_G$ displays the same domain motion as in PC2$_H$. The rotational motion is hinged on linker (184–194 and 196–197 [*L*-lp1]) and spine (374–379 [*S*-α2]) residues (Figure S5F). GtfA begins in octant V and ends in octant III transitioning over positive and negative eigenvalues once

for PC1, and two and three times for PC2 and PC3, respectively (Figure 4). GtfA occupies a larger proportion of the trajectory time in octants III and V (20.1 and 17.9%, respectively) and less time in octants I, IV, VI, and VII (7.1, 9.7, 6.1, 7.7%, respectively) (Table S1). By tracking the evolution of PC_G data clusters over time, the system's progression through 3D octants of PC space are: 5 to 7 to 8 to 4 to 1 to 3. Another commonality between the two enzymes is the presence of structures in octants other than that expected, given its corresponding time value, but matches with C_αRMSD and C_αRGYR. The key observations are HepI dynamics are relatively time-independent, with even distributions in time, suggesting it is dynamically oscillating about the average. However, for GtfA, there is an equilibration time (first 0.6–0.7 microseconds), followed by equilibrium distribution, which is narrower in range than HepI. HepI has a larger scale motion of two domains while GtfA is in closed configuration and fluctuates relatively less frequently.

2.4. Dynamics Restriction Affecting Catalytic Efficiency from Site-Directed Mutations

Examination of the dynamic cross-correlation matrices (Figure 5) shows that residues near each other in their primary sequence demonstrate strong positively correlated motions, as illustrated by the deep red coloring on and near the diagonal of the matrix for both HepI and GtfA. Additionally, off-diagonal regions within the same domain predominantly demonstrate strong positively correlated motions, while off-diagonal regions that represent the relationship between the two domains (i.e., HepI *N*-terminal domain residues 1–152 with C-terminal domain residues 180–322) demonstrate predominantly negatively correlated motions.
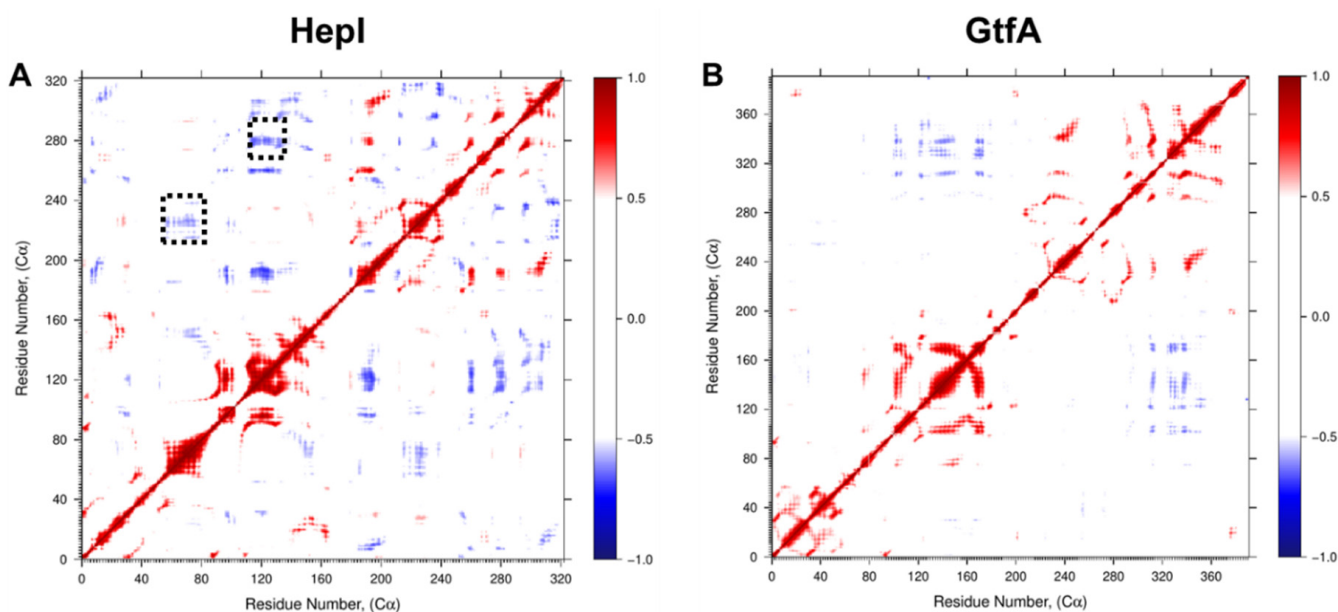

Figure 5. Dynamic C_α residue cross-correlation (DCC) of residue fluctuation over 1.2 μs MD simulation showing positively correlated (red), negatively correlated (blue), and uncorrelated (white) regions of (**A**) HepI and (**B**) GtfA. Dotted black boxes denote 60 and 120 s loop regions identified for mutagenesis.

Focusing on correlations to residues in HepI that are known to participate in conformational changes associated with binding of the sugar acceptor ligand (including Lys 64, Arg 63 and Arg 120) [17,18,36], we can identify a series of glycine and proline residues distant from the binding sites of either substrate (>25 Å away from the binding site of the sugar acceptor but within 15 Å of the sugar donor binding site) with strong negative correlated motions (Pro 216, Pro 240, Gly 280 and Gly 288) (Figure 6).
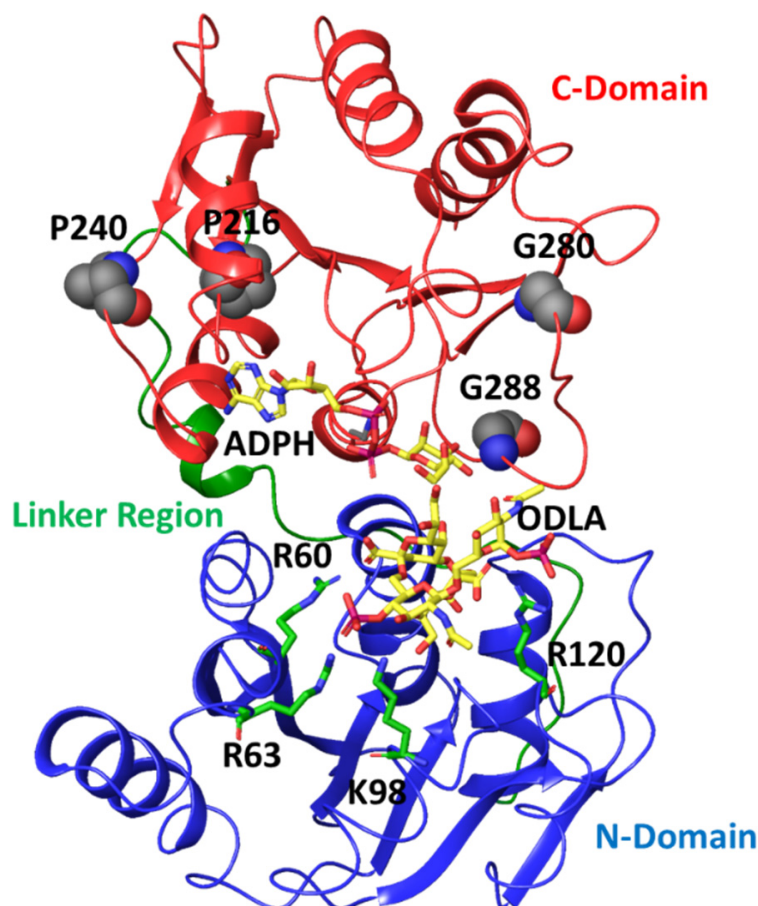
Figure 6. Crystal structure of HepI (PDB: 2H1H) with N domain, *C* domain and linker region colored in blue, red, green, respectively. The modeled ADPH and ODLA substrates (both in yellow) highlight the location of the catalytic site. The positively charged residues (R60, R63, K98 and R120) involved in FDLA binding are shown as stick representations. The four distal nonionizable residues (P216, P240, G280 and G288) located in the C-domain are shown in spacefill representation.

This lead us to hypothesize that these residues were necessary for sensing the binding of the sugar donor substrate (ADP-L-glycero-*D*-manno-heptose, ADPH), with Pro 216 and Pro 240 recognizing bind of the adenosine portion and Gly 280 and Gly 288 sensing the heptose portion. Due to the inherent flexibility of residues Gly 280 and Gly 288 (as well as the large degree of fluctuation observed for these residues over the course of the simulation; Figure 2C) and the loop in which they occupy lead to the hypothesis that these residues could play a role of differentiating between the bound sugar donor substrate and the nucleotide product, and critical for water exclusion from the active site. As part of our future directions, simulations of the liganded complex will be pursued to test this hypothesis. In the interim, experimental and computational studies were performed to investigate the role of these residues in catalysis. These glycine residues were individually mutated to prolines, and these proline residues were individually mutated to glycine, so as to yield mutant proteins with the greatest possible perturbation of their conformational flexibility at each position. The resultant proteins were then expressed, purified and kinetically characterized using high concentration of the sugar donor substrate (ADPH) while varying concentrations of the sugar acceptor substrate (ODLA which binds to the *N*-terminal residues that previously showed a strong degree of correlation to these Pro and Gly residues) to assess the impact of these residues on its binding (Table 1).

Table 1. Kinetic constants for the site-directed mutagenesis HepI.

HepI	k_{cat} (s^{-1})	Fold Change	K_M (M)	Fold Change	k_{cat}/k_M
Wild type	0.59 ± 0.09	-	8.5 ± 3.6	-	6.8×10^4
P216G	0.21 ± 0.02	$2.8 \downarrow$	1.6 ± 0.6	$5.6 \downarrow$	1.3×10^5
P240G	0.19 ± 0.01	$3.1 \downarrow$	0.9 ± 0.3	$10 \downarrow$	2.0×10^5
G280P	0.21 ± 0.02	$2.8 \downarrow$	1.5 ± 0.5	$6.0 \downarrow$	1.3×10^5
G288P	0.48 ± 0.07	$1.2 \downarrow$	10.2 ± 0.4	$1.1 \uparrow$	4.7×10^4

Each of the mutant HepI proteins exhibited modestly diminished k_{cat} values (with reductions of 3.1-fold or less as compared to wild-type HepI). Significantly, three of the mutants P216G, P240G and G280P all showed reduced K_m values (HepI•ODLA complex) to 1.6, 0.9 and 1.6 µM, respectively. The reduction in observed K_m in the mutant proteins leads to an overall enhanced catalytic efficiency (k_{cat}/K_m) relative to the wild-type enzyme HepI. To further establish how these mutants could have enhanced the observed catalytic efficiency, additional 0.5 ms MD simulations were carried out for P240G and G280P (the mutants with the largest experimental fold change of K_m). Since these residues exhibited motions that were strong and anticorrelated to the residues identified previously by us as essential to ligand binding [36], mutation that removes this anticorrelated interaction as observed in our DCC analyses (Figure 7) would be expected to result in direct enhancement to ligand binding affinity.
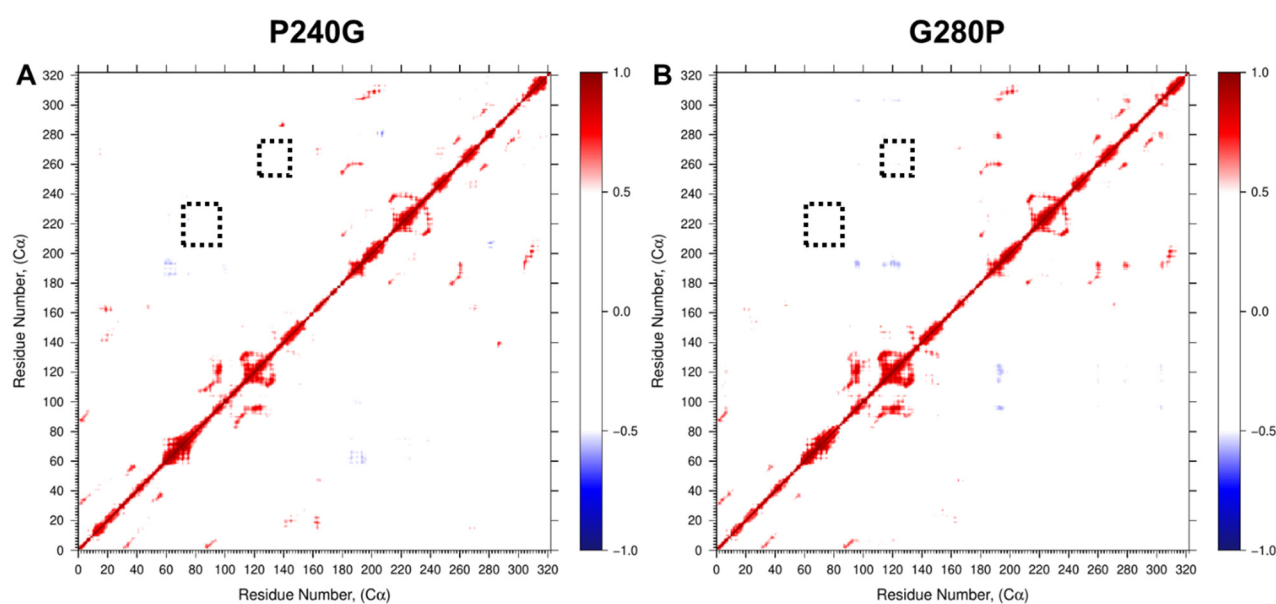

Figure 7. Dynamic cross-correlation (DCC) of residue fluctuation over 0.5 µs MD simulation of (**A**) P240G and (**B**) G280P HepI mutants showing loss of negatively correlated motion. Dotted black boxes denote 60 and 120 s loop regions identified for mutagenesis.

3. Discussion

We demonstrate that microsecond simulations of these GTs led to trajectories that encompass a robust sampling of conformational substates in both enzymes, regardless of whether the starting protein structure was from a ligand-bound or unbound state. We perform MD statistical analyses and employ 3D principal component analysis (PCA) to highlight structural retention (global and by domain) during the course of the simulation, as well as identification of the conformational substates and their transitions along the bound to unbound pathway. We show these substates to correspond to domain flexibility (domain repositioning) events that depict the same hierarchy in both enzymes regardless of MD starting state (GtfA: closed bound-ternary; HepI: open bound-binary). Our observation of a hierarchy of substates is akin to similar processes reported via in vitro and/or in

silico studies in myoglobin [37,38] and in adenylate kinase [39]. In those systems, the enzymes transition between substates with preferred directionality rather than by stochastic sampling. Whether the reported conformational directionality extends to familial enzymes of those systems remains to be fully determined.

For HepI, the predominance of per-domain RMSD values $\leq$ 2.0 Å highlight the structural permanence and stability of the domains. Structurally, the changes to the *C* domain starting at ~0.6 μs correspond to a slight outward displacement of the helical *N*-termini (*N*-cap and adjacent neighboring residues in both directions) of α-helices C-α2 (res: 219–231), C-α3 (res: 243–253), and C-α4 (res: 262–271). The *N*-termini of all three helices form part of the binding cavity. Specifically, the latter helix (C-α4) is positioned at the cavity midpoint, directly over the enzyme's catalytic core on *N*-β1, while the *N*-termini of the two former helices are situated on the noncatalytic side of the binding cavity. In the crystal structure (PDB: 2H1H), these three regions associate with the adenine base of the sugar cofactor. Residue E222 of C-α2 forms hydrogen bonds with the 2′ and 3′ hydroxyls of the ribose. Residue M242, which is the amino acid right before the start of the *N*-terminal end of C-α3, forms a hydrogen bond between its backbone carbonyl oxygen and the amine of the adenosine base. In C-α4, the *N*-cap T262 hydrogen bonds to an α-phosphate oxygen of the diphosphate group. The same small outward motion is also evident from ~0.6 μs onward in C-β5 (res: 275–279) and C-β6 (res: 294–298). Both β-sheets are located on the catalytic side of the binding cavity.

Ultimately, despite obtaining 1.2 μs MD simulations of both enzymes, we were only able to observe the partial interconversion between the open and closed conformations that provided a glimpses into the principle motions along the path for catalysis (Figures 3 and 4).Despite not realizing our full goal, the resulting trajectory data enabled us to (i) examine the root mean square fluctuation (RMSF) of residues, (ii) assess the dynamic cross-correlation (DCC) of the protein movements and (iii) identify the principal components of the motions of both HepI and GtfA. We noticed dramatic evolution of the C_αRMSD for GtfA (closed with substrates removed) and modest changes in HepI (open unbound), consistent with our hypothesis that removal of the substrates will return the enzyme into its unbound state with an "open" active site for substrate binding. The C_αRMSF analyses enabled the identification of dynamic loop regions of both proteins that were adjacent to the ligand binding sites that were suspected to be essential for ligand induced conformational changes. Subsequently, both the 60 and 120 s loops of the *N*-terminal domain (residues 1–152) of HepI were shown to be critical for substrate binding experimentally, which has also been subsequently confirmed by crystallography [2,3,40].

The DCC analyses (Figure 5) helped reveal that both proteins have a high degree of negatively correlated motions between the two domains, consistent with the enzyme practicing the open-to-closed motions even in the absence of sufficient time to perform the complete interconversion of those two states. To inquire whether dynamical motion in the unbound state plays any important role in catalysis, examination of the DCC matrices, specifically at the 60 and 120 s loops identified above, was carried out to reveal that there were numerous proline and glycine residues within the HepI C-terminal domain (residues 180–322; Table S1) that exhibited negatively correlated movements with these ligand binding residues. Since glycine and proline are known to exhibit dynamics that are distinct from other amino acids due to the lack of substitution at C-α (as with glycine) or the cyclical nature of the substitution at C-α (as with proline), and as others have noted their involvement in protein flexibility and protein folding [41,42], we mutated these residues to test the importance of their conformational variability on the overall behavior of the protein. We hypothesized that the rearrangement of the 60 and 120 s loops might communicate substrate occupancy of the *N*-terminal domain to the C-terminus (and visa-versa), and recent mutagenesis of these residues (where the glycines and prolines were interchanged; Table 1). Examination of the DCC for the P240G and G280P mutants reveals that the inter-domain negatively correlated motions observed in the WT for these residues are absent in the mutants (Figures 5 and 7), while the resulting proteins bind ~10 times more tightly to the sugar acceptor in the *N*-terminal domain. The most pronounced effect

was observed in the P240G mutant, which had nanomolar binding affinity for the sugar acceptor substrate, ODLA, and an overall catalytic efficiency of 2.0×10^5. P240G and the other mutated residues located in the opposing C-terminal domain are all >25 Å away from the ligand binding site for the sugar acceptor in the *N*-domain, yet their mutation to alter their conformational lability enhances the enzyme's ability to adopt the Michaelis complex with ODLA by an unknown mechanism. Our initial hypothesis was that these mutants yield a reduction in fluctuation in the 120, 240 and 280 s regions which may allow for a preordering of the enzyme into a conformation akin to that needed for the Michaelis complexes; however, this hypothesis requires further simulations of liganded complexes which is beyond the scope of this work. Interestingly, this finding in HepI supports the conclusions by Warshel and coworkers [43,44] that protein dynamics are primarily responsible for assisting the enzyme in adopting a state where chemistry can occur, and not in yielding a dynamical rate enhancement. Lastly, the analysis of principle components 1–3 of both HepI and GtfA (Figures 3 and 4) revealed that despite starting in different conformations (HepI open and GtfA closed), the two proteins maintain the similar primary principle components and have a shared conformational hierarchy. While further simulations at longer times and with liganded complexes are needed to fully understand the role of dynamics in catalyzing these glycosyltransfer reactions, the results to date suggest to us that the overall dynamics of these proteins are a conserved feature of the GT-B structural scaffold, even when the proteins are evolutionarily diverged with little sequence similarity.

4. Material and Methods

4.1. Multiple Sequence Alignment (MSA)

The MSA was generated by filtering the HepI (UniProtKB: P24173) and GtfA (UniProtKB: P96558) sequences on the European Bioinformatics Institutes (EMBLEBI) Clustal Omega MSA web server with default settings. The resulting MSA was visualized and sequence identity conservation was determined using the Unipro UGENE (v. 1.14.2) bioinformatics software.

4.2. Molecular Dynamics (MD) Simulations

Protein systems are built and equilibrated from their respective crystal structure coordinates (PDB Databank) using the CHARMM MD simulation program (v. c38a2) and the CHARMM c36 all-atom force field with CMAP corrections. All selected crystal structures met the following criteria: resolution ≤ 2.5 Å, R-Factor ≤ 0.2, and R-Free ≤ 0.30. Missing atoms and hydrogens were built using CHARMM. Protonation states of charged residues were determined by means of pK_a calculations (H++ server) and by examination of the surrounding hydrogen-bonding environment. For neutral histidine residues, proton positions (δ or ε) were assigned according to hydrogen-bond donating and accepting patterns. Unbound systems were set up by ligand exclusion from the respective open-binary (HepI, PDB: 2H1H: Chain A) and closed-ternary (GtfA, PDB:1PN3: Chain A) structures. Both systems were solvated in a cubic box of water molecules described by the TIP3P model; the MMTSB toolset was used. Crystal water molecules were retained, and depending on the system, sodium or chloride counter-ions were added to ensure electrostatic neutrality. The SHAKE algorithm [45] was applied to all bonds involving hydrogen atoms, which eliminates the vibrational frequencies of fast moving bonds consisting of hydrogen atoms to reduce the communication frequency of the computed forces in the parallelization of MD, thus, enabling the use of longer time steps to integrate the equations of motions [46,47]. Close contacts were removed by energy minimization using 25–50 steps of Steepest Descent minimization followed by 25–50 steps of adopted basis Newton-Raphson Method minimization, in which the coordinates of heavy atoms of the protein backbone, the ligand, and crystal water were held fixed. Careful consideration was placed on retention of the overall protein structure via RMSD of C_α atoms. Harmonic restraints on all protein heavy atoms (100 kcal/mol) and fixed constraints on ligand heavy atoms was maintained throughout

initial equilibration, which included: heating to 300 K with gradual scaling of temperature by 0.2 K every 100 femtoseconds (fs); application of the NPT ensemble (isobaric-isothermal) with Nosé-Hoover temperature control at 300 K for several ns until the reported pressure was consistently near 1 atm. Dynamics were propagated with the Leapfrog integrator with a time-step of 2 fs. Van der Waals forces were truncated with a switching function between 11 and 12 Å. Particle Mesh Ewald was used to model long-range electrostatic interactions with a real-space cutoff of 12 Å and a κ value of 0.333 Å; the grid space was set to be about 1 Å. Restraints on protein heavy atoms were then gradually released in a radial fashion (side-chains first, then backbone) from a defined point in the catalytic center using the NAMD simulation package (v. 2.8) for 10 ns. Final velocities and equilibrated coordinates were used to generate the necessary Anton1 input files (ark files) using the DESRES provided guesser scripts. Dynamics simulations were propagated using the RESPA integrator with a time-step of 2 fs. The NPT ensemble at 1 atm and 300 K was used with the Berendsen thermostat/barostat. Production was performed using 512 nodes of the Anton1 machine with structures saved in every 240 picoseconds during our allocation period. All simulation analyses were carried out using Bio3D [48] package in R. C_α root-mean-square deviation (C_αRMSD), C_α root-mean-square fluctuation (C_αRMSF), C_α radius of gyration (C_αRGYR) and inter-domain distances were evaluated to determine the conformational changes over the course of microsecond simulation. Dynamic cross-correlation (DCC) and principal component analysis (PCA) were performed to determine the intramolecular "cross-talks" between domains and principal motions observed throughout the simulation. The follow up 0.5 μs simulations of HepI P240G and G280P mutants, after our Anton1 allocation period ended, were performed with Gromacs simulation package with the same forcefield and MD simulation conditions on Linux workstations with NVIDIA GPU.

4.3. Site-Directed Mutagenesis of HepI from E. coli K12

All materials, solvents, competent cells were obtained as previously reported [17,18,36]. The Q5 High-Fidelity PCR kit from New England Biolabs was used according to its instructions to perform mutagenesis using *E. coli K-12* strain MB1760 960 bp HepI gene subcloned into pTOM-15b. New England Biolabs thermocycling conditions for routine PCR was used and the amplified DNA was then transformed into XL10-Gold competent cells, which were then incubated in a 5 mL LB/AMP overnight growth for plasmid DNA extraction by miniprep. Once the mutation was confirmed by sequence alignment with HepI wild-type DNA and the respective primers for each mutant, the purified plasmid was transformed into BL21-AI cells as per Agilent's instructions.

P216G Forward Primer: gcgcgccccacccaagtttaatccgtattcctg
P216G Reverse Primer: caggaatacggattaaacttgggtggggcgcgc
P240G Forward Primer: gttgaagtattgggcaagatgagtctggaaggcgttg
P240G Reverse Primer: caacgccttccagactcatcttgcccaatacttcaac
G280P Forward Primer: ggatagacccaatatcacggtttatccgccaaccgatccg
G280P Reverse Primer: cggatcggttggcggataaaccgtgatattgggtctatcc
G288P Forward Primer: ccgggattaattcctgggtatgggaagaatcagatggtatgtagggctcc
G288P Reverse Primer: ggagccctacataccatctgattcttcccatacccaggaattaatcccgg

4.4. Mutant Protein Expression and Purification

All buffers for SDS-PAGE gels and protein purification protocols are the same as for the wild-type HepI, as reported in the previous literature [29]. All mutants transformed into BL21-AI including WT were grown and expressed with the same conditions as described previously [17,18,36]. Briefly, cells were harvested, resuspended in 20 mL bind buffer per liter of growth (20 mM HEPES, 1 μM imidazole, 500 mM NaCl, pH = 7.5) on ice and lysozyme was added. The cells were then homogenized for 4–8 cycles and the lysate was clarified by centrifugation at 13,000 rpm for 1 h. The lysate was purified via FPLC with parameters described previously. Monitoring of 280 nm was used to identify fractions containing HepI and those fractions were then pooled and concentrated. The concentrated

protein was desalted into a storage buffer of 100 mM HEPES, 1 M KCl, pH 7.5, using a BioRad P6 polyacrylamide SEC column. The fractions containing protein were then concentrated and precipitated in an equal volume of saturated ammonium sulfate and stored in amber vials at 4 °C.

4.5. HepI Mutant Kinetics Characterization

HepI substrates were prepared as previously described [36] for kinetics analysis of HepI and its mutants. Steady-state kinetic data were collected using a UV–Vis spectrophotometric technique of a pyruvate kinase-lactate dehydrogenase coupled assay which monitors the reduction in NADH overtime upon release of ADP. Data collected were analyzed using the Michaelis–Menten approximation and reported in Table 1.

5. Conclusions

This is the first ever all atom, microsecond MD simulation of GT-B glycosytransferases, and the study suggests that C_αRMSD by itself is not a good test to determine structural stability at higher timeframes, which aligns with other work utilizing MD to simulate multidomain proteins connected by linker regions [49]. At these extended timescales, the simulations need to be examined for each individual domain and require domain-specific C_αRMSD analysis. The simulations of these two distantly related, but structurally conserved glycosyltransferases have given us insight into how the structural scaffold itself, and not the primary amino acid sequences, control the protein dynamic modes. The interconversion processes are a collection of smaller, localized motions, rather than a global, whole domain motion around a hinge region, highlighting the complexities of the protein dynamics. HepI and GtfA are evolutionarily diverged in their primary sequence yet maintain a conservation of protein dynamic modes required for substrate binding, an aspartate residue involved in catalysis and residues for stabilization of the conserved oxocarbenium ion intermediate. The conformational directionality and the chemical environment of the active site is a preserved feature of these structures and is encoded in the fold rather than the sequence, perhaps illustrating why evolution of this structural fold has been recapitulated to produce a series of enzymes catalyzing sugar transfer reactions. Additionally, the observation that mutagenesis of these distal nonionizable residues which exhibit anticorrelated motions to those residues that are important for substrate binding suggests that the enzyme dynamics for HepI are enabling the preordering of the enzyme into the Michaelis complex and not chemistry. While dynamic motions are often considered to be critical for traversing activation energy barriers, in this case this mutagenesis data are consistent with the theory that dynamic modes enable the electrostatic organization of the Michaelis complex and do not make a significant contribution to catalysis.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

References

1. Henrissat, B.; Surolia, A.; Stanley, P. A Genomic View of Glycobiology. In *Essentials of Glycobiology*; Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2009.

2. Albesa-Jové, D.; Romero-García, J.; Sancho-Vaello, E.; Contreras, F.-X.; Rodrigo-Unzueta, A.; Comino, N.; Carreras-González, A.; Arrasate, P.; Urresti, S.; Biarnés, X.; et al. Structural Snapshots and Loop Dynamics along the Catalytic Cycle of Glycosyltransferase GpgS. *Structure* **2017**, *25*, 1034–1044.e3. [CrossRef]

3. Grizot, S.; Salem, M.; Vongsouthi, V.; Durand, L.; Moreau, F.; Dohi, H.; Vincent, S.; Escaich, S.; Ducruix, A. Structure of the *Escherichia coli* Heptosyltransferase WaaC: Binary Complexes with ADP AND ADP-2-deoxy-2-fluoro Heptose. *J. Mol. Biol.* **2006**, *363*, 383–394. [CrossRef]

4. Vetting, M.W.; Frantom, P.A.; Blanchard, J.S. Structural and Enzymatic Analysis of MshA from *Corynebacterium glutamicum*: SUBSTRATE-ASSISTED CATALYSIS. *J. Biol. Chem.* **2008**, *283*, 15834–15844. [CrossRef] [PubMed]

5. Mulichak, A.M.; Losey, H.C.; Lu, W.; Wawrzak, Z.; Walsh, C.T.; Garavito, R.M. Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 9238–9243. [CrossRef] [PubMed]

6. Hurtado-Guerrero, R.; Zusman, T.; Pathak, S.; Ibrahim, A.F.M.; Shepherd, S.; Prescott, A.; Segal, G.; Van Aalten, D.M.F. Molecular mechanism of elongation factor 1A inhibition by a *Legionella pneumophila* glycosyltransferase. *Biochem. J.* **2010**, *426*, 281–292. [CrossRef] [PubMed]

7. Koç, C.; Gerlach, D.; Beck, S.; Peschel, A.; Xia, G.; Stehle, T. Structural and Enzymatic Analysis of TarM Glycosyltransferase from *Staphylococcus aureus* Reveals an Oligomeric Protein Specific for the Glycosylation of Wall Teichoic Acid. *J. Biol. Chem.* **2015**, *290*, 9874–9885. [CrossRef] [PubMed]

8. Schmidt, H.; Hansen, G.; Singh, S.; Hanuszkiewicz, A.; Lindner, B.; Fukase, K.; Woodard, R.W.; Holst, O.; Hilgenfeld, R.; Mamat, U.; et al. Structural and mechanistic analysis of the membrane-embedded glycosyltransferase WaaA required for lipopolysaccharide synthesis. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 6253–6258. [CrossRef] [PubMed]

9. Yao, Q.; Lu, Q.; Wan, X.; Song, F.; Xu, Y.; Hu, M.; Zamyatina, A.; Liu, X.; Huang, N.; Zhu, P.; et al. A structural mechanism for bacterial autotransporter glycosylation by a dodecameric heptosyltransferase family. *eLife* **2014**, *3*, e03714. [CrossRef] [PubMed]

10. Lairson, L.L.; Henrissat, B.; Davies, G.J.; Withers, S.G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **2008**, *77*, 521–555. [CrossRef]

11. Mulichak, A.M.; Losey, H.C.; Walsh, C.T.; Garavito, R.M. Structure of the UDP-Glucosyltransferase GtfB That Modifies the Heptapeptide Aglycone in the Biosynthesis of Vancomycin Group Antibiotics. *Structure* **2001**, *9*, 547–557. [CrossRef]

12. Coutinho, P.M.; Deleury, E.; Davies, G.J.; Henrissat, B. An Evolving Hierarchical Family Classification for Glycosyltransferases. *J. Mol. Biol.* **2003**, *328*, 307–317. [CrossRef]

13. Bourne, Y.; Henrissat, B. Glycoside hydrolases and glycosyltransferases: Families and functional modules. *Curr. Opin. Struct. Biol.* **2001**, *11*, 593–600. [CrossRef]

14. Breton, C.; Fournel-Gigleux, S.; Palcic, M.M. Recent structures, evolution and mechanisms of glycosyltransferases. *Curr. Opin. Struct. Biol.* **2012**, *22*, 540–549. [CrossRef]

15. Breton, C.; Šnajdrová, L.; Jeanneau, C.; Koča, J.; Imberty, A. Structures and mechanisms of glycosyltransferases. *Glycobiology* **2005**, *16*, 29R–37R. [CrossRef]

16. Vrielink, A.; Rüger, W.; Driessen, H.P.; Freemont, P.S. Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. *EMBO J.* **1994**, *13*, 3413–3422. [CrossRef]

17. Cote, J.M.; Ramirez-Mondragon, C.A.; Siegel, Z.S.; Czyzyk, D.J.; Gao, J.; Sham, Y.Y.; Mukerji, I.; Taylor, E.A. The Stories Tryptophans Tell: Exploring Protein Dynamics of Heptosyltransferase I from *Escherichia coli*. *Biochemistry* **2017**, *56*, 886–895. [CrossRef] [PubMed]

18. Czyzyk, D.J.; Sawant, S.S.; Ramirez-Mondragon, C.A.; Hingorani, M.M.; Taylor, E.A. *Escherichia coli* Heptosyltransferase I: Investigation of Protein Dynamics of a GT-B Structural Enzyme. *Biochemistry* **2013**, *52*, 5158–5160. [CrossRef] [PubMed]

19. Chen, L.; Coleman, W.G., Jr. Cloning and characterization of the *Escherichia coli* K-12 rfa-2 (rfaC) gene, a gene required for lipopolysaccharide inner core synthesis. *J. Bacteriol.* **1993**, *175*, 2534–2540. [CrossRef] [PubMed]

20. De Kievit, T.R.; Lam, J.S. Isolation and characterization of two genes, waaC (rfaC) and waaF (rfaF), involved in *Pseudomonas aeruginosa* serotype O5 inner-core biosynthesis. *J. Bacteriol.* **1997**, *179*, 3451–3457. [CrossRef] [PubMed]

21. Schnaitman, C.A.; Klena, J.D. Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol. Rev.* **1993**, *57*, 655–682. [CrossRef]

22. Sirisena, D.M.; Brozek, K.A.; MacLachlan, P.R.; Sanderson, K.E.; Raetz, C.R. The rfaC gene of *Salmonella typhimurium*. Cloning, sequencing, and enzymatic function in heptose transfer to lipopolysaccharide. *J. Biol. Chem.* **1992**, *267*, 18874–18884. [CrossRef]

23. Romero-García, J.; Francisco, C.; Biarnés, X.; Planas, A. Structure-Function Features of a Mycoplasma Glycolipid Synthase Derived from Structural Data Integration, Molecular Simulations, and Mutational Analysis. *PLoS ONE* **2013**, *8*, e81990. [CrossRef]

24. Righino, B.; Bozzi, M.; Pirolli, D.; Sciandra, F.; Bigotti, M.G.; Brancaccio, A.; De Rosa, M.C. Identification and Modeling of a GT-A Fold in the α-Dystroglycan Glycosylating Enzyme LARGE1. *J. Chem. Inf. Model.* **2020**, *60*, 3145–3156. [CrossRef]

25. Janos, P.; Kozmon, S.; Tvaroska, I.; Koca, J. Three-dimensional homology model of GlcNAc-TV glycosyltransferase. *Glycobiology* **2016**, *26*, 757–771. [CrossRef]

26. Panda, S.K.; Saxena, S.; Guruprasad, L. Homology modeling, docking and structure-based virtual screening for new inhibitor identification of *Klebsiella pneumoniae* heptosyltransferase-III. *J. Biomol. Struct. Dyn.* **2019**, *38*, 1887–1902. [CrossRef]

27. Ge, C.; Gómez-Llobregat, J.; Skwark, M.J.; Ruysschaert, J.-M.; Wieslander, Å.; Lindén, M. Membrane remodeling capacity of a vesicle-inducing glycosyltransferase. *FEBS J.* **2014**, *281*, 3667–3684. [CrossRef] [PubMed]

28. Ramírez, A.S.; Boilevin, J.; Mehdipour, A.R.; Hummer, G.; Darbre, T.; Reymond, J.-L.; Locher, K.P. Structural basis of the molecular ruler mechanism of a bacterial glycosyltransferase. *Nat. Commun.* **2018**, *9*, 445. [CrossRef] [PubMed]

29. 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. *Protein Eng. Des. Sel.* **2016**, *29*, 197–207. [CrossRef] [PubMed]

30. Šnajdrová, L.; Kulhánek, P.; Imberty, A.; Koča, J. Molecular dynamics simulations of glycosyltransferase LgtC. *Carbohydr. Res.* **2004**, *339*, 995–1006. [CrossRef]

31. Shaw, D.E.; Dror, R.O.; Salmon, J.K.; Grossman, J.P.; Mackenzie, K.M.; Bank, J.A.; Young, C.; Deneroff, M.M.; Batson, B.; Bowers, K.J.; et al. Millisecond-scale molecular dynamics simulations on Anton. In Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis, Portland, OR, USA, 14–20 November 2009; ACM: New York, NY, USA, 2009; pp. 1–11.

32. Gaieb, Z.; Morikis, D. Detection of Side Chain Rearrangements Mediating the Motions of Transmembrane Helices in Molecular Dynamics Simulations of G Protein-Coupled Receptors. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 131–137. [CrossRef]

33. Gur, M.; Zomot, E.; Bahar, I. Global motions exhibited by proteins in micro- to milliseconds simulations concur with anisotropic network model predictions. *J. Chem. Phys.* **2013**, *139*, 121912. [CrossRef]

34. Hu, X.; Hong, L.; Smith, M.D.; Neusius, T.; Cheng, X.; Smith, J.C. The dynamics of single protein molecules is non-equilibrium and self-similar over thirteen decades in time. *Nat. Phys.* **2016**, *12*, 171–174. [CrossRef]

35. Shaw, D.E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R.O.; Eastwood, M.P.; Bank, J.A.; Jumper, J.M.; Salmon, J.K.; Shan, Y.; et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341–346. [CrossRef]

36. Cote, J.M.; Hecht, C.J.S.; Patel, K.R.; Ramirez-Mondragon, C.A.; Sham, Y.Y.; Taylor, E.A. Opposites Attract: *Escherichia coli* Heptosyltransferase I Conformational Changes Induced by Interactions between the Substrate and Positively Charged Residues. *Biochemistry* **2020**, *59*, 3135–3147. [CrossRef]

37. Bourgeois, D.; Vallone, B.; Schotte, F.; Arcovito, A.; Miele, A.E.; Sciara, G.; Wulff, M.; Anfinrud, P.; Brunori, M. Complex Landscape of Protein Structural Dynamics Unveiled by Nanosecond Laue Crystallography. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8704–8709. [CrossRef]

38. Frauenfelder, H.; Sligar, S.G.; Wolynes, P.G. The energy landscapes and motions of proteins. *Science* **1991**, *254*, 1598–1603. [CrossRef] [PubMed]

39. Henzler-Wildman, K.A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M.A.; Petsko, G.A.; Karplus, M.; et al. Intrinsic motions along an enzymatic reaction trajectory. *Nature* **2007**, *450*, 838–844. [CrossRef] [PubMed]

40. Blaukopf, M.; Worrall, L.; Kosma, P.; Strynadka, N.C.J.; Withers, S.G. Insights into Heptosyltransferase I Catalysis and Inhibition through the Structure of Its Ternary Complex. *Structure* **2018**, *26*, 1399–1407.e5. [CrossRef] [PubMed]

41. Jacob, J.; Duclohier, H.; Cafiso, D.S. The Role of Proline and Glycine in Determining the Backbone Flexibility of a Channel-Forming Peptide. *Biochem. J.* **1999**, *76*, 1367–1376. [CrossRef]

42. Krieger, F.; Möglich, A.; Kiefhaber, T. Effect of Proline and Glycine Residues on Dynamics and Barriers of Loop Formation in Polypeptide Chains. *J. Am. Chem. Soc.* **2005**, *127*, 3346–3352. [CrossRef]

43. Warshel, A. Electrostatic Origin of the Catalytic Power of Enzymes and the Role of Preorganized Active Sites. *J. Biol. Chem.* **1998**, *273*, 27035–27038. [CrossRef]

44. Olsson, M.H.M.; Parson, W.W.; Warshel, A. Dynamical Contributions to Enzyme Catalysis: Critical Tests of A Popular Hypothesis. *Chem. Rev.* **2006**, *106*, 1737–1756. [CrossRef]

45. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341. [CrossRef]

46. Elber, R.; Ruymgaart, A.P.; Hess, B. SHAKE parallelization. *Eur. Phys. J. Spéc. Top.* **2011**, *200*, 211–223. [CrossRef] [PubMed]

47. Phillips, J.C.; Hardy, D.J.; Maia, J.D.C.; Stone, J.E.; Ribeiro, J.V.; Bernardi, R.C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130. [CrossRef] [PubMed]

48. Grant, B.J.; Rodrigues, A.P.; ElSawy, K.M.; McCammon, J.A.; Caves, L.S. Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics* **2006**, *22*, 2695–2696. [CrossRef] [PubMed]

49. Zhang, Y.; Wen, B.; Peng, J.; Zuo, X.; Gong, Q.; Zhang, Z. Determining structural ensembles of flexible multi-domain proteins using small-angle X-ray scattering and molecular dynamics simulations. *Protein Cell* **2015**, *6*, 619–623. [CrossRef] [PubMed]