Uniwersytet im. Adama Mickiewicza, Poznań, Polska

Wydział Biologii

Rozprawa doktorska

Rola tuneli przejściowych w funkcjonowaniu enzymów z głęboko osadzonymi miejscami aktywnymi

mgr inż. **Aravind Selvaram Thirunavukarasu**

Promotor

dr hab. Jan Brezovsky, prof. UAM

Uniwersytet im. Adama Mickiewicza, Poznań, Polska 2023

Adam Mickiewicz University in Poznań

Faculty of Biology

Doctoral thesis

Role of Transient Tunnels in Function of Enzymes with Buried Active Sites

Aravind Selvaram Thirunavukarasu, M. Tech.

Promotor

dr. hab. Jan Brezovsky, prof. UAM

Adam Mickiewicz University, Poznań, Poland 2023

Acknowledgements

My supervisor and mentor, **Prof. Jan Brezovsky**, deserves my heartfelt thanks for giving me the chance to join his research group and for guiding me with his wisdom and expertise throughout my Ph.D. journey.

I also appreciate the support and advice of **Prof. Agnieszka Ludwików**, the head of Ph.D., studies.

I am thankful to **dr. Karolina Cerbin, Ph.D**., and **Magdalena Walczak M.Sc**., who assisted me with the administrative matters related to my PhD program and residence permit.

I acknowledge the support of **Arleta Kucz, M.Sc.** and **Iwona Kanonik-Jędrzejak, M.Sc.** in administrative support related to travel grants and acquiring equipments for my research.

I am indebted to my **colleagues**, who offered me their moral and scientific support. They enriched my learning experience with their insights and feedback.

I would like to express my gratitude to my collaborators **Prof. Artur Góra, M.Sc. Katarzyna Szleper, dr. Gamze Tanriver, dr. Karolina Mitusińska** for their support and research assistance.

I thank the Department of Gene expression, **UAM** and International Institute of Molecular and Cell Biology (**IIMCB**) for creating a conducive research environment and for facilitating networking opportunities.

Lastly, I thank the **Poznan Supercomputing Center** for providing me with access to their supercomputer resources.

Dedicated to,

To my parents Dhanalakshmi and Thirunavukarasu for their motivation and being there for me when I needed them.

My wife Keerthi, for her constant support and encouragement.

Table of Contents

List of publications

1. Research article (**Impact Factor 2021: 6.931**, MNiSW points 2022: 200) 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

2. Research article (Impact Factor 2023: 0, MNiSW points 2023: 0) Sequeiros-Borja, C. E.; **Thirunavukarasu, A. S.**; Dongmo Foumthuim, C. J.;Brezovsky, J. Water will find a way: transport through narrow tunnels in hydrolases. bioRxiv 2023; doi: https://doi.org/10.1101/2023.05.24.542065

3. Research article (Impact Factor 2023: 0, MNiSW points 2023: 0) **Thirunavukarasu, A. S.,** Szleper, K., Tanriver, G., Mitusinska, K., Gora, A., Brezovsky, J., Water migration through enzyme tunnels is sensitive to choice of explicit water model. bioRxiv 2023, doi: https://doi.org/10.1101/2023.08.14.553223

Streszczenie

Enzymy z głęboko osadzonymi miejscami aktywnymi mają prowadzące do nich ścieżki, które nazywane są tunelami. Skupiając się na rodzinie enzymów hydrolaz i podkreślając znaczenie wody w tych tunelach, niniejsze badanie zapewnia cenny wgląd w ich funkcjonalne implikacje. Dzięki wykorzystaniu symulacji dynamiki molekularnej i adaptacyjnych technik próbkowania, stany konformacyjne systemów są dokładnie badane. W szczególności obliczenia tuneli prowadzących do miejsca aktywnego są przeprowadzane dla wszystkich ramek obszernej symulacji, co prowadzi do identyfikacji wcześniej nierozpoznanych tuneli. Śledząc ruch cząsteczek wody z objętości do określonego regionu (miejsca aktywnego) białka, określane są ścieżki transportu wykorzystywane przez enzym podczas symulacji. Aby ułatwić analizę ogromnych ilości danych związanych z tunelami i cząsteczkami wody, opracowano moduł Python o nazwie TransportTools. Moduł ten usprawnia analizę obszernych danych dotyczących tuneli i cząsteczek wody. Pomaga w ocenie wykorzystania cząsteczek wody, wydajności transportu i funkcjonalności w strukturach białkowych. Narzędzie to odkryło nowe tunele transportujące wodę w enzymach takich jak dehalogenaza haloalkanowa, zwiększając nasze zrozumienie dynamiki tuneli. Ponadto odkryto, że cząsteczki wody mogą skutecznie poruszać się w wąskich regionach tuneli poprzez interakcje wiązań wodorowych. Odkrycia te zostały potwierdzone poprzez porównanie hydrolazy epoksydowej typu dzikiego z jej mutantem E470G, w którym zaobserwowano wzrost wykorzystania alternatywnych tuneli w mutancie. Co więcej, wybór jawnych modeli wody, takich jak OPC, TIP3P i TIP4P-Ew, znacząco wpływa na wykorzystanie tuneli enzymatycznych przez cząsteczki wody. Zaobserwowane różnice między trzema modelami wody można przypisać różnicom w ich właściwościach dyfuzyjnych i rozkładach ładunku. W przypadku modelu TIP3P migracja wody była zauważalnie przyspieszona, transportując około 2,5- i 2,0-krotnie więcej cząsteczek wody niż w przypadku modeli OPC i TIP4P-Ew. Zjawisko to wynika przede wszystkim z zaobserwowanych krótszych czasów przejścia i większej liczby cząsteczek wody jednocześnie migrujących przez tunele przy zastosowaniu modelu TIP3P. Spójność tych wyników dla różnych enzymów, takich jak oksydaza alditolowa i cytochrom P450 2D6, wraz z geometrią tuneli, podkreśla ich szersze znaczenie. Podkreśla to znaczenie modelu wodnego dla dokładnych symulacji enzymów z głęboko osadzonymi miejscami aktywnymi, które mają kluczowe znaczenie dla zrozumienia ich cykli katalitycznych, projektowania enzymów i przewidywania czasów przebywania leków.

Słowa kluczowe

TransportTools, modele wody, tunele, enzymy, dynamika białek, wydajność transportu

Abstract

The enzymes with buried active sites have pathways leading to them which are called tunnels. By focusing on the hydrolase family of enzymes and emphasizing the importance of water within these tunnels, this study provides valuable insights into their functional implications. Through the utilization of molecular dynamics simulations and adaptive sampling techniques, the conformational states of the systems are thoroughly investigated. Notably, the calculation of tunnels leading to the active site is conducted for all frames of the extensive simulation, leading to the identification of previously unrecognized tunnels. By tracking the movement of water molecules from the bulk to the defined region (active site) of the protein, the transport pathways employed by the enzyme during the simulation are determined. To facilitate the analysis of vast amounts of data related to tunnels and water molecules, a Python module called TransportTools is developed. The module streamlines the analysis of extensive tunnel and water molecule data. It aids in assessing water molecule utilization, transport efficiency, and functionality within protein structures. This tool has uncovered new water-transporting tunnels in enzymes like haloalkane dehalogenase, enhancing our understanding of tunnel dynamics. Additionally, it is discovered that water molecules can effectively navigate narrow regions within the tunnels through hydrogen bonding interactions. These findings are validated through a comparison between wild type epoxide hydrolase and its mutant E470G, where an increase in the utilization of alternate tunnels is observed in the mutant. Moreover, the choice of explicit water models such as OPC, TIP3P, and TIP4P-Ew significantly impacts the utilization of enzyme tunnels by water molecules. The observed differences between the three water models can be attributed to variations in their diffusion properties and charge distributions. With TIP3P model, water migration was noticeably accelerated, transporting about 2.5– and 2.0-times more water molecules than with OPC and TIP4P-Ew models, respectively. This phenomenon primarily arises from the observed quicker transit times and more water molecules concurrently migrating through tunnels when employing the TIP3P model. The consistency of these findings across various enzymes, such as alditol oxidase and cytochrome P450 2D6, along with tunnel geometries, highlights their broader relevance. This emphasizes water model importance for accurate simulations of enzymes with buried active sites which are critical for understanding their catalytic cycles, enzyme design, and predictions of drug residence times.

Keywords

TransportTools, water models, tunnels, enzymes, protein dynamics, transport efficiency

Abbreviations

OPC	optimal point charge water model
TIP3P	three-site transferable intermolecular potential water model
TIP4P- Ew	four-site transferable intermolecular potential with Ewald correction water model
EC	enzyme commission number
PDB	protein data bank
MD	molecular dynamics
HTMD	high throughput molecular dynamics
MSM	Markov state model
RMSD	root-mean square deviation
TCG	tunnel conformation group

Summary of the doctoral research

Introduction to structure-function relationship of tunnels in enzymes:

Enzymes are biological molecules that act as catalysts, and they aid in chemical reactions in cells. They play a critical role in maintaining biochemical processes essential for life. The enzymatic conversion of the substrate to product occurs in the active site of the enzyme, where it can involve solvents or other cofactors for the catalytic conversion of the substrate to product.[1–3] When searching the active sites of the enzyme, it can be located on the surface where the substrate can bind on a surface cavity, or it can be between interfaces between domains[4] or buried deep within the core of the protein (**Figure 1**) and the substrates, solvents and the product must travel to the core of the protein to carry out its enzymatic activity,[5] which will be the focus of my doctoral research.

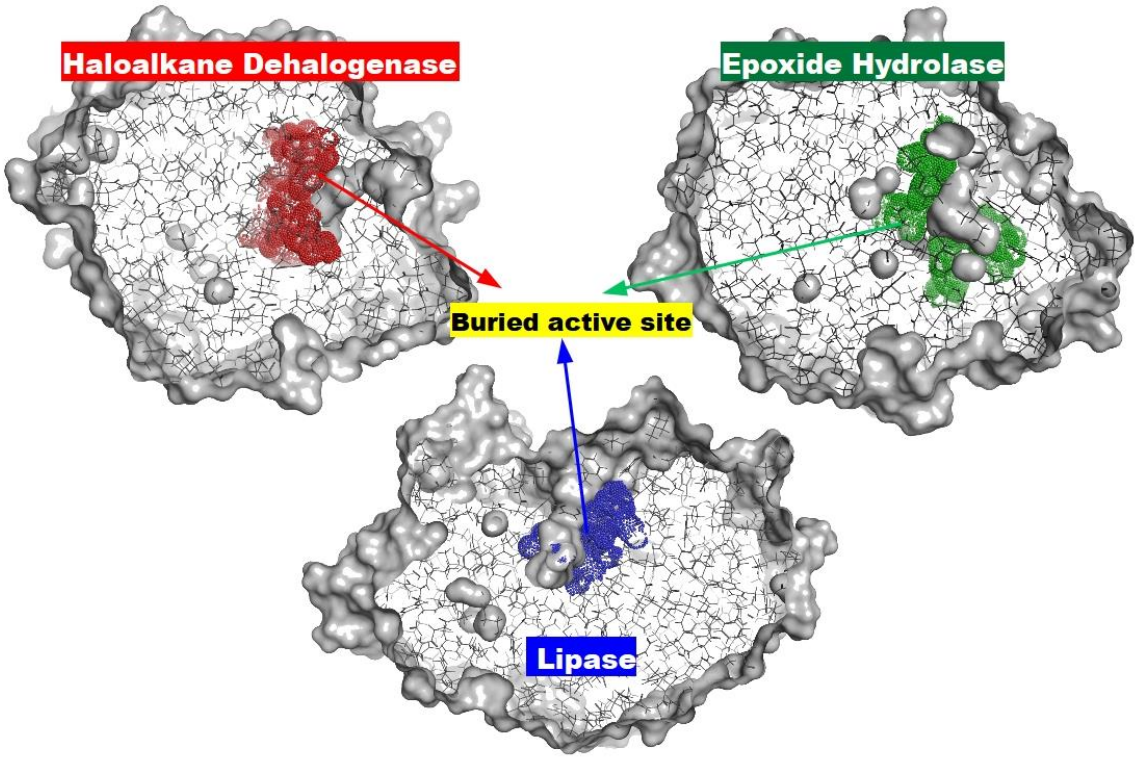


Figure 1: Surface representation of enzymes showing their buried active sites*. The protein is shown as surface, active site residues shown as dots and colored in red for haloalkane dehalogenase, green for epoxide hydrolase, and blue for lipase.*

The active site residues in such buried enzymes are found to be less flexible in terms of temperature factors or B-factors when compared to non-active site residues in apo-enzyme, and they are flexible to accommodate substrates or ligands upon their arrival.[5,6] Enzymes with buried active sites have been found to play critical roles in various cellular processes such as gene regulation, protein synthesis, and metabolic pathways.[7] In addition to their catalytic function, these enzymes have the capability of regulating their activity in response to different environmental conditions and they also have the feature of filtering out any unnecessary substrates that may enter the active site.[8]

Understanding how ligands migrate through proteins with buried active sites will help us to understand some characteristics of the protein like substrate specificity and hydropathy.[9]Such passages taken by the molecules to reach the active sites in the core of the protein are called tunnels and depending upon the enzyme, there can be multiple tunnels leading to the active site.[10] A study that characterized such pathways leading to the active site of the protein found out that over 64% of the analyzed enzyme structures from all the enzymatic classes had two or more tunnels on average in them. Also, the tunnels had different physiochemical properties in different parts of the tunnel respectively.[10] Which means, overall, the tunnels also can vary in terms of physiochemical properties and amino acid compositions which means depending on the amino acid composition, a tunnel can be hydrophilic or hydrophobic which leads to specific selectivity of the molecules it can transport.[7,11] As we know, the activity of the enzyme happens over time and the dynamic movement of the side chain and backbones of the amino acids in the enzyme, the voids in the proteins also vary. This leads to the statement that the tunnels are also observed to change over time as the surrounding environment changes.[7]

So now we know that tunnels are also an important component of enzymatic activity which makes it essential to study their structure-function relationship. The modifications of such enzymatic tunnels can also lead to changes in enzymatic activity, specificity, promiscuity, enantioselectivity, and stability.[11] The movement of the side chain residues in the enzyme can lead to obstruction in the pathways of tunnels and this obstruction can be called gating. During the dynamics, these gates can switch reversibly between open and closed states thereby controlling the traffic of molecules entering or leaving the enzyme through those tunnels, which makes the access pathway dynamic.[12] The ability to explore tunnels in protein is being developed over a long period of time. In the early days, when only X-ray crystallography and NMR spectroscopy were available, the large cavities and tunnels in the single snapshot of protein were only seen by visual methods. However, these methods were missing dynamic information about the pathways. At the same time, more advanced methods computational methods such as molecular dynamics simulations, were becoming popular with the availability of suitable hardware. This led to the development of complex algorithms to explore the voids

in the proteins from the snapshots obtained from molecular dynamics simulations and to explore the tunnels which are transiently occurring. The software tools designed to explore tunnels in proteins use Voronoi diagrams to define the skeleton of the voids within the protein structure.[13]

CAVER is a software tool used for identifying tunnels connecting buried protein cavities to the protein surface. CAVER 3.0, which employs Voronoi diagrams constructed from approximated protein structures. The software employs a two-step search process, optimizing the starting point and identifying pathways from the user-defined starting point to reachable points on the inner surface. Tunnels are ranked based on a cost function evaluating their length and distances to the closest atoms. CAVER is available as a plugin for PYMOL, a standalone Java application, and CAVER Analyst graphical interface.[14,15] It provides tunnel parameters, plotting data, visualization options, and compatibility with trajectory analysis and protein channels analysis. A tunnel found by CAVER 3.0 is illustrated in ***Figure 2A***, where the sphere representation of the Voronoi diagram is shown.[16]

MOLE is another software tool that uses Voronoi diagrams to identify tunnels in protein structures. The cost function in MOLE evaluates edges based on width and length, and the search is performed to find tunnels with the lowest overall cost. Identified tunnels are compared to avoid duplicates. MOLE is available as a stand-alone application, a plugin for PYMOL, and a web server. It provides tunnel parameters, profiles, surrounding atom lists, and visualization options.[17]

MOLAXIS is another software tool designed to identify both tunnels and channels in protein structures. It approximates protein atoms with spheres and constructs a Voronoi diagram based on the radius of the smallest atom. The search for tunnels is performed using a cost function that considers edge length and the average distance between vertices and their closest atoms. Tunnels are ranked and duplicates are discarded based on a user-defined forking threshold. MOLAXIS is available as a stand-alone application, a web server, and a VMD plugin, providing tunnel parameters, profiles, surrounding residues, and visualization options.[18]

CAVER clustering feature helps identify the most representative tunnels, providing a more concise output. Compatibility with popular visualization tools like PyMOL and support for trajectory analysis further enhance its utility, making it a valuable choice for analyses of tunnel dynamics compared to primarily static structure targets of MOLAXIS and MOLE.[16]

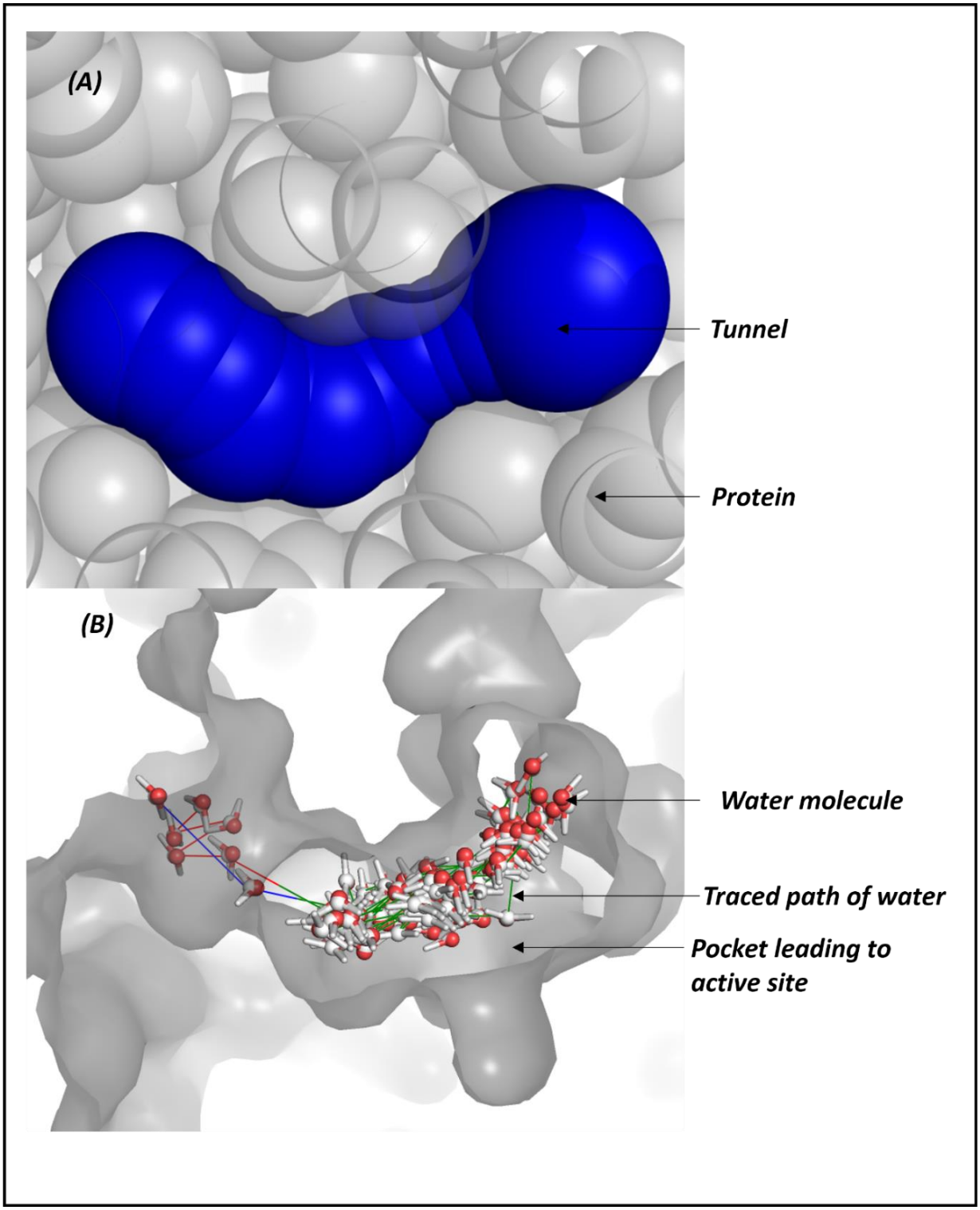


Figure 2: Detection of tunnels and water tracking. *(A) Determination of tunnels by CAVER for a single snapshot where a tunnel (in blue) is shown as a set of spheres connecting the continuous voids in protein (in grey spheres), (B) a raw path of AQUA-DUCT visualized as a line representing the successive positions of the water during the simulation. The blue line represents the entry of water, the red line represents the release of water and the green line represents that the water is in the active site.*

Over the past decade, there have been major improvements in computational power with simulation packages that are able to run simulations on GPU which made micro-second simulations accessible.[19] Also, there have been considerable improvements in machine learning methods for exploring more confirmatory states of the protein with fewer hardware requirements.[20,21] This now allows us to obtain massive simulations with millions of frames in the resulting trajectories. Protein dynamics studied at the atomic level using the current improvements in simulation techniques reveal side chain movements and structure-function relationships of proteins more comprehensively. One application of such a resolution would be to study the evolution of transient tunnels within proteins that appear and disappear over time. Transient tunnels are influenced by various factors, such as protein conformation, solvent, ligand, and environment.[22,23] Recent advancements in molecular dynamics simulations, particularly with enhanced sampling methods, have greatly improved our ability to study protein dynamics and explore the buried active sites of enzymes. Molecular dynamics simulations have been found to be the most suitable method to study drug binding and unbinding kinetics and to discover some rare molecular events which are not accessible to experimental methods.[24] By accurately predicting tunnels behavior over time, we can gain insights into the mechanism by which ligands move from the bulk solvent to the protein surface and subsequently migrate into the active site of the enzyme through these tunnels.[25] Moreover, in drug discovery and design, transient tunnels provide access to novel binding sites that can modulate protein function. Studying the properties of these tunnels, such as the types of amino acids they are composed of, can provide valuable information on their biochemical properties and the enzyme's catalytic mechanism. By targeting these transient tunnels, researchers can develop new drugs that selectively interact with the protein of interest, potentially leading to new treatments for various diseases.[12,26,27] Furthermore, transient tunnels can also be utilized in protein engineering by modifying them to alter the selectivity and efficiency of protein transport.[28,29] This can be achieved by introducing mutations or altering the properties of the amino acids lining the tunnels. For example, a single mutation to bulkier residue in a tunnel was shown to led to the decline in the overall efficiency of the product release through the tunnel which became a rate-limiting step in the biological process of that enzyme.[30] Conversely, another study improved the enzyme's activity to 32-fold higher than the wild type by modifying the key residues in the access tunnels leading to the active site, resulting in restricted accessibility.[31]

Developing a comprehensive methodology for analyzing transient tunnels in enzymes

With good motivation about the importance of studying enzymes with buried active sites, better hardware, and the advent of many improved sampling methods to explore a wide range of protein states, when looking at existing methods to study ligand or solvent transport through the channels, existing software such as CAVER, [16] MOLE [17] were available to analyze the tunnels in the simulations. Nevertheless, a consistent method for highlighting the equivalence between tunnel IDs in different simulations was absent. In addition, there was information about the physiochemical properties of the tunnels and information about the bottlenecks of the tunnels. But this was still not enough to say whether the tunnels were being used at all.

To establish structure–function relationship I focused on hydrolytic enzymes during my doctoral study. They are the class of enzymes that use water as a reaction component during the catalysis of the breakdown of complex molecules into smaller subunits. Hydrolytic enzymes are important enzymes that play a vital role in the human body such as gastrointestinal digestion of protein, cellular metabolism, and recycling of biological materials.[32] They are also used in sewage sludge treatment, where they degrade highly polymeric substrates.[33] A few examples are, lipases, which degrade lipids into fatty acids and glycerol, and haloalkane dehalogenases, which are one of the best-characterized families of enzymes that degrade halogenated compounds including chlorinated and brominated compounds.[34] Solvent, being an important component of a simulation, is dynamic and travels from the bulk to the core of the enzyme which is being simulated. Systematic analysis of such movements can give an insight into how the solvent is being utilized by the enzyme. Now, when we combine i) hydrolytic enzymes ii) enzymes with buried active sites and iii) the ability to detect tunnels systematically, lead to believe that if we systematically trace the utilization of waters by the enzyme, we will be able to characterize the function of the tunnels based upon the paths taken by such water molecules during the simulation time.

Upon looking up such tools, AQUA-DUCT represents one such tool capable of tracing the paths taken by the waters during simulations.[35,36] The software works in several stages to achieve its results. First, it identifies the residues of interest based on user-defined criteria, such as the active site of the protein and the boundaries within which the residues should be tracked. It then calculates the positions of these residues in each frame of the simulation, allowing paths to be generated that represent their movements. These paths represent the coordinates of the center of mass of a traceable residue at different frames of the simulation as shown in ***Figure 2B***. It captures the spatial movement of the residue over time, providing valuable information about its behavior within the molecular system. AQUA-DUCT further refines the analysis by separating the paths into distinct events, and tracking when residues

enter and leave the defined boundaries. This information helps to identify clusters of entry and exit points, which can be useful in understanding the pathways or tunnels within the molecular system.

Now, by combining all the tools and techniques mentioned above, a method was set to be developed to systematically analyze the utilization of water by a hydrolytic enzyme. This method begins with the results obtained from extensive simulations to explore numerous conformational states of the protein. Subsequently, by analyzing the simulation trajectory, tunnels can be identified, and the water molecules traced, providing valuable insights into the enzyme's utilization of water via these tunnels. Me with my colleagues developed this methodology using different hydrolytic enzymes which were i) haloalkane dehalogenase DhaA from *Rhodococcus rhodochrous* ii) lipase from *Diutina rugosa*, iii) epoxide hydrolase I from *Solanum tuberosum*, and iv) human epoxide hydrolase. All the enzymes have active sites buried in the core of the protein and were studied well in terms of simulation and characterized tunnels from previous studies.[37–39]

I focused on haloalkane dehalogenase as the system in which I conducted simulations and explored the transport pathways. Haloalkane dehalogenases (EC: 3.8.1.5), belongs to the α/β-hydrolase family of enzymes. Their function involves catalyzing the hydrolysis of carbon-halogen bonds in halogenated compounds. These toxic compounds are produced by bacteria, macroalgae, corals and terrestrial plants.[40] Haloalkane dehalogenases find application as industrial catalysts in the bioremediation of environmental pollutants and microbial dehalogenation processes. Structurally, DhaA (PDB ID: 4e46) consists of 293 residues forming two distinct domains: a rigid core domain and a flexible cap domain, between which the active site is housed, as illustrated in ***Figure 3***. The active site of DhaA contains a catalytic pentad consisting of the following residues: Asn 38, Trp 104, Asp 103, Glu 127, and His 269.

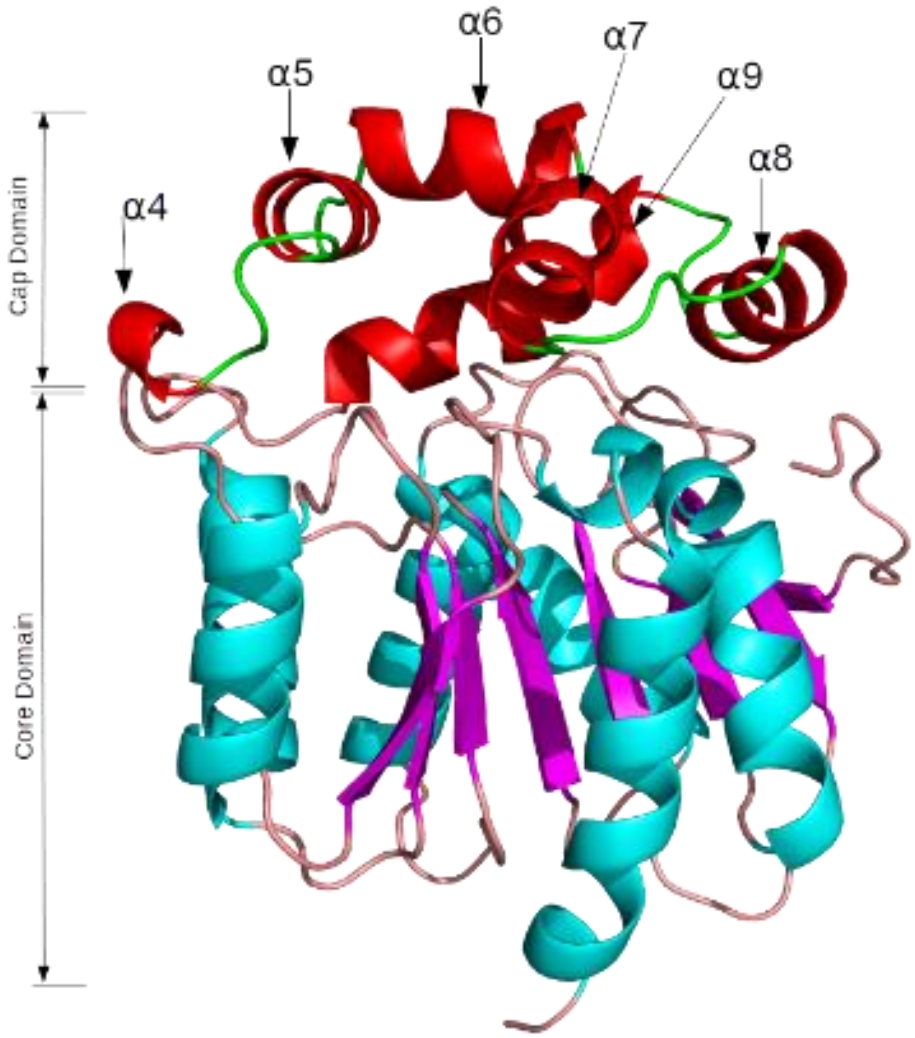


***FIGURE 3**: **Cartoon representation of domains in haloalkane dehalogenase DhaA** consisting of a core domain (in cyan and pink) and a flexible cap domain (red and green)*

The first stage in the developmental process was to select the simulation method to be used. When it comes to classical molecular dynamics (MD) simulations, exploration of the potential energy landscape occurs sequentially, often demanding extensive sampling time to surmount high-energy barriers, leading to substantial time and resource consumption.[41] Conversely, enhanced sampling techniques accelerate transitions over energy barriers, potentially reducing computational expenses, yet necessitating careful parameter tuning to ensure accurate representation of rare events, which is why adaptive sampling driven by Markov-state-models (MSM) was selected,[42] as implemented in HTMD package.[43] Previous studies employed this approach to investigate the DhaA enzyme, exploring its distinct conformational states and elucidating the impacts of introduced mutations on its tunnels.[44]

The HTMD module developed by Acellera provides a highly suitable Python-based programmable environment for conducting parallel simulations, constructing Markov State Models (MSMs) based on user-defined metrics, and iteratively exploring additional states throughout simulation iterations,[43] as illustrated in ***Figure 4***. I have used dihedral angles to effectively capture the changes in the enzyme's structure and explore of different states, including slower conversions. This approach allows the investigation of rare states, such as opened states of transient tunnels. Considering this, using the dihedral angles of residues surrounding the active site as a metric would enable the exploration of a wide range of states. By conducting this analysis across 10 iterations (epochs), an optimal and effective approach is employed to uncover the intricate conformational dynamics of the enzyme.

With simulations now in place, the subsequent phase involves identifying tunnels within the resultant simulation data. Based on our familiarity with the CAVER software,[16] we have established consistent starting points from which the tunnel search will initiated as the set of least mobile atoms across all simulations whose center of mass is located at the bottom of the active site cavity. Such consistency is crucial for achieving unified results. For tracking the movement of water molecules within the simulations, we employed AQUA-DUCT.[35] Specifically, the software was configured to track water molecules entering and exiting the region surrounding the active site residues, resulting in a comprehensive list of water molecules that migrate from the bulk solvent into and out of the active site region. This analysis allowed us to gain insights into the behavior and interactions of water molecules within the active site and surrounding regions. Upon completing the simulations and acquiring data regarding tunnels and water molecule trajectories for each distinct simulation, the task of directly comparing these elements across all simulations arose as a challenge. Consequently, we devised a comprehensive analytical approach, which eventually evolved into a software tool named TransportTools. This tool facilitated the integration and comparison of tunnel data and water dynamics from multiple simulations, providing a unified approach to analyzing and interpreting the results. The development of TransportTools proved to be a significant contribution and led to the publication of our work (**Publication 1**). TransportTools seamlessly integrates outputs from CAVER and AQUA-DUCT, enabling the consolidation of tunnels across simulations while accurately assigning water molecules to these unified tunnels. Detailed documentation for the technical information of its working and user manual is available at https://github.com/labbit-eu/transport_tools.

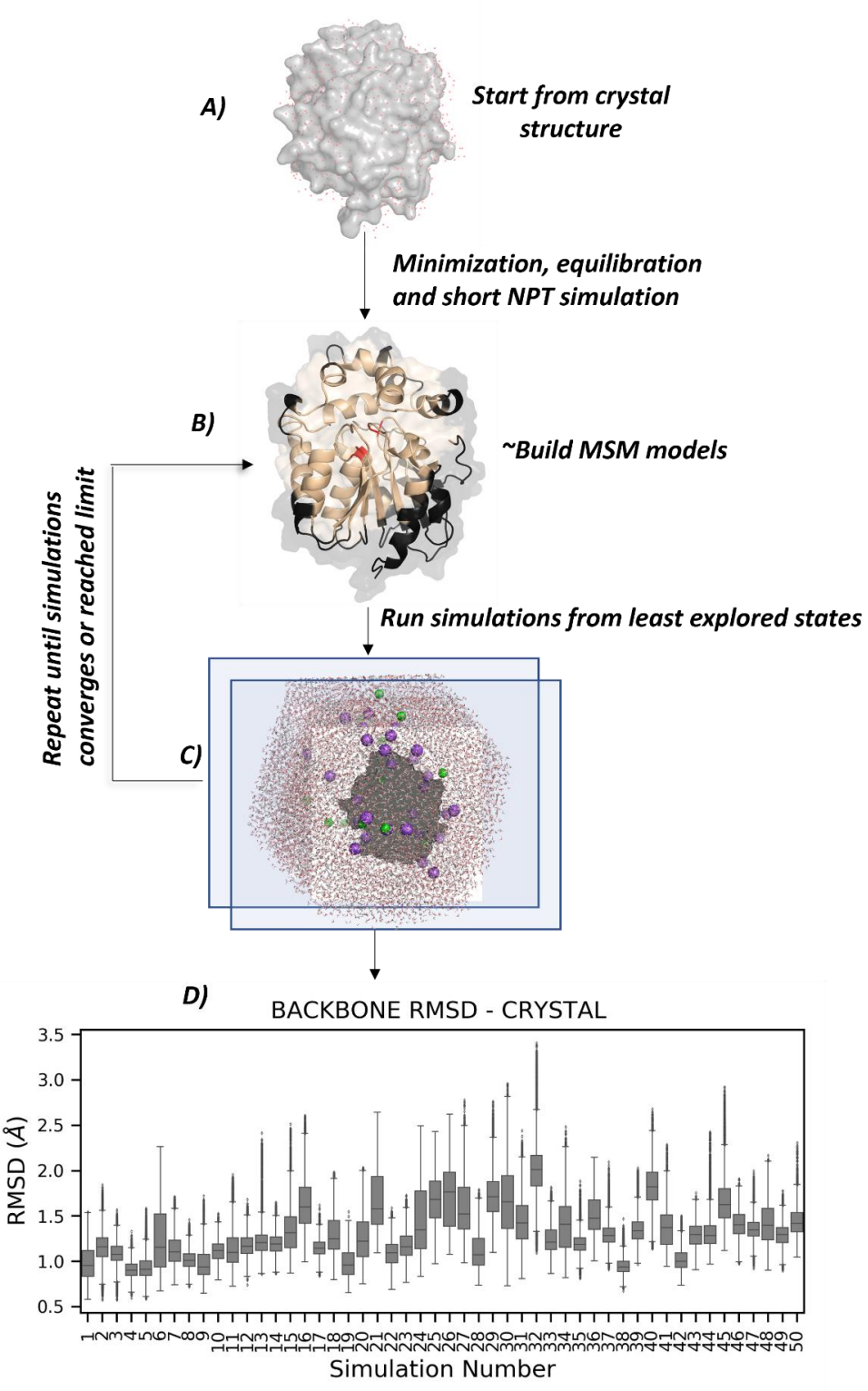


Figure 4: Exploring different conformational spaces of protein by adaptive sampling method. *Starting from the crystal structure (A), from initial simulations, MSMs are created based on the variation of the change in dihedral angles of the residues surrounding (shown as wheat color in B) the active site residues (shown in B as red color lines) where new simulations are started from least explored states in the MSM created. The various states explored are shown as backbone RMSDs to the crystal structure in (D).*

Using the TransportTools engine, end-users can access efficient analyses of molecular tunnels in extensive MD simulations, including those from massively parallel calculations or very long simulations. The tool provides information on molecular tunnels in biomolecules and their utilization by small molecules. It also enables a rigorous comparison of transport processes in different simulation sets, allowing for contrasting transport in the original system and system perturbed by mutations, different solvents, or the presence of ligands. A detailed illustration of the workflow of TransportTools is shown in ***Figure 5***.

Upon collecting the results of traced waters from individual simulations in DhaA, TransportTools was employed to identify the most suitable supercluster (a cluster of similar caver tunnels across various simulations) to which the water transport events can be assigned. This process enabled the determination of the transport pathways utilized by the tunnels. It is possible that certain novel tunnels, through which water transport occurs may not be initially present in the starting crystal structure. This highlights the dynamic nature of the tunnels and underscores the significance of sampling multiple conformational states of the protein. To resolve ambiguity, waters that are assigned to multiple superclusters are traced back to the precise simulation frame and reassigned to the corresponding tunnel that was present at that specific moment. This process ensures that, by the end of the calculations, each water molecule is assigned to a specific tunnel, allowing for the study of major transport pathways. This analysis is conducted based on the number of frames and the number of waters assigned to a particular supercluster. This comprehensive workflow was utilized to analyze the water transport usage of the other two selected enzymes (epoxide hydrolase and lipase) providing a holistic overview of their behavior in the transient tunnels. In-depth insights into the tunnels and transport events of epoxide hydrolase and lipase are elaborated upon in **Publication 2**. Subsequent sections of this thesis will focus on presenting results specifically pertaining to DhaA.

Characterizing the tunnels in DhaA

Haloalkane dehalogenase DhaA as discussed in the previous section, has a well-described structure, and characterized transport pathways. The names of tunnels are established in prior literature,[39] which are, the main tunnel (p1) and slot tunnels (p2 group of tunnels – p2a, p2b, p2c and p3 tunnel). The main tunnel is surrounded by the residues 145,176 and 172 which are part of helices $\alpha 4$ and $\alpha 5$. Building upon the pre-existing knowledge of characterized tunnels, the adaptive sampling simulations demonstrated a consistent pattern by successfully predicting both the well-established tunnels (***Figure 6A,C***) and certain transient ones (***Figure 6B,D***), attributed to the extensive and enhanced sampling techniques. The RMSD analysis in ***Figure 4D*** underscores the efficacy of adaptive sampling in exploring diverse protein states across various simulations.

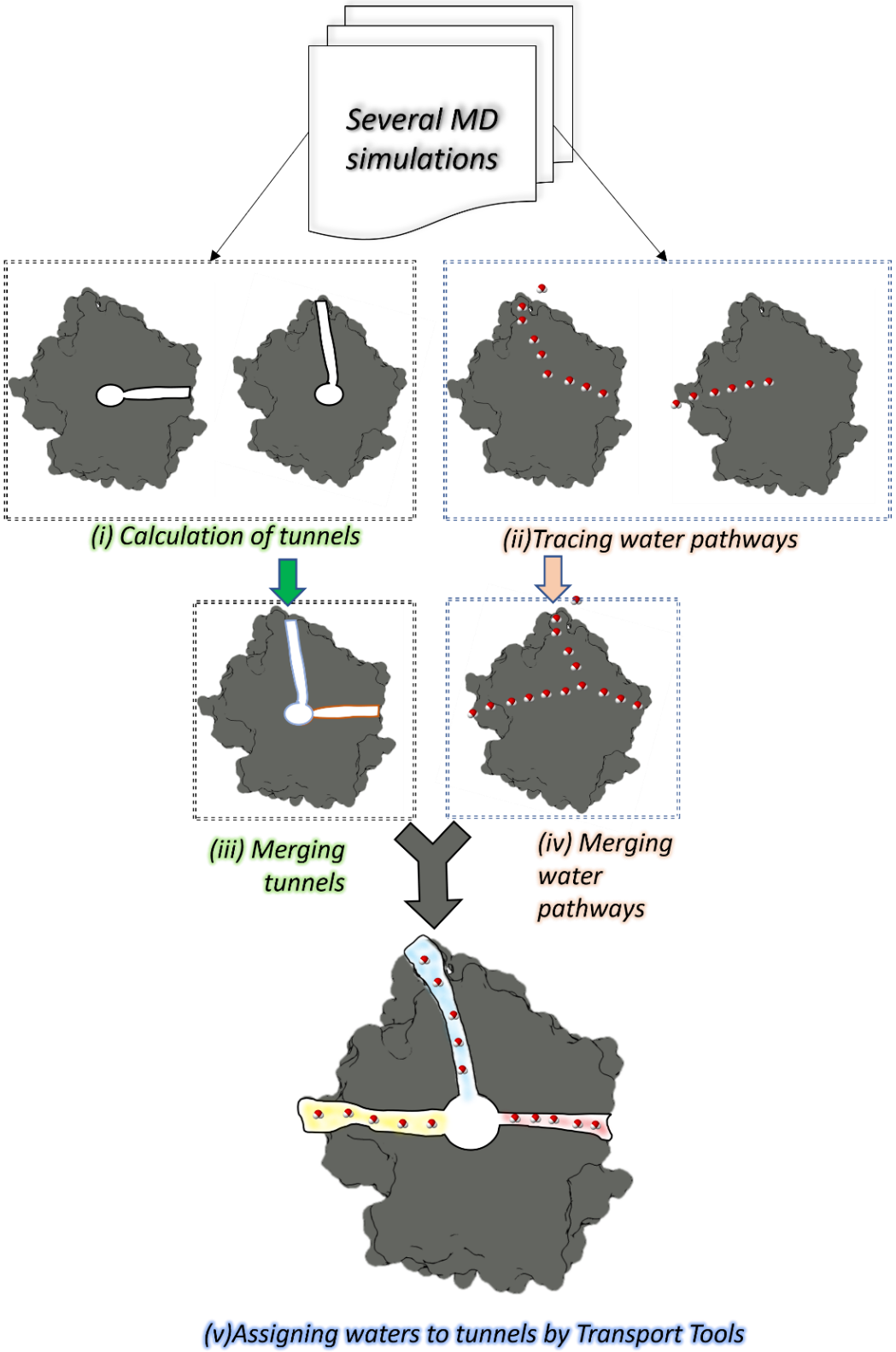


***FIGURE 5**: **TransportTools workflow** describing (i) The creation of superclusters from individual CAVER clusters (ii) Creating a network of waters from AQUA-DUCT results and (iii) Integrating the tunnel and water networks to form tunnel volumes (superclusters) and their associated water transport networks.*

From the results of TransportTools, the tunnels in all simulations were consolidated into a single set of results that revealed the presence of permanent and transient tunnels that were present in all 50 simulations. Looking only at the consolidated tunnels, or superclusters as those structures are called, there were 65 superclusters that were merged and clustered by hierarchical complete-linkage clustering of TransportTools. Many of the superclusters were novel and had not been seen in previous studies. The majority of the superclusters consisted of p1, all sub-tunnels of p2, and p3 tunnels (***Figure 6A, C***). But there were also other auxiliary tunnels that were predicted to exist in about 2% of the total simulation (***Figure 6B, D***). During this supercluster creation, some basic filters were applied to filter out irrelevant tunnels. The filters were such that the tunnels, which were too short (< 5 Å), had small bottleneck radius (< 0.75 Å), or were highly curved (the curvature > 2), were eliminated from consideration. The viability of predicted tunnels can be assessed by associating traced water molecules with superclusters, enabling the individual interpretation of each supercluster's functionality. By excluding tunnels without water transport events, only those tunnels facilitating water transportation to or from the active site would remain. After such filtering, the tunnels involved in water transport for DhaA were p1, p2a, p2b, p2c, and p3 (***Figure 6A, C***). There were two new tunnels named new_A and new_B ***(Figure 6B, D)***, which were also involved in water transport. The major water transport tunnel was p1, which was utilized by 2080 water molecules, while the least used was the p2b tunnel which transported 6 molecules throughout the simulation period. For epoxide hydrolase, there were 22 superclusters that were utilized out of which 3 were characterized tunnels and similarly, for Lipase, there were 24 superclusters out of which 2 were characterized tunnels. These new tunnels which were capable of water transport can be engineered to improve the overall efficiency of the enzyme, as highlighted by comparison of the tunnels used by human epoxide hydrolase and its mutant E470G. A single remote mutation can affect the utilization of the tunnels, notably the Tm5, Tside, Tcap1, and Tc/m by water molecule.[45]

Apart from the differences in water transport in the tunnels, the models used for the simulations can introduce differences in the interpreted simulations.[46–51] Due to the differences in the properties of waters, it might also be a case that the volume of waters migrating through such tunnels will be affected as well. To validate this, the final chapter of this thesis will focus on how these water models affect the transport of water molecules through such a tunnel.

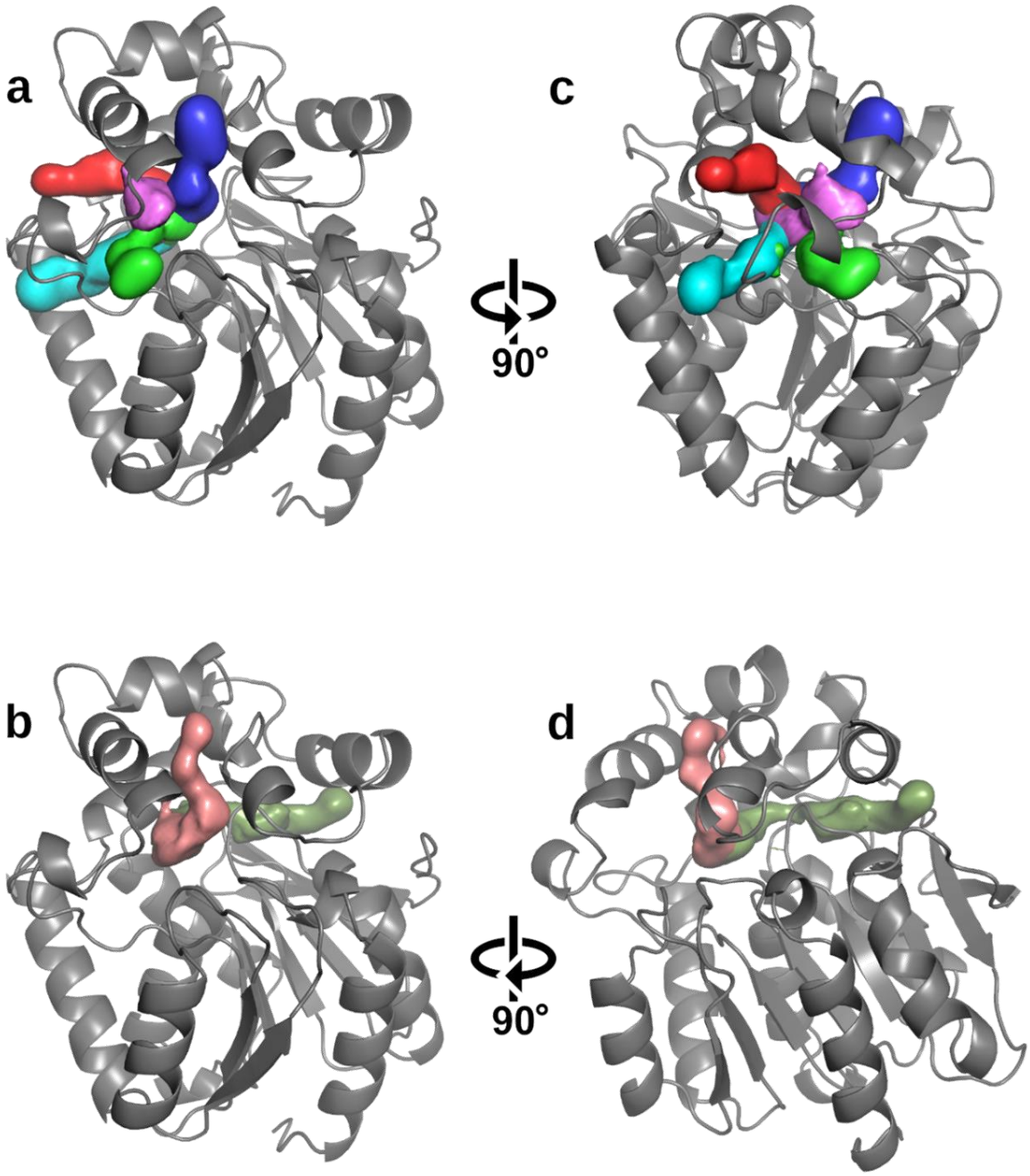


Figure 6: The tunnels which are described on the literature for DhaA. *(a) p1 (blue), p2a(green), p2b (pink),p2c (cyan)and p3 (red c) rotating the same view 90° counterclockwise in the Z-axis; and newly observed b) up (salmon) and back (light green) tunnels, d) rotating the same view 90° clockwise in the Z-axis. Adopted from Publication 2.*

Impact of water models on the routine analysis of tunnels

Water forms the base of life, and it is also an essential component in biomolecular simulations.[52] Water models play a vital role in accurately simulating the environment of the protein and its interaction with amino acids and itself.[47,53] The measurable properties of waters from experiments include physiochemical properties such as diffusion coefficient, dielectric constant, dipole moment, charge distribution, etc., which can define how a water behaves in biomolecular simulations.[48] The key differences in the different models of waters used in the simulations are based upon number of sites and parameters used to describe interaction between the sites and their physiochemical properties ***(Figure 7)***. The most commonly used water is the three-site transferrable intermolecular potential (TIP3P) model, which has two positive charges on the hydrogen atoms and one negative charge on the oxygen atom.[54] The improved version of TIP3P model is TIP4P where it is a four site model with one negative charge on an massless site called M instead of oxygen atom, and two positive charges on hydrogen atoms,[53] a reparametrized version of the TIP4P model called TIP4P-Ew was introduced to use with Ewald techniques.[55] 4-point Optimal Point Charge (OPC) model is another model which has been shown to provide more accurate results than of the older models. Also, 4-point models are computationally more expensive to simulate when compared to the TIP3P model.[56]

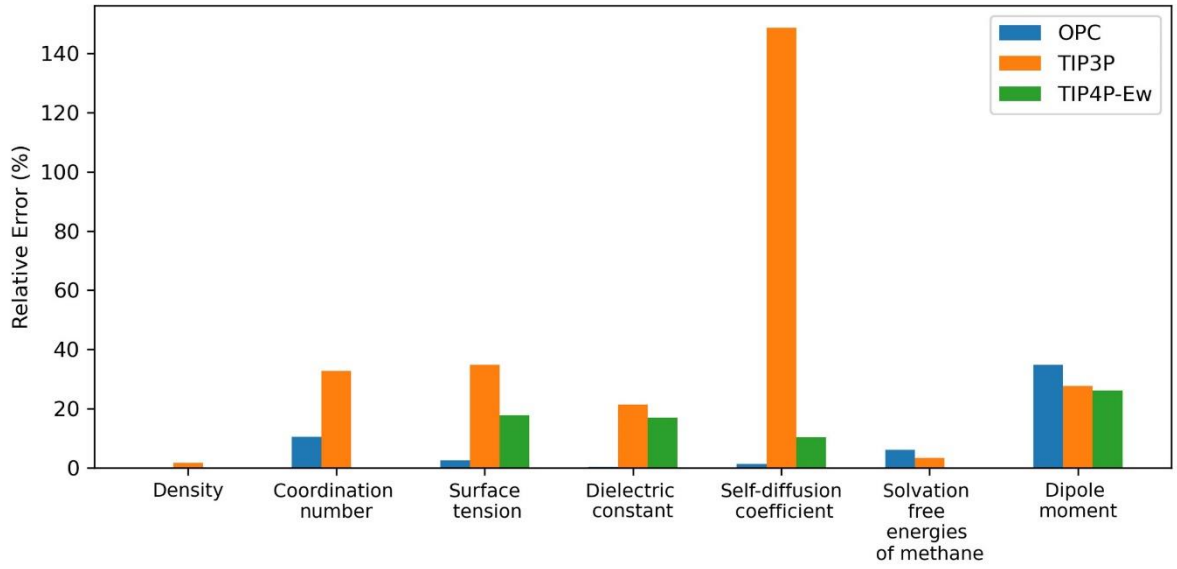


Figure 7: Key differences between the physiochemical properties of water models. *The figure shows the relative error with experimental waters for different tested parameters. Adopted from Publication 3.*

This part of the study focused on the differences in water use by tunnels introduced by different water models using haloalkane dehalogenase – DhaA as the test system. The simulations were organized into five distinct groups, distinguished by variations in the bottleneck radii of the p1 tunnel. The individual groups were designated as Tunnel Conformational Groups (TCGs). The starting structures representing the groups were chosen so that the bottlenecks became progressively wider from TCG0 to TCG4 ***(Figure 8)***. Each group had five replicates of 200 ns, totaling 1 µs per group, and this was repeated with three different water models (OPC, TIP3P, and TIP4P-Ew), accounting for 15 µs of cumulative simulation trajectories. Detailed methodological information is available in the methodology section of **Publication 3**.

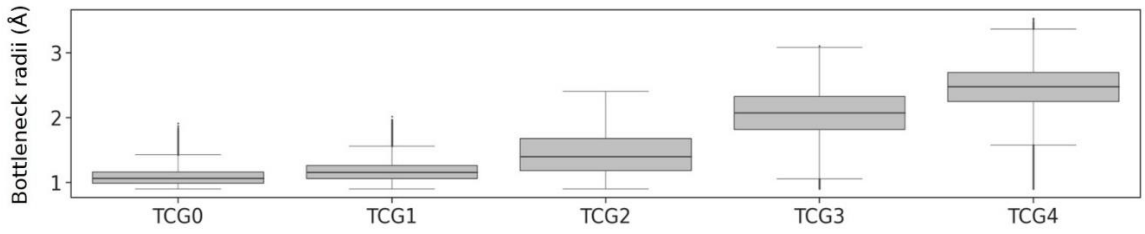


Figure 8: Average bottleneck radii of all the TCGs for p1 tunnel*. The plot shows average ±SD of 15 simulations. Adopted from Publication 3.*

Following the same protocol in the previous section to obtain transport events **(*Figure 5*)**, the transport events for TCGs in all water models were calculated and analyzed. Additionally, using the comparative analysis module of TransportTools, separate transport events results were availed for different groups and models ***(Figure 9)***. Upon inspecting the transport events implied by TransportTools, it was evident that there were several tunnel clusters present. However, most of the water transport events occurred in p1 and p2 tunnels and rarely in p3 and other tunnels.

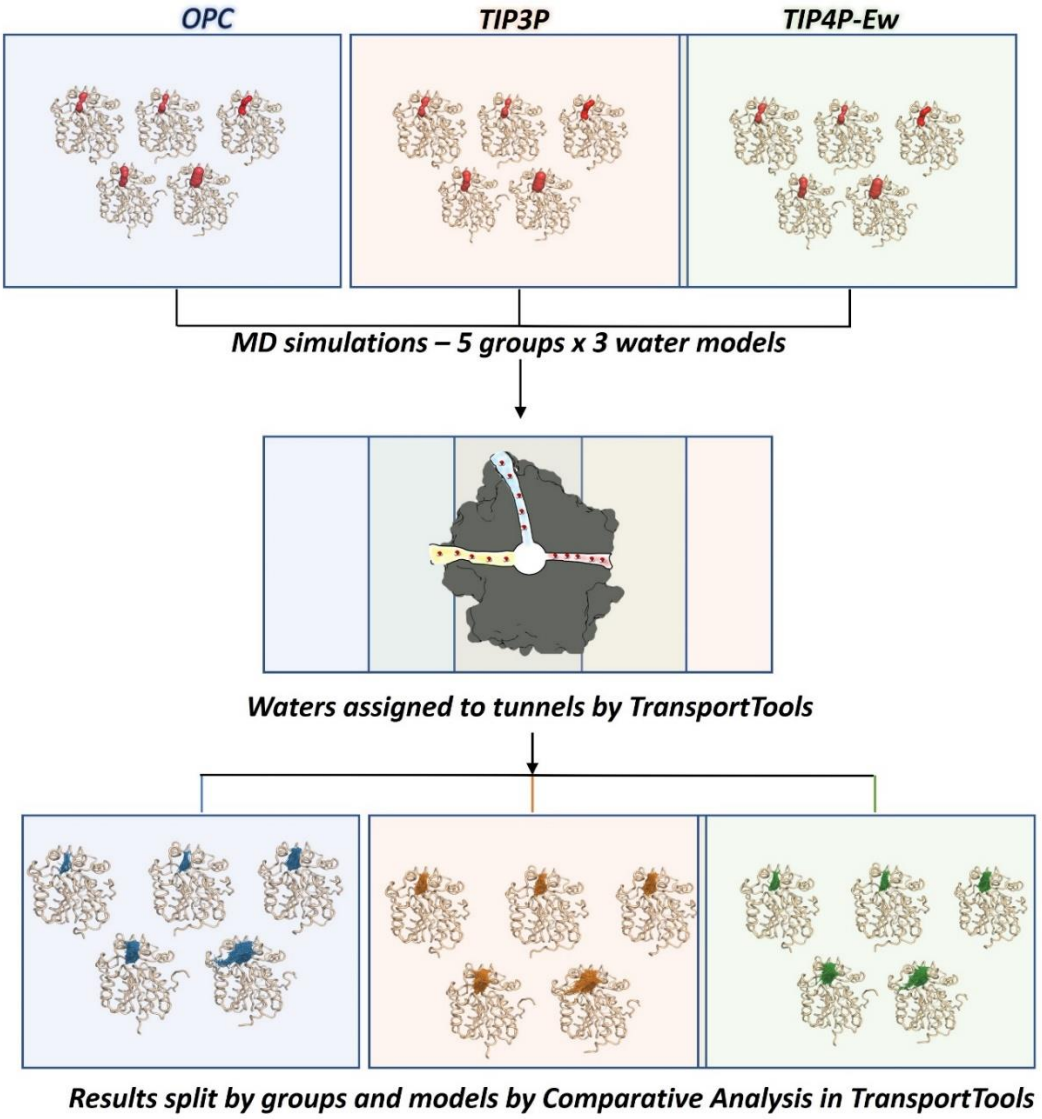


Figure 9: The workflow illustrates the effect of water models on water transport through p1 tunnel of DhaA. *First, MD simulations were conducted for three sets, each containing five groups with different tunnel geometry with five simulations, using distinct water models. Within the five groups, tunnel geometries remained constant, while progressively increasing their bottleneck radius across groups. This was followed by the merging of tunnels and water molecules using TransportTools, creating a unified dataset. Finally, the unified dataset is split based on water models and groups through the Comparative Analysis module of TransportTools, allowing for an in-depth analysis of tunnel-water interactions between water models and groups.*

To get a deep insight into the changes in water utilization by tunnels introduced by water models, an in-depth analysis was carried out for the p1 tunnel. This was done because it was the tunnel that was utilized for the selection of starting structures, and the bottleneck radii were progressively increasing along the TCGs which was confirmed by measuring them after the simulations. Also, this was the most prevalent and utilized tunnel as interpreted by

TransportTools. Now, when inspecting the transport of waters through the TCGs by different water models, TIP3P waters dominated in the number of waters entering and exiting the tunnels. This was followed by TIP4P-Ew waters and finally OPC waters ***(Figure 10A)***. This was true for all the TCGs regardless of the tunnel geometry which implied that the differences occurred only due to the properties of the water model itself. Viscosity and self-diffusion play a vital role in determining how fast water molecules can diffuse. OPC and TIP4P-Ew have values close to experimental methods and considerably higher than TIP3P. As a consequence, the overall number of transported water molecules was much lower compared to TIP3P, which could migrate faster due to its lower viscosity and higher self-diffusion coefficient.[55,57] To support this claim, upon measuring the time taken by the waters in the tunnel to enter/exit, it was revealed that TIP3P waters were the fastest to migrate within all TCGs, followed by TIP4P-Ew and OPC ***(Figure 10B)***. The significant advantage of the OPC model is that the optimized charge distribution due to accurate hydration free energy, in turn, enables the atomistic solute-solvent interactions to be more accurate,[55–57] as has been recently shown for water transport via aquaporin 1.[47] The findings were then also tested with two other systems, namely human cytochrome P450 2D6 (CYP2D6) and alditol oxidase from *Streptomyces coelicolor* (AldO) (***Figure 11***). Both the proteins had buried active sites and well-defined tunnels.[58,59] For cytochrome P450 2D6, the starting structures were systematically selected based upon the bottleneck radii of Ch2B-F, ChS, and Ch2C tunnels and T1 tunnel for alditol oxidase. The analysis was split into two different TCGs called TCGnarrow and TCGwide.[58,59]

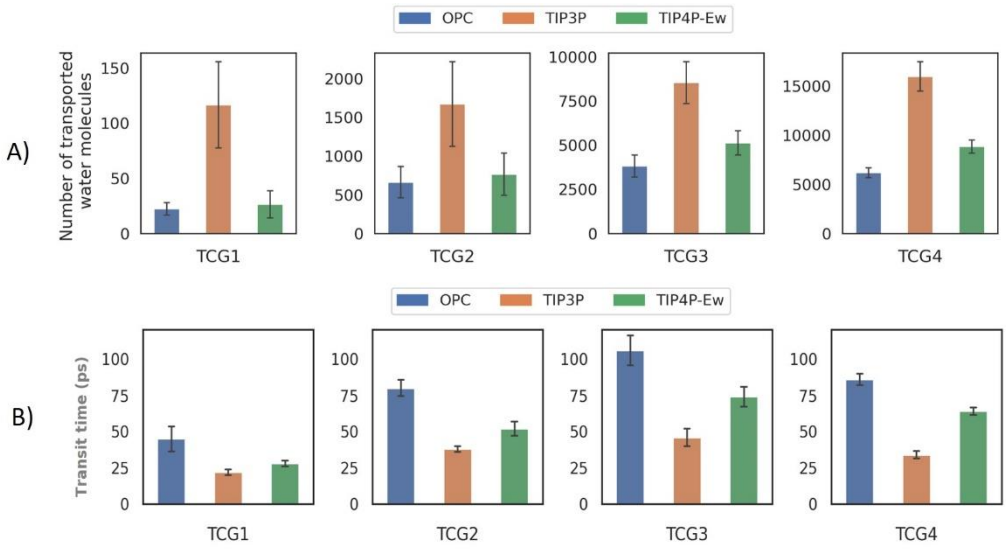


Figure 10: Impact of different water models on the transport of water molecules through P1 tunnels of DhaA in different TCGs. *(A) Number of traced water molecules in different TCGs with different water models in p1 tunnel. (B) The transit time of water molecules to enter/ release from p1 tunnel. Both plots shows average ±SE of 5 simulations. Adopted from Publication 3.*

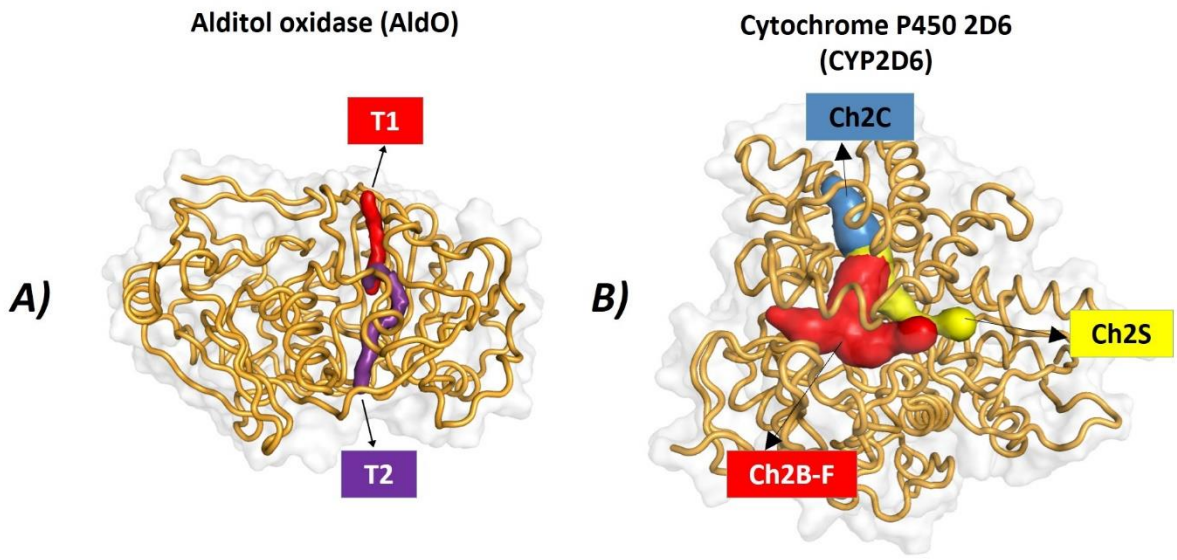


Figure 11: Spatial location of tunnels which are included in the analysis *for (A) alditol oxidase showing T1 and T2 tunnels, (B) cytochrome P450 2D6 showing Ch2B-F, Ch2C and Ch2S tunnels.*

The analysis of water transport across the above tunnels in both the enzymes revealed similar trend like DhaA where there were more number of waters transported by TIP3P waters followed by TIP4P-Ew and OPC waters. Similarly, TIP3P waters were the fastest to migrate across these tunnels. The above results strongly imply the differences in water utilization in tunnels are strongly affected by the choice of water models which are used during the simulation, regardless of the system being used and the geometry of tunnel in which the water is transported.

Conclusions and future perspectives

In summary, this study has effectively addressed the knowledge gap regarding the function and usage of tunnels in protein structures. Previous literature has acknowledged the existence of these tunnels; however, their specific roles and degree of utilization have not been extensively studied. By using the hydrolase family of enzymes as a test system and emphasizing the involvement of water as a key component within these tunnels, we have gained valuable insights into their significance and functional implications. By performing MD simulations and using adaptive sampling, it proved to be remarkably efficient in investigating the conformational states of the systems. Since the nature of tunnels is dynamic, the calculation of tunnels leading to the active site was performed for all frames of these massive simulation. This led to the discovery of new tunnels that had not been identified in the initial crystal structure. By tracking the movement of water moving from the bulk to the enzymes active sites, the paths taken by the water were determined, leading to an understanding of the water transport pathways used by the enzyme.

A Python module called TransportTools was developed based on the data obtained from the trajectories and traced paths of water in tunnels. This module serves as a valuable tool for analyzing large amounts of data related to tunnels and water migration. By using TransportTools, the process of determining the use of waters by tunnels, along with assessing their transport efficiency and functionality, has become a much easier task to perform. The development of this module has facilitated the exploration of tunnel dynamics and provided a practical means to evaluate the functional aspects of these critical components in protein structures. The analysis of water transport through these tunnels in haloalkane dehalogenase and other enzymes led to the discovery of several new tunnels capable of transporting water. It was also found that the waters can squeeze themselves into very narrow regions with the help of extra hydrogen bonds. These results were also validated by a comparison of wild-type epoxide hydrolase and its mutant E470G, which had an increased use of alternative tunnels in the mutant, potentially highlighting novel molecular bases for some diseases.

Finally, the effect of water model selection on water molecule transport through tunnels within three different enzymes—haloalkane dehalogenase, alditol oxidase, and cytochrome P450—was carefully investigated. A comparative analysis included three water models (TIP3P, OPC, and TIP4P-Ew) and five tunnel conformational groups (TCGs) characterized by different bottleneck radii. The results showed that TIP3P water molecules exhibited increased migration rates and frequencies within enzyme tunnels compared to OPC and TIP4P-Ew molecules, a trend that was consistent across tunnel geometries and different enzyme systems. These results highlight the critical role of the chosen water model in simulating complex enzymatic actions, informing enzyme design strategies, and predicting drug residence times. As water molecules play a pivotal role in these processes, the careful selection of a water model emerges as a critical determinant in achieving accurate insights and predictions.

Future research should explore a broader range of enzymes and protein systems to provide a comprehensive understanding of all biological systems. Advances in computational tools and algorithms, such as the incorporation of machine learning and high-performance computing, can enable more efficient analysis of tunnel dynamics. Investigating the relationship between tunnel properties and functionality would reveal structural determinants that control tunnel behavior, facilitating protein engineering and drug design. By combining computational and experimental approaches, a comprehensive understanding of tunnel-mediated processes can be achieved, leading to novel therapeutic strategies and biotechnological applications.

References

1. Madrona, Y., Hollingsworth, S. A., Khan, B. & Poulos, T. L. P450cin Active Site Water: Implications for Substrate Binding and Solvent Accessibility. *Biochemistry* **52**, 5039–5050 (2013).
2. Park, J., Czapla, L. & Amaro, R. E. Molecular Simulations of Aromatase Reveal New Insights Into the Mechanism of Ligand Binding. *J Chem Inf Model* **53**, 2047–2056 (2013).
3. Prejanò, M., Alberto, M. E., Russo, N., Toscano, M. & Marino, T. The Effects of the Metal Ion Substitution into the Active Site of Metalloenzymes: A Theoretical Insight on Some Selected Cases. *Catalysts* **10**, 1038 (2020).
4. Chowdhury, R. & Maranas, C. D. From directed evolution to computational enzyme engineering—A review. *AIChE Journal* **66**, e16847 (2020).
5. Prokop, Z. *et al.* Engineering of protein tunnels: Keyhole-lock-key model for catalysis by the enzymes with active sites. in Lutz, S. & Bornscheuer, U.T. *Protein Engineering Handbook*. Weinheim: Wiley-VCH, 421–464 (2012).
6. Yuan, Z., Zhao, J. & Wang, Z.-X. Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Engineering, Design and Selection* **16**, 109–114 (2003).
7. Kingsley, L. J. & Lill, M. A. Substrate tunnels in enzymes: Structure-function relationships and computational methodology. *Proteins: Structure, Function, and Bioinformatics* **83**, 599–611 (2015).
8. 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* **12**, 2101–2109 (2016).
9. Kingsley, L. J. & Lill, M. A. Including ligand-induced protein flexibility into protein tunnel prediction. *J Comput Chem* **35**, 1748–1756 (2014).
10. Pravda, L. *et al.* Anatomy of enzyme channels. *BMC Bioinformatics* **15**, 379 (2014).
11. Kokkonen, P., Bednar, D., Pinto, G., Prokop, Z. & Damborsky, J. Engineering enzyme access tunnels. *Biotechnol Adv* **37**, 107386 (2019).
12. Marques, S. M. *et al.* Enzyme Tunnels and Gates As Relevant Targets in Drug Design. *Med Res Rev* **37**, 1095–1139 (2017).
13. Brezovsky, J. *et al.* Software tools for identification, visualization and analysis of protein tunnels and channels. *Biotechnol Adv* **31**, 38–49 (2013).
14. Kozlikova, B. *et al.* CAVER Analyst 1.0: graphic tool for interactive visualization and analysis of tunnels and channels in protein structures. *Bioinformatics* **30**, 2684–2685 (2014).
15. Jurcik, A. *et al.* CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics* **34**, 3586–3588 (2018).
16. Chovancova, E. *et al.* CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput Biol* **8**, e1002708 (2012).
17. Sehnal, D. *et al.* MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *J Cheminform* **5**, 39 (2013).
18. Yaffe, E., Fishelovitch, D., Wolfson, H. J., Halperin, D. & Nussinov, R. MolAxis: Efficient and accurate identification of channels in macromolecules. *Proteins: Structure, Function, and Bioinformatics* **73**, 72–86 (2008).
19. Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald. *J Chem Theory Comput* **9**, 3878–3888 (2013).
20. Han, R. & Luber, S. Trajectory-based machine learning method and its application to molecular dynamics. *Mol Phys* **118**, e1788189 (2020).
21. Sidky, H., Chen, W. & Ferguson, A. L. Machine learning for collective variable discovery and enhanced sampling in biomolecular simulation. *Mol Phys* **118**, (2020).

22. Sharma, N. *et al.* Role of allosteric switches and adaptor domains in long-distance cross-talk and transient tunnel formation. *Sci Adv* **6**, (2020).
23. Gora, A., Brezovsky, J. & Damborsky, J. Gates of Enzymes. *Chem Rev* **113**, 5871–5923 (2013).
24. Decherchi, S. & Cavalli, A. Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation. *Chem Rev* **120**, 12788–12833 (2020).
25. Brezovsky, J., Kozlikova, B. & Damborsky, J. Computational Analysis of Protein Tunnels and Channels. in *Methods in Molecular Biology* vol. 1685 25–42 (Humana Press Inc., 2018).
26. Bzówka, M., Mitusińska, K., Hopko, K. & Góra, A. Computational insights into the known inhibitors of human soluble epoxide hydrolase. *Drug Discov Today* **26**, 1914–1921 (2021).
27. Liu, J. *et al.* Multivalent Drug Design and Inhibition of Cholera Toxin by Specific and Transient Protein–Ligand Interactions. *Chem Biol Drug Des* **71**, 408–419 (2008).
28. Ahn, M. *et al.* Modulating co-translational protein folding by rational design and ribosome engineering. *Nature Communications 2022 13:1* **13**, 1–14 (2022).
29. Meng, S. *et al.* Tunnel engineering for modulating the substrate preference in cytochrome P450BsβHI. *Bioresour Bioprocess* **8**, 26 (2021).
30. Biedermannová, L. *et al.* A Single Mutation in a Tunnel to the Active Site Changes the Mechanism and Kinetics of Product Release in Haloalkane Dehalogenase LinB. *Journal of Biological Chemistry* **287**, 29062–29074 (2012).
31. Pavlova, M. *et al.* Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat Chem Biol* **5**, 727–733 (2009).
32. Miguéns-Gómez, A. *et al.* Glucagon-like peptide-1 regulation by food proteins and protein hydrolysates. *Nutr Res Rev* **34**, 259–275 (2021).
33. Burgess, J. E. & Pletschke, B. I. Hydrolytic enzymes in sewage sludge treatment: A mini-review. *Water SA* **34**, 343 (2018).
34. Kokkonen, P. *et al.* Structure-Function Relationships and Engineering of Haloalkane Dehalogenases. in *Aerobic Utilization of Hydrocarbons, Oils, and Lipids* 367–387 (Springer International Publishing, 2019). doi:10.1007/978-3-319-50418-6_15.
35. Magdziarz, T. *et al.* AQUA-DUCT 1.0: structural and functional analysis of macromolecules from an intramolecular voids perspective. *Bioinformatics* **36**, 2599–2601 (2020).
36. Magdziarz, T. *et al.* AQUA-DUCT: a ligands tracking tool. *Bioinformatics* **33**, 2045–2046 (2017).
37. Mitusińska, K., Magdziarz, T., Bzówka, M., Stańczak, A. & Gora, A. Exploring Solanum tuberosum Epoxide Hydrolase Internal Architecture by Water Molecules Tracking. *Biomolecules* **8**, 143 (2018).
38. Roterman, I. Identification of tunnels as in potato hydrolases. *Bioinformation* **16**, 21–25 (2020).
39. Klvana, M. *et al.* Pathways and mechanisms for product release in the engineered haloalkane dehalogenases explored using classical and random acceleration molecular dynamics simulations. *J Mol Biol* **392**, 1339–1356 (2009).
40. Bidleman, T. F. *et al.* A review of halogenated natural products in Arctic, Subarctic and Nordic ecosystems. *Emerg Contam* **5**, 89–115 (2019).
41. Adcock, S. A. & McCammon, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem Rev* **106**, 1589–1615 (2006).
42. Doerr, S. & De Fabritiis, G. On-the-fly learning and sampling of ligand binding by high-throughput molecular simulations. *J Chem Theory Comput* **10**, 2064–2069 (2014).
43. Doerr, S., Harvey, M. J., Noé, F. & De Fabritiis, G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J Chem Theory Comput* **12**, 1845–1852 (2016).
44. Kokkonen, P., Bednar, D., Dockalova, V., Prokop, Z. & Damborsky, J. Conformational changes allow processing of bulky substrates by a haloalkane dehalogenase with a small and buried active site. *Journal of Biological Chemistry* **293**, 11505–11512 (2018).

45. Sequeiros-Borja, C., Thirunavukarasu, A. S., Foumthuim, C. J. D. & Brezovsky, J. Water will find a way: transport through narrow tunnels in hydrolases. *bioRxiv* 2023.05.24.542065 (2023) doi:10.1101/2023.05.24.542065.
46. Kadaoluwa Pathirannahalage, S. P. *et al.* Systematic Comparison of the Structural and Dynamic Properties of Commonly Used Water Models for Molecular Dynamics Simulations. *J Chem Inf Model* **61**, 4521–4536 (2021).
47. Gonzalez, M. A., Zaragoza, A., Lynch, C. I., Sansom, M. S. P. & Valeriani, C. Influence of water models on water movement through AQP1. *Journal of Chemical Physics* **155**, (2021).
48. Kadaoluwa Pathirannahalage, S. P. *et al.* Systematic Comparison of the Structural and Dynamic Properties of Commonly Used Water Models for Molecular Dynamics Simulations. *J Chem Inf Model* **61**, 4521–4536 (2021).
49. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods 2016 14:1* **14**, 71–73 (2016).
50. Celebi, A. T., Nguyen, C. T., Hartkamp, R. & Beskok, A. The shear viscosities of common water models by non-equilibrium molecular dynamics simulations. *Mol Simul* **151**, (2010).
51. Agrawal, N. R. & Brezovsky, J. Impact of water models on structure and dynamics of ligand-transport tunnels in enzymes derived from molecular dynamics simulations. *bioRxiv* (2023) doi:10.1101/2023.04.19.537534.
52. Rhee, Y. M., Sorin, E. J., Jayachandran, G., Lindahl, E. & Pande, V. S. Simulations of the role of water in the protein-folding mechanism. *Proc Natl Acad Sci U S A* **101**, 6456–6461 (2004).
53. Onufriev, A. V. & Izadi, S. Water models for biomolecular simulations. *WIREs Computational Molecular Science* **8**, e1347 (2018).
54. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79**, 926–935 (1983).
55. Horn, H. W. *et al.* Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J Chem Phys* **120**, 9665–9678 (2004).
56. Izadi, S., Anandakrishnan, R. & Onufriev, A. V. Building Water Models: A Different Approach. *J Phys Chem Lett* **5**, 3863–3871 (2014).
57. Mark, P. & Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J Phys Chem A* **105**, 9954–9960 (2001).
58. Wang, A., Stout, C. D., Zhang, Q. & Johnson, E. F. Contributions of ionic interactions and protein dynamics to cytochrome P450 2D6 (CYP2D6) substrate and inhibitor binding. *Journal of Biological Chemistry* **290**, 5092–5104 (2015).
59. Forneris, F. *et al.* Structural Analysis of the Catalytic Mechanism and Stereoselectivity in *Streptomyces coelicolor* Alditol Oxidase. *Biochemistry* **47**, 978–985 (2008).

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: 5

Main text:

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

Supplementary information:
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

Water will find a way: transport through narrow tunnels in hydrolases

Impact Factor 2023: 0

MNiSW points 2023: 0

Citations: 0

Main text:

https://doi.org/10.1101/2023.05.24.542065

Supplementary information:
https://www.biorxiv.org/content/biorxiv/early/2023/05/25/2023.05.24.542065/DC1/embed/media-1.pdf

Repository:

https://doi.org/10.5281/zenodo.7966082

https://doi.org/10.5281/zenodo.7966059

https://doi.org/10.5281/zenodo.7966092

https://doi.org/10.5281/zenodo.7965661

https://doi.org/10.5281/zenodo.7965881

Publication 3

Water migration through enzyme tunnels is sensitive to choice of explicit water model

Impact Factor 2023: 0

MNiSW points 2023: 0

Citations: 0

Main text:

https://doi.org/10.1101/2023.08.14.553223

Supplementary information:
https://www.biorxiv.org/content/biorxiv/early/2023/08/17/2023.08.14.553223.full.pdf

Co-author statements

TransportTools: a library for high-throughput analyses of internal voids in biomolecules and ligand transport through them

1. Brezovsky, J.
2. Thirunavukarasu, A. S.
3. Surpeta, B.
4. Sequeiros-Borja, C. E.
5. Mandal, N.
6. Sarkar, D. K.

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

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

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.

.....
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

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
Wydział Biologii
Instytutu Biologii Molekularnej i Biotechnologii UAM

Nishita Mandal, M.Sc.
Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland
Email: nishita.mandal@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 2 in Supplementary File 7 and wrote the draft of this document.

Nmandal 14/03/2023
Signature

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

Water will find a way: transport through narrow tunnels in hydrolases

1. Sequeiros-Borja, C. E.
2. Thirunavukarasu, A. S.
3. Dongmo Foumthuim C. J.
4. Brezovsky J.

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, May 26, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the first author of the following publication (pre-print):

Sequeiros-Borja CE, Thirunavukarasu AS, Dongmo Foumthuim CJ, Brezovsky J, 2023: Water will find a way: transport through narrow tunnels in hydrolases. bioRxiv doi doi.org/10.1101/2023.05.24.542065

I declare the following contribution to this publication:

contributed to preparation of hEpx and E470G (human epoxide hydrolase and one mutant respectively) systems for molecular dynamics simulations; ran molecular dynamics simulations and perform its initial analysis; performed the tunnel network analysis for all the proteins; performed the overall analysis of transport tunnels and transport of water for all systems; performed a detailed water transport and interactions analyses for all systems; preparation of the figures; writing, revising and editing the paper.

.....
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, May 26, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication (pre-print):

Sequeiros-Borja CE, **Thirunavukarasu AS**, Dongmo Foumthuim CJ, Brezovsky J, 2023: Water will find a way: transport through narrow tunnels in hydrolases. bioRxiv doi doi.org/10.1101/2023.05.24.542065

I declare the following contribution to this publication:

contributed to preparation of Hal (haloalkane dehalogenase) system for molecular dynamics simulations; ran molecular dynamics simulations and perform its initial analysis; revising the paper.

.....
Signature

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

Cedrix Dongmo Foumthuim, Ph. D.
Campus scientifico
via Torino 155
30172 Venice
Italy
Email: cedrix.dongmo@roma2.infn.it

Venice, May 26, 2023

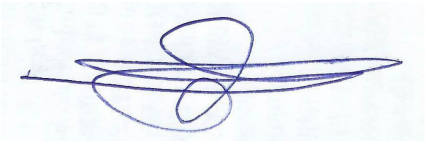
CO-AUTHOR CONTRIBUTION STATEMENT

Hereby, I certify that I am the co-author of the following publication (pre-print):

Sequeiros-Borja CE, Thirunavukarasu AS, **Dongmo Foumthuim CJ**, Brezovsky J, 2023: Water will find a way: transport through narrow tunnels in hydrolases. bioRxiv doi https://doi.org/10.1101/2023.05.24.542065

I declare the following contribution to this publication:

contributed to the preparation of Lip (lipase) system for molecular dynamics simulations; ran molecular dynamics simulations and performed its initial analysis; revising and editing the paper.



…….
Signature

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

Poznan, May 25, 2023

Jan Brezovsky, Ph.d.
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:

Sequeiros-Borja C, Thirunavukarasu AS, Dongmo Foumthuim CJ, Brezovsky J, bioRxiv 2023.05.24.542065; doi: https://doi.org/10.1101/2023.05.24.542065

I declare the following contribution to this publication:

conceived and coordinated the project; designed the calculations, set up, performed, and analyzed simulations of Epx; interpreted data; co-wrote the manuscript

.

.....
Signature

Water migration through enzyme tunnels is sensitive to choice of explicit water model

1. Thirunavukarasu A. S.
2. Szleper K.
3. Tanriver G.
4. Mitusinska K.
5. Gora A.
6. Brezovsky J.

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, August 24, 2023

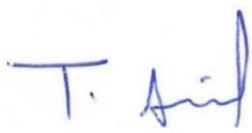
CO-AUTHOR CONTRIBUTION STATEMENT

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

Aravind Selvaram Thirunavukarasu, Katarzyna Szleper, Gamze Tanriver, Karolina Mitusinska, Artur Gora, Jan Brezovsky. Water migration through enzyme tunnels is sensitive to choice of explicit water model. bioRxiv 2023.08.14.553223; doi: https://doi.org/10.1101/2023.08.14.553223

I declare the following contribution to this publication:

contributed to methodology development for the project, data curation, formal analysis, investigation of DhaA; wrote the manuscript.



.....
Signature

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

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

Katarzyna Szleper, M.Sc.
Biotechnology Centre,Silesian University of Technology
ul. Krzywoustego 8, 44-100 Gliwice, Poland
Email: katarzyna.szleper@polsl.pl

Gliwice, August 24, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

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

Aravind Selvaram Thirunavukarasu, **Katarzyna Szleper**, Gamze Tanriver, Karolina Mitusinska, Artur Gora, Jan Brezovsky. Water migration through enzyme tunnels is sensitive to choice of explicit water model. bioRxiv 2023.08.14.553223; doi: https://doi.org/10.1101/2023.08.14.553223

I declare the following contribution to this publication:

contributed to data curation, formal analysis, investigation and writing original draft for Alditol oxidase (AldO).

Katarzyna Szleper
.....
Signature

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

Gamze Tanriver, Ph.D.
Biotechnology Centre,Silesian University of Technology
ul. Krzywoustego 8, 44-100 Gliwice, Poland
Email: gamze.tanriver@polsl.pl

Gliwice, August 24, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

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

Aravind Selvaram Thirunavukarasu, Katarzyna Szleper, **Gamze Tanriver**, Karolina Mitusinska, Artur Gora, Jan Brezovsky. Water migration through enzyme tunnels is sensitive to choice of explicit water model. bioRxiv 2023.08.14.553223; doi: https://doi.org/10.1101/2023.08.14.553223

I declare the following contribution to this publication:

contributed to data curation, formal analysis, investigation and writing original draft for Cytochrome P450 2D6(CYP2D6).

.....
Signature

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

Karolina Mitusińska, Ph.D.
Biotechnology Centre,Silesian University of Technology
ul. Krzywoustego 8, 44-100 Gliwice, Poland
Email: k.mitusinska@tunnelinggroup.pl

Gliwice, August 24, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

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

Aravind Selvaram Thirunavukarasu, Katarzyna Szleper, Gamze Tanriver, **Karolina Mitusinska**, Artur Gora, Jan Brezovsky. Water migration through enzyme tunnels is sensitive to choice of explicit water model. bioRxiv 2023.08.14.553223; doi: https://doi.org/10.1101/2023.08.14.553223

I declare the following contribution to this publication:

contributed to resources, software setup and reviewing & editing of the manuscript.

Karolina Mitusińska
Signature

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

Artur Gora, Ph.D.
Biotechnology Centre,Silesian University of Technology
ul. Krzywoustego 8, 44-100 Gliwice, Poland
Email: Artur.Gora@polsl.pl

Gliwice, August 24, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

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

Aravind Selvaram Thirunavukarasu, Katarzyna Szleper, Gamze Tanriver, Karolina Mitusinska, **Artur Gora**, Jan Brezovsky. Water migration through enzyme tunnels is sensitive to choice of explicit water model. bioRxiv 2023.08.14.553223; doi: https://doi.org/10.1101/2023.08.14.553223

I declare the following contribution to this publication:

contributed to conceptualization, administration, supervision, validation of the project, reviewing & editing of the manuscript

.....
Signature

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, August 28, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

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

Aravind Selvaram Thirunavukarasu, Katarzyna Szleper, Gamze Tanriver, Karolina Mitusinska, Artur Gora, **Jan Brezovsky**. Water migration through enzyme tunnels is sensitive to choice of explicit water model. bioRxiv 2023.08.14.553223; doi: https://doi.org/10.1101/2023.08.14.553223

I declare the following contribution to this publication:

contributed to conceptualization of the project, its supervision, data interpretation, and editing of 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