Uniwersytet im. Adama Mickiewicza, Poznań, Polska
Wydział Biologii
Instytut Biologii Molekularnej i Biotechnologii
Zakład Ekspresji Genów
Laboratorium Biomolekularnych Interakcji i Transportu

Rozprawa doktorska

Spojrzenie na molekularny mechanizm leżący u podstaw rzadkich procesów transportowych w enzymach z ukrytymi miejscami aktywnymi

mgr Nishita Mandal

Promotor: dr hab. Jan Brezovsky, prof. UAM

2024

Adam Mickiewicz University, Poznań, Poland
Faculty of Biology
Institute of Molecular Biology and Biotechnology
Department of Gene Expression
Laboratory of Biomolecular Interactions and Transport

Doctoral thesis

Insights into Molecular Mechanism behind Rare Transport Processes in Enzymes with Buried Active Sites

M.Sc. Nishita Mandal

Promotor: dr hab. Jan Brezovsky, prof. UAM

2024

Dedicated to,

My parents, ***Purnima Mandal and Nirjhar Kumar Mandal,*** *for their unwavering support and encouragement throughout my journey.*

My brother, ***Nilesh Kumar Mandal,*** *for his boundless support and for being there whenever I needed him.*

Acknowledgments

dr hab. Jan Brezovsky, Prof. UAM
I am deeply grateful for welcoming me as a member of the team and for providing valuable guidance, understanding, and patience.
For providing a scientific and supportive atmosphere.

dr hab. Agnieszka Ludwików, Prof. UAM
for the professional support and management of the doctoral school.

Prof. dr. Siewert-Jan Marrink
for hosting me at the University of Groningen and exciting collaboration.

mgr Malgorzata Klimorowska and mgr Magdalena Walczak
for the help in all matters related to doctoral school.

mgr Arleta Kucz and mgr Iwona Kanonik-Jędrzejak
for the administrative support and patience.

Colleagues from the Laboratory of Biomolecular Interactions and Transport
for their friendly support, insightful scientific discussions, and constructive criticism during the PhD journey.

Bartłomiej Surpeta
for inspiring discussions regarding research and life

My Parents and my Brother
For their faith in me and mental support

Anand Manaparambil
For endless scientific discussions and friendly support

Contents

List of publications

Publications included in the thesis

1. **Research article**

   Jan Brezovsky, Aravind Selvaram Thirunavukarasu, Bartlomiej Surpeta, Carlos Eduardo Sequeiros-Borja, **Nishita Mandal**, Dheeraj Kumar Sarkar, Cedrix J Dongmo Foumthuim, Nikhil Agrawal, TransportTools: a library for high-throughput analyses of internal voids in biomolecules and ligand transport through them, *Bioinformatics*, Volume 38, Issue 6, March 2022, Pages 1752–1753, **https://doi.org/10.1093/bioinformatics/btab872.**

   Impact Factor 2021: 6.931

   MNiSW points 2022: 200

2. **Research article**

   Nishita Mandal, Bartlomiej Surpeta, and Jan Brezovsky, Reinforcing Tunnel Network Exploration in Proteins Using Gaussian Accelerated Molecular Dynamics, *Journal of Chemical Information and Modeling*, Volume 64, Issue 16, August 2024, Pages 6623–6635, **https://doi.org/10.1021/acs.jcim.4c00966.**

   Impact Factor 2023: 5.6

   MNiSW points 2024: 100

3. **Research article**

   Nishita Mandal, Jan A. Stevens, Adolfo B. Poma, Bartlomiej Surpeta, Carlos Sequeiros-Borja, Aravind Selvaram Thirunavukarasu, Siewert-Jan Marrink and Jan Brezovsky, Benchmarking coarse-grained simulation methods for investigation of transport tunnels in enzymes, *bioRxiv,* 2024.09.16.613244, September 2024, **https://doi.org/10.1101/2024.09.16.613244.**

   Impact Factor 2023: N/A

   MNiSW points 2024: N/A

Publications not included in the thesis

1. **Research article**

   Debadrita Basu, Sarmistha Majumdar, **Nishita Mandal**, Shubhra Ghosh Dastidar, Mechanisms of influence of the microtubule over-stabilizing ligands on the structure and intrinsic dynamics of α,β-Tubulin, *Computational Biology and Chemistry*, Volume 96, 2022, **https://doi.org/10.1016/j.compbiolchem.2021.107617**

   Impact Factor 2022: 2.6

   MNiSW points 2022: 70

Streszczenie

Enzymy są jednymi z najsilniejszych biologicznych katalizatorów w przyrodzie. Prawie ~50-60% enzymów ma głęboko osadzone miejsce aktywne, a mechanizm funkcjonalny w tych enzymach jest kontrolowany przez tunele transportowe. Tunele odgrywają znaczącą rolę w transporcie biomolekuł, które są kluczowe dla procesów takich jak kataliza i transdukcja sygnału. Ze względu na dynamiczny charakter tych tuneli i enzymów, odpowiednią metodą do ich badania są symulacje dynamiki molekularnej (MD), znane również jako mikroskopia obliczeniowa. Klasyczne symulacje MD (*ang – classical molecular dynamics - cMD*) cechuje ograniczona możliwość próbkowania rzadkich zdarzeń, takich jak przejściowe otwarcia tuneli, ograniczając naszą zdolność do pełnego zbadania tych ścieżek. W moich badaniach doktoranckich skupiłam się na opracowaniu zaawansowanych metod, które adresują ograniczenia próbkowania cMD w celu zbadania stałych i rzadko występujących tuneli przejściowych w enzymach. Praca doktorska składa się z trzech kluczowych sekcji. W pierwszej części omówiłam rozwój własnego narzędzia, TransportTools, które może obsługiwać ogromne zbiory danych generowane przez symulacje sieci tuneli, umożliwiając wydajną analizę i ujednolicenie danych sieci tuneli z szerokiej gamy enzymów i różnych metod symulacji. W drugiej sekcji wykorzystałam Gaussowsko przyspieszoną MD (*ang. Gaussian accelerated molecular dynamics - GaMD*) do zbadania dynamicznego otwierania tuneli w modelowym enzymie dehalogenazie haloalkanowej LinB. Praca ta wykazała zdolność metody GaMD do uchwycenia rzadkiej dynamiki tuneli, której cMD nie jest w stanie uchwycić, zapewniając nowy wgląd w transport ligandów i funkcjonalność enzymu. GaMD wykorzystuje potencjał wzmacniający (*ang. boosting potential)* do obniżenia bariery energii potencjalnej między stanami metastabilnymi enzymu i pomaga enzymom eksplorować rozległą przestrzeń konformacyjną. Przy użyciu GaMD obliczeniowo odkryłam nowy rzadki tunel, który umożliwiał transport wody i ligandów. Omówiłam mechanizm otwierania wspomnianego rzadkiego tunelu w LinB i jego mutantach. Na koniec oceniłam metody gruboziarniste (*ang. coarse-grained - CG*), używając modeli Martini i SIRAH, aby porównać ich skuteczność w badaniu sieci tuneli. Te metody CG wykazały cenne wyniki dla zbadania otwierania tuneli w długiej skali czasowej i rozróżnienia między tunelami przejściowymi i trwałymi. Metody CG oferują równowagę między zgrubną naturą a wydajnością obliczeniową. Do analizy porównawczej metod CG wykorzystałam modelowy system LinB wraz z 9 enzymami należących do klas EC 1-3, biorąc pod uwagę głęboko umiejscowione miejsca aktywne i zróżnicowaną sieć tuneli. Podsumowując, moja rozprawa przyczynia się do rozwoju technik symulacyjnych w celu zrozumienia rzadkich zdarzeń tunelowych w białkach. Metody te otwierają nowe możliwości zrozumienia mechanizmów enzymatycznych, transportu ligandów majacych potencjał w odkrywaniu leków i inżynierii białek poprzez pokonanie ograniczeń cMD.

Słowa kluczowe: Tunele, rozszerzone symulacje MD, symulacje gruboziarniste, bramki, enzymy

Abstract

Enzymes are one of the nature's strongest biological catalysts. Almost ~50-60% of the enzymes have buried active site and functional mechanism in these enzymes are controlled by transport tunnels. Tunnels play significant role for the transport of biomolecules, which are crucial for processes such as catalysis and signal transduction. Due to dynamic nature of these tunnels and enzymes, the method of choice to study them is molecular dynamics (MD) simulations, also known as computational microscopy. Classical MD (cMD) simulations struggle to sample rare events like transient tunnel openings, limiting our ability to fully explore these pathways. In my Ph.D. research, I have focused on developing advanced methods that address the sampling limitations of cMD to explore permanent and rare transient tunnels in enzymes. The thesis consists of three key sections. In the first section, I have discussed about the development of an in-house tool, TransportTools, that can handle the massive datasets generated by tunnel network simulations, enabling efficient analysis and unification of tunnel network data from wide array of enzymes and different simulation methods. Second section, I have used Gaussian accelerated MD (GaMD) to investigate the dynamic opening of tunnels in the model enzyme haloalkane dehalogenase LinB. This work demonstrated the ability of GaMD method to capture the rare tunnel dynamics that cMD is not able to capture, providing new insights into ligand transport and enzyme functionality. GaMD uses boost potential to lower the potential energy barrier between metastable states of the enzyme and help enzymes to explore vast conformational space. With the use of GaMD, I discovered novel rare tunnel computationally, that was also effective in transporting water and ligands. I have discussed about the opening mechanism of this rare tunnel in LinB and its mutants. Lastly, I evaluated coarse-grained (CG) methods, using Martini and SIRAH, to benchmark their effectiveness in studying tunnel networks. These CG methods showed valuable results for capturing long-timescale tunnel opening and distinguish between transient and permanent tunnels. The CG methods offers a strong balance between coarse nature and computational efficiency. For benchmarking CG methods, I have used model system LinB along with 9 enzymes from EC class 1-3, considering buried active site and diverse tunnel network. Overall, my thesis contributes to the aspects of advancing simulation techniques for understanding rare tunnel events in proteins. These methods opens new possibilities for understanding enzyme mechanisms, ligand transport with potential in drug discovery and protein engineering by overcoming the limitations of cMD.

Keywords : Tunnels, enhanced MD simulations, coarse grained simulations, gates, enzymes

Abbreviations

ABF	Adaptive Biasing Force
aMD	Accelerated Molecular Dynamics
BE	2-bromoethanol
Br^-	bromide ion
CG-MD	Coarse Grained Molecular Dynamics
cMD	Classical Molecular Dynamics
Cryo-EM	Cryogenic Electron Microscopy
CV	Collective Variables
DBE	1,2-dibromoethane
EC	Enzyme Commission
Elastic	Elastic Network Models
GaMD	Gaussian accelerated Molecular Dynamics
Gō	GōMartini3
LiGaMD	Ligand Gaussian accelerated MD
MD	Molecular Dynamics
NADP	Nicotinamide Adenine Dinucleotide Phosphate
NMR	Nuclear Magnetic Resonance
PC1	Principal Component 1
PC2	Principal Component 2
PCA	Principal Component Analysis
RAMD	Random Accelerated Molecular Dynamics
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
SIRAH	Southamerican Initiative for a Rapid and Accurate Hamiltonian
ST	Side Tunnel
TT	Transport Tools
VMD	Visual Molecular Dynamics

Objectives of the thesis

1. Co-developing of methodology to efficiently calculate transport tunnels and to perform their quantitative comparisons.
2. Evaluating usage of accelerated simulation methods for exploring rare tunnels; use the developed methods to provide a comprehensive description of transport mechanisms in dehalogenases.
3. Benchmarking coarse-grained methods for their ability to capture intricate details of transport tunnels in enzymes.

Introduction to structural and functional significance of rare enzymatic tunnels

Enzymes are biological catalysts, that carry out essential biochemical reactions in nature and have been long standing subject of scientific inquiry. These complex proteins carry out the conversion of substrate to product with remarkable efficiency and efficacy. Due to huge diversity of enzymatic structural and functional mechanism, it is not easy to understand them fully. Many enzymes have their active sites deep within their structural cores[1,2], making it necessary to the existence of intricate pathways or tunnels[2–6] through which these molecules must travel. The journey of these substrates, solvents, and products to and from these active sites is often less straightforward than it appears. These tunnels are mainly dynamic in nature and can adapt and change in response to various environmental factors, result in influencing enzymatic activity, specificity and stability [7].

Protein tunnels are very important for managing the entry of ligands to the active site of enzyme, ensuring that only correct molecules pass through. They are also quite widespread appearing in ~50% of enzymes from all Enzyme Commission (EC) classes[8]. This selective control provided by tunnels is key to preserve enzyme efficiency by avoiding formation of nonproductive complexes that could affect negatively to the enzyme's function[7,9,10]. Additionally, these tunnels act as a protective barrier, preventing potentially harmful compounds from reaching the active site, which is crucial for enzymes like metalloenzymes[11], where these tunnels help preventing the complete inactivation of the catalyst. Furthermore, these tunnels enable the precise coordination of reactions involving multiple substrates or cofactors, also regulate the sequence of multistep reactions, and create a suitable environment for reactions that require an absence of water or vice-versa. These functions are important to maintain homeostasis in a living cell, where thousands of protein-ligand interaction simultaneously[7,10,12].

The dynamic nature of tunnels within enzymes is crucial in understanding the structure-function relationship of these biological catalysts. Gates are present in tunnels, which are defined as dynamic element such as residues, loops, secondary structure or domain within tunnel and can access open and closed states of the protein[7,10,13]. Gates considerably effect the movement of molecules through tunnels. Because of presence of gates, the tunnels are transient in nature means, tunnel might appear for short time and then due to movement of gate it closes

again (Figure 1)[7]. These transient tunnels are rarely opening and mostly hidden from experimental methods.[7,10]

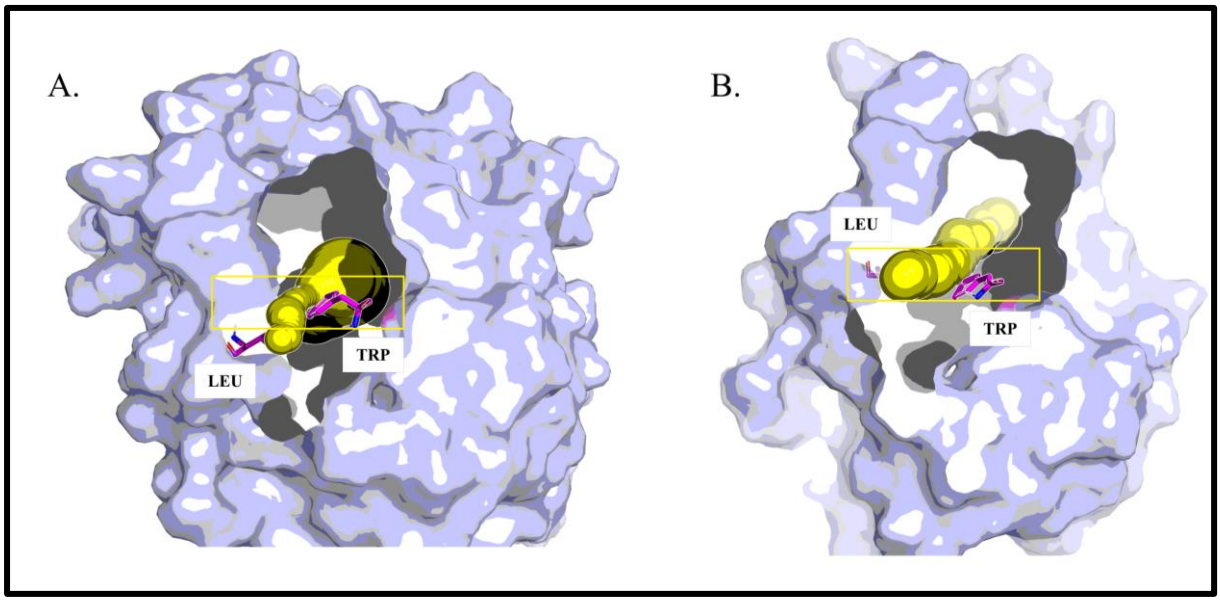


Figure 1. Representation of gates and tunnel geometry. A) Tunnel in yellow color, and box showing narrowest part of tunnel governed by gates (TRP and LEU) and **B)** shows change in tunnel geometry due to movement of gates and more opening of bottleneck part. The surface representation in sky blue color shows the protein.

Experimental methods such as X-ray crystallography[7], cryo-EM[14], and NMR spectroscopy[15] can be used to generate single snapshot of the protein, hence it is difficult to analyze tunnels in them. Nowadays, due to advancement of these experimental methods, there is increase in availability of multiple 3D protein structures. Computational tools such as AlphaFold[16], gained popularity to predict protein structure but still it is difficult to get multiple snapshot of the protein structure from these methods that is enough to study transient tunnels.

To study tunnel dynamics, classical MD (cMD) simulations are often employed, though they struggle with limitations[12,17,18]. The main one is related to sampling as cMD is often limited to capture tens of microsecond, which is not enough to capture higher timescale structural change in proteins. cMD often get stuck in certain minima of the energy surface, hence again limit its efficiency to capture wide conformational space[12,17,18]. Recently, there has been a lot advancement in the field of MD, different enhanced sampling methods have emerged, that are able to overcome sampling limitations of cMD. We can analyse structural changes and

movement of side-chains that the proteins undergoes over time and can get insight into how tunnel dynamics changes overtime. This knowledge not only enhances our understanding of enzyme tunnel function but also opens up new ways, where targeting or modifying tunnels can lead to the development of novel drugs[10]. Biased enhanced sampling methods such as adaptive biasing force (ABF)[19], umbrella sampling[20] and metadynamics[21] are able to overcome limitation of cMD and capture molecule transport through tunnels effectively. The main limitation for these methods is definition of collective variable (CVs), that require knowledge of the system, knowledge of tunnels, etc. The definition of CV for effective capturing of tunnels, is hence not straightforward and easy[19,22–26]. Apart from the methods mentioned here, there are computational methods for selective study of ligand movement and does not need defining of tunnel such as Random Accelerated MD (RAMD)[27,28], Ligand Gaussian accelerated MD (LiGaMD)[29,30]. From these methods the permanent tunnels might be captured, but the transient tunnels are rather difficult to capture[31–35]. From these perspective, accelerated MD (aMD)[36] and GaMD[37,38] are useful for characterization of tunnels as they do not require to define any CVs. Out of the two, GaMD gives more opportunity to reconstruct the original free energy landscape from their applied boost potential. During my PhD studies, I have studied how GaMD can overcome sampling limitation of cMD and increase the sampling of rare tunnels without definition of CV. Also, I have come across many cases when GaMD is not applied in full potential on proteins due to boost potential related instability. I have ran multiple tests to find the perfect boost potential for my system before exploration of tunnels (I have discussed more about it in summary).

However, studying these tunnels computationally requires specialized geometry-based tools like CAVER3.0[4] and MOLE 2.5[39], as static structures do not account for protein conformational changes. CAVER3.0 is written in java and have three steps procedure. First, the tool detect tunnels or pathways in every snapshot of MD simulation. Second stage clusters the tunnels and finally in third stage it calculates and generates the output results. CAVER uses Voronoi diagram[40] that consider atoms of the protein and it divide regions in protein as internal and external space. The starting point is user defined and typically active site of the enzyme. The software then identify possible pathways by tracing the movement of small probe (0.7 – 0.9 nm) along the edges of the Voronoi diagram[40]. For each detected tunnels, CAVER calculates various properties such as bottleneck radius (narrowest part of the tunnel), length and curvature. Finally, the software clusters tunnels based on similar properties, called tunnel cluster. Each

cluster's average length and bottleneck radius is generated by the tool. Also CAVER generated 3D visualization, which is very useful to visualize tunnels using tools like PyMOL[41] or VMD[42].

Another way to overcome the limitations of cMD is by reducing the atomic definitions, such as coarse graining methods[43], where the atoms are represented by coarse grained beads (Figure 2A). By reducing the atomic definition, the time step can be increased by 10-20 folds (20 – 40 fs timestep)[43–45], hence rapidly decreasing the computational time required to run the simulations (Figure 2B). Running coarse grained MD (CG-MD)[44] does not require definition of CV as well. The main concern here is stability of CG model and capture of details using coarser model. During my doctoral studies, I have also tested if CG-MD is effective in capturing tunnels (I have discussed more about it in summary).

As a model system in my research , I have used haloalkane dehalogenase LinB along with its mutants for understanding the tunnel mechanism. They belong to α/β-hydrolase fold superfamily and have catalytic pentad (a nucleophile, a catalytic acid, a base and two halide stabilizing residues) for their reaction mechanisms[46]. Transport mechanisms in dehalogenases play a critical role in influencing the complex kinetic behaviors observed in these enzymes. Haloalkane dehalogenase plays important role in the biodegradation of halogenated compounds[34], these are found usually in environmental pollutants due to their use in agriculture, industrial waste and synthetic chemical production. These enzymes catalyse carbon-halogen bonds in alkanes, and convert them to alcohol, halide ions and protons. This reaction is significant for the detoxification of harmful halogenated organic compounds. Dehalogenases often exhibit very complex kinetic patterns due to the involvement of multiple transport tunnels for substrates, intermediates, and products[34,47]. Here the enzyme have specific selectivity towards particular tunnel for transport of substrate and different tunnel for release of products.

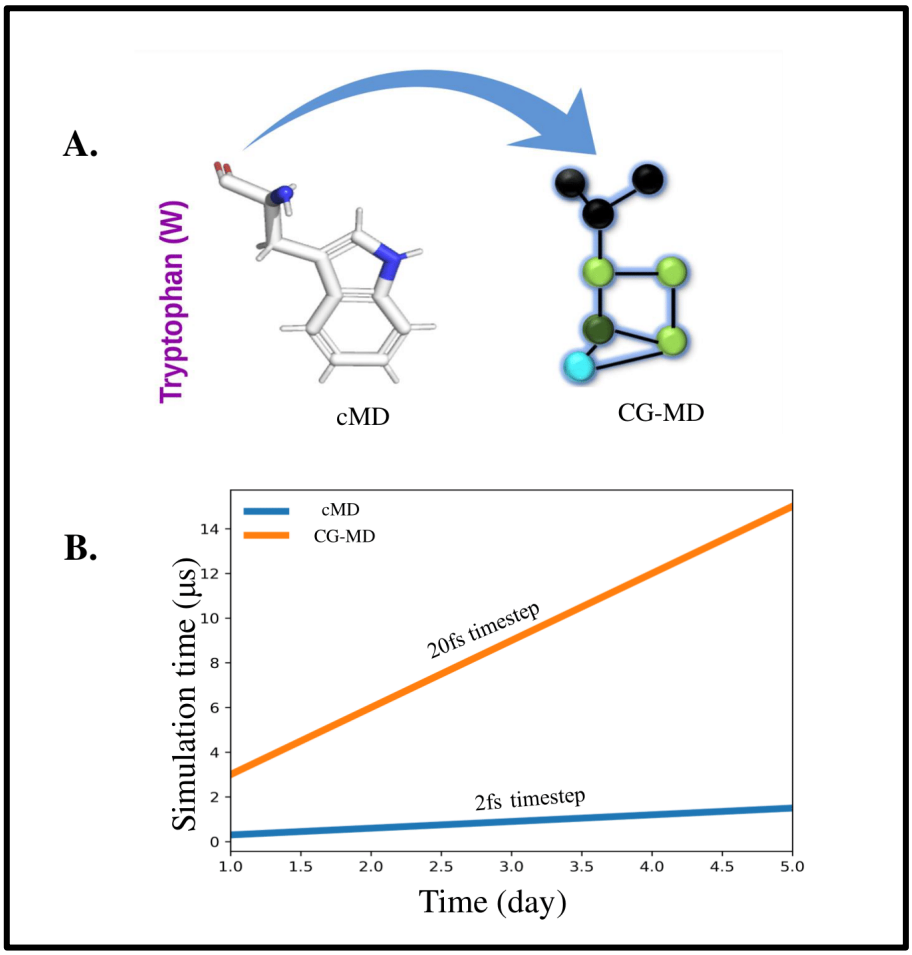


Figure 2. Coarse grained representations. A) All atom and coarse grained representation of amino acid Tryptophan showing difference between number of atoms in all-atom (27 atoms) and number of beads in CG (8 beads) and B) shows high timestep that can be used for CG-MD in respective to cMD simulations.

For LinB as well, one key factor contributing to these complex kinetics is the presence of gates on the tunnels within this enzyme structure[31]. The conformational changes associated with these movement of gates can create multiple kinetic phases, as the enzyme shifts between different conformational states that either can facilitate or obstacle the molecular movement through tunnels[47]. Moreover, the interaction of dehalogenases with cofactors or ligands can also impact the transport mechanisms, altering the enzyme's kinetic behavior[31,46]. The binding and release of these additional components may introduce additional steps in the catalytic cycle, contributing to the overall complexity[33,48]. Tunnel network can be seen in (Figure 3) showing known tunnels (p1, p2, and p3) in LinB-Wt and it's two mutant, LinB-Closed and LinB-Open. From *de novo tunnel engineering* work[31], we can gain understanding about the basic functionality of tunnel in LinB and its variants. In LinB-Wt, the p1 acts as main tunnel and p2

as auxiliary. In LinB-Closed, due to tunnel engineering (L177W) at the mouth of p1 tunnel, this tunnel become less prevalent, but still acts as main tunnel. Further tunnel engineering (W140A, F143L, and I211L) to open the p3 tunnel in LinB-Open, makes transient tunnel p3 similarly relevant as p2 tunnel in LinB-Open mutant. Additionally, to benchmark CG methods for tunnel study, I have used multiple enzymes from EC class 1, 2 and 3 (discussed in detail in summary).

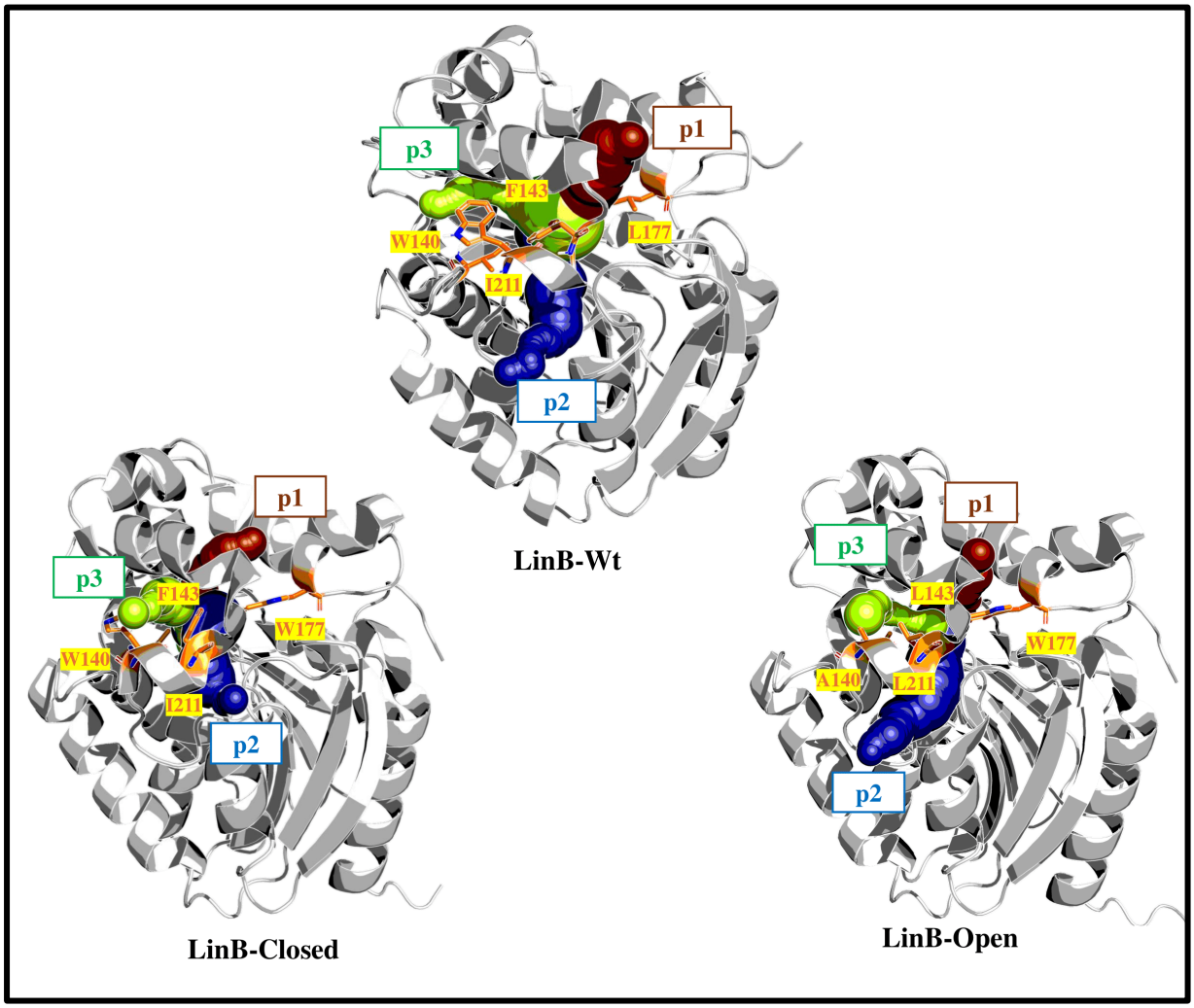


Figure 3. Tunnels in LinB enzyme variant and mutational information for LinB mutants. The main tunnel p1 along with transient tunnels p2 and p3. The mutation at p1 tunnel mouth and p3 tunnel mouth are highlighted.

I have discussed summary and main findings during my dissertation preparation in the next three sections, followed by conclusion and future plans. **Publications 1-3** contains details for the publications, at the end of the thesis sections.

Summary of the doctoral research

Quantitatively comparing huge dataset of tunnel networks

By handling a huge dataset of tunnel network from the enzyme calculations, we realized the requirement of tool that can process much massive data and can unify tunnel network for easy comparison. Colleagues from Laboratory of Biomolecular Interaction and Transport, collectively developed TransportTools (TT) software[49], that unifies the tunnel networks derived by CAVER[50] from many different MD simulations and enriches them with data on ligand tracing provided by AQUADUCT[51] (Figure 4). I have contributed to the development of TT by utilizing the tool for comparing LinB-Wt and mutants, providing critical ideas required for the development of the tool and performed its testing. I have performed short cMD simulations (totaling 300 ns) of LinB-Wt along with its mutant to represent use-case example. Finally, the tunnels were unified and tested using comparative analysis between LinB-Wt and mutants. The results of the analysis shows that TT superclusters were able capture and differentiate between main and auxiliary tunnels (**Supplementary file 7 - use case 2 of publication 1**). Combining the information of the number of frames that the tunnel occurred, and observed water transport events through them, was possible only thanks to TT's unification, which cumulatively can provide insights into significance of the particular tunnel. From the study, I found the effect of mutation at p3 tunnel in LinB-Open, which can be seen by analyzing number of frames, priority and bottleneck residues as an output of TT. Similarly, effect of mutation at the mouth of p1 tunnel that has been closed by mutation L176W can be seen in LinB-Closed.

From the package, the user can get a complete picture of not only permanent and transient pathways but also transport of waters through them, which is important to determine priority of the tunnel. TT users, can perform high-throughput simulations by leveraging state-of-the-art algorithms. It generates detailed in-depth picture into the dynamic movement of tunnels as well as interactions between protein and their ligands. The tool is quite versatile and suitable for wide range of studies from membrane transport studies to enzyme catalysis, providing useful information that are required to understand tunnels and biological functions. The interesting feature, is its capability of quantitatively comparing the tunnels using comparative analysis. It's important for my doctoral research work to understand the difference between tunnels in different LinB variants as well as to probe for discrepancies in tunnels description when using different simulation methods. The full description of TT can be found

in **(publication 1)**. It was very efficient for comparing the data on tunnels from GaMD and CG-MD simulations and comparing with tunnels from cMD simulations, which identified critical aspects of difference exists between them. In general, I have used TT for comparison of tunnel network between different variants of the protein captured through different methods.

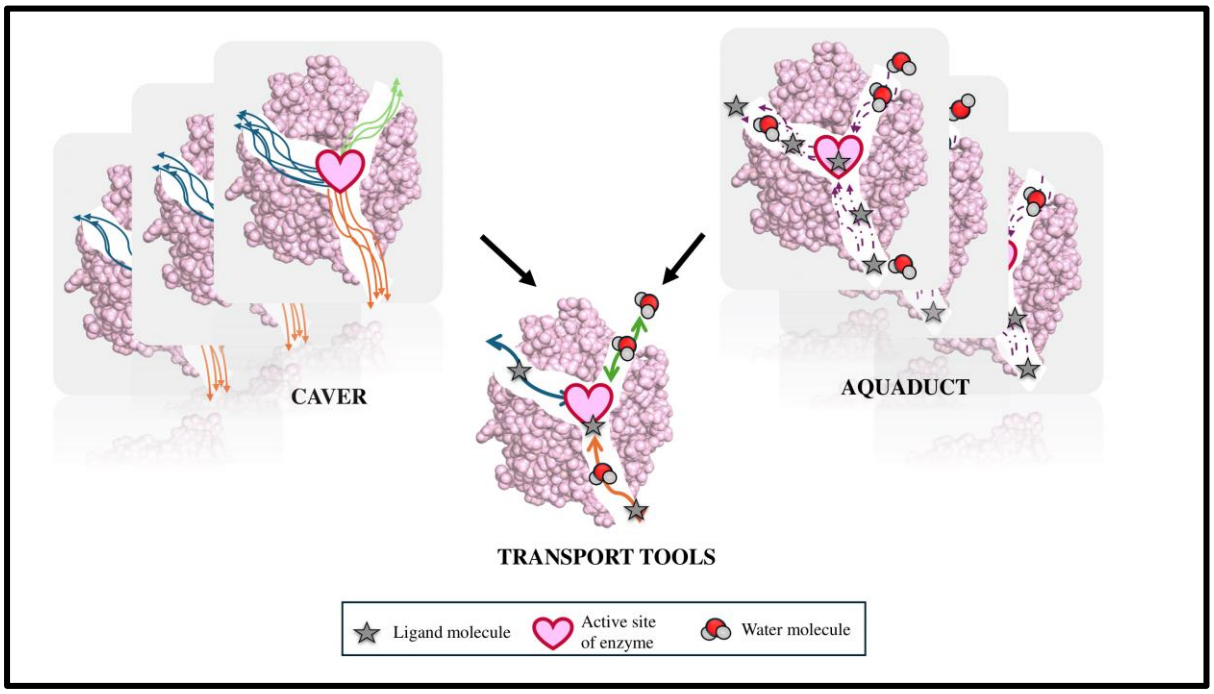


Figure 4. Summary Workflow of Transport Tools. CAVER provided tunnel geometry details, while AQUA-DUCT tracked molecule movement. This combined approach mapped transport events, identifying the pathways used for entry and exit through individual tunnels or shared routes. The frames of the structure comes from multiple simulations.

Now, having a powerful tool to study tunnels, I moved on to address the limitations of cMD for sampling rare tunnels.

Gaussian accelerated MD simulations for Rare Tunnel exploration

Rare tunnels in enzymes are the main focus of my Doctoral research. As I mentioned in the introduction part, in-order to study rare tunnel, it is important to consider dynamic nature of the protein and dynamic nature of the tunnels. Proteins very often face great energy barriers between their metastable states and require long sampling times in order to achieve these rare structural changes like rare opening of tunnel[18], hence cMD is not powerful enough.
GaMD[37] provides a promising solution by enhancing sampling without explicitly defined CVs. GaMD also allows the reconstruction of the original free energy profiles by accurate reweighting procedure, which adjusts the configurations based on boost applied. The reweighting factor considers the probability of a state in boosted potential relative to original potential. While GaMD has been explored in ligand migration studies, its capability to study tunnel networks in the absence of ligands is not yet well explored. In this publication, I explored the tunnel network using GaMD.[52]

GaMD considers proteins potential energy and calculates harmonic boost potential, which is being adaptively added to proteins potential energy surface. GaMD also provides parameters to adjust the power of boost potential. There are two module available within GaMD, single boost and dual boost[37]. For single boost option, the boost potential is added to only the total potential energy or only dihedral potential of the protein. Whereas in dual boost, both potential energy and dihedral boost can be applied and it provides highest acceleration[37]. The boost energy is strongest in the low energy regions, where the protein tend to get stuck. This approach helps the protein to escape these low energy wells and visit high energy states more frequently. This facilitate visiting conformational changes that are rather rare.

This study investigates the efficiency of using the GaMD method to explore the tunnel networks in model system haloalkane dehalogenase LinB. LinB has a flexible cap and a stable core domain encompassing three so far known-tunnels, one permanent and two auxiliary[31]. By focusing on three variants of LinB, namely LinB-Wt, LinB-Closed, and LinB-Open, this study evaluates the capability of GaMD to capture tunnel dynamics and further compares the performance of GaMD to cMD simulations. To get extensive simulation data to study these processes, I have performed 5 x 5 µs replicas of cMD for LinB-Wt, LinB-Closed and LinB-Open and 5 x 5 µs replicas of same set using GaMD, totaling 150 µs of simulation data. To get

diverse starting structure for these simulations, clustering analysis was performed using cpptraj[53] module, on short starting simulation of 200 ns to obtain five diverse clusters. I have performed numerous testing of different combinations of boost potentials to obtain the final boost used for LinB variants, i.e., σ0P = 1.3 (total potential) and σ0D = 2.5 (dihedral potential). Here, I was able to get the maximum boost for dihedral potential resulting in *k*0D of 1.0, but for the total potential energy, I was able to reach only *k*0P of 0.09. The low values of total potential boost can be due to more flexible regions in LinB. The flexible cap domain makes the enzyme quite sensitive to boost potentials. Even though I could not reach full potential of GaMD, which is rather difficult and tricky to set for protein systems. But, the obtained results were rather good considering the rare sample space.

Our systems were relatively stable showing Root Mean Square Deviation (RMSD) below ~2.5 Å, which shows that there is nothing wrong with the GaMD simulations. As for comparison, cMD simulations also remained below ~2.5 Å. I did observed some boost potential effect on the catalytic pentad residue His272 and Glu132, but still the RMSD remained between 2.0 - 2.5 Å, indicating the catalytic system to be preserved during GaMD simulations. From the Root Mean Square Fluctuation (RMSF) plot, I noticed large fluctuation of cap domain of LinB and also tunnel network in that region. The tunnel networks were calculated using CAVER 3.0 and to handle huge dataset of tunnel network from cMD and GaMD, I have used out in-house tool TT, which was very effective for unification of tunnels from cMD and GaMD, required for comparison between them. Excitingly, I have discovered a novel tunnel here using GaMD that opens rarely and still can be seen scarcely in cMD proving it is not artifact[52]. It was possible due to improved sampling space in GaMD simulation. I named the tunnel as side tunnel (ST), as it opens at the side helix region. This tunnel was not known until now for LinB. Along with ST, I also focused on previously known auxillary tunnel p3. In LinB-Open mutant, p3 was de novo engineered, hence the opening of p3 is highest in LinB-Open mutant out of all LinB variants. Through cMD, p3 was studied in LinB-Wt and mutants. Comparing the percentage of p3 opening in LinB-Wt and its mutant through GaMD, I observed higher sampling of this transient tunnel for GaMD in comparison to cMD. Also, GaMD retain the trends between mutant tunnel occurrences, proving its efficiency for tunnel study **(publication 2)**.

More interesting part for me was to study the functional relevance of newly discovered ST. for that I have considered pocket residues at the mouth of ST, as these are the first contact

residues with ligand and can significantly influence transport through the tunnel. At this time, I came across experimental study by Raczyńska et al. [54] focusing on transient binding sites on enzyme surface, as potential protein engineering sites. Interestingly their study was on LinB-Wt and they found mouth of ST at the surface as highest importance. They showed experimentally, ST mouth pocket to bind drug 1-chlorohexane. Further they experimentally, mutated residue at ST mouth pocket A189F, which significantly increased the enzyme activity by 21.4% for 1-chlorohexane and 26.2% for 1-bromocyclohexane[54]. This highlights the importance of ST mouth pocket. Of course, the ligands needs to travel from the surface of the LinB to the active site for enzyme catalysis, that it travel through newly discovered ST which connects ST mouth pocket and active site. To test the significance of ST mouth pocket, I considered several cryptic pocket tools such as PASSer 2.0[55], FTMove[56] and DeepSite[57]. From the results, I observed ST mouth pocket is one of the major ligand binding pocket from all of the tools, suggesting its importance and also potential allosteric connection between active site pocket and ST mouth pocket.

For understanding functional viability of biomolecule transport in ST, I probed the ST along with primary p1 tunnel and auxiliary p3 tunnel. For this, I have used a ligand transport tool called CaverDock[58], which have been quite successful for studying ligand transport rates, predicting unbinding rates. The tool calculates energy barrier profiles of ligand transport for every snapshot of simulation that contain the tunnel of interest. So, how does CaverDock works?

CaverDock simulates dynamic movement of ligand efficiently through tunnel[58]. It starts with considering the input receptor or protein structure, 3-D ligand structure and tunnel of interest that can be obtained using CAVER 3.0. It divides the tunnel of interest into series of segments and each segment perform iterative docking of the ligand, ensuring ligand maintains reasonable orientation and avoid steric clashes along the path. The docking algorithm is customized version of AutoDock Vina[59], which is well known for molecular docking. Unlike conventional docking where the tool usually finds best binding pose at static structure, CaverDock docks the ligand along the tunnel. During docking process, it calculates interaction energy profile, which identifies energy barriers or resistance points. As an output of this tool, we can get energy profile of the ligand along the path, ligand trajectory that can be visualized to study tunnel dynamics and potential binding points through binding free energies.

To understand transport through ST, I performed docking calculations using CaverDock on the ensemble of top 100 conformations of tunnels with highest throughput using GaMD trajectories. I have considered four molecules, 2-bromoethanol (be), 1,2-dibromoethane (dbe), bromide ion (Br^-), and water (H_2O). These molecules are required for catalytic cycle of LinB. The transport efficiency was compared between ST , p3 and p1. From the results, I observed all four molecules were able to transport successfully through ST, at level similar for auxillary known tunnel p3. Further, it transports H_2O with lowest energy barrier (<5 kcal/mol). This further strengthen the significance of ST equivalent to p3 **(publication 2)**. After studying the significant relevance of ST in LinB-Wt, I wanted to study the opening mechanism of ST. Only because of GaMD, I was able to capture enough sample space for studying the mechanism of opening. For this, distance based Principal component analysis (PCA)[60] was performed along with Hydrogen bond (H-bond) analysis. PCA is a strong tool to determine overall dynamic pattern of a system[61]. It reduces the number of variables in a dataset and highlights the greatest variance along the coordinates. For this study, I have considered the distance between residues of side-helix region that shows highest fluctuation from RMSF with catalytic residue His272. The first two principal components captured at least 80% variance suggesting that they are enough to understand the system. From PCA plots, I observed two different states in case of LinB-Wt in GaMD in comparison to cMD, shows higher sampling achieved by GaMD, The states mostly confer to the movement of side helix in LinB-Wt and its mutants (**publication 2**). Still it was not enough to understand how it is happening and what the difference is between LinB-Wt ST opening as compared to mutants.

Further I performed distance analysis between the side helix residue L176 and D146, along with movement of helix in LinB-Wt comparing with ST opening. The analysis was performed for all 25000 frames (5 replicates) obtained through the GaMD simulations, capturing movement of helix and opening of ST simultaneously. Here, I found the movement of side helix is accompanied by the opening of ST in LinB-Wt, depicting the mechanism of opening (Figure 5A). The residue L176 is mutated for LinB-Closed and LinB-Open mutant, forming H-bond with D146, making it more difficult for ST opening in mutants. For mutants, I performed H-bond analysis and distance analysis for movement of helix simultaneously with ST opening. It shows for mutant, there should be breakage of H-bond, that facilitates movement of side helix and opening of ST (Figure 5B). Hence for mutants, it is energetically less favorable and rarer to see ST through GaMD simulations. Overall, GaMD improved the sampling space for rare ST and provided more insight for the mechanism of opening.

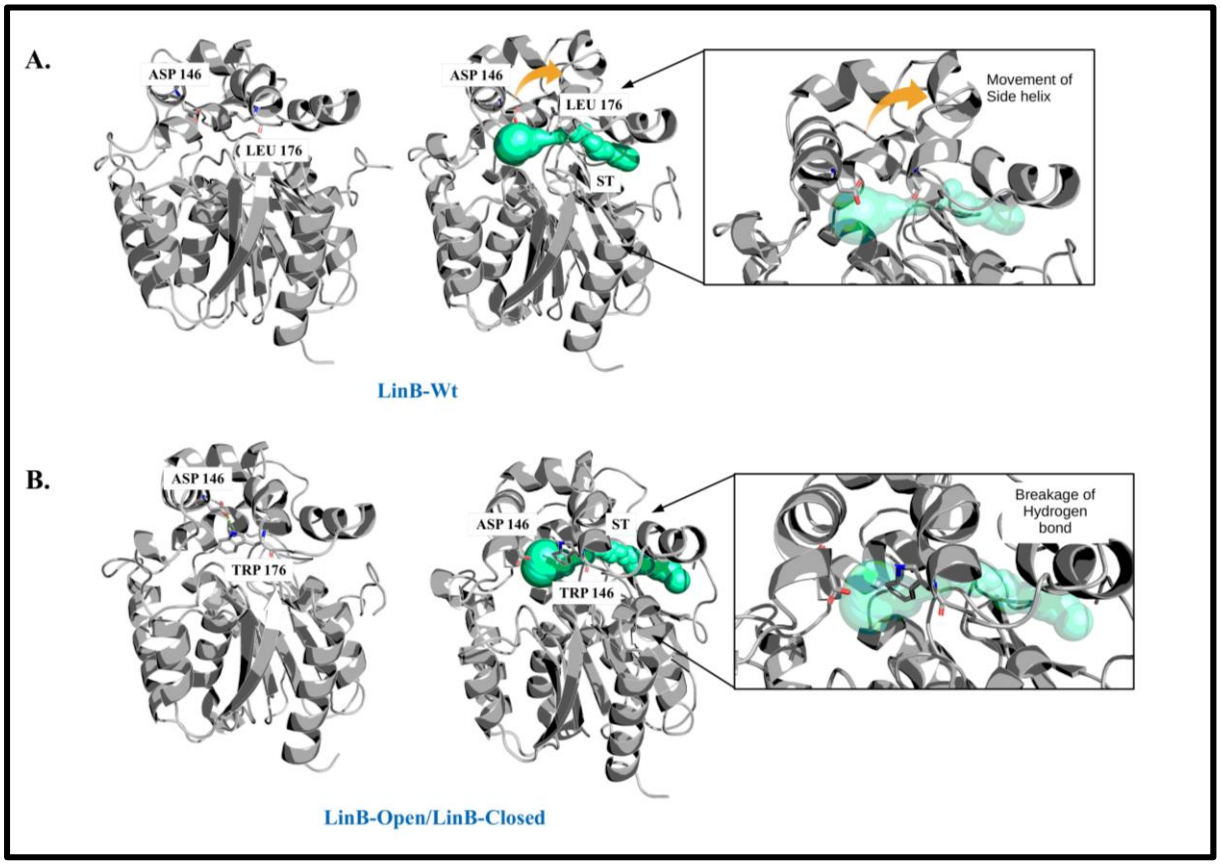


Figure 5. Mechanism of ST opening in LinB-Wt and LinB mutants. A) In LinB-Wt, the requirement for opening of ST is movement of side helix away from the protein core and **B)** In LinB-mutants, due to mutation of L176W, it forms H-bond with D146. For opening of ST in mutant, the breakage of this H-bond facilitates movement of side helix.

From this study, I found the capabilities of GaMD simulations to capture rare opening of transient tunnels. GaMD proves important to tunnel investigation and overcoming sampling limitation of cMD. Further, it opens up new possibilities for efficient sampling of rare tunnels in enzymes that can become druggable site, in addition to active sites.

Coarse Grained methods to study Tunnels

Finally, The last part of my Doctoral work was to understand if CG methods can be effective and accurate enough for study of tunnel network in enzymes. As CG is very effective in overcoming the sampling limitation of cMD by reducing the complexity of atomic definition by CG beads[43,44]. This reduction in complexity allows CG methods to explore larger systems and capture long-timescale processes, such as protein folding, conformational changes, and protein-protein interactions, that are often inaccessible to all-atom simulations[62]. Also due to reduction in complexity, CG methods can be run using time step of ~20-40 fs, which is 10-20 times faster than cMD[44,45]. I have also received a competitive national grant, **PRELUDIUM-22** for carrying out this work in collaboration of Prof. Siewert J Marrink from University of Groningen, The Netherlands and Dr. Adolfo Poma from IPPT-PAN, Warsaw, Poland. The main challenge for this project was ability of CG models to distinguish important amino acids differences that make up tunnel lining residues in rare tunnels between variants.

Now a days, there are many famous CG methods like the famous Martini[63] and SIRAH force-fields[64], that have been developed and validated across numerous protein systems, for understanding conformational dynamics, protein folding pathways, and intermolecular interactions[45]. These methods, effectively balance the compromise between resolution and efficiency, allowing to ask questions without compromising on computational resources that are required for cMD. As a result, CG-MD have become indispensable in the study of protein mechanism[62,64,65]. In **publication 2**, I have discussed how reducing the energy barrier between the metastable states of the enzyme can be effective way to capture rare transport tunnels in enzymes. Here, I will discuss how effective it is to reduce atomic complexity to study tunnel dynamic (**publication 3**).

For this study, I have considered CG methods SIRAH and Martini for studying tunnel. Martini 3.0 uses a 2-4 atoms to describe the CG bead mapping. Martini 3.0, captures the chemical essence of all atom structures, with residue level coarse graining and it have shown to align well experimental methods. New Martini 3.0, also have introduced new dedicated water beads that improves water properties and prevent freezing[62]. For stabilizing the structure under Martini 3.0, Elastic Network Models (Elastic)[66] and GōMartini3 (Gō)[67] can be used. Martini have been shown to work well with protein-ligand binding predictions but it still lacks high chemical specificity for the interaction between ligand and protein[62]. Although, for my

study I am not considering ligands, but there are still presence of co-factors, that need to be in coarse nature. In this study, I have used both of these structure-bias method Elastic and Gō model, to keep the secondary and tertiary structure of the proteins stable. SIRAH, is another alternative CG force field, that maps the atomic position of all-atom model to CG beads[64]. Although, SIRAH faces limitation to generalize across all amino acids. SIRAH use its own water, that is shown to match experimental condition[68]. SIRAH does not use any structural bias restraint to keep the structure stable and hence faces difficulty for stabilizing big systems[64]. SIRAH have been shown to work well with protein-peptide interactions but still it faces struggle in modeling conformational changes[64].

The restraints applied are different between Elastic model and Gō model. Elastic model uses harmonic potential or springs between protein backbone and the strength of the model is pretty strong to keep the model stable. It also faces with limitation for making the system overstable or "sticky" hindering dynamic conformational changes[66,67]. Elastic is applied typically distance 0.5 – 0.9 nm. Whereas Gō model uses contact map to apply modified LJ potentials. Gō model does not over-stabilize the system and retain some of its native flexibility[67]. For this study, I have used Elastic network with bond force constant 700 $kJ.mol^{-1}.nm^{-2}$, and the default value of upper and lower elastic bond cutoff of 0.5 and 0.9 nm. For Gō model, distance cutoff were kept as 0.3-1.1 nm and Gō potentials were kept 9.414 kJ/mol. In our system, I have observed Gō model to have ~50% less number of restraints, again keeping the model flexible enough to explore conformational changes.

For this work, I have considered Haloalkane Dehalogenase as model system as I already have the cMD simulation of LinB-Wt and the mutants. I have performed 5 x 5 μs replicas for Elastic and Gō model along with 5 x 5 μs replicas for SIRAH, totaling 300 μs of simulations. The simulations of Martini were performed using GROMACS[69] simulation engine and for SIRAH and cMD, I have used AMBER[70] simulation engine. For, SIRAH as it does not have any structural bias to keep the model stable, I developed in-house restraint protocol to keep my CG model stable. I have used dihedral angles restraints on residues in α-helices (4 kcal/mol/$Å^2$) and distance restraints on residues in β-sheets (20 kcal /mol/$Å^2$) to strengthen the backbone hydrogen bonds in between the strands (Figure 6). I observed significant improvement after applying the restraints on SIRAH models (**publication 3**).

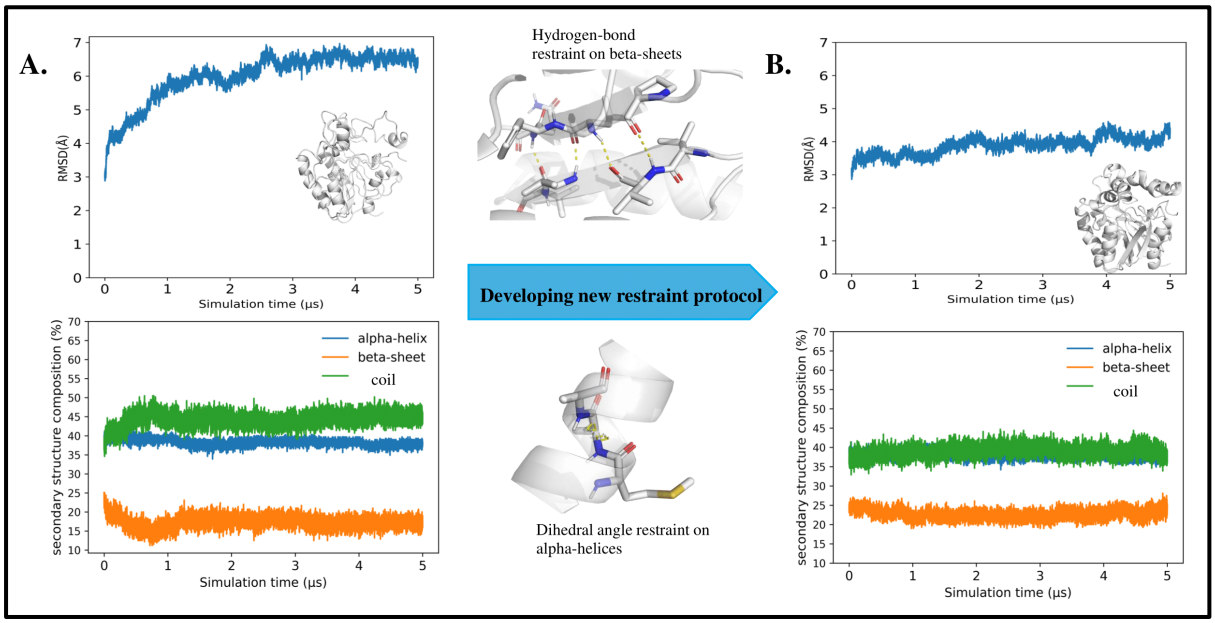


Figure 6. Effect of new restraint protocol in SIRAH. A) shows RMSD of LinB CG model using SIRAH and secondary structure elements without the restraints and **B)** shows stability of the model after using the restraint protocol and reduction of secondary structure loss.

Additionally, I have also benchmarked the CG methods by testing with three EC classes of enzymes. I have considered three proteins from each EC class, already capturing huge structural and functional diversity for the tunnel network exploration across the enzyme family. From EC1, I have considered three enzymes with buried active site, nitric oxide reductase that have a HEME cofactor (pdb id: 1jfb; all-alpha), anthocyanidin synthase (pdb id: 1gp4; all-beta), and delta(4)-3-ketosteroid 5-beta-reductase that have NADP cofactor (pdb id: 3bur; alpha/beta). From EC2, I have considered, arginine kinase (pdb id: 1m15; all-alpha), Bacillus subtilis levansucrase (pdb id: 1oyg; all-beta), and human cholesterol sulfotransferase SULT2B1b (pdb id: 1q20; alpha/beta) and Finally from EC3, I have analyzed phosphodiesterase 10 (pdb id: 2oup; all-alpha), Salmonella typhimurium LT2 neuraminidase (pdb id: 1dim; all-beta), and bacterial lipase (pdb id: 1cvl; alpha/beta). Structurally these proteins are shown in (Figure 7).

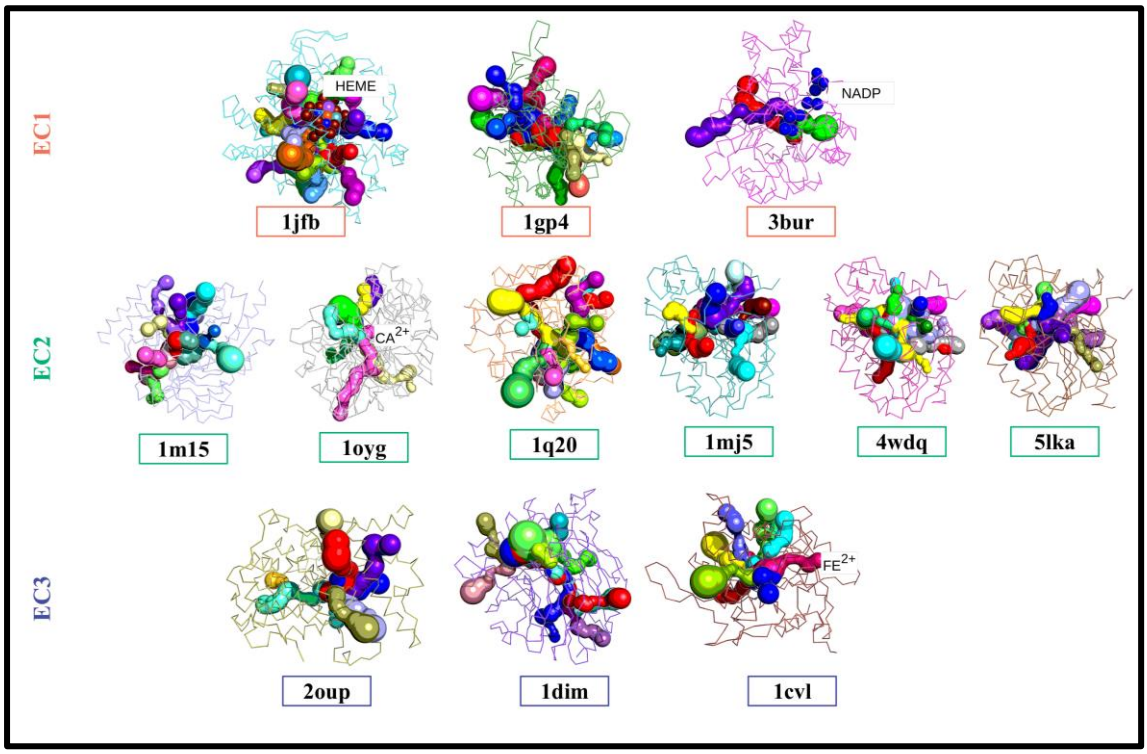


Figure 7. Enzymes from EC1-3 class captured through GōMartini3 CG model, showing structural and functional diversity of EC classes enzymes used for the study. The tunnels shown here are the representative of the average length and bottleneck radius, that occur at least in 10% of simulation frames.

I have performed 2μs single replica for these 9 enzymes, separately for each of these proteins captured through three methods. The Tunnels were calculated using CAVER and using divide-and-conquer[71] approach with updated CG bead size[62,64]. Again, The methods of choice to handle this huge dataset of Tunnel network was TT for unification and comparison between cMD and CG-MD. As of results, I found satisfactory performance of Elastic, Gō and SIRAH for identifying functional known and transient tunnels[72]. I observed many additional transient tunnels captured through CG method, that have little or no comparable counterparts from cMD, observed likely due to sampling enhancement. Although some of the tunnels might be false positive, but this rarely affect the tunnel network analysis. Also, through Elastic model, the comparison between permanent and transient tunnels were not accurate, showcasing the limitation possibly due to overstabilization. Overall, all CG methods reflect similar bottleneck radius and length and Gō excelled in replicating detailed tunnel engineering effects in LinB enzyme between LinB-Wt and mutants. Gō offered faster simulations and exploring broader conformational space, making it ideal for studying large protein systems. The detailed analysis is provided in **(publication 3).**

Conclusions and future perspectives

In conclusion, the thesis showcase significant insights into rare tunnel mechanisms in enzymes with buried active sites. Using advanced MD techniques such as GaMD along with CG-MD simulations, I have successfully captured varied tunnel network dynamic that were inaccessible through cMD simulations. GaMD proved effective in identifying rare tunnel opening, while CG methods provided faster alternative to study tunnel network dynamics for large protein systems or screening of tunnels in many proteins effectively. The thesis also shows development of powerful tool, TT, which is very effective for quantitative comparison between all the tunnels captured through GaMD and CG methods.

Through in depth comparison of various CG methods, the work demonstrated that CG Gō model, in particular can replicate the intricate detail of mutational tunnel effect. This has been possible due to effective coarse graining of amino acids in Martini 3.0 and also application of Gō model, that keep the model stable yet flexible enough to capture rare tunnels[73]. Despite its coarse nature, the method successfully captured both permanent p1, p2 and transient p3, ST tunnels. Furthermore, the important tunnel property such as the bottleneck radius and tunnel length were similar between CG Gō and cMD. CG methods can significantly accelerate simulations while accessing diverse conformational spaces, making them promising tools for studying complex biological systems.

This work help us advancing our understanding for enzyme tunnel networks. For future studied I would like to devise methodology that can improve detection of false tunnels from CG methods. Also, it would be interesting to study ligand transport simulations using CG methods, although to get the ligand structure is bit complicate but growing ligand library using Martini seems promising. This can be an natural extension for the study.

References

(1) 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.

(2) Krone, M.; Kozlíková, B.; Lindow, N.; Baaden, M.; Baum, D.; Parulek, J.; Hege, H. -C.; Viola, I. Visual Analysis of Biomolecular Cavities: State of the Art. *Comput. Graph. Forum* **2016**, *35* (3), 527–551. https://doi.org/10.1111/cgf.12928.

(3) 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.

(4) Petřek, M.; Otyepka, M.; Banáš, P.; Košinová, P.; Koča, J.; Damborský, J. CAVER: A New Tool to Explore Routes from Protein Clefts, Pockets and Cavities. *BMC Bioinformatics* **2006**, *7* (1), 316. https://doi.org/10.1186/1471-2105-7-316.

(5) Pravda, L.; Berka, K.; Svobodová Vařeková, R.; Sehnal, D.; Banáš, P.; Laskowski, R. A.; Koča, J.; Otyepka, M. Anatomy of Enzyme Channels. *BMC Bioinformatics* **2014**, *15* (1), 379. https://doi.org/10.1186/s12859-014-0379-x.

(6) Brezovsky, J.; Chovancova, E.; Gora, A.; Pavelka, A.; Biedermannova, L.; Damborsky, J. Software Tools for Identification, Visualization and Analysis of Protein Tunnels and Channels. *Biotechnol. Adv.* **2013**, *31* (1), 38–49. https://doi.org/10.1016/j.biotechadv.2012.02.002.

(7) Gora, A.; Brezovsky, J.; Damborsky, J. Gates of Enzymes. *Chem. Rev.* **2013**, *113* (8), 5871–5923. https://doi.org/10.1021/cr300384w.

(8) 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.

(9) Mitusińska, K.; Wojsa, P.; Bzówka, M.; Raczyńska, A.; Bagrowska, W.; Samol, A.; Kapica, P.; Góra, A. Structure-Function Relationship between Soluble Epoxide Hydrolases Structure and Their Tunnel Network. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 193–205. https://doi.org/10.1016/j.csbj.2021.10.042.

(10) 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.

(11) Banerjee, R.; Lipscomb, J. D. Small-Molecule Tunnels in Metalloenzymes Viewed as Extensions of the Active Site. *Acc. Chem. Res.* **2021**, *54* (9), 2185–2195. https://doi.org/10.1021/acs.accounts.1c00058.

(12) Marques, S. M., Brezovsky, J., Damborsky, J. *UNDERSTANDING ENZYMES: FUNCTION, DESIGN, ENGINEERING, AND ANALYSIS*; Pan Stanford Publishing, 2016.

(13) 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.

(14) Peplow, M. Cryo-Electron Microscopy Reaches Resolution Milestone. *ACS Cent. Sci.* **2020**, *6* (8), 1274–1277. https://doi.org/10.1021/acscentsci.0c01048.

(15) 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.
(16) 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.
(17) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450* (7172), 964–972. https://doi.org/10.1038/nature06522.
(18) 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.
(19) Sarkar, D. K.; Surpeta, B.; Brezovsky, J. Incorporating Prior Knowledge in the Seeds of Adaptive Sampling Molecular Dynamics Simulations of Ligand Transport in Enzymes with Buried Active Sites. *J. Chem. Theory Comput.* **2024**, acs.jctc.4c00452. https://doi.org/10.1021/acs.jctc.4c00452.
(20) 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.
(21) Capelli, R.; Carloni, P.; Parrinello, M. Exhaustive Search of Ligand Binding Pathways via Volume-Based Metadynamics. *J. Phys. Chem. Lett.* **2019**, *10* (12), 3495–3499. https://doi.org/10.1021/acs.jpclett.9b01183.
(22) Kingsley, L. J.; Lill, M. A. Including Ligand-Induced Protein Flexibility into Protein Tunnel Prediction. *J. Comput. Chem.* **2014**, *35* (24), 1748–1756. https://doi.org/10.1002/jcc.23680.
(23) Paloncýová, M.; Navrátilová, V.; Berka, K.; Laio, A.; Otyepka, M. Role of Enzyme Flexibility in Ligand Access and Egress to Active Site: Bias-Exchange Metadynamics Study of 1,3,7-Trimethyluric Acid in Cytochrome P450 3A4. *J. Chem. Theory Comput.* **2016**, *12* (4), 2101–2109. https://doi.org/10.1021/acs.jctc.6b00075.
(24) Fu, H.; Bian, H.; Shao, X.; Cai, W. Collective Variable-Based Enhanced Sampling: From Human Learning to Machine Learning. *J. Phys. Chem. Lett.* **2024**, *15* (6), 1774–1783. https://doi.org/10.1021/acs.jpclett.3c03542.
(25) Rydzewski, J.; Nowak, W. Ligand Diffusion in Proteins via Enhanced Sampling in Molecular Dynamics. *Phys. Life Rev.* **2017**, *22–23*, 58–74. https://doi.org/10.1016/j.plrev.2017.03.003.
(26) Sohraby, F.; Nunes-Alves, A. Advances in Computational Methods for Ligand Binding Kinetics. *Trends Biochem. Sci.* **2023**, *48* (5), 437–449. https://doi.org/10.1016/j.tibs.2022.11.003.
(27) Kokh, D. B.; Amaral, M.; Bomke, J.; Grädler, U.; Musil, D.; Buchstaller, H.-P.; Dreyer, M. K.; Frech, M.; Lowinski, M.; Vallee, F.; Bianciotto, M.; Rak, A.; Wade, R. C. Estimation of Drug-Target Residence Times by τ-Random Acceleration Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2018**, *14* (7), 3859–3869. https://doi.org/10.1021/acs.jctc.8b00230.
(28) Lüdemann, S. K.; Lounnas, V.; Wade, R. C. How Do Substrates Enter and Products Exit the Buried Active Site of Cytochrome P450cam? 1. Random Expulsion Molecular Dynamics Investigation of Ligand Access Channels and Mechanisms 1 1Edited by J. Thornton. *J. Mol. Biol.* **2000**, *303* (5), 797–811. https://doi.org/10.1006/jmbi.2000.4154.
(29) Miao, Y.; Bhattarai, A.; Wang, J. Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD): Characterization of Ligand Binding Thermodynamics and Kinetics. *J. Chem. Theory Comput.* **2020**, *16* (9), 5526–5547. https://doi.org/10.1021/acs.jctc.0c00395.

(30) Wang, J.; Miao, Y. Ligand Gaussian Accelerated Molecular Dynamics 2 (LiGaMD2): Improved Calculations of Ligand Binding Thermodynamics and Kinetics with Closed Protein Pocket. *J. Chem. Theory Comput.* **2023**, *19* (3), 733–745. https://doi.org/10.1021/acs.jctc.2c01194.
(31) 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.
(32) Muvva, C.; Murugan, N. A.; Kumar Choutipalli, V. S.; Subramanian, V. Unraveling the Unbinding Pathways of Products Formed in Catalytic Reactions Involved in SIRT1–3: A Random Acceleration Molecular Dynamics Simulation Study. *J. Chem. Inf. Model.* **2019**, *59* (10), 4100–4115. https://doi.org/10.1021/acs.jcim.9b00513.
(33) Klvana, M.; Pavlova, M.; Koudelakova, T.; Chaloupkova, R.; Dvorak, P.; Prokop, Z.; Stsiapanava, A.; Kuty, M.; Kuta-Smatanova, I.; Dohnalek, J.; Kulhanek, P.; Wade, R. C.; Damborsky, J. Pathways and Mechanisms for Product Release in the Engineered Haloalkane Dehalogenases Explored Using Classical and Random Acceleration Molecular Dynamics Simulations. *J. Mol. Biol.* **2009**, *392* (5), 1339–1356. https://doi.org/10.1016/j.jmb.2009.06.076.
(34) Marques, S. M.; Dunajova, Z.; Prokop, Z.; Chaloupkova, R.; Brezovsky, J.; Damborsky, J. Catalytic Cycle of Haloalkane Dehalogenases Toward Unnatural Substrates Explored by Computational Modeling. *J. Chem. Inf. Model.* **2017**, *57* (8), 1970–1989. https://doi.org/10.1021/acs.jcim.7b00070.
(35) Biedermannová, L.; Prokop, Z.; Gora, A.; Chovancová, E.; Kovács, M.; Damborský, J.; Wade, R. C. A Single Mutation in a Tunnel to the Active Site Changes the Mechanism and Kinetics of Product Release in Haloalkane Dehalogenase LinB. *J. Biol. Chem.* **2012**, *287* (34), 29062–29074. https://doi.org/10.1074/jbc.M112.377853.
(36) 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.
(37) 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.
(38) Mandal, N.; Surpeta, B.; Brezovsky, J. Deciphering Rare Opening of Gates in Ligand-Transport Tunnels of Enzymes Using Enhanced Molecular Dynamics Simulations. In *FEBS Open Bio*; 2022; Vol. 12, pp 67–336. https://doi.org/10.1002/2211-5463.13440.
(39) 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.
(40) Aurenhammer, F. Voronoi Diagrams—a Survey of a Fundamental Geometric Data Structure. *ACM Comput. Surv.* **1991**, *23* (3), 345–405. https://doi.org/10.1145/116873.116880.
(41) The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. http://www.pymol.org.
(42) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.
(43) Ingólfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; De Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. The Power of Coarse Graining in Biomolecular Simulations. *WIREs Comput. Mol. Sci.* **2014**, *4* (3), 225–248. https://doi.org/10.1002/wcms.1169.

(44) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116* (14), 7898–7936. https://doi.org/10.1021/acs.chemrev.6b00163.
(45) Borges-Araújo, L.; Patmanidis, I.; Singh, A. P.; Santos, L. H. S.; Sieradzan, A. K.; Vanni, S.; Czaplewski, C.; Pantano, S.; Shinoda, W.; Monticelli, L.; Liwo, A.; Marrink, S. J.; Souza, P. C. T. Pragmatic Coarse-Graining of Proteins: Models and Applications. *J. Chem. Theory Comput.* **2023**, *19* (20), 7112–7135. https://doi.org/10.1021/acs.jctc.3c00733.
(46) 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.
(47) Hasan, K.; Gora, A.; Brezovsky, J.; Chaloupkova, R.; Moskalikova, H.; Fortova, A.; Nagata, Y.; Damborsky, J.; Prokop, Z. The Effect of a Unique Halide-Stabilizing Residue on the Catalytic Properties of Haloalkane Dehalogenase DatA from *Agrobacterium Tumefaciens* C58. *FEBS J.* **2013**, *280* (13), 3149–3159. https://doi.org/10.1111/febs.12238.
(48) Stepankova, V.; Khabiri, M.; Brezovsky, J.; Pavelka, A.; Sykora, J.; Amaro, M.; Minofar, B.; Prokop, Z.; Hof, M.; Ettrich, R.; Chaloupkova, R.; Damborsky, J. Expansion of Access Tunnels and Active-Site Cavities Influence Activity of Haloalkane Dehalogenases in Organic Cosolvents. *ChemBioChem* **2013**, *14* (7), 890–897. https://doi.org/10.1002/cbic.201200733.
(49) 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.
(50) 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.
(51) 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.
(52) Mandal, N.; Surpeta, B.; Brezovsky, J. Reinforcing Tunnel Network Exploration in Proteins Using Gaussian Accelerated Molecular Dynamics. *J. Chem. Inf. Model.* **2024**, *64* (16), 6623–6635. https://doi.org/10.1021/acs.jcim.4c00966.
(53) 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.
(54) 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.
(55) 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.
(56) 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.
(57) 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.
(58) Filipovic, J.; Vavra, O.; Plhak, J.; Bednar, D.; Marques, S. M.; Brezovsky, J.; Matyska, L.; Damborsky, J. CaverDock: A Novel Method for the Fast Analysis of Ligand Transport. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2020**, *17* (5), 1625–1638. https://doi.org/10.1109/TCBB.2019.2907492.
(59) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.
(60) Lever, J.; Krzywinski, M.; Altman, N. Principal Component Analysis. *Nat. Methods* **2017**, *14* (7), 641–642. https://doi.org/10.1038/nmeth.4346.
(61) Ernst, M.; Sittel, F.; Stock, G. Contact- and Distance-Based Principal Component Analysis of Protein Dynamics. *J. Chem. Phys.* **2015**, *143* (24), 244114. https://doi.org/10.1063/1.4938249.
(62) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: A General Purpose Force Field for Coarse-Grained Molecular Dynamics. *Nat. Methods* **2021**, *18* (4), 382–388. https://doi.org/10.1038/s41592-021-01098-3.
(63) Periole, X.; Marrink, S.-J. The Martini Coarse-Grained Force Field. In *Biomolecular Simulations*; Monticelli, L., Salonen, E., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2013; Vol. 924, pp 533–565. https://doi.org/10.1007/978-1-62703-017-5_20.
(64) Machado, M. R.; Barrera, E. E.; Klein, F.; Sóñora, M.; Silva, S.; Pantano, S. The SIRAH 2.0 Force Field: Altius, Fortius, Citius. *J. Chem. Theory Comput.* **2019**, *15* (4), 2719–2733. https://doi.org/10.1021/acs.jctc.9b00006.
(65) Alessandri, R.; Barnoud, J.; Gertsen, A. S.; Patmanidis, I.; de Vries, A. H.; Souza, P. C. T.; Marrink, S. J. Martini 3 Coarse-Grained Force Field: Small Molecules. *Adv. Theory Simul.* **2022**, *5* (1), 2100391. https://doi.org/10.1002/adts.202100391.
(66) Periole, X.; Cavalli, M.; Marrink, S.-J.; Ceruso, M. A. Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J. Chem. Theory Comput.* **2009**, *5* (9), 2531–2543. https://doi.org/10.1021/ct9002114.
(67) Souza, P. C. T.; Borges-Araújo, L.; Brasnett, C.; Moreira, R. A.; Grünewald, F.; Park, P.; Wang, L.; Razmazma, H.; Borges-Araújo, A. C.; Cofas-Vargas, L. F.; Monticelli, L.; Mera-Adasme, R.; Melo, M. N.; Wu, S.; Marrink, S. J.; Poma, A. B.; Thallmair, S. GōMartini 3: From Large Conformational Changes in Proteins to Environmental Bias Corrections. *bioarxiv*, 2024.04.15.589479, April 2024. https://doi.org/10.1101/2024.04.15.589479.
(68) Darré, L.; Machado, M. R.; Pantano, S. Coarse-Grained Models of Water: Coarse-Grained Models of Water. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2* (6), 921–930. https://doi.org/10.1002/wcms.1097.
(69) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism

from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

(70) D. A. Case; I.Y. Ben-Shalom; S.R. Brozell; D.S. Cerutti; Cheatham, T. E.; V.W.D. Cruzeiro; T.A. Darden; R.E. Duke; D. Ghoreishi; M.K. Gilson; H. Gohlke; A.W. Goetz; D. Greene; Harris, R. C.; N. Homeyer; Yandong Huang; S. Izadi; A. Kovalenko; T. Kurtzman; T.S. Lee; S. LeGrand; P. Li; C. Lin; J. Liu; T. Luchko; R. Luo; D.J. Mermelstein; K.M. Merz; Y. Miao; G. Monard; C. Nguyen; H. Nguyen; I. Omelyan; A. Onufriev; F. Pan; R. Qi; D.R. Roe; A. Roitberg; C. Sagui; S. Schott-Verdugo; Shen, J.; C.L. Simmerling; J. Smith; R. Salomon- Ferrer; J. Swails; R.C. Walker; J. Wang; H. Wei; R.M. Wolf; X. Wu; L. Xiao; D.M. York; P.A. Kollman. Amber 2018, 2018. https://doi.org/10.13140/RG.2.2.31525.68321.

(71) 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.

(72) Mandal, N.; Stevens, J. A.; Poma, A. B.; Surpeta, B.; Sequeiros-Borja, C.; Thirunavukarasu, A. S.; Marrink, S.-J.; Brezovsky, J. Benchmarking Coarse-Grained Simulation Methods for Investigation of Transport Tunnels in Enzymes. *bioRxiv,* 2024.09.16.613244, September 2024. https://doi.org/10.1101/2024.09.16.613244.

(73) Poma, A. B.; Cieplak, M.; Theodorakis, P. E. Combining the MARTINI and Structure-Based Coarse-Grained Approaches for the Molecular Dynamics Studies of Conformational Transitions in Proteins. *J. Chem. Theory Comput.* **2017**, *13* (3), 1366–1374. https://doi.org/10.1021/acs.jctc.6b00986.

PUBLICATIONS

PUBLICATION 1

TransportTools: A Library for High-Throughput Analyses of Internal Voids in Biomolecules and Ligand Transport through Them

Impact Factor 2021: 6.931

MNiSW points 2022: 200

Citations: 11

Main text (available online):

https://doi.org/10.1093/bioinformatics/btab872

Supplementary information (available online):

https://academic.oup.com/bioinformatics/article/38/6/1752/6491234#supplementary-data

Source code:

https://github.com/labbit-eu/transport_tools

Repository:

https://doi.org/10.5281/zenodo.5642954

PUBLICATION 2

Reinforcing Tunnel Network Exploration in Proteins Using Gaussian Accelerated Molecular Dynamics

Impact Factor 2023: 5.6

MNiSW points 2024: 100

Citations: 1

Main text (available online):

https://doi.org/10.1021/acs.jcim.4c00966

Supplementary information (available online):

https://pubs.acs.org/doi/10.1021/acs.jcim.4c00966

Repositories:

https://zenodo.org/doi/10.5281/zenodo.11092891

https://zenodo.org/doi/10.5281/zenodo.11093856

PUBLICATION 3

Benchmarking coarse-grained simulation methods for investigation of transport tunnels in enzymes

Impact Factor 2023: N/A

MNiSW points 2024: N/A

Citations: 0

Main text (available online):

https://www.biorxiv.org/content/10.1101/2024.09.16.613244v1

Supplementary information (available online):

https://www.biorxiv.org/content/10.1101/2024.09.16.613244v1.supplementary-material

CO-AUTHOR STATEMENTS

PUBLICATION 1

TransportTools: A Library for High-Throughput Analyses of Internal Voids in Biomolecules and Ligand Transport through Them

1. Mandal, N.;

2. Brezovsky, J.;

3. Thirunavukarasu, A. S.;

4. Surpeta, B.;

5. Sequeiros-Borja, C.;

6. Sarkar, D.K.

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytutu Biologii Molekularnej i Biotechnologii UAM

Poznan, September 20, 2024

Nishita Mandal, M.Sc.
Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland
Email: nishita.mandal@amu.edu.pl

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:

Brezovsky J, Thirunavukarasu AS, Surpeta B, Sequeiros-Borja CE, Mandal N, Sarkar DK, Dongmo Foumthuim CJ, Agrawal N, 2021: TransportTools: A Library for High-Throughput Analyses of Internal Voids in Biomolecules and Ligand Transport Through Them. *Bioinformatics 38:* 1752-1753, DOI: 10.1093/bioinformatics/btab872.

I declare the following contribution to this publication:

contributed to user-testing of the developed tools; generated data for Use-case 2 in Supplementary File 7 and wrote the draft of this document.

.....
Signature

.....
Promotor
dr hab. Jan Brezovsky, prof. UAM

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytutu Biologii Molekularnej i Biotechnologii UAM

Jan Brezovsky, Ph.D.
Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland
Email: janbre@amu.edu.pl

Poznan, March 14, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:

Brezovsky J, Thirunavukarasu AS, Surpeta B, Sequeiros-Borja CE, Mandal N, Sarkar DK, Dongmo Foumthuim CJ, Agrawal N, 2021: TransportTools: A Library for High-Throughput Analyses of Internal Voids in Biomolecules and Ligand Transport Through Them. *Bioinformatics 38:* 1752-1753, DOI: 10.1093/bioinformatics/btab872.

I declare the following contribution to this publication:

conceived the research project and the new computational tool; designed and implemented a large majority of algorithms and tests, and prepared the user and technical documentation for the tool; coordinated the work of the project team; analyzed and interpreted the data; and wrote the manuscript.

.....
Signature

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytutu Biologii Molekularnej i Biotechnologii UAM

Aravind Selvaram Thirunavukarasu, M.Tech.
Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland
Email: arathi@amu.edu.pl

Poznan, March 14, 2023

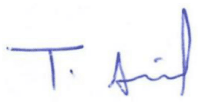
CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:

Brezovsky J, Thirunavukarasu AS, Surpeta B, Sequeiros-Borja CE, Mandal N, Sarkar DK, Dongmo Foumthuim CJ, Agrawal N, 2021: TransportTools: A Library for High-Throughput Analyses of Internal Voids in Biomolecules and Ligand Transport Through Them. *Bioinformatics 38:* 1752-1753, DOI: 10.1093/bioinformatics/btab872.

I declare the following contribution to this publication:

contributed to user-testing of the developed tools; generated data for performance evaluation of the TransportTools and Use-case 1 summarized in Supplementary File 5 and 6 and wrote drafts of these two documents.



.....
Signature

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytutu Biologii Molekularnej i Biotechnologii UAM

Bartłomiej Surpeta, M.Sc.
Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland
Email: bartlomiej.surpeta@amu.edu.pl

Poznan, March 14, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:

Brezovsky J, Thirunavukarasu AS, Surpeta B, Sequeiros-Borja CE, Mandal N, Sarkar DK, Dongmo Foumthuim CJ, Agrawal N, 2021: TransportTools: A Library for High-Throughput Analyses of Internal Voids in Biomolecules and Ligand Transport Through Them. *Bioinformatics 38:* 1752-1753, DOI: 10.1093/bioinformatics/btab872.

I declare the following contribution to this publication:

contributed to user-testing of the developed tools; wrote the draft of state of the art overview in Supplementary File 1; conceived and written the tutorial included in the user guide.

Bartłomiej Surpeta

.....
Signature

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytutu Biologii Molekularnej i Biotechnologii UAM

Carlos Eduardo Sequeiros-Borja, M.Sc.
Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland
Email: carseq@amu.edu.pl

Poznan, March 14, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:

Brezovsky J, Thirunavukarasu AS, Surpeta B, Sequeiros-Borja CE, Mandal N, Sarkar DK, Dongmo Foumthuim CJ, Agrawal N, 2021: TransportTools: A Library for High-Throughput Analyses of Internal Voids in Biomolecules and Ligand Transport Through Them. *Bioinformatics 38:* 1752-1753, DOI: 10.1093/bioinformatics/btab872.

I declare the following contribution to this publication:

contributed to user-testing of the developed tools; implemented algorithms for structural alignment and surface visualization of tunnel clusters; edited the manuscript.

Carlos Eduardo Sequeiros Borja

.....
Signature

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU

Dheeraj Kumar Sarkar, M.Sc.
Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland
Email: dhesar@amu.edu.pl

Poznan, March 14, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:

Brezovsky J, Thirunavukarasu AS, Surpeta B, Sequeiros-Borja CE, Mandal N, Sarkar DK, Dongmo Foumthuim CJ, Agrawal N, 2021: TransportTools: A Library for High-Throughput Analyses of Internal Voids in Biomolecules and Ligand Transport Through Them. *Bioinformatics 38:* 1752-1753, DOI: 10.1093/bioinformatics/btab872.

I declare the following contribution to this publication:

contributed to user-testing of the developed tools; generated data for Use-case 3 in Supplementary File 8 and wrote the draft of this document.

Dheeraj kumar Sarkar

.....
Signature

ul. Uniwersytetu Poznańskiego 6, Collegium Biologicum, 61-614 Poznań
NIP 777 00 06 350, REGON 000001293
tel. +48 61 829 59 50, ibmib@amu.edu.pl

www.ibmib.amu.edu.pl

PUBLICATION 2

Reinforcing Tunnel Network Exploration in Proteins Using Gaussian Accelerated Molecular Dynamics

1. Mandal, N.;

2. Surpeta, B.;

3. Brezovsky, J.

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytutu Biologii Molekularnej i Biotechnologii UAM

Poznan, September 20, 2024

Nishita Mandal, M.Sc.
Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland
Email: nishita.mandal@amu.edu.pl

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:

Mandal N, Surpeta B, Brezovsky J, 2024: Reinforcing Tunnel Network Exploration in Proteins using Gaussian Accelerated Molecular Dynamics. Journal of Chemical Information and Modeling, 64: 6623-6635. https://doi.org/10.1021/acs.jcim.4c00966

I declare the following contribution to this publication:

setup all systems for classical and Gaussian accelerated MD, performed all MD simulations, calculations of tunnels, cryptic pocket analysis, migration analysis, H-bond analysis for ST mechanism, calculated distance based PCA for side helix movement; analyzed, interpreted and visualized the data; and co-wrote the manuscript.

Signature

Promotor
dr hab. Jan Brezovsky, prof. UAM

Poznan, September 19, 2024

dr inż. Bartłomiej Surpeta, Ph.D.
International Institute of Molecular and Cell Biology in Warsaw, 02-109 Warsaw
Email: bsurpeta@iimcb.gov.pl

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am a co-author of the following publication:

Mandal N, **Surpeta B**, Brezovsky J, 2024: Reinforcing Tunnel Network Exploration in Proteins using Gaussian Accelerated Molecular Dynamics. Journal of Chemical Information and Modeling, 64: 6623-6635. https://doi.org/10.1021/acs.jcim.4c00966

I declare the following contribution to this publication:

devised the tunnel re-weighting, principal component analysis (PCA), and clustering analysis protocols; co-devised and co-supervised the project; co-wrote the manuscript

Bartłomiej Surpeta

.....

Signature

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytutu Biologii Molekularnej i Biotechnologii UAM

Poznan, September 19, 2024

dr hab. Jan Brezovsky, Ph.D., prof. UAM
Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland
Email: janbre@amu.edu.pl

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:

Mandal N, Surpeta B, Brezovsky J, 2024: **Reinforcing Tunnel Network Exploration in Proteins using Gaussian Accelerated Molecular Dynamics.** *Journal of Chemical Information and Modeling*, 64: 6623-6635. https://doi.org/10.1021/acs.jcim.4c00966

I declare the following contribution to this publication:

co-devised and co-supervised the project; interpreted the results; co-wrote the manuscript

.....
Signature

ul. Uniwersytetu Poznańskiego 6, Collegium Biologicum, 61-614 Poznań
NIP 777 00 06 350, REGON 000001293
tel. +48 61 829 59 50, ibmib@amu.edu.pl

www.ibmib.amu.edu.pl

PUBLICATION 3

Benchmarking coarse-grained simulation methods for investigation of transport tunnels in enzymes

1. Mandal, N.;

2. Stevens, J. A.;

3. Poma, A. B.;

4. Surpeta, B.;

5. Sequeiros-Borja, C. ;

6. Thirunavukarasu, A.S.;

7. Brezovsky, J.

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytutu Biologii Molekularnej i Biotechnologii UAM

Poznan, September 20, 2024

Nishita Mandal, M.Sc.
Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland
Email: nishita.mandal@amu.edu.pl

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:

Mandal N, Stevens JA, Poma AB, Surpeta B, Sequeiros-Borja C, Thirunavukarasu AS, Marrink SJ, Brezovsky J, 2024: Benchmarking coarse-grained simulation methods for investigation of transport tunnels in enzymes. BioRxiv 2024.09.16.613244. https://doi.org/10.1101/2024.09.16.613244

I declare the following contribution to this publication:

setup LinB-Wt and mutants All-atom models, LinB mutants Coarse Grained (CG) models, EC class 1-3 enzyme CG models, performed all the MD simulations from AA and CG models using AMBER and GROMACS simulation engine, implemented CG tunnels calculation using beads from CG models, performed calculations of tunnels, correlation analysis, co-devised restraint protocol for SIRAH CG model; analyzed, interpreted and visualized the data; prepared the figures; co-devised the project and co-acquired the funding; and co-wrote the manuscript.

.....
Signature

.....
Promotor
dr hab. Jan Brezovsky, prof. UAM

university of groningen

Groningen, September 20, 2024

M.Sc. Jan A. Stevens
Molecular Dynamics Group, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 7, 9747 AG Groningen, the Netherlands
Email: j.a.stevens@rug.nl

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:

Mandal N, Stevens JA, Poma AB, Surpeta B, Sequeiros-Borja C, Thirunavukarasu AS, Marrink SJ, Brezovsky J, 2024: Benchmarking coarse-grained simulation methods for investigation of transport tunnels in enzymes. bioRxiv 2024.09.16.613244.
https://doi.org/10.1101/2024.09.16.613244

I declare the following contribution to this publication:

contributed to setup and optimization of CG simulation protocol with Martini Elastic model, and edited the manuscript.

Jan A. Stevens - 20/09/24

Signature

INSTYTUT PODSTAWOWYCH PROBLEMÓW TECHNIKI

POLSKIEJ AKADEMII NAUK

02-106 Warszawa, ul. A. Pawińskiego 5B, Fax (+48)22 8269815, Telefon centr. (+48)22 8261281, Dyrektor (+48)22 8268911
NIP:525 000 89 79 REGON: 326026 http://www.ippt.pan.pl

Warsaw, September 19, 2024

Adolfo B. Poma, Ph.D.
Biosystems and Soft Matter Division,
Institute of Fundamental Technological Research, Polish Academy of Sciences
ul. Pawińskiego 5B, 02-106 Warsaw, Poland
Email: apoma@ippt.pan.pl

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:
Mandal N, Stevens JA, Poma AB, Surpeta B, Sequeiros-Borja C, Thirunavukarasu AS, Marrink SJ, Brezovsky J, 2024: **Benchmarking coarse-grained simulation methods for investigation of transport tunnels in enzymes**. *bioRxiv 2024.09.16.613244.* https://doi.org/10.1101/2024.09.16.613244
I declare the following contribution to this publication:
helped setup and optimization of CG simulation protocol with Gō-Martini model for LinB-Wt, supervised its application to other systems, and edited the manuscript.

.....

Signature

Poznan, September 19, 2024

dr inż. Bartłomiej Surpeta, Ph.D.
International Institute of Molecular and Cell Biology in Warsaw, 02-109 Warsaw
Email: bsurpeta@iimcb.gov.pl

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:

Mandal N, Stevens JA, Poma AB, Surpeta B, Sequeiros-Borja C, Thirunavukarasu AS, Marrink SJ, Brezovsky J, 2024: **Benchmarking coarse-grained simulation methods for investigation of transport tunnels in enzymes**. *bioRxiv 2024.09.16.613244.* https://doi.org/10.1101/2024.09.16.613244

I declare the following contribution to this publication:

jointly devised restraint protocol for stabilization of SIRAH CG model, selected diverse EC2 class enzymes with suitable tunnel networks, and performed their parametrization and optimization of their AA simulation protocol, and edited the manuscript.

Bartłomiej Surpeta
.....
Signature

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytutu Biologii Molekularnej i Biotechnologii UAM

Poznan, September 19, 2024

Carlos Sequeiros-Borja, Ph.D.
Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland
Email: carseq@amu.edu.pl

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:

Mandal N, Stevens JA, Poma AB, Surpeta B, Sequeiros-Borja C, Thirunavukarasu AS, Marrink SJ, Brezovsky J, 2024: **Benchmarking coarse-grained simulation methods for investigation of transport tunnels in enzymes**. *bioRxiv 2024.09.16.613244.* https://doi.org/10.1101/2024.09.16.613244

I declare the following contribution to this publication:

selected diverse EC1 class enzymes with suitable tunnel networks, and performed their parametrization and optimization of their AA simulation protocol, and edited the manuscript.

Carlos Eduardo Sequeiros Borja

.....
Signature

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytutu Biologii Molekularnej i Biotechnologii UAM

Poznan, September 19, 2024

Aravind Selvaram Thirunavukarasu, Ph.D.
Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland
Email: arathi@amu.edu.pl

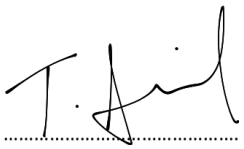
CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:

Mandal N, Stevens JA, Poma AB, Surpeta B, Sequeiros-Borja C, Thirunavukarasu AS, Marrink SJ, Brezovsky J, 2024: **Benchmarking coarse-grained simulation methods for investigation of transport tunnels in enzymes**. *bioRxiv 2024.09.16.613244.* https://doi.org/10.1101/2024.09.16.613244

I declare the following contribution to this publication:

selected diverse EC3 class enzymes with suitable tunnel networks, and performed their parametrization and optimization of their AA simulation protocol, and edited the manuscript.



Signature

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Wydział Biologii
Instytutu Biologii Molekularnej i Biotechnologii UAM

Poznan, September 19, 2024

dr hab. Jan Brezovsky, Ph.D., prof. UAM
Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland
Email: janbre@amu.edu.pl

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication:

Mandal N, Stevens JA, Poma AB, Surpeta B, Sequeiros-Borja C, Thirunavukarasu AS, Marrink SJ, Brezovsky J, 2024: **Benchmarking coarse-grained simulation methods for investigation of transport tunnels in enzymes**. *bioRxiv 2024.09.16.613244.* https://doi.org/10.1101/2024.09.16.613244

I declare the following contribution to this publication:

jointly conceived the project, acquired funding, analyzed the data, and interpreted the results; co-wrote the draft of the manuscript; and coordinated the project.

.....
Signature

CURRICULUM VITAE

Nishita Mandal

PhD Student
Laboratory of Biomolecular Interactions and Transport Adam Mickiewicz University, 61-614, Poznan, Poland & International Institute of Molecular and Cell Biology in Warsaw, Poland
Phone number: (+48) 511619393,
Email address: nisman@amu.edu.pl ,www.linkedin.com/in/nishita-mandal-2b9b93276 , https://x.com/MandalNishita, https://orcid.org/0000-0002-9114-0281, https://scholar.google.com/citations?user=xmXoBA8AAAAJ&hl=en&oi=ao

Proficient in structural bioinformatics and molecular dynamics with ~5 years of experience. Currently pursuing PhD in computational structural biology.

SKILLS

- Molecular dynamics of proteins (AMBER and GROMACS)
- Enhanced sampling methods – Gaussian accelerated MD and Coarse grained simulations
- Docking of small molecules
- Cryptic pocket analysis
- Protein tunnels analysis – CAVER, Transport tools, CaverDock
- Protein structure analysis and visualization: • 3D Structure prediction / Homology modelling: AlphaFold2, Modeller, I-TASSER • Mutagenesis analysis tools: dbSNP, I-Mutant 3.0, HotSpot Wizard 3.0 • Molecular interactions: cpptraj, pytraj • Visualizer: PyMol, VMD, Blender, Chimera
- Programming / Scripting: Bash, Python

RESEARCH EXPERIENCE AND EDUCATION

Principal Investigator
Adam Mickiewicz University, Poznan, Poland — Jan 2024 – Present

Grant: National Science Center (NCN), Preludium, Poland

- Project title: "Optimization of coarse-grained molecular dynamics simulations for tunnel exploration of large proteins with buried active sites leveraging an encouraging speed-accuracy compromise"
- Responsible for performing the calculations of coarse-grained (CG-MD) simulations for enzyme tunnels.
- Managing the allocated funds for the project, presentations in scientific conferences, and publishing the work in peer-reviewed scientific journals.

PhD Student
Adam Mickiewicz University, Poznan, Poland — Nov 2019 – Present

Supervisor: Dr. Jan Brezovsky
Fellowship: National Science Center (NCN), Poland

- To perform Molecular Dynamics (MD) simulations using enhanced sampling method like Gaussian accelerated MD (GaMD) simulations and coarse-grained (CG-MD) simulations for enzymes and analysing transport pathways from buried active sites to bulk solvent.
- Contributed in development of a python library for massive analyses of biomolecular voids and utilization of ligands.
- To study biological relevance of rare transport processes in haloalkane dehalogenase using MD simulations.

Research Internship

University of Groningen, Groningen, The Netherlands June 2022 – July 2022

- 1 month internship in University of Groningen, Netherlands under Prof. Dr. Siewert-Jan Marrink to setup coarse grained model for haloalkane dehalogenase and it's mutants using Martini Elastic Network model.

DBT-COE Project trainee

Bose institute, Division of Bioinformatics, Kolkata, India Nov 2018 – Nov 2019

Supervisor: Dr. Shubhra Ghosh Dastidar
Fellowship: DBT-COE, India

- Investigated molecular mechanisms of the Taxol and Taxotere to influence the intrinsic dynamics of α, β-Tubulin.
- The microtubule over-stabilizing ligands that arrest the mitotic cell division by preventing spindle dissociation, is a promising strategy to fight against cancers.
-

Master of Science (H) Biochemistry and Master's Dissertation

Banaras Hindu University, Varanasi, India June 2016 – July 2018

Supervisor: Dr. S. C. Gupta

- Examined the effects of curcuminoid on the C6 glioma cells.
- The efficacy of curcuminoid on cell proliferation, migration, apoptosis and NFκB (p65) activation was measured using clonogenic assay, cell migration assay, AO/ PI staining, and immunofluorescence analysis on C6 cells.

Summer Trainee

CCMB, Hyderabad, India June 2017 – August 2017

Supervisor: Dr. Tushar Vaidya

- Development of the biochemical aspects and function of 3'UTR of DRG1 and DRG2 genes of *Leishmania*

Bachelors of Science (H) Biochemistry

Delhi University, New Delhi, India June 2013 – July 2016

HONOURS AND AWARDS

- Awarded the prestigious PRELUDIUM grant from National Science Centre Poland, to carry out coarse grained simulations and tunnel analysis (~11% acceptance rate / year), nr - 2023/49/N/NZ2/02567 in collaboration with Prof. Siewert Jan Marrink, University of Groningen.
- Best Research grant – Uniwersytetu im. Adama Mickiewicza w Poznaniu Best Research grant from ID-UB "02 - Wsparcie udziału naukowców i doktorantów w prestiżowych konferencjach naukowych" nr 037/02/POB2/0011.
- Internship grant – Uniwersytetu im. Adama Mickiewicza w Poznaniu Internship grant from ID-UB "13 - Wsparcie umiędzynarodowienia badań naukowych prowadzonych w Szkole Doktorskiej UAM" nr 04 8/13/UAM/0016 for 2 weeks internship in University of Groningen, Netherlands.
- FEBS Bursary award – IUBMB–FEBS–PABMB Congress FEBS Bursary award to attend the conference as young scientist.
- Travel grant – Uniwersytetu im. Adama Mickiewicza w Poznaniu Travel grant from ,,Uniwersytet Jutra – zintegrowany program rozwoju Uniwersytetu im. Adama Mickiewicza w Poznaniu", nr POWR. 03.05.00-00-Z303/17" to attend the BioSB conference in June, 2021

PUBLICATIONS

- N Mandal, J A. Stevens, A B. Poma, B Surpeta, C Sequeiros-Borja, A S Thirunavukarasu, S J Marrink and J Brezovsky*, Benchmarking coarse-grained simulation methods for investigation of transport tunnels in enzyme *bioRxiv, 2024.09.16.613244,* September 2024, https://doi.org/10.1101/2024.09.16.613244

- N Mandal, B Surpeta* and J Brezovsky*. Reinforcing Tunnel Network Exploration in Proteins using Gaussian Accelerated Molecular Dynamics, *Journal of Chemical Information and Modeling* (2024). https://doi.org/10.1021/acs.jcim.4c00966

- J Brezovsky*, AS Thirunavukarasu, B Surpeta, CE Sequeiros-Borja, N Mandal, DK Sarkar, CJ Dongmo Foumthuim, N Agrawal. TransportTools: a library for high-throughput analyses of internal voids in biomolecules and ligand transport through them, Bioinformatics (2022). https://doi.org/10.1093/bioinformatics/btab872

- N Mandal, B Surpeta, J Brezovsky. Deciphering rare opening of gates in ligand-transport tunnels of enzymes using enhanced molecular dynamics simulations, Febs open bio (2022). ***Conference poster proceedings***, https://doi.org/10.1002/2211-5463.13440

- D Base, S Majumdar, N Mandal, SG Dastidar. Mechanisms of Influence of the Microtubule Over - Stabilizing Ligands on the Structure and Intrinsic Dynamics of α,β-Tubulin, Computational Biology and chemistry (2021). https://doi.org/10.1016/j.compbiolchem.2021.107617

WORKSHOPS AND CONFERENCES

- Two contributed talks at ISQBP 2024: Meeting of the International Society of Quantum Biology and Pharmacology, Athens, Greece, May 2024.
- Poster presentation at EMBO Workshop: Workshop on Computational structural biology, Heidelberg, Germany, December 2023.
- Poster presentation at the Young Modellers conference at Toulouse, France, May 2023.
- Invited Talk at PRACE-LAB Summit 2022, Poznan, Poland, October 2022.
- Poster presentation at FEBS congress as young scientist, Lisbon, Portugal, July 2022.
- Poster presentation at the APFED-22 conference at Bayreuth, Germany, May 2022.
- Invited Talk at Bose Institute, India, January 2022.
- Oral talk under young scientific investigator's session, 4th CEBC, Krakow, Poland, June 2021.

TEACHING EXPERIENCE

- 3 years teaching experience to B.Sc. students (Structural Bioinformatics Practical course) as main and assistant tutor.

“What I cannot create, I do not understand.”

– Richard Feynman