ABSTRACT: We applied the transition path sampling (TPS) method to study the translocation step of the catalytic mechanism of galactofuranosyl transferase 2 (GlfT2). Using TPS in the field of enzymatic reactions is still relatively rare, and we show its effectiveness on this enzymatic system. We decipher an unknown mechanism of the translocation step and, thus, provide a complete understanding of the catalytic mechanism of GlfT2 at the atomistic level. The GlfT2 enzyme is involved in the formation of the mycobacterial cell wall and transfers galactofuranose (Galf) from UDP-Galf onto a growing acceptor Galf chain. The biosynthesis of the galactan chain is accomplished in a processive manner, with the growing acceptor substrate remaining bound to GlfT2. The glycosidic bond formed by GlfT2 between the two Galf residues alternates between β-(1→6) and β-(1→5) linkages. The translocation of the growing galactan between individual additions of Galf residues is crucial for the function of GlfT2. Analysis of unbiased trajectory ensembles revealed that the translocation proceeds differently depending on the glycosidic linkage between the last two Galf residues. We also showed that the protonation state of the catalytic residue Asp372 significantly influences the translocation. Approximate transition state structures and potential energy reaction barriers of the translocation process were determined. The calculated potential reaction barriers in the range of 6−14 kcal/mol show that the translocation process is not the rate-limiting step in galactan biosynthesis.

INTRODUCTION

Galactofuranosyl transferase 2 (GlfT2) is a bacterial glycosyltransferase from Mycobacterium tuberculosis. The enzyme catalyzes the transfer of galactofuranose (Galf) from a donor substrate, UDP-Galf, onto a growing galactan chain formed from Galf units. The structure of GlfT2 with both donor and acceptor substrates is shown in Figure 1. GlfT2 is responsible for the galactan biosynthesis of the bacterial cell wall, which is essential for survival. Compounds that can inhibit galactan biosynthesis are therefore potential therapeutics for tuberculosis. Knowledge of the catalytic mechanism of GlfT2 is a prerequisite for the rational design of such compounds and has been of increasing focus recently.1−6 The enzyme is a bifunctional glycosyltransferase, which catalyzes the formation of alternate β-(1→6) and β-(1→5) linkages between Galf residues. GlfT2 is the processive enzyme with the growing glycan chain that remains bound through multiple Galf transfers. It has been suggested3,6 that the reaction proceeds at a single catalytic center. The reaction is terminated once the Galf chain reaches 25−30 units, and then galactan dissociates from GlfT2.

The GlfT2 processive cycle can be described as a sequence of three elementary steps. In the first chemical step, GlfT2 catalyzes the formation of a new glycosidic (β-(1→6) or β-(1→5)) linkage. The chemical step is followed by a translocation step, which reorients the nonreducing end of the assembled...
galactan to be available for the second chemical step, the formation of another glycosidic (β-(1–5) or β-(1–6)) linkage. This cycle is illustrated in Figure 2.

We have recently investigated both chemical reactions leading to the β-(1–6) and β-(1–5) linkages, respectively. A quantum mechanics/molecular mechanics (QM/MM) molecular dynamics study described the chemical steps of the catalytic mechanism at the atomistic level and revealed that β-(1–6) and β-(1–5) proceed via the same S_n2-like reaction mechanism with a dissociative oxocarbenium transition state. The calculated reaction barrier was found to be similar for both reactions. It has been postulated that the ability of the catalytic Asp372 to form a functioning orientation with the O6H and O5H hydroxyl groups is crucial for the bifunctional activity of GlfT2. However, we still lack atomistic details of this process, which would provide insight into the mechanism of the formation of alternating glycosidic linkages. Key questions are how the nucleophilic hydroxyl groups (O6H or O5H) switch between the two distinct catalytically productive orientations in a single catalytic site and how this translocation step affects the reaction rate.

In this study, we apply the transition path sampling (TPS) method to address the role of the translocation step in the entire catalytic process. The use of TPS is especially suitable for the examination of the translocation step since the geometrical features of the starting structure, as well as the final structure of the translocation step, are known, but the translocation mechanism is unknown. This study represents a continuation of our recent QM/MM molecular dynamics investigation of the chemical steps of the GlfT2 reaction mechanism. Two different acceptor substrates were used in our TPS simulations (Figure 3). The first substrate represented the acceptor substrate after a β-(1–6) reaction (labeled Acc1), and the second represented the acceptor substrate after a β-(1–5) reaction (labeled Acc2). The translocation step reorients the Acc1 and Acc2 substrates to the proper position for the β-(1–5) and β-(1–6) reactions, respectively. Each acceptor substrate was placed in three different enzymatic environments of the active site. In the first, the catalytic Asp372 was protonated and UDP (charge −3) was deprotonated (labeled AspH); in the second, UDP (charge −2) was protonated and Asp372 was deprotonated (labeled UDPH); and in the third, UDP was removed from the active site (labeled noUDP). Thus, in total, we simulated six different systems. TPS results offer an insight into the translocation step at the atomistic level. Also, our study provides transition state structures and estimates reaction barriers for the translocation step.

### RESULTS AND DISCUSSION

**Spring Shooting Statistics.** We carried out TPS simulations using the spring shooting algorithm for the translocation between the START and END state for the following six systems: Acc1-AspH, Acc1-UDPh, Acc1-noUDP, Acc2-AspH, Acc2-UDPh, and Acc2-noUDP. The studied systems were constructed to explore the role of the substrate structure in the translocation as well as to study the effect of different protonation states and the absence of UDP on the translocation process. Thus, the studied complexes differ in the acceptor substrate bound to GlfT2 and in the environment of the active center. Acc1 and Acc2 denote a system that, after the translocation step, is oriented ready for the β-(1–5) and β-(1–6) reactions, respectively. AspH and UDPh indicate whether the proton from the donor nucleophilic hydroxyl group is on the catalytic acid Asp372 or the UDP after the transfer reaction. The label Acc2-AspH, for example, describes a system in which the acceptor substrate is Acc2, and the catalytic acid Asp372 is in the protonated state (with charge −1) while UDP is in the deprotonated state (with charge −3). Close-ups of Acc1-AspH and Acc2-AspH systems, together with the four collective variables (CVs) used in TPS simulations, are shown in Figure 4. The CVs describe the orientation of the nonreducing terminal Galβ residue with respect to Asp371, catalytic Asp372, and side binding Glu300. The reducing end of the acceptor tetrasaccharides was not monitored since...
epitope mapping of known trisaccharide substrates indicated that the interactions of the reducing end with protein are smaller compared to the other end.\(^7\)

Initial TPS simulations were performed using the standard flexible length one-way shooting with a uniform selector. The simulations were initiated from trajectories obtained by RMSD driving as detailed in the Methods section. Simulation details and the results of the uniform one-way shooting are shown in the SI. This approach has proven to be ineffective for the system under study. The acceptance rate was low, and the decorrelation of paths was unsatisfactory. For this reason, the spring shooting algorithm was chosen for producing TPS simulations. The spring shooting algorithm is expected to be suitable for a biomolecular system with asymmetric barriers.\(^10\)

The final trajectory from the uniform one-way shooting attempt was used as an initial trajectory for the spring shooting. The exceptions are the Acc1-noUDP and Acc2-noUDP systems, which were initiated directly from the RMSD driving trajectory.

The path tree diagrams of the ensemble evolution of four systems are shown in the SI (Figures S5−S8). Acceptance rate statistics are shown in Table 1, and the selected path length histograms are shown in Figure 5. For the Acc1 systems, the acceptance rate of systems with protonated catalytic Asp372 (Acc1-AspH) is highest at 45%. The acceptance rate of Acc1-UDPh systems is lower at 33%, and for the Acc1-noUDP, the acceptance rate is 27%. For the Acc2 systems, the acceptance rate for the Acc2-AspH system of 43% is similar to the Acc1-AspH case.

The systems without UDP (noUDP) and with protonated UDP (UDPh) had the lowest number of generated trajectories among simulations with both acceptor substrates. For the Acc1-UDPh system, nine decorrelated trajectories, and for the Acc1-noUDP system, 14 decorrelated trajectories were obtained. In the case of the Acc2 system, results are dramatically different. An analysis of these simulations revealed that in the Acc2-UDPh system, none of the forward shooting events reached an endpoint during the five ns long simulation and were therefore not accepted. In the case of Acc2-noUDP, only one of the forward shooting events was accepted. As a result, the TPS simulation of the Acc2-UDPh system was
unable to generate a single reactive trajectory of the translocation process and the Acc2-noUDP system generated only two trajectories. Thus, these two systems do not provide a sufficient amount of productive translocations and were therefore excluded from further analyses. The low frequency of dihedral angle motions and buffeting by the enzyme environment in the translocation process of tetrasaccharide requires a relatively long time for initiated trajectories. The shooting algorithm for generating trajectories might be ineffective in such cases.\(^9\) We suspect that the Acc2-noUDP and Acc2-UDPh systems represent such cases. Moreover, a comparison of the starting structures of Acc1 and Acc2 systems revealed that their interactions with the enzyme differ. In the Acc2 system, the O3H hydroxyl group of the second-to-last Galf residue of the tetrasaccharide interacts with Glu300. This interaction is not seen in the Acc1 systems (Figure 4). This indicates that the O3H−Glu300 interaction hinders the translocation process in the Acc2-noUDP and Acc2-UDPh systems and might also contribute to the lower rate of accepted forward shooting moves in the Acc2-AspH system.

Looking at the statistics of the forward and backward shooting moves, we can see a difference between the Acc1-AspH and Acc2-AspH systems. Forward shooting moves in the Acc1-AspH systems had an over 50% acceptance rate, while backward shooting moves had a 38% acceptance rate. In contrast, the situation is reversed in the Acc2-AspH systems: forward shooting moves had 30% acceptance, and backward shooting moves had a 54% acceptance rate. The Acc1-UDPh system had approximately the same acceptance rate for both shooting directions. Finally, the Acc1-noUDP system had the lowest acceptance rate of the forward shooting moves (18%) among the successful systems.

The comparison of the path length histograms shown in Figure 5 demonstrates a difference between the Acc1-AspH and the other three systems. With the Acc1-AspH system, the distribution of path lengths peaks at short trajectories under one ns. The Acc1-UDPh system features a more expanded distribution of path lengths. There is still a small peak of short trajectories, though the distribution is more or less even and includes the longest trajectories of all four systems. As in the Acc1-AspH systems, the Acc2-AspH system has the distribution shifted to short trajectories, but the effect is less pronounced, and there are longer trajectories. Interestingly, the system without UDP (Acc1-noUDP) features no short trajectories. The minimal trajectory length, in this case, is barely under 1 ns. The distribution of the path lengths is otherwise uniform beyond this point.

The total number of shooting events is the same for the Acc1-UDPh and Acc2-AspH systems (220), but higher with the Acc1-AspH system (265). This is caused by the generally shorter trajectories of the Acc1-AspH systems compared to the other two systems. We were therefore able to perform more shooting events using the same computational resources. The number of decorrelated trajectories of Acc1-AspH and Acc2-AspH is essentially the same (18 vs 20) but lower for the Acc1-UDPh system. Tree diagrams showing the evolution of the path ensembles are shown in the SI (Figures S5−S8).

Overall, the Acc1-AspH system features the highest total acceptance rate, the highest acceptance rate of forward shooting moves, and the shortest trajectories. The Acc2-AspH system features a comparable overall acceptance rate, but a much higher acceptance rate of backward shooting moves at the expense of the forward shooting moves. The trajectories are shifted toward shorter lengths, but there are also longer trajectories. The Acc1-UDPh system has its total acceptance rate split equally among the backward and forward shooting moves, and also has more evenly distributed path lengths and the longest trajectories of all four systems.

**Path Density Histograms.** Differences between the transition path ensembles of the four different systems can be visualized using path density histograms. Path density histograms show what portion of trajectories visit a given point in conformational space. Path density histograms of all CVs used to define the START and END states are shown in the SI (Figures S9 and S10). The most significant difference can be seen in the path density histograms of CV3 and CV4 shown in Figure 6. We note that the combination of the two collective variables CV3 and CV4 describes the orientation of the substrate with respect to the catalytic Asp372 and the side binding amino acid Glu300. Since different distances were used as CV3 and CV4 for the Acc1 and Acc2 systems,
histograms showing both sets of distances are provided in Figure 6.

Figure 6 shows that with the Acc2-AspH system, the histograms are very centralized and shifted closer to Asp372. This implies that all of the trajectories of the Acc2-AspH system are confined to a limited conformational subspace close to Asp372. A comparison of Acc2-AspH to Acc1-AspH also revealed that the O5–Glu300 distance (CV4) in Acc2 is smaller than the corresponding distance O6–Glu300 in Acc1.

With the Acc1-AspH system, the density peak is moved toward higher values of CV3. This implies that the system occupies a conformational subspace further away from Asp372. The peaks of both histograms are broader than with the Acc2-AspH system. This indicates that the Acc1-AspH system is more diffuse and can access a more extensive section of the conformational space than the Acc2-AspH system.

Comparing histograms of the Acc1-AspH and Acc1-UDP system, the effect of different active site protonation on the translocation process. In the Acc1-UDP system, with deprotonated Asp372 and protonated UDP, the path density is shifted toward smaller distances between Asp372 and both Galf hydroxyl groups, O5 and O6. This is accompanied by an increase in the O5–Glu300 distance. Interestingly, the Acc1-UDP system also explored part of the conformational space that was unexplored by the Acc1-AspH system. Specifically, the region of long O5–Asp372 distances (>5 Å) and long O6–Glu300 distances (>9 Å). The tendency to explore this part of the conformational space that is not relevant to the translocation could explain the lower acceptance rates of the Acc1-UDP system. The Acc1-noUDP system, also with deprotonated Asp372 but without UDP, behaves similarly to the Acc1-UDP system. We can see a similar shift to lower Asp372 distances and higher Glu300 distances. The protonation state and/or the presence of UDP also lead to differences in the distributions of the torsion angles related to the glycosidic bond of the terminal nonreducing Galf unit (Figure S15).

As the saccharide chain grows by one Galf unit in the course of the catalytic reaction, it should move continually outside the catalytic site. Figure 7 shows path density histograms of the acceptor substrate movement using CV5, which measures the distance of the Galf chain (disregarding the nonreducing terminal Galf) to the active site. With the Acc1-UDP system, this distance grows by ∼3 Å as the translocation proceeds. The Acc1-noUDP system exhibits a similar trend but reaches a lower distance at the END state. Acc1-AspH and Acc2-AspH, on the other hand, start at higher CV5 and maintain this distance through the translocation process. This behavior is consistent with the effect that the protonated Asp372 has on the increased distance of the nonreducing terminal Galf to Asp372, as shown in Figure 6. Here, we see that this effect propagates to the increased distance of the whole Galf chain. The flexibility of the Galf chain can be seen in the distributions of glycosidic bond torsion angles (Figure S15). In short, the Acc2 tetrasaccharide is more flexible than Acc1, and Acc1-UDP has more flexibility around Galβ3(1→5)Galβ2 and Galβ2(1→6)Galβ1 bonds. The Acc1-AspH and Acc1-noUDP systems show comparatively little flexibility. The relatively low movement of the Galf chain with respect to the active site is consistent with the assumed nondissociative mechanism of GlfT2. The full dissociation of the acceptor substrate is prevented by a flexible loop covering the active site (shown in Figure 1 in pink). The native GlfT2 substrate also contains a lipid group binding, which would further disfavor substrate dissociation.

Transition State. One feature of the spring shooting algorithm is the biasing of the shooting point selection toward the transition state region. This allows us to take the frames selected as shooting points and obtain a structure of the transition state ensemble. Shooting points before reaching the transition state region were discarded; the rest were used to analyze the transition state.

Figure 8 shows the transition state ensemble of the translocation process for the four analyzed systems. Table 2 shows the average values of CV1–CV4 in the transition state ensemble. An analysis of the structures in the transition state ensemble revealed that in all three cases, the nonreducing terminal Galf residue occupied a relatively small, well-defined region. This is reflected in the very similar values of CV1 (Table 2) for all four systems. The remaining part of the tetrasaccharide substrate is more flexible. We can also see a different orientation of the nonreducing terminal Galf unit, and the entire tetrasaccharide, in the Acc1-UDP system. This manifests itself in the effect of the different protonation patterns. In contrast, with the Acc1-AspH system, in the Acc1-UDP, the interaction of the Galf unit with the phosphate group of UDP is weakened, while the interaction with Asp372 strengthened. As a result, the Galf unit rotates and interacts with Asp372 via the O2H and O3H hydroxyl groups. This orientation of Acc1-UDP results in increased distances of O5 and O6 atoms to Glu300 (CV4 in Table 2) and the difference in glycosidic bond torsion angles (Figure S15). Acc1-noUDP lacks the phosphate group of the UDP, and the nonreducing Galf unit is much closer to Asp372, which can be seen in the low values of CV3 (Table 2). From Figure 8, we can see three hydroxyl groups (OH3, OH5, and OH6) interacting with Asp372.

The orientations of the nonreducing terminal Galf unit in Acc2-AspH resemble the orientation in the Acc1-AspH system, except for a tilt toward Asp371. This can be seen in a lower value of CV2 in Table 2. The difference between the Acc1-
AspH system and the Acc2-AspH system originates from the difference in the linkage to the previous Galf unit. The β-(1→6) linkage in Acc1 permits the O5H hydroxyl group of the preceding Galf unit to interact with the phosphate group. Lys369 partially mediates this interaction. In contrast, the β-(1→5) linkage in Acc2 exposes the O6H hydroxyl group of the preceding Galf unit to Glu300. The interaction with Glu300 shifts and tilts the Acc2 tetrasaccharide, and as a result, the nonreducing terminal Galf approaches Asp371. Compared to Acc1-AspH, the transition state ensemble of Acc2-AspH also features a lower distance between the O5 atom and Glu300 and a higher distance between the O6 atom and Glu300. This is in agreement with the different final orientations of the two acceptor substrates in the END state. The distance between the O5 atom and Asp372 in Acc2-AspH is much lower than in Acc1-AspH. The decrease in the O5−Asp372 distance in Acc2-AspH is only accompanied by a moderate increase in the O6−Asp372 distance. This indicates a stronger interaction between the nonreducing terminal Galf and Asp372, which could hinder the translocation process in Acc2-AspH. Another factor might be the higher flexibility of the Acc2 tetrasaccharide glycosidic bonds (Figure S15).

In the studied systems, the position of Lys369 also differs. In the Acc1-AspH system, Lys369 interacts with the phosphate group of UDP and with the tetrasaccharide via the O2H hydroxyl group of the nonreducing terminal Galf unit and the O5H hydroxyl group of the previous Galf unit. The interaction of Lys369 with UDP disappears in the Acc1-UDPh system. Instead, Lys369 uniquely interacts with the tetrasaccharide substrate, specifically with the O4 atom of the nonreducing terminal Galf residue and with the O6 atom of the previous Galf residue. A similar situation is seen in the Acc1-noUDP system, where the UDP is not present in the active site. In the Acc2-AspH system, the interaction between Lys369 and the second-to-last Galf unit disappeared due to the different glycosidic linkages. Therefore, Lys369 is more flexible. The different behavior of Lys369 is also illustrated in the path density histograms in Figure S12.

Potential Energy. The information about the transition state ensemble allowed us to use the transition state optimization technique, precisely the NEB method, to obtain a single optimized transition state structure and its potential energy. Figure 9 shows the relative potential energy of the optimized transition state and the optimized END state compared to an optimized START state.

We can see that systems with deprotonated Asp372, Acc1-UDPh, and Acc1-noUDP have a higher potential energy barrier, 14.0 and 13.9 kcal/mol, respectively. These systems also have a higher relative potential energy of the END state. In the Acc1-noUDP case, it is 2.0 kcal/mol, and in the Acc1-AspH system, Lys369 interacts with the phosphate group of UDP and with the tetrasaccharide via the O2H hydroxyl group of the nonreducing terminal Galf unit and the O5H hydroxyl group of the previous Galf unit. The interaction of Lys369 with UDP disappears in the Acc1-UDPh system. Instead, Lys369 uniquely interacts with the tetrasaccharide substrate, specifically with the O4 atom of the nonreducing terminal Galf residue and with the O6 atom of the previous Galf residue. A similar situation is seen in the Acc1-noUDP system, where the UDP is not present in the active site. In the Acc2-AspH system, the interaction between Lys369 and the second-to-last Galf unit disappeared due to the different glycosidic linkages. Therefore, Lys369 is more flexible. The different behavior of Lys369 is also illustrated in the path density histograms in Figure S12.

Figure 8. Snapshots of transition state structures of the substrate translocation step in four systems (Acc1-AspH, Acc1-UDPh, Acc1-noUDP, and Acc2-AspH). Thick licorice representation shows average structures; thin lines show all structures in the transition state ensemble. Water molecules are omitted for clarity; analysis of the presence of water around the substrate is shown in Figure S13. Figure S14 shows the full-length acceptor substrate and all surrounding Glft2 residues.

Table 2. Average Values of CV1−CV4 and Distances d3 and d4 in the Transition State Ensemble

|          | CV1      | CV2      | CV3      | d3^a    | CV4      | d4^b    |
|----------|----------|----------|----------|---------|----------|---------|
| Acc1-AspH| 5.70 ± 0.12 | 5.09 ± 0.13 | 4.92 ± 0.35 | 4.10 ± 0.21 | 9.89 ± 0.10 | 9.39 ± 0.26 |
| Acc1-UDPh| 5.71 ± 0.20 | 4.89 ± 0.06 | 5.16 ± 0.52 | 3.99 ± 0.27 | 10.29 ± 0.25 | 11.32 ± 0.29 |
| Acc1-noUDP| 5.89 ± 0.15 | 4.86 ± 0.04 | 4.23 ± 0.29 | 3.65 ± 0.11 | 10.58 ± 0.25 | 11.18 ± 0.22 |
| Acc2-AspH| 5.61 ± 0.08 | 4.12 ± 0.09 | 4.35 ± 0.15 | 3.20 ± 0.19 | 8.21 ± 0.17  | 10.74 ± 0.18  |

^aDistance between the nonreacting oxygen atom of the nonreducing Galf and Asp372. ^bDistance between the reacting oxygen atom of the nonreducing Galf and Glu300.
AspH case, it is 1.2 kcal/mol. On the other hand, the two systems with protonated Asp372, Acc1-AspH, and Acc2-AspH have a lower potential energy barrier and also lower potential energy of the END state. Despite the difference in the substrate in these two systems, the values of their END state potential energy are very similar, but they differ in their barrier heights. The potential energy barrier is 6.1 kcal/mol, and the END state potential energy is −2.4 kcal/mol for the Acc1-AspH system. For the Acc2-AspH system, the potential energy barrier is higher at 10.3 kcal/mol, and the END state potential energy is −2.3 kcal/mol. The higher barrier and END state potential energies of Acc1-UDPPh and Acc1-noUDP might help to explain the higher TPS failure rate of these systems, while the lower potential energy barrier of the Acc1-AspH system might help explain its higher TPS success rate.

**Cluster Analysis.** To gain further insight, we performed cluster analysis on the whole ensemble of reactive trajectories to identify the dominant conformations of the translocating substrate and the active site. All relevant details are shown in the SI. Figure 10 shows the representative structures of the most occupied clusters for the four analyzed systems. In all cases, the top cluster represents over 65% of all trajectory snapshots, and for the Acc1-UDPPh system over 90%. A simple analysis of the evolution of reactive trajectories (shown in Figures S20–S23) shows that in all cases, the top cluster serves as a “central hub” conformation, through which (almost) all trajectories pass. This conformation likely represents a metastable state important for the translocation process.

Figure 10 again highlights the effect of the protonation pattern on the translocation process. The two cases with the deprotonated Asp372 (Acc1-UDPPh and Acc1-noUDP) feature close contacts of the Galf hydroxyl groups with the deprotonated Asp372. A total of three hydroxyl groups (OH3, OH5, and OH6) are orientated toward Asp372 in the Acc1-UDPPh system, and the situation is similar in Acc1-noUDP. In both these cases, the conformation of the nonreducing Galf residue is closer to the START state than the END state.

The two systems featuring Asp372 in the protonated state (Acc1-AspH and Acc2-AspH) exhibit a different behavior of Asp372 toward the translocation Galf unit. In the Acc2-AspH system, Asp372 interacts with the OH5 group of Galf via a hydrogen bond but acts as a hydrogen bond donor and not an acceptor. The conformation of the Galf unit in the dominant cluster of these two systems is also closer to the transition state structure or even the END state than to the START state.

**DISCUSSION**

Most glycosyltransferases catalyze a single glycosyl residue transfer, after which the elongated acceptor and UDP leave the enzyme. However, GlfT2 is a processive enzyme that does not release the galactan chain until it reaches 30–35 Galf residues. Therefore, the GlfT2 catalytic process requires galactan translocation and regeneration of the active site by releasing the remaining part of the donor, UDP. This is a complex process, and understanding this is further complicated by the lack of knowledge about the order of all of the required steps. Therefore, several issues need to be addressed. Among these

---

**Figure 9.** Potential energy of the optimized transition state (TS) and the optimized END state relative to the optimized START state.

**Figure 10.** Representative structures of the translocating substrate of the most populated clusters identified by the cluster analysis of all of the reactive trajectories for the four studied systems.

---

ACS Omega, 2020, 5, 21374–21384
are understanding whether a proton transfer from the protonated catalytic base Asp372 to UDP, which facilitates UDP leaving the active site, occurs before or after the translocation of the acceptor substrate step. Another question is whether the UDP-Galf donor enters the active site before or after the translocation step. The results of this study do not answer these questions directly, but they do provide detailed information on this process at the microscopic level.

Comparison of the results for Acc1-AspH and Acc1-UDPh suggests that the deprotonation of the catalytic base leads to increased interactions between the hydroxyl groups of the nonreducing terminal Galf unit and the catalytic base Asp372. This is reflected in path density histograms (Figure 6) that are shifted toward lower CV3 values in the systems with deprotonated Asp372. These interactions also affect the structures of the transition state ensemble (Figure 8) and the dominant structure obtained from cluster analysis (Figure 10).

Both structures show multiple hydroxyl groups of the Galf interacting with Asp372 via either hydrogen bonds or through electrostatic interactions. With protonated Asp372, interactions with Asp372 are fundamentally different. First, the catalytic base Asp372 is neutral and can also function as a hydrogen bond donor. The change in interactions between the nonreducing terminal Galf residue and Asp372 is seen in the structures of the transition state ensemble (Figure 8) and dominant structures of the cluster analysis (Figure 10) as well as in the path density histogram, which is shifted toward the higher CV3 values (Figure 6). This movement of the nonreducing terminal Galf unit further from Asp372 is responsible for a relocation movement of the whole tetrasaccharide chain away from the active site, as shown in Figure 7. The system without UDP and with deprotonated Asp372 (Acc1-noUDP) behaves similarly to the system with the protonated UDP and deprotonated Asp372 (Acc1-UDPh).

The difference in the translocation of Acc1 and Acc2 substrates is caused by the difference in the linkage between the nonreducing terminal and second-to-last Galf residue of a tetrasaccharide. The β-(1→5) linkage in the Acc2 substrate leaves the OH6 hydroxyl group of the preceding Galf unit exposed to the environment. The structure of the transition state ensemble (Figure 8) shows the nonreducing terminal Galf residue tilted toward Asp371 and the preceding OH6 hydroxyl group oriented toward Glu300. The situation is reversed in Acc1, where the β-(1→6) linkage exposes the OH5 hydroxyl group of the preceding Galf unit. The OH5 hydroxyl group is then in a position to interact with the phosphate group of UDP. The dominant structure of the Acc1-AspH translocation (shown in Figure 10) shows the nonreducing terminal Galf residue tilted the other way, toward the side binding Glu300. The alternation of created linkages is based on these results, likely caused by the interactions of the preceding OH5 and OH6 groups and Glu300 and the phosphate group of the leaving UDP.

For both substrates, the systems with the shortest translocation trajectories (Figure 5) and the highest rate of accepted forward shooting moves (Table 1) are Acc1-AspH and Acc2-AspH, respectively. This may suggest that the translocation step proceeds with the protonated catalytic base Asp372 and with the UDP present in the active site. The calculated barriers are, due to the methodology used, only estimates, but they support the TPS results and the proposed mechanism. The predicted barrier for the Acc1-AspH translocation step in GlfT2 is comparable with the barrier of 4 kcal/mol predicted for the processive motion in celllobiohydrolase. These barriers are significantly lower than the calculated barriers for the chemical steps (~29 kcal/mol), suggesting that the translocation step is not the rate-limiting process in GlfT2.

The results also shed light on the two Acc2 systems that failed to produce successful translocation trajectories, Acc2-UDPh and Acc2-noUDP, which both features deprotonated Asp372. We can see that translocation of Acc1 in the presence of deprotonated Asp372 is more complicated than with the protonated Asp372. We can also see that even with protonated Asp372, the translocation of Acc2 is more complicated than with the Acc1 translocation. Following these trends, it makes sense that the translocation of Acc2 in the presence of deprotonated Asp372 would fail.

Altogether, our results, especially for the favorable systems Acc1-AspH and Acc2-AspH, support a processive mechanism of GlfT2 and may suggest that the remaining part of the donor UDP leaves the active site after translocation. However, additional experiments will be necessary to establish details of the kinetics.

## CONCLUSIONS

In this study, we carried out extensive TPS simulations of the translocation step of Galf tetrasaccharide inside the active site of a processive glycosyltransferase GlfT2. The transition path sampling methodology, specifically the spring shooting algorithm, was used to generate transition trajectories of the translocation process. In total, six systems were constructed and investigated. The studied systems differed in the glycosidic linkages of the translocating tetrasaccharide substrate (Acc1 vs Acc2), in the protonation pattern of residues in the active site (AspH vs UDPh), or they lacked the UDP substrate (noUDP). The analysis of the generated ensembles of unbiased translocation trajectories revealed atomistic details of the galactan-GlfT2 interactions that drive the processive motion and how these circumstances affect the translocation process. This analysis indicates that after the first chemical step (the formation of the β-(1→5) or β-(1→6) glycosidic linkage), the translocation step with the protonated catalytic base and UDP present in the active site follows. UDP then leaves the active site, clearing a place for the entry of the donor UDP-Galf, after which the second chemical step occurs (the formation of the β-(1→6) or β-(1→5) glycosidic linkage). The calculated barrier of the translocated process is comparable with other processive systems, such as celllobiohydrolase. Moreover, the translocation barrier is significantly lower than the transfer barrier, suggesting that the rate-limiting step in the biosynthesis of the galactan chain by GlfT2 is the glycosylation reaction, not the chain translocation.

## METHODS

TPS is an unbiased powerful simulation technique for studying rare transition processes. It has several advantages over other enhanced sampling methods, such as metadynamics or umbrella sampling. First, it does not require detailed insight into the process. There is no need for a collective variable that correctly describes the entire process. TPS simulations only need the definition of the initial (START) and final (END) state to be defined. Second, TPS does not bias the underlying dynamics of the system. The result of a TPS simulation is an ensemble of unbiased reactive trajectories connecting two predefined stable states.
The starting structures of a substrate–enzyme complex for TPS, each containing one of the tetrasaccharides Acc1 or Acc2 (shown schematically in Figure 3) located in the active site of GlfT2, were taken from our previous study of the GlfT2 glycosyltransferase reaction mechanism. The system was built from the crystallographic structure of GlfT2 with bound UDP (PDB ID: 4FYI), the rest of the donor substrate was modeled manually and acceptor substrate docked. Classical MD simulations of the ternary complexes were performed to relax the systems and judge the stability of the substrate binding. More details are provided in ref 8. The β-(1→6) and β-(1→5) reactions were then simulated using the CPMD method coupled with the string method. The final structures from these simulations represent the START states for TPS simulations in the present study. The tetrasaccharide Galβp[(1→6)]-Galβp[(1→5)]-Galβp[(1→6)]-Galβp (Acc1) represents the substrate for the reaction that yields a pentasaccharide with a β-(1→5)-linked nonreducing terminal Galf. The END state of the Acc1 TPS simulations corresponds to the productive location of the nonreducing terminal Galf with the nucleophilic OSH hydroxyl group oriented toward Asp372. The tetrasaccharide Galβp[(1→6)]-Galβp[(1→5)]-Galβp (Acc2) is the substrate for the formation of a pentasaccharide with a β-(1→6) linkage at the nonreducing end. At the end state of this translocation, the nucleophilic O6H hydroxyl group is oriented toward Asp372. The initial structures taken from the reaction mechanism study contained catalytic Asp372 in a protonated state and UDP in a deprotonated state. These systems are labeled AspH in this study. In addition to this, systems with the altered protonated state were prepared. Systems labeled UDPH contained catalytic Asp372 in a deprotonated state and UDP in a protonated state. Systems without UDP and with deprotonated As372 were labeled noUDP.

Force field parameters for the protein and saccharide part were the same as in the previous study: FF14SB12 and Glycam06,13 respectively. The UDP, in both protonation states, was treated using parameters from the project F90 data (shown schematically in Figure 3) located in the active site of GlfT2, were taken from our previous study of the GlfT2 glycosyltransferase reaction mechanism. The systems were driven using the path driving functionality of PMFlib coupled with AMBER14. The systems were driven from the starting conformation to the target RMSD value of 0.05 Å for 5 ns of simulation.

TPS was performed using the software package Open-Path-Sampling (OPS) version 0.9.412,24 coupled with OpenMM version 7.2.214 as the MD engine. Simulations were run on GPUs using the CUDA interface of OpenMM. Particle Mesh Ewald was used to treat long-range electrostatic interactions with a cutoff of 10 Å. The time step used was 2 fs. Bonds containing hydrogen were constrained to their equilibrium lengths. Velocity Verlet with velocity randomization integrator24 was used to propagate equations of motion. The temperature was set to 300 K, and the collision rate to 1/ps. The pressure was kept at 1 atm using a Monte Carlo barostat. The trajectory was saved every 1000 MD steps (2 ps).

The START state of the translocation process is based on the structure of the ternary complex after the glycosylation reaction simulated in our previous work.8 The END state of the translocation process is based on the ternary complex structure created using acceptor substrate docking and classical MD in our earlier work.8 In the TPS simulations, the START and END states are defined using four collective variables (CV1→CV4), schematically shown in Figure 4. The CVs provide a robust definition of the end states independent of the nature/origin of the initial structures. They characterized the distance between relevant groups of atoms represented by their center of mass (CoM). CV1 describes the position of the nonreducing terminal Galf unit with respect to the active center defined by Glu300 and Asp372 residues; CV2 describes the orientation of OSH and O6H hydroxyl groups of the Galf unit relative to Asp371; CV3 describes the proper orientation (and interaction) of the reacting nucleophilic hydroxyl groups (OSH in Acc1 and O6H in Acc2) with respect to the catalytic amino acid Asp372, and CV4 describes the orientation (and interaction) of the nonreacting hydroxyl groups (OSH in Acc1 and O6H in Acc2) with respect to the side binding Glu300. In addition to CV3 and CV4, we defined the distances d3 and d4: d3 describes the distance of the nonreacting hydroxyl group to Asp372, and d4 describes the distance of the reacting hydroxyl group to Glu300. Thus, CV1 is defined as the distance between the CoM of the four atoms Glu300_OE1, Glu300_OE2, Asp372_OD1, and Asp372_OD2 and the CoM of the nonreducing terminal Galf unit; CV2 is defined as the distance between the CoM of the two atoms Asp371_OD1 and Asp371_OD2 and the CoM of the two atoms Galf_OSH and Galf_O6H. Since the END state of the Acc1 and Acc2 system differs in the orientation of the nucleophilic hydroxyl group OSH and O6H with respect to Asp371, the definition of CV3 and CV4 for Acc1 and Acc2 differs. CV3 is defined as the distance between the CoM of the two atoms Asp372_OD1 and Asp372_OD2 and the CoM of Galf_OSH for Acc1 and Galf_O6H for Acc2, respectively. The other distance not used as CV3 for a given system is labeled d3. Similarly, CV4 is defined as the distance between the CoM of two atoms, Glu300_OE1 and Glu300_OE2, and the CoM of Galf_O6H for Acc1 and Galf_OSH for Acc2, respectively. Similarly, the other distance not used as CV4 for a given system is labeled d4. The CV3 and CV4, and their values, were defined in accordance with the suggested acceptor substrate orientation.
CVSs and CVs were used to analyze the movement of the whole tetrasaccharide chain with respect to the catalytic center. The CVS (sugar-center move) was defined as the distance between the CoM of the protein residues Tyr344, Asp371, and Asp372 and the CoM of the first three sugar units from the reducing end of the Galf tetrasaccharide.

Initial trial TPS simulations were performed using a standard flexible length one-way shooting algorithm. The production TPS simulations were performed using the spring shooting algorithm.10 The spring constant used was 0.1, and the range used was 10 frames. The exceptions were the noUDP systems, for which the spring shooting simulation was initiated directly from the RMSD driving trajectory, and the range was changed to 50 frames; the spring constant was kept at 0.1. The starting position of the shooting point was kept at the default 0.5 of the trajectory length. The maximum length of the shooting attempt was 5 ns, after which the shooting move would be rejected. This was done to avoid excessive use of the available computational resources.

To estimate the energy barriers of the translocation step, a minimum potential energy path optimization was performed using the nudged elastic band (NEB) method25,26 implemented in the package ASE27 coupled with AMBER. The NEB simulations were performed at the MM level using the same force field parameters as the TPS simulations described previously. The last structure of the transition state ensemble was selected and quenched using restrained MD. The active center (last two Galf units, Asp372, Glu300, UDP) was restrained, and an MD simulation with the temperature gradually decreasing to 0 K was performed. Water molecules were then removed, and the implicit solvent model28 was used instead in the NEB optimization. A similar procedure was used for the optimization of the START and END structures. For these structures, standard minimization in AMBER was performed using S000 steps of steepest descent and 2500 steps of the conjugate gradient method to obtain their potential energy. Additional two structures, each at either side of the approximate TS structure, were created using linear interpolation for a total of five images used in the NEB optimization. The two endpoints were kept fixed, and only the other three beads were optimized. The tetrasaccharide substrate, Glu300, Lys369, Asp371, and Asp372, and the last phosphate group of UDP were allowed to move freely during the NEB optimization. The rest of the system was restrained using positional restraints with a force value of 25 kcal/mol. The optimization was terminated when a point of maximum force lower than 0.05 eV·Å⁻¹ (1.153 kcal·mol⁻¹·Å⁻¹) was reached.

Cluster analysis of the transition trajectory ensembles was performed using the cpptraj program from the AMBER suite. The clustering was performed based on the RMSD of the two nonreducing terminal Galf units, Glu300, Lys369, Asp371, and Asp372. The hierarchical agglomerative algorithm was used for the clustering, and it was terminated once the minimum distance between clusters was larger than 1.5 Å. This resulted in four to nine total clusters being identified, depending on the system.

**ASSOCIATED CONTENT**

**Supporting Information**

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.0c01434.

Initial one-way shooting (Table S1, Figures S1–S8); path density histograms of CVs 1–4 (Figures S9 and S10); path density histograms showing the position of Lys369 (Figures S11 and S12); water molecules around TS (Figure S13); full-length substrates at TS (Figure S14); distribution of glycosidic bond torsion angles (Figure S15); clustering analysis results (Table S2, Figures S16–S19); evolution of reactive trajectories (Figures S20–S23) (PDF)

NAT-optimized translocation transition states (ZIP)

**AUTHOR INFORMATION**

**Corresponding Author**

Jaroslav Koča — Central European Institute of Technology (CEITEC) and Faculty of Science, National Centre for Biomolecular Research, Masaryk University, Brno 601 77, Czech Republic; orcid.org/0000-0002-2780-4901; Email: jkoca@ceitec.cz

**Authors**

Pavel Janoš — Central European Institute of Technology (CEITEC) and Faculty of Science, National Centre for Biomolecular Research, Masaryk University, Brno 601 77, Czech Republic; orcid.org/0000-0002-9068-2652

Igor Tvaroska — Central European Institute of Technology (CEITEC), Masaryk University, Brno 601 77, Czech Republic; Institute of Chemistry, Slovak Academy of Sciences, Bratislava 84536, Slovak Republic; orcid.org/0000-0003-2740-944X

Christoph Dellago — Faculty of Physics, University of Vienna, Wien 1090, Austria; orcid.org/0000-0001-9166-6235

Complete contact information is available at: https://pubs.acs.org/doi/10.1021/acsomega.0c01434

**Author Contributions**

This manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

This work was financially supported by the Ministry of Education, Youth and Sports of the Czech Republic under the project CEITEC 2020 (LQ1601) and the program INTER COST (LTC17076). This financial support is gratefully acknowledged. This work was also supported by The Ministry
of Education, Youth and Sports from the Large Infrastructures for Research, Experimental Development and Innovations project "IT4Innovations National Supercomputing Center – LM2015070". Additional computational resources were provided by CESNET LM2015042 and CERIT Scientific Cloud LM2015085, provided under the program "Projects of Large Research, Development, and Innovations Infrastructures." I.T. would like to thank the Scientific Grant Agency of the Ministry of Education of the Slovak Republic and the Slovak Academy of Sciences (Grant VEGA-02/0024/16) for their support. P.J. would like to thank the BISON project for funding his research stay.

**REFERENCES**

(1) May, J. F.; Splain, R. A.; Brotchi, C.; Kiessling, L. L. A Tethering Mechanism for Length Control in a Processive Carbohydrate Polymerization. *Proc. Natl. Acad. Sci. USA* 2009, 106, 11851–11856.

(2) Levgood, M. R.; Splain, R. A.; Kiessling, L. L. Monitoring Processivity and Length Control of a Carbohydrate Polymerase. *J. Am. Chem. Soc.* 2011, 133, 12758–12766.

(3) May, J. F.; Levgood, M. R.; Splain, R. A.; Brown, C. D.; Kiessling, L. L. A Processive Carbohydrate Polymerase That Mediates Bifunctional Catalysis Using a Single Active Site. *Biochemistry* 2012, 51, 1148–1159.

(4) Rose, N. L.; Complete, G. C.; Lin, S.-J.; McNeil, M.; Palic, M. M.; Lowary, T. L. Expression, Purification, and Characterization of a Galactofuranosyltransferase Involved in Mycobacterium tuberculosis Arabinogalactan Biosynthesis. *J. Am. Chem. Soc.* 2006, 128, 6721–6729.

(5) Complete, G. C.; Lowary, T. L. Synthesis of Galactofuranose-containing Acceptor Substrates for Mycobacterial Galactofuranosyltransferases. *J. Org. Chem.* 2008, 73, 4513–4525.

(6) Wheatley, R. W.; Zheng, R. B.; Richards, M. R.; Lowary, T. L.; Ng, K. K. S. Tetrameric Structure of the Glt2 Galactofuranosyltransferase Reveals a Scaffold for the Assembly of Mycobacterial Arabinogalactan. *J. Biol. Chem.* 2012, 287, 28132–28143.

(7) Szczepańska, M. G.; Zheng, R. B.; Complete, G. C.; Lowary, T. L.; Pinto, B. M. STD-NMR Studies of Two Acceptor Substrates of Glt2, a Galactofuranosyltransferase from Mycobacterium Tuberculosis: Epitope Mapping Studies. *Bioorg. Med. Chem.* 2010, 18, 5123–5128.

(8) Janoś, P.; Ruzom, S.; Tvaroska, I.; Koča, J. How Mycobacterium Tuberculosis Galactofuranosyl Transferase 2 (Glt2) Generates Alternating β-1(6) and β-1(5) Linkages: A QM/MM Molecular Dynamics Study of the Chemical Steps. *Chem. – Eur. J.* 2018, 24, 7051–7059.

(9) Dellago, C.; Bolhuis, P.; Geissler, P. L. Transition path sampling. *Adv. Chem. Phys.* 2002, 123, 1–78.

(10) Brotzakis, Z. F.; Bolhuis, P. G. A One-way Shooting Algorithm for Transition Path Sampling of Asymmetric Barriers. *J. Chem. Phys.* 2016, 145, No. 164112.

(11) Knott, B. C.; Crowley, M.; Himmel, M. E.; Ståhlberg, J.; Beckman, G. T. Carbohydrate–Protein Interactions That Drive Processive Polysaccharide Translocation in Enzymes Revealed from a Computational Study of Celllobiohydrolyase Processivity. *J. Am. Chem. Soc.* 2014, 136, 8810–8819.

(12) Maijer, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* 2015, 11, 3696–3713.

(13) Kirschner, K. N.; Yonge, A. B.; Tschampel, S. M.; González-Outirriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. *Carbohydrates. J. Comput. Chem.* 2008, 29, 622–655.

(14) Dupradeau, F. Y.; Cézard, C.; Lelong, R.; Stanisławski, E.; Pécher, J.; Delepine, J. C.; Cieplak, P. RE DD. B: A Database for RESP and ESP Atomic Charges, and Force Field Libraries. *Nucleic Acids Res.* 2007, 36, D360–D367.