Reinforcing Tunnel Network Exploration in Proteins using Gaussian Accelerated Molecular Dynamics

Nishita Mandal[1,2], Bartlomiej Surpeta[*1,2] and Jan Brezovsky[*1,2]

1. Laboratory of Biomolecular Interactions and Transport, Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Uniwersytetu Poznanskiego 6, 61-614 Poznan, Poland
2. International Institute of Molecular and Cell Biology in Warsaw, Ks Trojdena 4, 02-109 Warsaw, Poland

Corresponding authors

Bartlomiej Surpeta, E-mail: bartlomiej.surpeta@amu.edu.pl

Jan Brezovsky, E-mail: janbre@amu.edu.pl

Abstract

Tunnels are structural conduits in biomolecules responsible for transporting chemical compounds and solvent molecules to and from the active site. Tunnels have been shown to be present in a wide variety of enzymes across all their functional and structural classes. However, the study of such pathways is experimentally challenging, as they are typically transient. Alternatively, computational methods such as molecular dynamics (MD) simulations have been successfully proposed to explore tunnels. Conventional MD (cMD) has been shown to provide structural details to characterize tunnels, but suffers from sampling limitations to capture rare tunnels opening on longer timescales. Therefore, in this study, we explored the potential of Gaussian Accelerated MD (GaMD) simulations to improve exploration of complex tunnel networks in enzymes. Here, we used the haloalkane dehalogenase LinB with its two variants with engineered transport pathways, which is not only well known for its application potential, but in particular has been extensively studied experimentally and computationally from the perspective of its tunnel network and their importance in multi-step catalytic reaction. Our study demonstrates that GaMD efficiently improves tunnel sampling and allows identification of all known tunnels for LinB and its two mutants. Furthermore, the improved sampling provided insight into previously unknown transient side tunnel (ST). The extensive conformational landscape explored by GaMD simulations allowed us to investigate in detail the mechanism of ST opening. We were able to determine variant-specific dynamic properties of ST opening, which were previously inaccessible due to limited sampling of cMD. Our comprehensive analysis supports multiple indicators of the functional relevance of the ST, emphasizing its potential significance beyond structural considerations. In conclusion, our research proves that the GaMD method can be used to overcome the sampling limitations of cMD also for effective study of tunnels in enzymes, providing further means for the identification of rare tunnels in enzymes with potential for drug development, precision medicine, and rational protein engineering.

Keywords: transport tunnel, molecular dynamics, haloalkane dehalogenase, enhanced sampling methods, mechanism

Introduction

Enzymes work as natural catalysts to enable execution of complex reactions inside living cells. Because of their complexity to carry out specific biochemical reactions, the structural basis of their efficiency and sensitivity is still not fully understood. Enzymatic reactions take place at the active site, which can be on the surface or buried inside the protein's core. In the enzymes with buried active sites, the transport of molecules from the bulk solvent to the active site and their release is carried out through transport pathways called tunnels[1]. These tunnels are found to be widespread, comprising over 50% of enzymes occurring in all six enzyme commission (EC) classes[2]. In order to understand the structural basis of activity related to the tunnels, in-depth investigation is required to study the dynamic behavior of these pathways, which determines the exchange rate of a particular substrates entrance to the active site or a product release to the bulk solvent[3]. The tunnels provide additional regulation of the movement of solvents and products[4].

The tunnels of certain enzymes may have narrow parts, called bottlenecks, which determine the selectivity of compounds passing to/from the active site. Bottlenecks are mostly controlled through gates represented by residues, loops, secondary structures, or domains[5]. Due to the presence of gates, the tunnels are transient in nature, hence their characterization is non-trivial task, sometimes even impossible in a single static structure[3–5], while frequently critical to regulate the transport, especially in multi-step reactions[4,5]. Gating residues often act as a selectivity filter allowing a certain size of substrates/products to travel through the tunnel and hence they are crucial for catalytic reaction of the enzyme[5]. Given that many enzymes contain molecular tunnels that have been associated with the emergence of different diseases and that the inhibitors binding to these tunnels can function as effective medications, we can comprehend the biological significance of tunnels[4]. The mutation in the tunnel residues can result in variants with markedly modulated properties[1,6]. Recent research shows the enzymes' catalytic functions and its prospects can be altered by the dynamics, geometry and physicochemical properties of the tunnel[4]. Therefore, dynamics and flexibility of enzymes and their tunnels have to be considered as an important factor while in-depth studying structural biology in the context of structure-based drug design[4].

Due to development in experimental methods such as X-ray crystallography[7], cryo-EM[8], various types of NMR spectroscopy[9] and cutting-edge computational technologies such as deep learning-based protein structure prediction with AlphaFold2[10], increasing number of high quality three-dimensional static structures are available for in-depth investigation including potential analysis of tunnels. While the experimental methods result in structures of the proteins, the study of tunnels in those structures can be achieved by dedicated geometry-based tools exploring free van der Waals volume such as CAVER 3.0[11], Mole 2.5[12] and MolAxis[13]. As static structures derived from experimental techniques or predicted computationally usually do not account for multiple protein conformational states, these techniques are frequently followed by molecular dynamics (MD) simulations, to explore the protein conformational landscape and therefore capture dynamics of tunnels and study their transient nature[4,7]. Additionally, exploration of tunnels by geometry-based methods can be supplemented by ligand-tracking methods using tools such as Streamline tracing[14], Visual abstraction of Solvent Pathlines[15], Watergate[16], AQUA-DUCT[17] and trj_cavity[18].

Due to the difficulty of the biomolecules to easily cross large energy barriers, milliseconds to seconds or even more sampling is required to visit rare structural changes[7,19]. Opening of transient tunnels represents such challenging task for Conventional MD methods (cMD), which are typically limited to tens of microseconds[20]. To tackle this limitation, various biased enhanced sampling methods were proposed, such as adaptive biasing force (ABF)[21], umbrella sampling[22], metadynamics[23], and others. Definition of the collective variables (CV) represents their main limitation, as it requires thorough knowledge of the system and frequently extensive optimization before finding the optimal solution. Considering the systems containing complex tunnel networks, multiple CVs would be required for efficient exploration of desired transport pathways. On the other hand, accelerated MD (aMD) and Gaussian accelerated MD (GaMD) were proposed to enhance sampling of conformational space without explicitly defined CVs[24], which represent a promising solution for exploration of tunnels. Furthermore, the GaMD[25] provides an opportunity to reconstruct the original free energy landscape due to appropriately controlled boosting potential added. Although GaMD has been investigated on numerous enzymes, its utilization from the perspective of reliable sampling of the tunnel dynamics remains unknown.

In this study, we examine the efficiency of GaMD to explore stable and transient tunnels in haloalkane dehalogenase's (LinB) tunnel network. LinB belongs to the α/β hydrolase superfamily which catalyze the hydrolytic dehalogenation of the halogenated compounds[26], with a high application potential usage prospective in bioremediation, biosensing or biocatalysis[27,28]. Importantly, due to its multi-step reaction and utilization of water during the reaction, this class of enzymes was shown to require multiple tunnels crucial to synchronize particular steps and hence renders it as a perfect model for our investigation[29]. Structurally, LinB has a flexible cap domain (Figure 1A) and stable core domain[26], while its active site consists of catalytic pentad (E132 – a catalytic acid, D108 – a nucleophile, H272 – a base, W109 and N38 – two halide-stabilizing residues)[30]. LinB was found to present three known tunnels: one permanent tunnel (p1) and other auxiliary tunnels (p2 and p3)[6]. Therefore, in the present study we focused on its three variants to thoroughly test GaMD capabilities to explore conformational space of known LinB tunnel network. These included: (i) LinB wild-type (LinB-Wt) in which p1 represents the main tunnel, while p2 tunnels have auxiliary role (Figure 1B); (ii) designed LinB-Closed carrying L177W mutation resulting in drop of p1 tunnel occurrence (Figure 1C)[6]; and (iii) LinB-Open variant, in which the main p1 tunnel was closed by the same mutation L177W as seen in the former variant, therefore reduction in p1 functioning as the main tunnel. Introduced mutations W140A, F143L, and I211L result in opening of additional auxiliary p3 tunnel (Figure 1D)[6]. As the tunnel dynamics and function vary significantly across selected LinB variants, they represent suitable models for exploration of GaMD capabilities to capture rare tunnel dynamics, to validate its efficiency and sensitivity with the broad knowledge from literature as well as to compare it with cMD simulations.

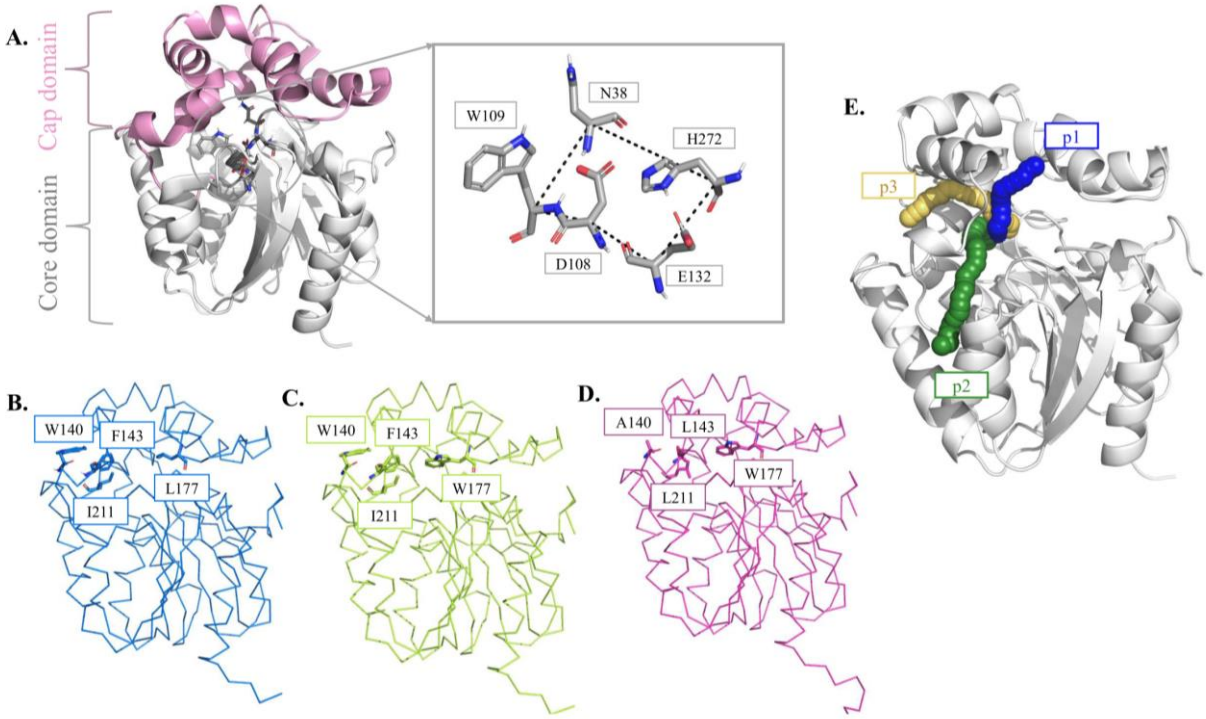

Figure 1. Structural representation of LinB and its variants. **A.** LinB has a more flexible cap domain (pink cartoon) and more stable core domain (grey cartoon). The catalytic residues of LinB forming catalytic pentad are shown as grey sticks, **B.** LinB-Wt, **C.** LinB-Closed mutant, carrying one mutation L177W, **D.** LinB-Open mutant with four mutations: L177W, F143L, W140A & I211L and **E.** known p1, p2 and p3 tunnels of LinB.

Methods

System setup and Conventional molecular dynamics simulations. The cMD simulation was performed on three LinB variants: LinB-Wt (PDB code: 1MJ5)[26], LinB-Closed (PDB code: 4WDQ)[6], and LinB-Open (PDB code: 5LKA)[6]. The systems were first protonated using the H++ server[31] at pH 8.5[6] and salinity of 0.1 M. Systems were solvated using 4-point OPC water models[32] and then were neutralized with counterions (Na+ and Cl-) to achieve the final NaCl concentration of 0.1 M. Initially, systems were energy minimized using 500 steps of steepest descent followed by 500 steps of conjugate gradient with decreasing harmonic restraints in 5 rounds using PMEMD[33] module of AMBER18[34] with ff14SB force field[35]. The following restraints were applied to the system: 500 kcal mol^{-1}Å^{-2} on all heavy atoms of enzyme, proceed with 500, 125, 25, and 0 kcal mol^{-1}Å^{-2} on backbone atoms of the enzyme only. Minimization was followed by 2 ns equilibration MD simulation, with gradual heating to 310 K under a constant volume using the Langevin thermostat[36] with collision frequency of 1.0 ps^{-1} and harmonic restraints of 5.0 kcal mol^{-1}Å^{-2} on the positions of all enzyme atoms, periodic boundary conditions with the particle mesh Ewald method[37,38]. Berendsen barostat was used in order to control the pressure of the system. The systems were run using 4 fs time step enabled by SHAKE and hydrogen mass repartitioning algorithm[39]. Finally, these simulations were continued with an unrestrained 200 ns production simulation performed with PMEMD.CUDA at constant pressure and temperature, storing the frames every 20 ps.

Clustering analysis was performed using the hierarchical agglomerative (HierAgglo) algorithm in cpptraj[40] based on 200 ns trajectory separately for each of the three systems. Cutoff of 4.5 was used with nclusters 5 and linkage method as average linkage, to obtain five most diverse conformations of the system. These five the most diverse conformations were used as a seed structure for the production cMD continued to unrestrained 5 μs simulation, storing the frames every 200 ps, under constant pressure and temperature for all three LinB variants.

Gaussian accelerated molecular dynamics. The five most diverse structures obtained using clustering analysis in cMD were used as seeds for GaMD as well, in order to diversify the sampling space, in all three LinB variants. After preparing the structures for GaMD simulation, a short 8 ns cMD was run to collect the potential statistics of the system – maximum (Vmax), minimum (Vmin), average (Vav), and standard deviation (σV)[25]. GaMD equilibration of 8 ns was performed after adding the boost potential[25]. The first 16 ns of each run was further considered as equilibration and not taken into account in final analyses. Parameters controlling the strength of applied boosting potential: σ0P, which is standard deviation on the first potential boost on total potential energy of the system was set to 1.3 kcal/mol; and σ0D, which is standard deviation of the second potential boost on dihedral angle of the system was set to 2.5 kcal/mol. The parameters for σ0P and σ0D were tested iteratively using a range of values aiming at reaching the value of the coefficients k0D and k0P as close as possible to maximum 1.0, which provides the highest acceleration to the system, or the highest possible whenever the system was unstable before reaching maximum value. Parameters used for testing are described in detail in (Table S1). The system threshold energy was set to lower bound ($E = Vmax$). Simulations were run using 4 fs time-step analogously to cMD with hydrogen mass repartitioning applied[39]. Finally, these dual-boost GaMD simulations were continued for 5 μs of unconstrained production MD. The simulations were performed analogously to cMD settings, namely under constant pressure and temperature using AMBER18 with ff14SB force field, storing the frames every 200 ps for the analyses.

Basic analysis. Generated trajectories resulting from both methods, for all three LinB variants and corresponding replicates (for a total 30 repetitions of 5 μs sampling) were processed using cpptraj[40] implemented in AmberTools17[34]. The root mean square deviation (RMSD) and root mean square fluctuation (RMSF) were calculated with the initial structure as a reference considering the backbone heavy atoms of the protein (N, CA, C). RMSD was calculated without the N-terminal tail (composed of first 11 amino acids), due to its overall high flexibility. The protein's radius of gyration (Rg)[41] and solvent accessible surface area (SASA)[42] were calculated considering all heavy atoms of the protein. Distances between residues were calculated using the cpptraj considering alpha carbons (Cα) of respective residues.

Tunnel Analysis. The tunnels were calculated using the CAVER 3.0.2[11] software with 6 Å shell radius, and 4 Å shell depth. The starting point residues for tunnel calculation were specified as Asn38, Asp109 and His272 corresponding to the numbering from the crystal structure. A probe radius of 0.9 Å was used to calculate the potential pathways in the enzymes with time sparsity of 1. Finally, the pathways were clustered using a clustering threshold of 3.0. The tunnels were calculated in 5 μs simulation of cMD and GaMD along with their initial fractions, namely: 2.5, 1 and 0.5 μs GaMD simulation for the purpose of sampling evaluation and comparison between GaMD and cMD methods. Subsequently, TransportTools[43] (TT)

software v0.9.2 was used to get a unified tunnel network across all variants. Transport tunnels from 5, 2.5, 1 and 0.5 μs GaMD and cMD in the enzyme variants were compared using the comparative analysis module of TT. The clustering method was set to average and clustering cutoff of 1 was used.

Reweighting of GaMD tunnel profiles. Reweighting protocol was implemented to obtain properties of investigated tunnels from original GaMD trajectories reweighted according to boost potential applied during simulation. For this purpose, the output files of individual GaMD runs containing boost potential were parsed along with corresponding TT profiles in CSV format for all superclusters, giving the access to tunnel characteristics reweighted to original free energy landscape without bias. Top 100 tunnels based on reweighted TT throughput from each variant were processed further to study the migration of ligands through them.

Cryptic pocket detection. To analyze the cryptic and allosteric pockets, three different tools were used. *DeepSite*[44] tool which identify viable druggable binding site on the target protein and pockets likely to bind a small molecules, *PASSer 2.0*[45] detecting probable allosteric site pockets, and *FTMove*[46] searching for important cryptic binding hotspots using all known conformers of the protein. Web-servers for all three tools where used using LinB-Wt crystal structure.

Migration of ligands through GaMD tunnels. To evaluate the efficacy of GaMD tunnels in transporting ligands, CaverDock software v1.1[47] was used to perform molecular docking across top tunnels. Four ligands: Bromide ion (br-), 2-Bromoethanol (be), 1,2-Dibromoethane (dbe) and water (h2o), were used to study the transport events in the tunnels, representing the substrate, products, and water important for catalytic activity of LinB[6]. The MGLTools v1.5.6[48] was utilized to prepare the inputs using prepare_receptor4.py and prepare_ligand4.py scripts with default settings. Finally, the upper-bound trajectories results from CaverDock were considered to study energy barrier and successful migration of ligand through particular tunnels using in-house python script[49].

Distance-based principal component analysis. Principal component analysis (PCA)[50] was calculated from the distances of side helix (crystal structure, residue IDs: 166 – 179) to catalytic residue His272 (Figure S1). Python Scikit-learn package[51] was used for the calculation of PCA, while the distances were calculated using the distance module in cpptraj of AmberTools17. PCA was calculated separately for all three enzyme variants. Cluster analysis was done using HDBscan[51]. The parameters for clustering were: min_cluster_size and min_samples were kept 60, allow_single_cluster as True and cluster_selection_epsilon was set to 0.5.

Results and Discussion

GaMD overcomes cMD sampling limitations and enables exploration of more diverse space. To achieve the optimal acceleration parameters, short GaMD simulations (50 ns) were tested on LinB-Wt, followed by extended simulations (1000 ns). For LinB-Wt, the total potential boost σ0P was set to 1.3, resulting in *k*0P 0.09 without prohibitive instability in the system that was observed with higher σ0P values. For the dihedral potential boost, the value of σ0D was set to 2.5 resulting in *k*0D value of 1.0, maximum possible boost. Therefore, the

σ0P=1.3 and σ0D=2.5 dual-boost GaMD LinB-Wt was performed for all three variants of LinB (Table S1). The low value of *k*0P for our model system LinB can be due to its two domains, a more flexible cap domain and stable core domain. The flexible cap domain provides the possibility of opening of multiple tunnels in the protein which can make the enzyme quite sensitive to boost potential energies in GaMD.

To analyze the stability of the enzymes across the simulations, we calculated the RMSD and RMSF for all replicates of simulations from investigated LinB variants, in both cMD and GaMD simulations. The time evolution of RMSD for wild-type in cMD and GaMD oscillated below ~2.5 Å, indicating sufficient equilibration and convergence of the simulations (Figure 2A). Analogously, the LinB mutants indicate even more stable behavior of the systems (Figures S2 and S3), and do not display significant increases in RMSD suggesting more rigid internal dynamics due to introduced mutations, even in the cap domain regardless the applied boosting potential. Further, to verify the compactness of the protein during simulations, especially upon the boosting, we evaluated the Rg and SASA from GaMD trajectories and compared those with cMD, which did not elucidate any significant differences (Figures S4-S6).

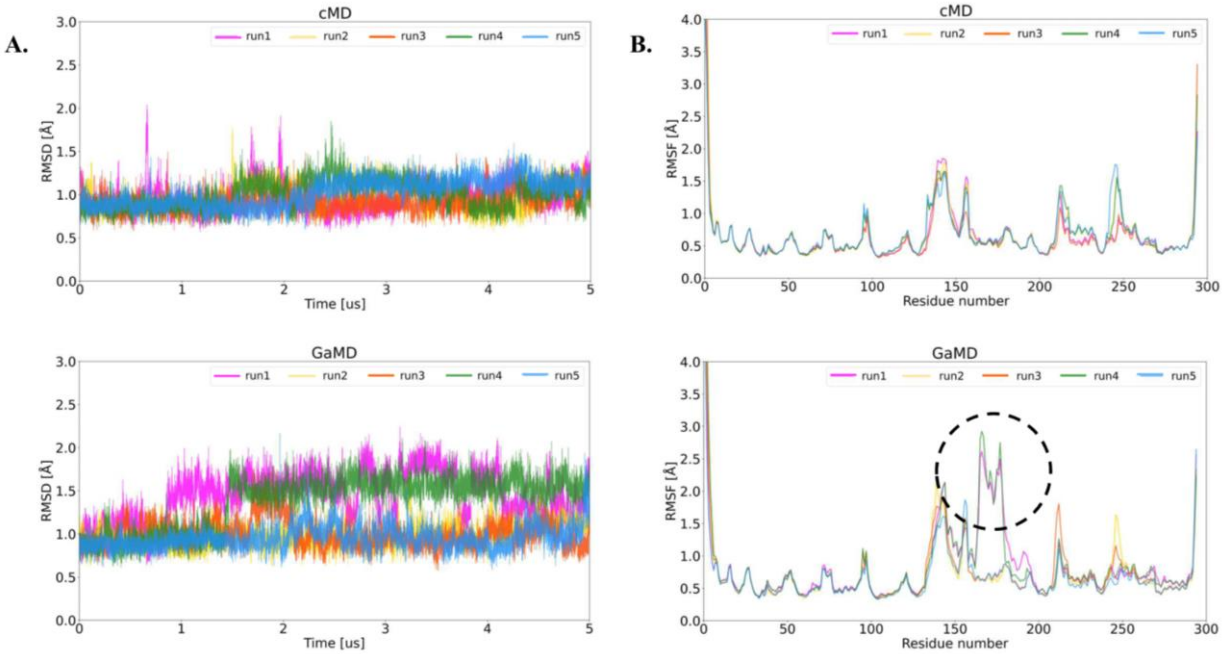

Figure 2. Analysis of protein stability. **A.** The RMSD time-evolution of LinB-Wt enzyme without the 11-residues long N-terminal tail during cMD and GaMD simulations and **B.** The RMSF of LinB-Wt from cMD and GaMD simulations, the most fluctuating residues from the GaMD simulation are highlighted by black dashed circle.

Having verified that the core of the protein is stable, we further evaluated the behavior of the residues forming the catalytic machinery. Therefore, we examined the time-evolution RMSD of individual residues Asn38, Trp109, Asp108, Glu132, and His272 forming the catalytic pentad and all of them together, which may have been affected by the additional boost potential in GaMD, thus sampling inactive/unphysical conformations (Figures S7-S11). The most flips are seen in case of residues Asp108, Glu132 and His272. In case of His272 and Glu132, His272 backbone forms hydrogen bond with Glu132. During the simulation, the hydrogen bond breaks and these residues enter into different conformational state and later they return back to the initial conformation or sometimes the His272 ring enters into different conformational state,

hence the fluctuation in the RMSD plot (Figures S7-S11). In case of Asp108, the side chain often fluctuates back and forth in different conformation, ~1.25 – 1.50 Å RMSD. The values ranged between 2.0 – 2.5 Å while considering them separately and cumulatively in both cMD and GaMD, which shows that the enzymes' catalytic machineries are not distorted due to added boost potential in GaMD.

GaMD enabled proper detection of tunnel network known for LinB and led to elucidate new side tunnel (ST). We calculated the tunnels, to investigate the internal dynamics of ligand transport pathways due to enhanced sampling. The tunnel networks were detected by combined approach including CAVER 3.0.2 calculations, and further tunnel unification across all simulated systems and replicates in TransportTools, which revealed the presence of all known branches of p1 and p2 tunnels, namely p1a, p1b and p2a, p2b, p2c, respectively. Two rare tunnels – known p3 and newly discovered side tunnel (ST) – were found additionally (Figure 3A). The newly discovered ST opening corresponds to the region of high RMSD fluctuations displaying increased values in two replicates of GaMD (especially replicates 1 and 4), due to boosting potential helping the enzyme to explore broader sampling space. This observation is also in line with the RMSF profiles with increased fluctuations in the cap region (residue number 166 – 179) for these two corresponding replicates, suggesting improved sampling (Figure 2B). Further, in order to compare the sampling efficiency between cMD and GaMD in the time evolution of the simulation, we considered various fractions of the trajectories in our analysis, including: full length (5 μs); first half (2.5 μs); initial 20% (1 μs), and 10 % (500 ns). Based on the fraction of the ensembles with the particular tunnel present, we noted that at least 1 μs of GaMD is required to observe enhanced exploration of tunnels, which was in particular noticeable for transient tunnel ST (Figures S12-S14). GaMD enhances the exploration of the sampling space to capture the rare tunnels more effectively than in cMD (Figure 3A). Besides the tunnel occurrence as defined by the probe radius used for the CAVER calculations (0.9 Å), we considered also other, mostly geometrical properties of the tunnels: average length, average bottleneck radius and maximum bottleneck radius. Interestingly, although tunnels exhibit consistent geometrical properties within various groups, it can be noticed that GaMD simulations have a tendency to open tunnels more, resulting in increased maximum bottleneck radius exceeding 3Å (Figures S12-S14).

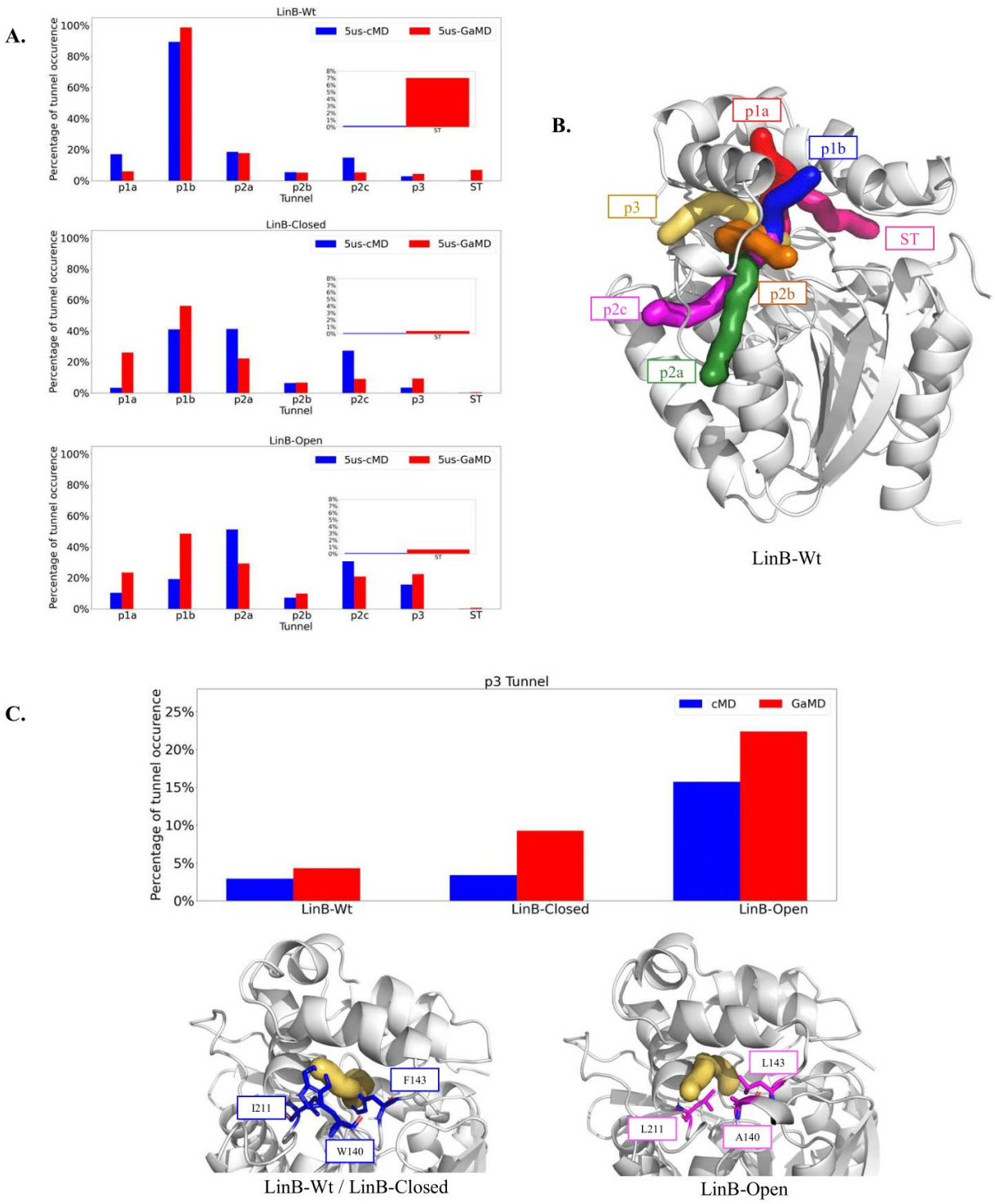

Figure 3. Occurrence of tunnels in LinB variants. **A.** Percentage of tunnel being open in the ensemble for LinB-Wt, LinB-Closed and LinB-Open variants. Inset plots present the increased sampling of ST captured by GaMD in LinB-Wt. **B.** Representation of the LinB-Wt crystal structure (grey cartoon) and the tunnel network in the LinB family (colored spheres). **C.** Plot of p3 tunnel from TT analysis in respective Wt and mutants with cMD and GaMD method (top) and representation of protein structure region where p3 tunnel opens in Wt and LinB-Closed mutant, where it contains bulky residues TRP, PHE and ILE also, region of p3 tunnel opening in LinB-Open, which contain mutation W140A+F143L+I211L and leads to the widening of this region (bottom).

Results obtained from GaMD simulations were compared with the tunnel network in three variants with already published data on de novo tunnel engineering[6]. We found that the p3 tunnel opening covers a similar trend as in previous studies, namely the characteristic tunnel network known for each variant is reproduced in the GaMD trajectories. In LinB-Wt, the p1 tunnel branches represent the primary conduits, which changes in mutants due to closure of p1 tunnels, resulting in significant decrease in its occurrence. Furthermore, we observed significant increase of the p3 tunnel presence in LinB-Open variant due to p3-opening mutations. Interestingly, the GaMD tunnels have the tendency of more frequent and broader tunnel opening (Figure 3C) in case of p3 tunnel (high bottleneck radius shown in Figures S12-S14). Additionally, we have observed increased sampling of the ST in the wild-type protein and its occurrence comparable to rare p3 tunnel in GaMD simulations (Figure 3A), which was not that pronounced in mutants. Furthermore, although it is sampled by cMD, we noted its increased sampling mostly for GaMD in LinB-Wt, while the other two variants do not show such prominent ST openings (Figure 3A) due to large structural perturbation in the helix caused by boosted GaMD simulation, the opening of a transient ST is enhanced. We monitored the RMSF of the protein throughout the simulation and found that there is increased fluctuation in the cap domain in two out of five GaMD simulation replicas, indicating that GaMD visited broader conformational space.

Insights into functional relevance of discovered side tunnel (ST). In addition to tunnels, ligand interaction with enzyme surface pockets or cryptic pockets has been considered important and pocket residues can influence the ligand transport through tunnels[52]. Interestingly, Raczyńska et al[52] published a study focused on finding transient binding sites on enzyme surface as potential sites to engineer enzyme activity. Their study conducted on LinB-Wt, indicated the ability of ST entrance pocket to bind 1-chlorohexane. Moreover, they have shown that the experimentally tested mutation of ST entrance pocket residue A189F increased the enzyme activity by 21.4% for 1-chlorohexane and 26.2% for 1-bromocyclohexane[52], which highlights the importance of ST pocket and the newly discovered ST path which is connecting the active site or active site pocket to the ST pocket for LinB (Figure 4). To study the opening site of ST at the enzyme surface and its connectivity with the active site, we used cryptic and allosteric pocket detecting tools such as *FTMove*, *PASSer 2.0* and *DeepSite tool*. We found the cryptic pocket at the mouth of ST as one of the major ligand binding cryptic pocket, which indicates there can be an allosteric communication by ST pocket. *DeepSite tool* prediction of cryptic pocket shows ST pocket to be in second highest score after active site p1 tunnel pocket (Figure S15). Similar results were obtained by PASSer 2.0 allosteric site prediction tool (Figure S16) and FTMove where the p1 tunnel pocket was ranked 3 and ST pocket was ranked 2 among all the different cryptic pockets detected by the tool (Figure S17).

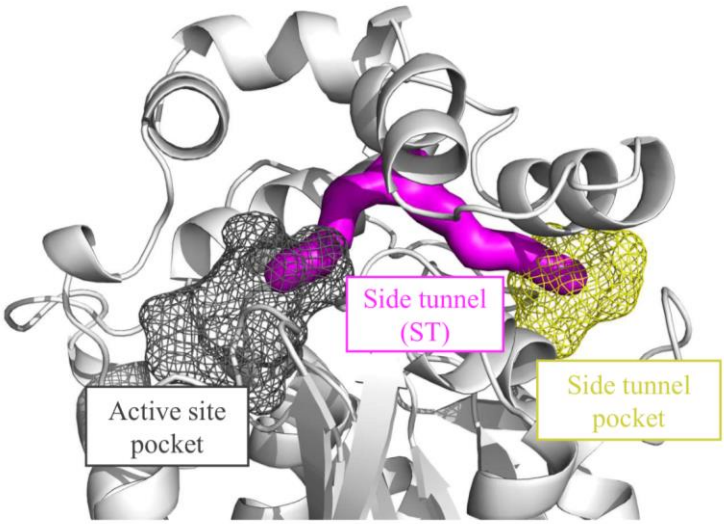

Figure 4. Cryptic pockets detected in LinB-Wt structure by FTMove. showing active site pocket (in dark grey color) which is located at active site and ST pocket (in yellow color) which is located at mouth of ST tunnel. The ST connects these two pockets.

ST displays ability to transport ligands similar to p3. We observed that the ST is connected to cryptic pockets. To study the efficiency of transient ST for transporting ligands between p3 and p1b, we performed CaverDock calculations using an ensembles of top 100 tunnels with highest throughput with four substrate and product molecules (Figure 6, top). Tunnels used for calculations were obtained from the GaMD trajectories and using CaverDock we calculated the energy profile by assessing binding energy of the ligand to the protein along the entire tunnel length. Furthermore, we calculated energy barriers, that have to be overcome by the ligands for successful transport (SI-2). Such energetic evaluation of transport proved that ST was able to transport all four ligands and interestingly, energetically these transports were comparable with energy barriers sampled for auxiliary tunnel p3 (Figure 6). Our evaluation indicated the importance of ST in pair with the known p3 tunnel. Corresponding observations were obtained for haloalkane dehalogenase DhaA, where similar side tunnel was capable of transport water molecules at levels comparable to p3[53]. Additionally, the energy barrier of water (h2o) was found to lower as compared to others (< 5 kcal/mol) in all three tunnels, which shows that all the tunnels were effectively able to transport water through the enzyme. Furthermore, the best performing tunnel found to be p1b from Open mutant, which was in-line with the study on de novo tunnel engineering[6].

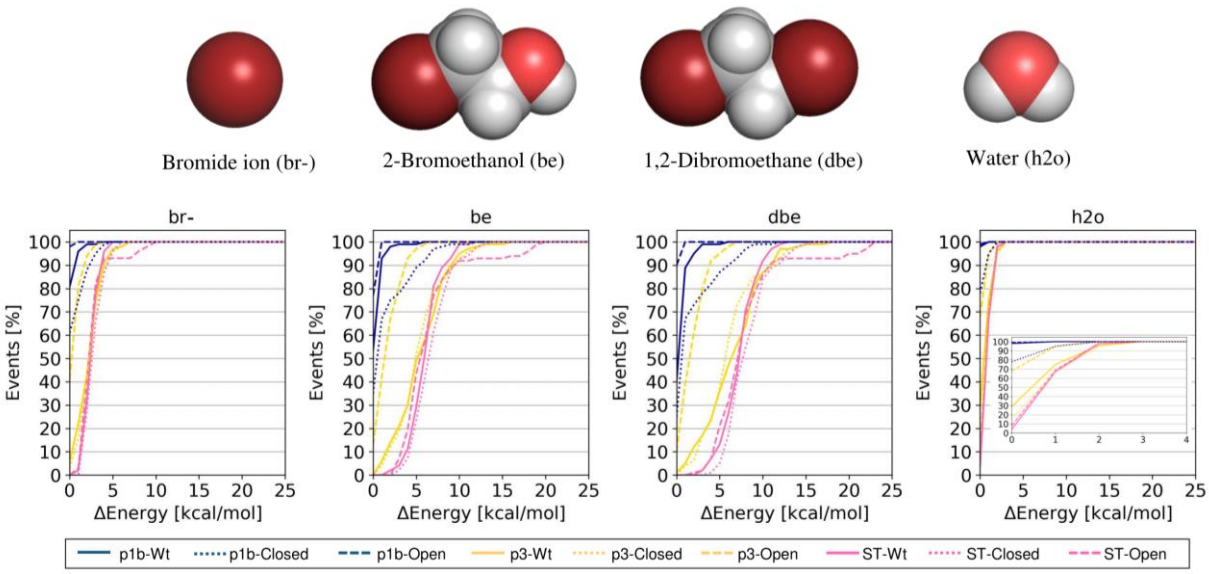

Figure 6. Energetic analysis of ligand transport through transient ST. The structure of four ligands: 1,2-Dibromoethane (dbe), 2-Bromoethanol (be), Bromide ion (br-) and Water (h2o), that have been employed for transport analysis (top). Plots showing successful transport events in all three variants, LinB-Wt, LinB-Open and LinB-Closed, blue describes transport events of p1b, yellow represents transport events of p3 and pink shows transport event of ST(bottom).

ST opening mechanism is different in LinB-Wt and its mutants. The importance and validity of ST is confirmed by our results. Therefore, to understand the conformational changes respective to ST opening, we calculated the distance-based PCA in LinB-Wt and its mutants corresponding to the most dynamic region of protein confirmed through RMSF. In LinB-Wt, we found that GaMD simulations sample two distinct conformational states in contrast to cMD simulations (Figure 6), which indicate that this side helix region provide possibility for the protein to visit different conformational state (Figure S18). In cMD simulations, we also detected the ST occurrence but less frequently, which confirms that the opening is not forced by the biasing potential, but highlights that the system requires broad sampling to capture it efficiently. Major conformational states in all variants were further investigated by HDBscan clustering of distance-based PCA data. We found two predominant states for LinB-Wt in GaMD (Figure 6, bottom), which were also partially seen for mutants but only sporadically (Figures S19-S21). Importantly, this analysis further highlights improved sampling in GaMD trajectories as compared to cMD.

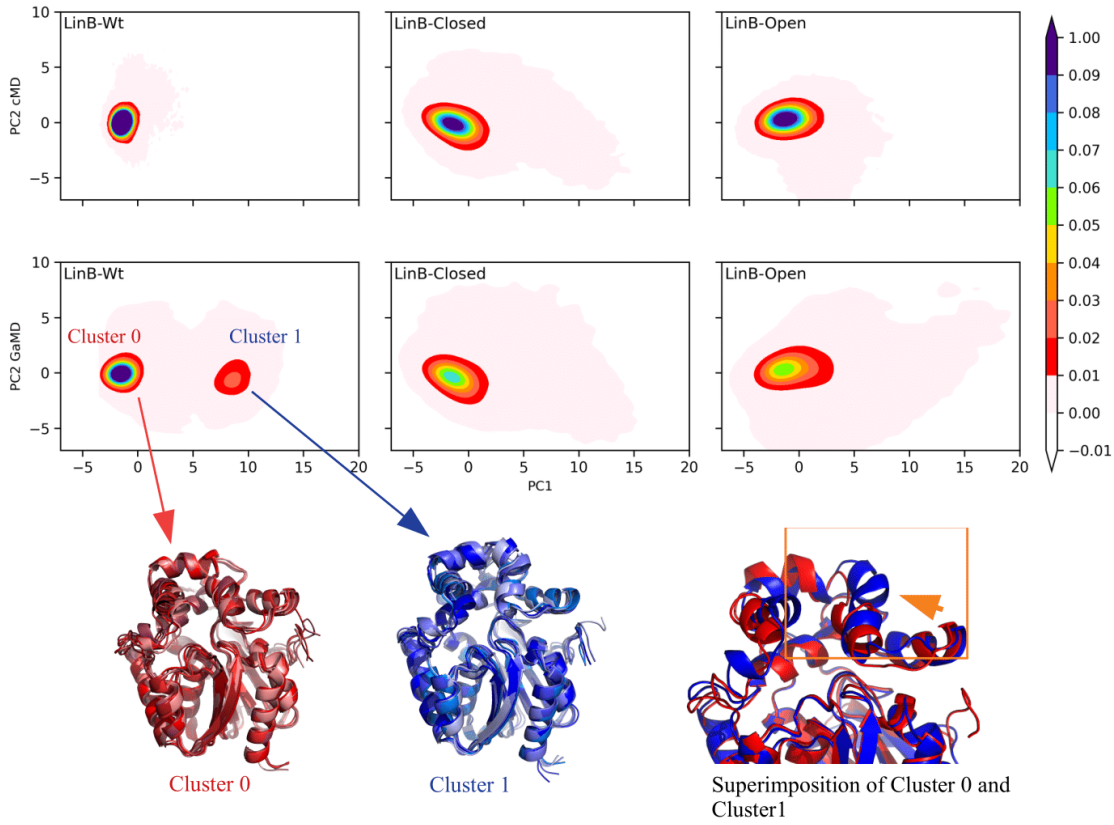

Figure 6. Detection of distinct conformational states sampled by LinB variants. Distance based PCA analysis: PC1 and PC2 were plotted for all three variants of LinB and representation of cluster 0 and cluster 1 (all representative frames) from LinB-Wt GaMD simulation, showing two different states of the protein, shown in the superimposed structures, where red represents cluster 0 and blue color represents cluster 1, showing different conformational state captured by LinB-Wt.

We investigated the mechanism of the rare ST opening in-depth in LinB-Wt and mutants. In LinB-Wt, the prerequisite requirement for opening of ST is movement of the side helix away from the protein cap domain. Due to high fluctuation of the side helix of the cap domain, the ST opens in LinB-Wt frequently (Figure 7A and S22), especially in GaMD boosted simulations. In case of mutants (LinB-Closed and LinB-Open), there are two factors involved in the limited opening of ST. First, the side helix of the cap domain has to move further from the cap domain of the protein. Second, there is mutation at the bottleneck residue of the ST (L177W), which introduces hydrogen bonding with D147 and breakage of this hydrogen bond promotes the opening of ST but is notably more energetically unfavorable, resulting in the ST being less frequently open in mutants when compared to wild-type enzyme (Figure 7B and S23-S24). We had already observed that LinB-Wt samples more ST in comparison to mutants, because in case of LinB-Wt due to lack of H-bond donor, it is easier for the movement of the side helix apart from the protein contributing to the opening of the side tunnel, while in mutants due to the hydrogen bonding the opening is less favorable therefore infrequent.

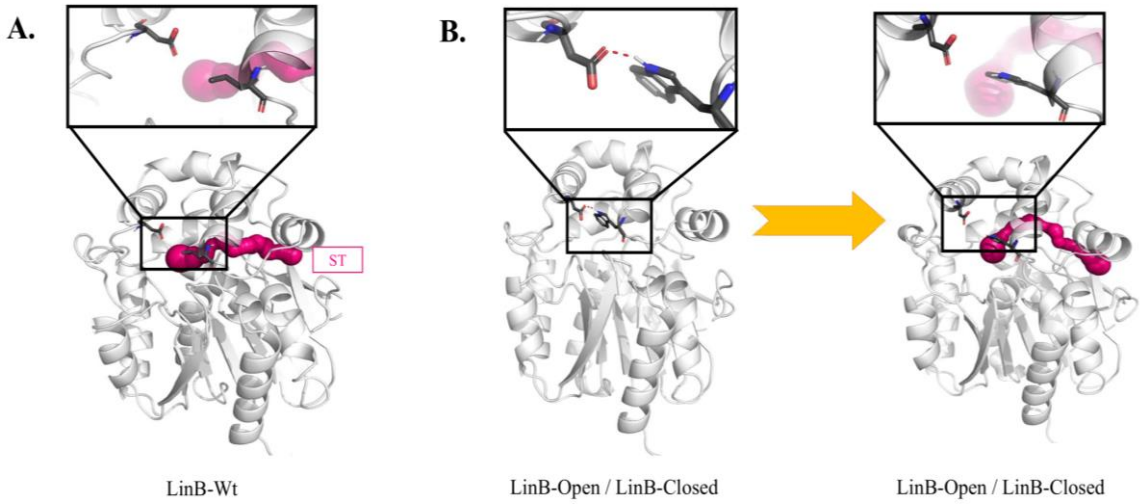

Figure 7. Mechanism of side tunnel opening. (A) Representation of ST in LinB-Wt protein showing necessary movement of side helix for opening of ST. (B) Representation of ST in mutants, showing necessary movement of side helix accompanied by H-bond breakage between Trp176 and Asp146.

Discussion

In this work, we discussed the application potential of GaMD enhanced sampling method to study tunnels and ligand transport pathways in LinB dehalogenases. Our study proved the GaMD method is not only able to identify all known tunnels (p1a, p1b, p2a, p2b and p2c) comparable to cMD method but is also capable to enhance sampling of the conformational space of enzymes and visit rare tunnels (ST and p3 tunnel) efficiently. Improved sampling provided by GaMD simulation enabled us to identify new, not explored ST and to further understand the mechanism of its opening in three LinB variants. This opening mechanism turned out to be clear and straight-forward in case of LinB-Wt enzyme and more complex in the case of mutants. Importantly, such thorough investigation on newly discovered ST path would not be possible without sufficient conformational exploration provided by GaMD. What is critically important, this method provided better picture of conformational dynamics for three LinB variants, but also enabled for correct and consistent exploration of the already know tunnels from broad computational and experimental data for all of them. Explored mechanism revealed that in LinB-Wt, the movement of the side helix is the main contributing factor for the opening of the ST pathway. For mutants, the sampling of ST pathways is less pronounced because of the L177W mutation resulting in hydrogen bond between introduced W177 and D147. This prohibits the opening of the helix observed in the wild-type enzyme and proves the ST is mostly unfavorable for mutants and therefore rarely seen in standard and enhanced MD simulations. Furthermore, this ST pathway is proven to be functionally important as indicated by the exploration of druggable, cryptic and allosteric pockets, all consistently pointing the ST pathway mouth as the functional one. It is further strengthened by recent experimental study presenting the ST pathway mouth as able to bind drugs. Additionally, we performed migration analysis using CaverDock with GaMD tunnels, confirming the capacity of transient ST to

transport the substrate and products molecules efficiently, at levels comparable to auxiliary tunnel p3, which supports the auxiliary role of ST in LinB.

In the case of the p3 tunnel, we have observed, that the GaMD provides better sampling in comparison to standard MD simulations. Furthermore, we noted the trend expected for the LinB-Wt and mutant enzymes from the previous literature data. The p3 tunnel is detected more frequently in LinB-Open mutant due to the mutation W140A+F143L+I211L, carrying mutations of bulky residues to smaller ones aiming at providing more space for the opening of p3 tunnel. On the other hand, LinB-Wt and LinB-Closed mutant, the amino acids at the mouth of p3 tunnels are not modified, hence the opening of p3 tunnel is more restricted as compared to LinB-Open variant.

Conclusions

Taken all into account, our study proves the GaMD as a practically useful approach to investigate tunnels in proteins, which by overcoming the sampling limitations of standard MD simulations, can explore better the conformational dynamics of the protein under study, while limiting the computational costs, and hence open up new possibilities for the identification of tunnels by efficiently sampling rare events and overcoming unfavorable energy barriers. Therefore, as shown for LinB enzyme in our study, this method is suitable for effective exploration of new pathways which can become a druggable sites to target, besides the conventional targeting of the active site itself.

Associated content

Supporting_Information

The Supporting Information of the following files are available free of charge.

- Testing parameters for GaMD simulations; region used for distance based PCA calculation; RMSD and RMSF of mutant enzymes; Rg and SASA of LinB-Wt and it's mutants; RMSD of catalytic residues; comparison between cMD and GaMD tunnel properties from TransportTools; result of cryptic and allosteric pocket detecting tools; plots showing PCA comparison between cMD and GaMD; HDBscan cluster analysis of LinB-Wt and its mutants; mechanism of side tunnel opening using helix movement in LinB-Wt; mechanism of side tunnel opening using helix movement and h-bond analysis in the mutant enzymes (PDF)
- Upper bound energy profiles using caverdock from individual 100 tunnels in all variants with all four ligands (PDF)
- Underlying data is available on Zenodo repository: (i) https://zenodo.org/doi/10.5281/zenodo.11092891 containing input, output and analysis files; and (ii) https://zenodo.org/doi/10.5281/zenodo.11093856 containing the parameter files and stripped MD trajectory files;

Author Information

Corresponding Authors

* janbre@amu.edu.pl (J.B.) & bartlomiej.surpeta@amu.edu.pl (B.S.)

Author Contributions

N.M. performed all MD simulations, calculations of tunnels, cryptic pocket, PCA and migration analysis; and prepared the manuscript draft. B.S. devised tunnel reweighting, PCA and clustering analysis protocol. J.B. and B.S. devised and supervised the project. N.M. and B.S. analyzed the data. J.B, N.M. and B.S. interpreted the results. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

This work was supported by the National Science Centre, Poland (Grant Number 2017/26/E/NZ1/00548).

Notes

The authors declare no competing financial interest.

Acknowledgements

The computations were performed at the Poznan Supercomputing and Networking Center.

Abbreviations

MD, Molecular Dynamics; cMD, Conventional MDs; GaMD, Gaussian Accelerated MD; RMSD, Root Mean Square Deviation; RMSF, Root Mean Square Fluctuation; TT, TransportTools; ST, side Tunnel; EC, Enzyme Commission; ABF, Adaptive Biasing Force; CV, Collective Variable; NMR, Nuclear Magnetic Resonance; aMD, Accelerated MD; PC1, Principal Component 1; PC1, Principal Component 2; cryo-EM, Cryogenic electron microscopy.

References

(1) Prokop, Z., Gora, A., Brezovsky, J., Chaloupkova, R., Stepankova, V., Damborsky, J. *Engineering of Protein Tunnels: Keyhole-Lock-Key Model for Catalysis by the Enzymes with Buried Active Sites*; 421-464; Protein Engineering Handbook, Wiley-VCH, Weinheim, 2012.

(2) Monzon, A. M.; Zea, D. J.; Fornasari, M. S.; Saldaño, T. E.; Fernandez-Alberti, S.; Tosatto, S. C. E.; Parisi, G. Conformational Diversity Analysis Reveals Three Functional Mechanisms in Proteins. *PLOS Comput. Biol.* **2017**, *13* (2), e1005398. https://doi.org/10.1371/journal.pcbi.1005398.

(3) Kingsley, L. J.; Lill, M. A. Substrate Tunnels in Enzymes: Structure-Function

Relationships and Computational Methodology: Protein Tunnel Structure-Function Relationship. *Proteins Struct. Funct. Bioinforma.* **2015**, *83* (4), 599–611. https://doi.org/10.1002/prot.24772.
(4) Marques, S. M.; Daniel, L.; Buryska, T.; Prokop, Z.; Brezovsky, J.; Damborsky, J. Enzyme Tunnels and Gates As Relevant Targets in Drug Design: TUNNELS AND GATES IN DRUG DESIGN. *Med. Res. Rev.* **2017**, *37* (5), 1095–1139. https://doi.org/10.1002/med.21430.
(5) Gora, A.; Brezovsky, J.; Damborsky, J. Gates of Enzymes. *Chem. Rev.* **2013**, *113* (8), 5871–5923. https://doi.org/10.1021/cr300384w.
(6) Brezovsky, J.; Babkova, P.; Degtjarik, O.; Fortova, A.; Gora, A.; Iermak, I.; Rezacova, P.; Dvorak, P.; Smatanova, I. K.; Prokop, Z.; Chaloupkova, R.; Damborsky, J. Engineering a de Novo Transport Tunnel. *ACS Catal.* **2016**, *6* (11), 7597–7610. https://doi.org/10.1021/acscatal.6b02081.
(7) Marques, S. M., Brezovsky, J., Damborsky, J. *UNDERSTANDING ENZYMES: FUNCTION, DESIGN, ENGINEERING, AND ANALYSIS*; Pan Stanford Publishing, 2016.
(8) Peplow, M. Cryo-Electron Microscopy Reaches Resolution Milestone. *ACS Cent. Sci.* **2020**, *6* (8), 1274–1277. https://doi.org/10.1021/acscentsci.0c01048.
(9) Puthenveetil, R.; Vinogradova, O. Solution NMR: A Powerful Tool for Structural and Functional Studies of Membrane Proteins in Reconstituted Environments. *J. Biol. Chem.* **2019**, *294* (44), 15914–15931. https://doi.org/10.1074/jbc.REV119.009178.
(10) Yang, Z.; Zeng, X.; Zhao, Y.; Chen, R. AlphaFold2 and Its Applications in the Fields of Biology and Medicine. *Signal Transduct. Target. Ther.* **2023**, *8* (1), 115. https://doi.org/10.1038/s41392-023-01381-z.
(11) Chovancova, E.; Pavelka, A.; Benes, P.; Strnad, O.; Brezovsky, J.; Kozlikova, B.; Gora, A.; Sustr, V.; Klvana, M.; Medek, P.; Biedermannova, L.; Sochor, J.; Damborsky, J. CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput. Biol.* **2012**, *8* (10), e1002708. https://doi.org/10.1371/journal.pcbi.1002708.
(12) Berka, K.; Sehnal, D.; Bazgier, V.; Pravda, L.; Svobodova-Varekova, R.; Otyepka, M.; Koca, J. Mole 2.5 - Tool for Detection and Analysis of Macromolecular Pores and Channels. *Biophys. J.* **2017**, *112* (3), 292a–293a. https://doi.org/10.1016/j.bpj.2016.11.1585.
(13) Yaffe, E.; Fishelovitch, D.; Wolfson, H. J.; Halperin, D.; Nussinov, R. MolAxis: A Server for Identification of Channels in Macromolecules. *Nucleic Acids Res.* **2008**, *36* (suppl_2), W210–W215. https://doi.org/10.1093/nar/gkn223.
(14) Vassiliev, S.; Comte, P.; Mahboob, A.; Bruce, D. Tracking the Flow of Water through Photosystem II Using Molecular Dynamics and Streamline Tracing. *Biochemistry* **2010**, *49* (9), 1873–1881. https://doi.org/10.1021/bi901900s.
(15) Bidmon, K.; Grottel, S.; Bös, F.; Pleiss, J.; Ertl, T. Visual Abstractions of Solvent Pathlines near Protein Cavities. *Comput. Graph. Forum* **2008**, *27* (3), 935–942. https://doi.org/10.1111/j.1467-8659.2008.01227.x.
(16) Vad, V.; Byška, J.; Jurcík, A.; Viola, I.; Gröller, E.; Hauser, H.; Marques, S. M.; Damborský, J.; Kozlíková, B. Watergate: Visual Exploration of Water Trajectories in Protein Dynamics. *Eurographics Workshop Vis. Comput. Biol. Med.* **2017**, 10 pages. https://doi.org/10.2312/VCBM.20171235.
(17) Magdziarz, T.; Mitusińska, K.; Gołdowska, S.; Płuciennik, A.; Stolarczyk, M.; Ługowska, M.; Góra, A. AQUA-DUCT: A Ligands Tracking Tool. *Bioinformatics* **2017**, *33*

(13), 2045–2046. https://doi.org/10.1093/bioinformatics/btx125.
(18) Paramo, T.; East, A.; Garzón, D.; Ulmschneider, M. B.; Bond, P. J. Efficient Characterization of Protein Cavities within Molecular Simulation Trajectories: *Trj_cavity*. *J. Chem. Theory Comput.* **2014**, *10* (5), 2151–2164. https://doi.org/10.1021/ct401098b.
(19) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450* (7172), 964–972. https://doi.org/10.1038/nature06522.
(20) Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **2009**, *5* (6), 1632–1639. https://doi.org/10.1021/ct9000685.
(21) Darve, E.; Pohorille, A. Calculating Free Energies Using Average Force. *J. Chem. Phys.* **2001**, *115* (20), 9169–9183. https://doi.org/10.1063/1.1410978.
(22) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199. https://doi.org/10.1016/0021-9991(77)90121-8.
(23) Laio, A.; Gervasio, F. L. Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Rep. Prog. Phys.* **2008**, *71* (12), 126601. https://doi.org/10.1088/0034-4885/71/12/126601.
(24) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120* (24), 11919–11929. https://doi.org/10.1063/1.1755656.
(25) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11* (8), 3584–3595. https://doi.org/10.1021/acs.jctc.5b00436.
(26) Okai, M.; Ohtsuka, J.; Imai, L. F.; Mase, T.; Moriuchi, R.; Tsuda, M.; Nagata, K.; Nagata, Y.; Tanokura, M. Crystal Structure and Site-Directed Mutagenesis Analyses of Haloalkane Dehalogenase LinB from Sphingobium Sp. Strain MI1205. *J. Bacteriol.* **2013**, *195* (11), 2642–2651. https://doi.org/10.1128/JB.02020-12.
(27) Swanson, P. E. Dehalogenases Applied to Industrial-Scale Biocatalysis. *Curr. Opin. Biotechnol.* **1999**, *10* (4), 365–369. https://doi.org/10.1016/S0958-1669(99)80066-4.
(28) Bidmanova, S.; Chaloupkova, R.; Damborsky, J.; Prokop, Z. Development of an Enzymatic Fiber-Optic Biosensor for Detection of Halogenated Hydrocarbons. *Anal. Bioanal. Chem.* **2010**, *398* (5), 1891–1898. https://doi.org/10.1007/s00216-010-4083-z.
(29) Liskova, V.; Bednar, D.; Prudnikova, T.; Rezacova, P.; Koudelakova, T.; Sebestova, E.; Smatanova, I. K.; Brezovsky, J.; Chaloupkova, R.; Damborsky, J. Balancing the Stability-Activity Trade-Off by Fine-Tuning Dehalogenase Access Tunnels. *ChemCatChem* **2015**, *7* (4), 648–659. https://doi.org/10.1002/cctc.201402792.
(30) Pavlová, M.; Klvaňa, M.; Jesenská, A.; Prokop, Z.; Konečná, H.; Sato, T.; Tsuda, M.; Nagata, Y.; Damborský, J. The Identification of Catalytic Pentad in the Haloalkane Dehalogenase DhmA from Mycobacterium Avium N85: Reaction Mechanism and Molecular Evolution. *J. Struct. Biol.* **2007**, *157* (2), 384–392. https://doi.org/10.1016/j.jsb.2006.09.004.
(31) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: A Server for Estimating pKas and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res.* **2005**, *33* (Web Server), W368–W371. https://doi.org/10.1093/nar/gki464.
(32) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different

Approach. *J. Phys. Chem. Lett.* **2014**, *5* (21), 3863–3871. https://doi.org/10.1021/jz501780a.
(33) Mermelstein, D. J.; Lin, C.; Nelson, G.; Kretsch, R.; McCammon, J. A.; Walker, R. C. Fast and Flexible Gpu Accelerated Binding Free Energy Calculations within the Amber Molecular Dynamics Package. *J. Comput. Chem.* **2018**, *39* (19), 1354–1358. https://doi.org/10.1002/jcc.25187.
(34) Case DA, Betz RM, Botello-Smith W, Cerutti DS, Cheatham, III TE, Darden TA, et al. AMBER18, 2018.
(35) Maier, 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* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.
(36) Zwanzig, R. Nonlinear Generalized Langevin Equations. *J. Stat. Phys.* **1973**, *9* (3), 215–220. https://doi.org/10.1007/BF01008729.
(37) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.
(38) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593. https://doi.org/10.1063/1.470117.
(39) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11* (4), 1864–1874. https://doi.org/10.1021/ct5010406.
(40) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/ct400341p.
(41) Lobanov, M. Yu.; Bogatyreva, N. S.; Galzitskaya, O. V. Radius of Gyration as an Indicator of Protein Structure Compactness. *Mol. Biol.* **2008**, *42* (4), 623–628. https://doi.org/10.1134/S0026893308040195.
(42) Shrake, A.; Rupley, J. A. Environment and Exposure to Solvent of Protein Atoms. Lysozyme and Insulin. *J. Mol. Biol.* **1973**, *79* (2), 351–371. https://doi.org/10.1016/0022-2836(73)90011-9.
(43) Brezovsky, J.; Thirunavukarasu, A. S.; Surpeta, B.; Sequeiros-Borja, C. E.; Mandal, N.; Sarkar, D. K.; Dongmo Foumthuim, C. J.; Agrawal, N. TransportTools: A Library for High-Throughput Analyses of Internal Voids in Biomolecules and Ligand Transport through Them. *Bioinformatics* **2022**, *38* (6), 1752–1753. https://doi.org/10.1093/bioinformatics/btab872.
(44) Jiménez, J.; Doerr, S.; Martínez-Rosell, G.; Rose, A. S.; De Fabritiis, G. DeepSite: Protein-Binding Site Predictor Using 3D-Convolutional Neural Networks. *Bioinformatics* **2017**, *33* (19), 3036–3042. https://doi.org/10.1093/bioinformatics/btx350.
(45) Xiao, S.; Tian, H.; Tao, P. PASSer2.0: Accurate Prediction of Protein Allosteric Sites Through Automated Machine Learning. *Front. Mol. Biosci.* **2022**, *9*, 879251. https://doi.org/10.3389/fmolb.2022.879251.
(46) Egbert, M.; Jones, G.; Collins, M. R.; Kozakov, D.; Vajda, S. FTMove: A Web Server for Detection and Analysis of Cryptic and Allosteric Binding Sites by Mapping Multiple Protein Structures. *J. Mol. Biol.* **2022**, *434* (11), 167587. https://doi.org/10.1016/j.jmb.2022.167587.

(47) Vavra, O.; Filipovic, J.; Plhak, J.; Bednar, D.; Marques, S. M.; Brezovsky, J.; Stourac, J.; Matyska, L.; Damborsky, J. CaverDock: A Molecular Docking-Based Tool to Analyse Ligand Transport through Protein Tunnels and Channels. *Bioinformatics* **2019**, *35* (23), 4986–4993. https://doi.org/10.1093/bioinformatics/btz386.
(48) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791. https://doi.org/10.1002/jcc.21256.
(49) Pakuła, K.; Sequeiros-Borja, C.; Biała-Leonhard, W.; Pawela, A.; Banasiak, J.; Bailly, A.; Radom, M.; Geisler, M.; Brezovsky, J.; Jasiński, M. Restriction of Access to the Central Cavity Is a Major Contributor to Substrate Selectivity in Plant ABCG Transporters. *Cell. Mol. Life Sci.* **2023**, *80* (4), 105. https://doi.org/10.1007/s00018-023-04751-6.
(50) Lever, J.; Krzywinski, M.; Altman, N. Principal Component Analysis. *Nat. Methods* **2017**, *14* (7), 641–642. https://doi.org/10.1038/nmeth.4346.
(51) Scikit-Learn: Machine Learning in Python. *the Journal of machine Learning research* **2011**, No. pp. 2825-2830.
(52) Raczyńska, A.; Kapica, P.; Papaj, K.; Stańczak, A.; Shyntum, D.; Spychalska, P.; Byczek-Wyrostek, A.; Góra, A. Transient Binding Sites at the Surface of Haloalkane Dehalogenase LinB as Locations for Fine-Tuning Enzymatic Activity. *PLOS ONE* **2023**, *18* (2), e0280776. https://doi.org/10.1371/journal.pone.0280776.
(53) Sequeiros-Borja, C.; Surpeta, B.; Thirunavukarasu, A. S.; Dongmo Foumthuim, C. J.; Marchlewski, I.; Brezovsky, J. Water Will Find Its Way: Transport through Narrow Tunnels in Hydrolases. *J. Chem. Inf. Model.* **2024**, acs.jcim.4c00094. https://doi.org/10.1021/acs.jcim.4c00094.
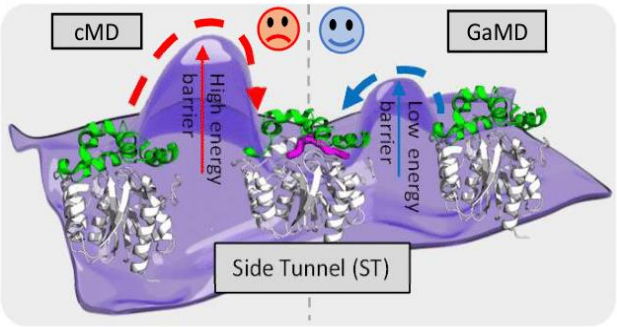
Table of Contents Graphic