Uniwersytet im. Adama Mickiewicza, Poznań, Polska

Wydział Biologii

Rozprawa doktorska

Symulacje procesów wiązania ligandów w białkach

mgr. Dheeraj Kumar Sarkar

Promotor

dr hab. Jan Brezovsky, prof. UAM

ADAM MICKIEWICZ
UNIVERSITY
POZNAŃ

2023

Adam Mickiewicz University, Poznań, Poland

Faculty of Biology

Doctoral thesis

Simulations of ligand binding processes in proteins

Dheeraj Kumar Sarkar, M.Sc.

Promotor

dr hab. Jan Brezovsky, prof. UAM

ADAM MICKIEWICZ
UNIVERSITY
POZNAŃ

2023

ACKNOWLEDGEMENTS

With all my gratitude, I would like to thank,

My supervisor and mentor, **dr hab. Jan Brezovsky, Prof. UAM,** *for being such a fantastic guide, for his supervision and insightful feedback, and providing me with all resources.*

Head of my PhD studies, **dr hab. Agnieszka Ludwików, Prof. UAM,** *for all support during my doctoral studies.*

Prof. Dibyendu Mondal *and* **Prof. Ariane Nunes-Alves,** *for the valuable discussions, suggestions, and insights in my collaborative projects.*

dr Karolina Cerbin *and* **mgr Magdalena Walczak,** *for the assisting with the administrative matters related to my doctoral studies.*

mgr Arleta Kucz *and* **mgr Iwona Kanonik-Jędrzejak,** *for the invaluable support with the processing of co-finances for conferences and workshops.*

My **friends** *and* **colleagues,** *for their unwavering support, insightful talks, and encouragement.*

My **Parents**, *for the constant support and advice.*

Poznan Supercomputing Center, *for the computational resources and infrastructures during my doctoral studies.*

Dedicated to my Parents

TABLE OF CONTENTS

STRESZCZENIE

Procesy biologiczne są wewnętrznie regulowane przez małe cząsteczki i ich interakcje, szczególnie w postaci kompleksów białko-ligand. Transport ligandów w białkach odgrywa kluczową rolę w wielu procesach biologicznych, w tym w transdukcji sygnału, katalizie enzymatycznej oraz transporcie składników odżywczych i metabolitów. Dlatego zrozumienie procesów wiązania ligandów ma ogromne znaczenie dla opartego na strukturze projektowania leków i inżynierii ulepszonych katalizatorów enzymatycznych. W szczególności znaczna część enzymów ma swoje aktywne miejsce znajdujące się w głęboko osadzonych przestrzeniach, a wyzwania polegają na udanym uchwyceniu procesów wiązania i uwalniania ligandów, ponieważ dostęp do tych głęboko osadzonych miejsc przez małe cząsteczki jest ogólnie ograniczony przez tunele białkowe i swego rodzaju bramki molekularne. Zmiany w tunelach białkowych często mogą prowadzić do zmiany aktywności, selektywności, niespecyficzności i stabilności. Aby zaradzić tym przeciwnościom, w swoich badaniach doktorskich przyczyniłem się do zbadania właściwości termicznych i kinetycznych poprzez ocenę wykorzystania małych cząsteczek przez tunele transportowe w białkach z miejscami aktywnymi znajdującymi się w głęboko osadzonych przestrzeniach. Zastosowano metody dynamiki molekularnej (MD) z ulepszonym próbkowaniem, aby skutecznie ocenić termodynamikę i kinetykę procesów wiązania ligandów, jednocześnie rozumiejąc leżące u podstaw mechanizmy molekularne. Praca składa się z trzech manuskryptów. Pierwsza część pracy koncentruje się na opracowaniu metody poprzez zaprojektowanie opartych na wiedzy schematów umiejscawiania ligandów w celu skutecznego badania procesów interakcji ligandów w dehalogenazie haloalkanowej. Metoda wykorzystuje adaptacyjne symulacje próbkowania do wysokoprzepustowego próbkowania zjawisk wiązania kierując się modelami stanów Markowa (*ang. Markov State Models – MSM*) w celu generowania znaczących modeli kinetycznych opisujących długotrwałe stany białko-ligand w formie związanej i niezwiązanej, oraz procesy ich interkonwersji. W drugiej części rozprawy masowe wykorzystanie tuneli molekularnych w celu ułatwienia transportu ligandów zostało ocenione i określone ilościowo, aby uzyskać bardziej szczegółowy wgląd w procesy transportu ligandów, pokazując zastosowanie wewnętrznie opracowanego narzędzia. Wreszcie trzecia część pracy dotyczy dostępności aktywnego miejsca cytochromu C dla zatłoczonych hydrotropów oraz roli ich wiązania na stabilność termiczną tego enzymu. Proces badano w dwóch temperaturach obejmujących trzy różne kompozycje hydrotropów, stosując adaptacyjne symulacje próbkowania kierowane modelami Markowa. Ujawniono kilka spostrzeżeń na temat roli rozpuszczalników hydrotropowych w zapewnianiu stabilności funkcjonalnym częściom cytochromu C i regulacji dynamiki stanów otwartych i zamkniętych. Ogólnie rzecz biorąc, teza reprezentuje zastosowanie, ocenę, oraz opracowanie wysokoprzepustowego protokołu dynamiki molekularnej wykorzystującego kompleksy białko-ligand do badania procesów wiązania ligandów i jego roli w lepszym zrozumieniu sprzężenia między dynamiką a funkcją enzymów.

Słowa kluczowe

Adaptacyjne wysokowydajne symulacje, procesy wiązania ligandów, modele stanów Markowa, enzymy, tunele, bramy

ABSTRACT

Biological processes are intrinsically regulated by small molecules and their interactions, particularly in the form of protein-ligand complexes. The transport of ligands in proteins plays a crucial role in many biological processes, including signal transduction, enzyme catalysis and the transport of nutrients and metabolites. Therefore, understanding ligand binding processes is of great importance for the structure-based design of drugs and the engineering of improved enzyme catalysts. Notably, a significant fraction of enzymes have their active site buried in deep cavities, and the challenges lie in successfully capturing the ligand binding and unbinding processes, as access to those cavities by small molecules is generally restricted by protein tunnels and gates. Changes in protein tunnels can often lead to altered activity, selectivity, promiscuity, and stability. To address these shortcomings, in my Ph.D. research, I contributed to investigating thermal and kinetic properties by assessing the utilization of small molecules via transport tunnels in proteins with buried active sites. Enhanced molecular dynamics (MD) methods were applied to effectively assess the thermodynamics and kinetics of ligand binding processes while understanding the underlying molecular mechanisms. The thesis consists of three manuscripts. The first part of the thesis focuses on developing a method by designing knowledge-based seeding schemes to study ligand interaction processes in haloalkane dehalogenase effectively. The method employs adaptive sampling simulations for high-throughput sampling of binding phenomena, guided by Markov State Models (MSMs) to generate meaningful kinetic models describing protein-ligand bound and unbound long-lived states and their interconversion processes. In the second part of the thesis, the massive use of molecular tunnels to facilitate ligand transport was evaluated and quantified to gain more detailed insights into ligand transport processes, showcasing the applicability of the in-house developed software tool. Finally, the third part of the thesis deals with the accessibility of the Cytochrome c active site for crowded hydrotropes and the role of their binding on the thermal stability of this enzyme. The process was studied at two temperatures involving three different compositions of the hydrotropes using adaptive sampling simulations guided by Markov models. Several insights were revealed into the role of hydrotropic solvents in providing stability to functional parts of Cytochrome c and regulating the dynamics of the open and closed states. Overall, this thesis represents the application, evaluation, and development of a high-throughput simulations protocol using protein-ligand complexes to study ligand binding processes and its role in improving the understanding of the coupling between the dynamics and function of enzymes.

Keywords

Adaptive high-throughput simulations, ligand binding processes, Markov state models, enzymes, tunnels, gates

LIST OF PUBLICATIONS

Publications included in the thesis

1. **Preprint**

   Sarkar, D.K.; Surpeta, B.; Brezovsky J. Incorporating prior knowledge to seeds of adaptive sampling molecular dynamics simulations of ligand transport in enzymes with buried active sites. *bioRxiv* 2023.09.21.558608. DOI: https://doi.org/10.1101/2023.09.21.558608.

   Impact Factor 2022: NA

   MNiSW points 2023: NA

2. **Research article**

   Brezovsky, J.; Thirunavukarasu, A. S.; Surpeta, B.; Sequeiros-Borja, C. E.; Mandal, N.; **Sarkar, D. K.**; Foumthuim, C. J. D.; 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.

   Impact Factor 2021: 6.931

   MNiSW points 2022: 200

3. **Research article**

   Bharadwaj, P.[#]; **Sarkar, D. K.**[#]; Bisht, M.; Shet, S. M.; Nataraj, S. K.; Lokesh, V.; Franklin, G.; Brezovsky, J.; Mondal, D. Nano-Structured Hydrotrope-Caged Cytochrome c with Boosted Stability in Harsh Environments: A Molecular Insight. *Green Chemistry* **2023**. https://doi.org/10.1039/d3gc01704d.

   Impact Factor 2022: 9.8

   MNiSW points 2023: 200

[#] - shared first author

Publications not included in the thesis

1. **Research article**

   Bharadwaj, P.; Shet, S. M.; Bisht, M.; **Sarkar, D. K.;** Franklin, G.; Nataraj, S. K.; & Mondal, D. (2023). Suitability of Adenosine Derivatives in Improving the Activity and Stability of Cytochrome c: Insights into the Effect of Phosphate Group. (submitted to the *Journal of Physical Chemistry B*)

   Impact Factor 2022: NA

   MNiSW points 2023: NA

2. **Research article**

   Jayaprakash, N. G.; **Sarkar, D. K.**; Surolia, A. Atomic Visualization of Flipped-back Conformations of High Mannose Glycans Interacting with Cargo Lectins: An MD Simulation Perspective. *Proteins* **2023**. https://doi.org/10.1002/prot.26556.

   Impact Factor 2022: 2.9

   MNiSW points 2023: 100

[#] - shared first author

ABBREVIATIONS

DBE	1,2-dibromoethane
HLD	Haloalkane dehalogenase
MD	Molecular dynamics
HTMD	High-throughput molecular dynamics
MSM	Markov state models
CV	Collective variables
WE	Weighted ensemble
SEEKR	simulation-enabled estimation of kinetics rates
RAMD	Random accelerated molecular dynamics
GaMD	Gaussian molecular dynamics
ATP	adenosine-5’-triphosphate
IL	Ionic liquid
CytC	Cytochrome c

INTRODUCTION TO LIGAND TRANSPORT PROCESSES IN PROTEINS

Biological processes are governed by biomolecular interactions, particularly protein-small molecule, protein-protein, and protein-nucleic acid complexes.[1,2] Also, transport of such small molecules or ligands in proteins plays a crucial role in several biological processes, including signal transduction, enzyme catalysis, and the transporting nutrients and metabolites.[3–5] The understanding of ligand binding processes is therefore essential in the structure-based drug design[6,7] and development of enzyme catalysts.[8] Due to the inherent dynamic and often volatile nature of protein-ligand (un)binding processes, there is a long-standing quest to capture high-resolution ligand binding processes in proteins. Notably, a significant fraction of proteins (~60%) have their active site buried in deep cavities and are accessible by small molecules or water via molecular tunnels or channels.[7–10] Tunnels can facilitate the transport of ligands to buried active site. Often changes in protein tunnels may result in altered activity, selectivity, promiscuity, and stability.[11,12] Therefore, it is imperative to understand the role of tunnels in modulating or regulating such vital life processes, for which several critical questions have been assessed in the current doctoral thesis work (**Figure 1**).

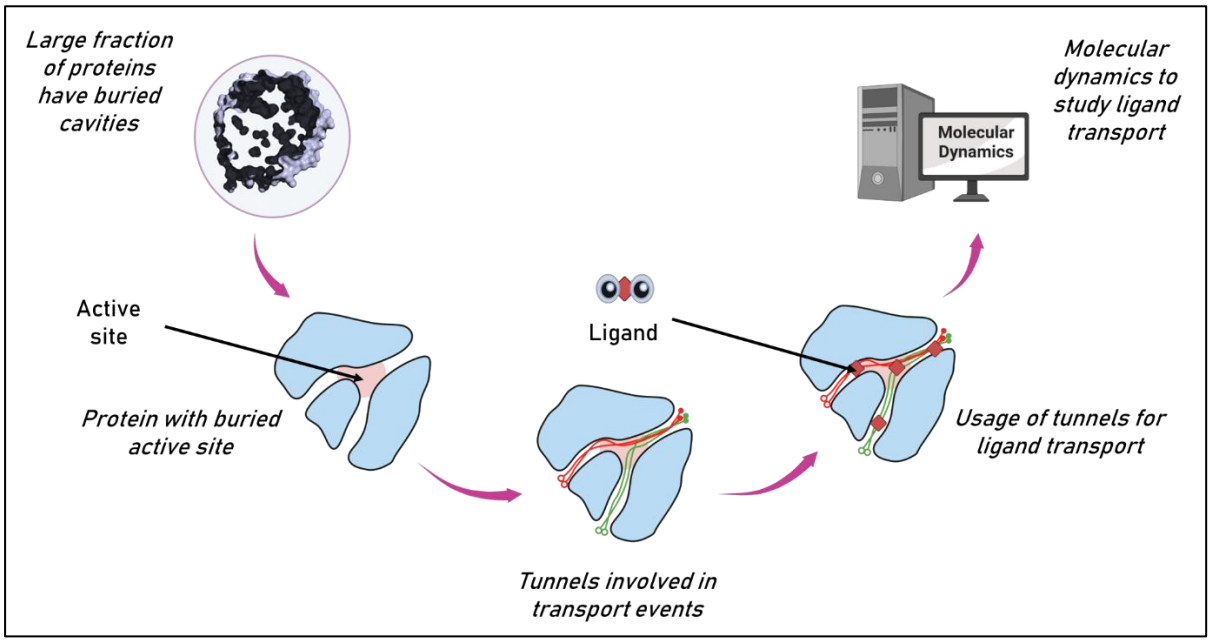


Figure 1. Schematic representation critical questions in assessing protein tunnels. *The usage of molecular dynamic simulations in investigating ligand binding processes in proteins with buried active sites is highlighted.*

Despite significant advancements in the drug design field, several challenges still need to be addressed. One such challenge is effectively investigating high-resolution binding processes in proteins with buried active sites, where protein tunnels and their gating residues often restrict access to the active site. Small molecules like water and others usually utilize these tunnels and gates for accomplishing enzymatic catalysis and reactions.[11]

Several experimental and molecular dynamics (MD) simulations have captured binding processes in proteins with and without buried active sites. However, these methods have limitations, including low temporal resolution and limited conformational sampling. Hybrid-enhanced sampling methods have been developed to overcome these limitations and capture binding processes with higher temporal resolution and increased conformational sampling.[2,13,14] There is a considerable need to correctly resolve relevant protein-ligand binding kinetics in structure-based drug design and rational protein engineering due to their importance in lead compound optimization and drug screening.[15] For example, classical MD simulations have been implemented over the years in deriving the molecular level understanding of some pharmaceutically relevant proteins FKBP506, Src kinase, G protein-coupled receptors to study the protein-ligand binding processes and the associated kinetics.[16,17] However, atomistic simulations often encounter problems in sampling effectively slow biological processes in milliseconds to microseconds timescales.[18] Thus, enhanced sampling methods are very effective for studying such rare events.[19] Of such implementations, the methods that focus on molecular dynamics are weighted ensemble (WE) methods,[20–22] milestoning,[23] simulation-enabled estimation of kinetics rates (SEEKR),[24] Gaussian accelerated molecular dynamics (GaMD)[25] Metadynamics,[26,27] Markov state model (MSM),[28,29] Random accelerated Molecular dynamics (RAMD),[30] scaled MD[31] and many more. Among those enhanced sampling methods, Weighted ensemble, MSM, Metadynamics, GaMD, and SEEKR are the most often utilized improved sampling methods for studying biomolecular association and dissociation kinetics. Methods like metadynamics simulations can be used for high-definition sampling with predefined collective variables (CVs) but are often challenged by hidden energy barriers and slow sampling convergence in case low-quality CVs are defined. CV-free methods, like WE and MSM, are biased-free simulations but often suffer from exploiting the binding-pose and proper utilization of transport pathways for sampling ligand binding processes.[15,16]

Recently, hybrid methods of GaMD, such as LiGaMD and DeepWest, have been markedly successful in sampling the attractive conformational space in protein-ligand complexes.[19,32] In the LiGaMD variant of GaMD, selective boosts were applied to non-bonded interactions to enable ligand dissociations and with the advantages of GaMD in having significantly reduced energy barriers associated with ligand binding processes. On the other hand, variant DeepWest combines GaMD and WE to effectively reach the convergence by providing a well-sampled initial distribution or conformations of protein-ligand complexes followed by unbiased short WE simulations and finally metastable states guided by deep learning were resolved using MSM to tackle proper equilibrium distribution in the simulations to remove the initial bias. Markov state models (MSMs) are advantageous as such it does not require long MD trajectories for sampling equilibrium distributions, enabling their efficient execution in high-throughput parallel manner.[33,34] MSMs are useful for predicting the kinetics of protein-ligand (un)binding events and identifying the key conformational states involved in these slow processes or rare events.[35–37]

The utility of adaptive sampling simulations using MSMs (Figure 2) has motivated me to implement MSMs in my research to gain a deeper understanding of the transport of ligands in proteins. Additionally, I have evaluated a methodological approach to design effective seed conformations for adaptive simulations to boost exploration of regions with high energy barriers that occur in ligand binding processes in protein with buried active sites. The approach is discussed in the first part of the thesis, shedding light on the effect of using additional knowledge to generate the initial conformations of protein-ligand complexes to streamline the adaptive sampling of unfavorable regions earlier reported in adaptive sampling protocols.[38] In the second section, I contributed to the development of new software tool, TransportTools, and showcased its application to effectively quantify the utilization of transport tunnels used for ligand mitigations from the active site to the bulk solvent environment. Finally, I tested the applicability of adaptive sampling in investigating the effects of binding of hydrotope solvents, such as ATP and Choline salicylate, to the cytochrome c metalloprotein on its stability and function.

Overall, the thesis aims to contribute towards effective exploration the ligand transport processes in proteins with buried active sites and investigate the use of adaptive sampling

methods based on MSMs to capture and understand the thermodynamics and kinetics of association/dissociation of small molecules in proteins.

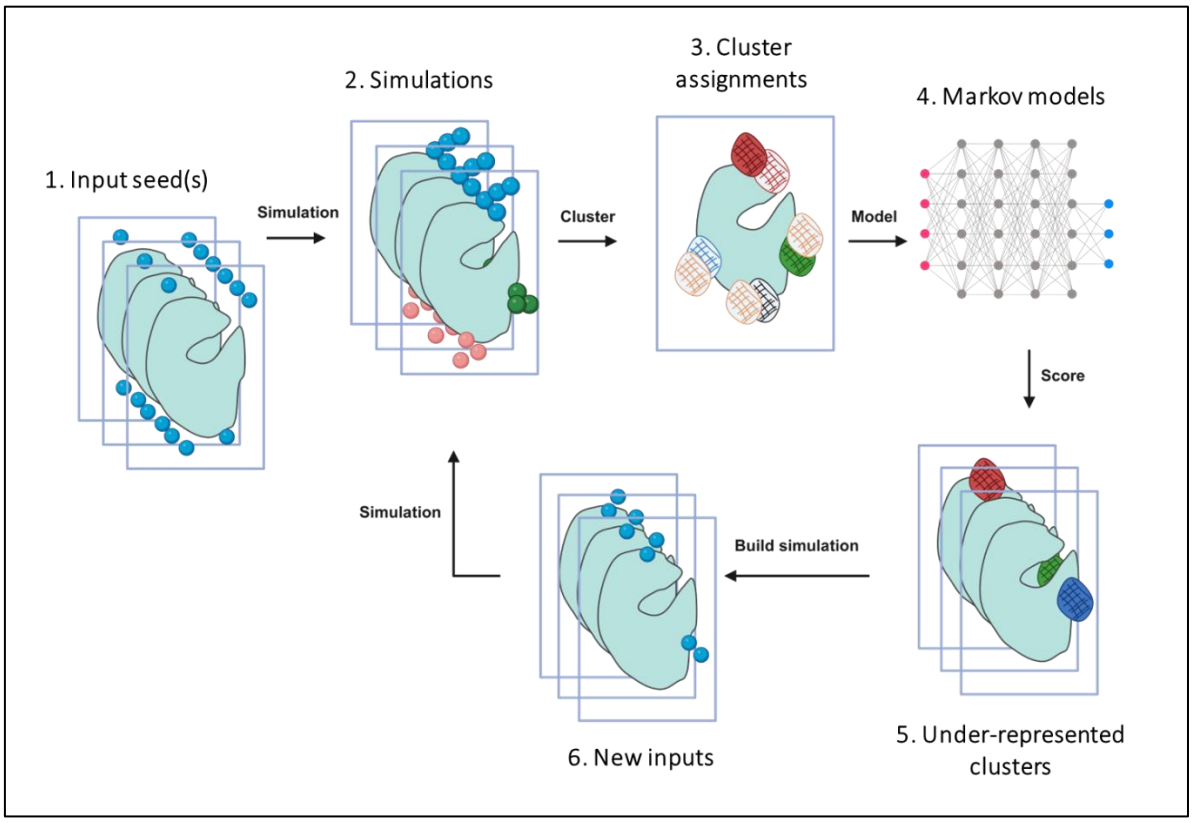


Figure 2. Schematic representation of the adaptive sample protocol. *Short simulations are respawned while generating Markov models from sampling the most interesting regions in ligand binding processes.*

SUMMARY OF DOCTORAL RESEARCH

Using knowledge-based seeding of high-throughput molecular dynamics for ligand binding studies

A large fraction of proteins has their active sites located in deep cavities within their structure which are connected to the bulk environment by tunnels or molecular pathways. The characteristics of such pathways that link the active site to the surrounding solvent impact the binding and release of ligands, as well as the enzyme's catalytic capabilities[12]. As a result, tunnels play a crucial role in altered activity, substrate specificity, enantioselectivity, and stability.[10] Haloalkane dehalogenases (HLDs), LinB (EC 3.8.1.5) are one such enzyme model belonging to the α/β-hydrolases family. HLDs are suitable models for basic enzymology studies due to the demonstrated significance of the amino acid residues impacting the size and shape

of the tunnels connecting the active site and protein surface for substrate specificity and enzymatic activity.[39–43] This enzyme has two halide stabilizing residues Asn-38 and Trp-109 and the charged residues forming the catalytic triad are Asp-108, His272, and Glu-132. This is well supported by structural and biochemical data available for HLDs.[44] HLDs are known to catalyse the hydrolysis of halogenated compounds by formation of corresponding alcohol, a halide ion and a proton.[45] Additionally, the transport routes in LinB are well characterized with three major variants viz. LinB_Wt (wild type), LinB32 (closed form) and LinB86 (open form).[45,46] Among these variants, LinB86 was engineered as the most potential variants with de novo ***p3*** functional tunnels that can potentially utilize both alcohol and halide ion products along with water molecules for its catalysis.[40,41] Tunnel gates and bottleneck residues play critical roles in facilitating significant control over the entry and exit of small molecules,[47] as well as the involvement of multiple transient states during the process of small molecule association and dissociation.[29] One of the key goals of my thesis work was to construct seed conformations utilizing tunnel bottleneck information to avoid kinetic traps that are frequently countered in MD simulations studies of protein-ligand binding complexes.

This is addressed in the current part of the thesis, where I focused on the development of the methodological approach to investigate the role of applying more knowledge in generating initial conformation or seed conformations **(Figure 3)** for adequate sampling and systematically assess ligand (un)binding processes in LinB86 and its substrate 1,2-dibromoethane (DBE). To accomplish this goal, we devised four schemes **(Figure 3)**, starting from positioning the substrate in more random places in the protein active site and bulk solvent environments. The more knowledge-based approaches included seed conformations of DBE at the bottleneck portions of the tunnels to overcome high energy barriers. This knowledge-based strategy was based on coupling tunnel information in LinB86 using the CAVER3 tool[48] tunnels and then leveraging tunnel paths and dynamics from MD simulations. Finally, the high-affinity binding areas were calculated using a docking tool, CaverDock[49], based on the upper bound energies and tunnel bottleneck information. Overall, the generated seed conformations were used as initial conformations for adaptive sampling simulations. To examine the capability of various schemes in sampling relevant parts of LinB86 and developing more meaningful kinetic models, I ran three replicate adaptive sampling high-throughput simulations (~900 simulations) with a total of 45 μs duration for each replica.

In this study, we systematically studied the slow binding process of LinB86-DBE while assessing the role of applying more knowledge in generating initial inputs for adaptive sampling simulations guided by Markov models.[33,50] We implemented the HTMD protocol[50] which consists of several batches of short MD simulations that are iteratively run in a more intelligently, and only the interesting regions or rare events are respawned for next batch of simulations (Figure 2). Interestingly, the study indicated that applying more knowledge could significantly influence the binding processes in sampling protein conformational space leading to generation of more meaningful kinetic models. For example, scheme "Bulk" (randomly placed DBE seeds in bulk solvent around the protein) was very inefficient in sampling binding processes pertaining to DBE migration to the depth of buried active site. Hence, such a scheme was insufficient in investigating DBE association processes. Moving to more knowledge-based seeding schemes, "Cavity" could significantly sample bound poses and egress of DBE to the bulk but frequently failed in sampling DBE binding to the active site. Such limitation was overcome in the "Cavity&Bulk" approach, which was found capable of sampling a substantial number of bound poses and release from cavity as well as ligand access events, since the initial seeds had both bound and bulk locations. Finally, the information rich schemes, "Cavity&Bulk" and "Tunnels" could sample more effectively the complex transport pathways of LinB86 including the auxiliary pathways of ***p2*** and ***p3*** than any other schemes. Additionally, "Tunnels" scheme had more control over the sampling of (un)binding pathways, whereas it lacked sampling of bound as well as bulk regions as the initial seeds were comprised of DBE positioned in the bottlenecks of LinB86 tunnels. Therefore, schemes like "Tunnels" could be utilized in exploring more meaningful states to connected with use of complex transport network in proteins with buried cavities. Considering the ability of the MSM models to predict k_{off}/k_{on}, both "Cavity&Bulk" as well as "Tunnels" seeding schemes consistently reached values on the same order of magnitude with the experimental observations. Overall, the current work argues that using prior information to generate seeds can systematically escape the kinetic traps that protein-ligand simulations can encounter while accessing all the intriguing conformational space of protein-ligand complexes. Also, seeding with information on tunnel ensemble could be beneficial for initiation of MD methods which requires prior knowledge for example, umbrella sampling[51] or metadynamics[26] to study ligand binding processes. A full description of the study is available in Publication 1.

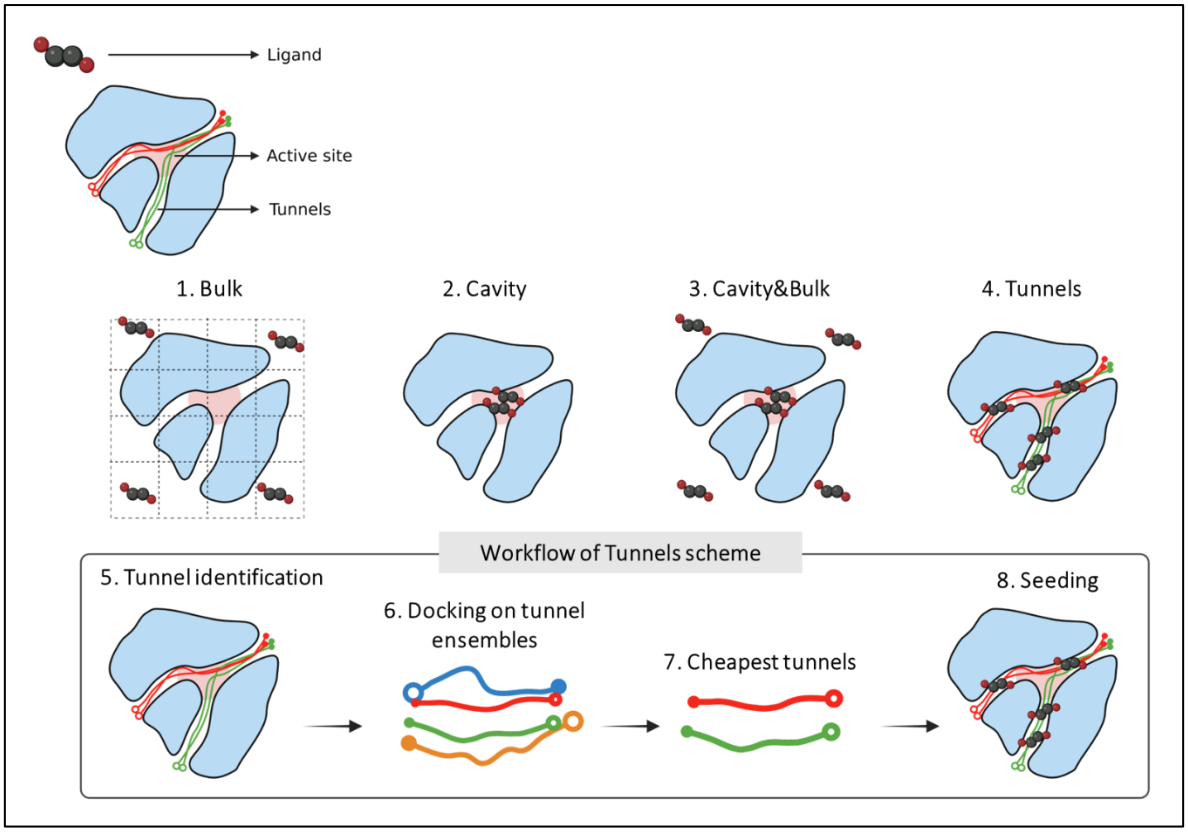


Figure 3. Schematic representations of designed schemes for seeding. *The studied schemes are from random to more knowledge-based seeding of ligands (1. Bulk to 4. Tunnels). The workflow of 4. Tunnels is depicted with 5. Identification of tunnels from MD trajectories by CAVER, 6. Ligand trajectory-based docking using CaverDock on tunnel ensembles, 7. Profiling for cheapest tunnel parts or chunks with lower energy barriers, 8. Ligand positioning at the cheapest chunks of tunnels and preparation of seed conformations.*

Identification and utilization of ligand transport processes in proteins

Since the first part of my thesis is primarily focused on study of substrate (un)binding processes with the LinB86 variant of *haloalkane dehalogenase*, the assessment of ~45us of adaptive sampling simulations coupled with four different schemes and three replicates, altogether ~10,000 simulations made it clear that we need a tool to robustly couple the rare transport events with utilization of individual transport tunnels in LinB86 by its DBE substrate. The complexity of the characterized tunnel networks for utilization of small molecules in LinB86 and other variants have been featured by Brezovsky *et al.,* 2016.[46] Additionally, LinB86's vast sampling of substrate DBE led by Markov models provided a vivid picture of the mitigation of DBE, the long-live metastable states and associated transition networks. It was clearly evident that most of the characterized paths ***p1a, p1b, p2*** and ***p3*** were utilized for the transport of DBE

in/out of the active site to bulk. Given the complex network of the metastable states which involved the transient states of the ligand in preferred pathways of LinB86, it would be very intuitive to properly assessing which tunnels are prioritized by the substrate for traversing to and from active site to the bulk environment. However, having the massively sampled dataset of LinB86-DBE complexes, the challenge lies in systematically tracking the mitigation of ligands from active site to bulk environment and answering which tunnels are relevant or mostly preferred during transport of the small molecules. To address the issue, TransportTools[52], a python-based library **(Figure 4)** was developed to properly investigate the entry/release of small molecules in protein with buried cavities or active sites. The utility and workflow of the tool have been tested and discussed in **Publication 2.** TransportTools relies on the CAVER tool[48] for calculating the geometry of tunnels and the AQUA-DUCT tool[53] for tracking the movement of ligands in proteins with buried cavities to assess the composite tunnel networks. The preferred selectivity of DBE was analyzed, and the ***p1b*** transport pathway was found to be the most preferred then other pathways. It is noteworthy that the proportion of ligand-tracking pathway samples in the whole dataset was relatively minimal due to the higher saving frequency (100 ps) used in the adaptive simulations of LinB86 with DBE, as in the first part of my research work, the focus was to develop a cost-effective methodology to effectively investigate the ligand dynamics and kinetics from protein-ligand sampling space. However, a lower saving frequency of ~10-20 ps would improve the resolution of those sampling processes at a higher computational cost. Also, the nature of adaptive sampling simulations, which do not always exploit ligand migration in and out of the active site to bulk solvent and are sensitive to initial poses, as observed in previous studies as well as in **Publication 1**, clearly indicated that the addition of knowledge-based seeds could be very useful in tracing ligand mitigation events by AQUA-DUCT[53] and inferring to the utilization of pathways by TransportTools.[52]

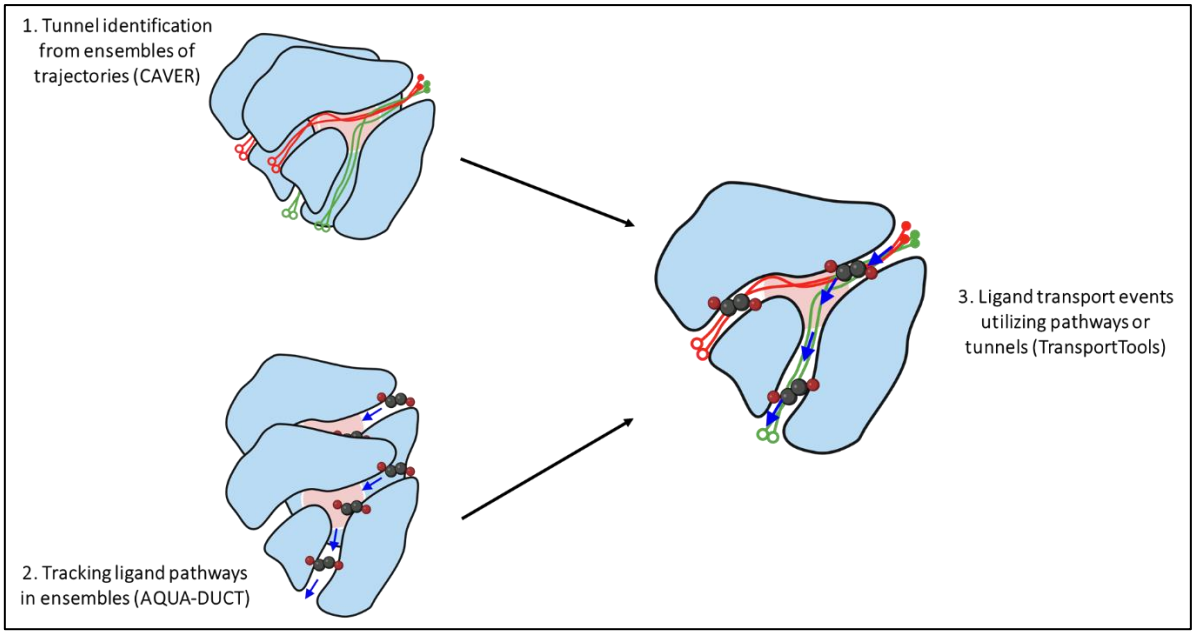


Figure 4. Brief workflow of TransportTools. *The massive dataset is assessed for utilization of individual tunnel by small molecules or ligands based by using CAVER for information on tunnel geometry and AQUA-DUCT for tracking the movement of small molecules and finally unifying the transport events mitigating in and out of individual tunnels or those utilizing the transport pathways.*

Effect of hydrotropes' binding on accessibility and stability of Cyt c active site

To investigate the applicability of the study of protein-ligand binding processes and its role in thermal stability and activity, I studied the solvation effects of adenosine-5'-triphosphate (ATP) and Choline Salicylate ionic liquid (IL) hydrotrope molecules in Cytochrome c (Cyt c) metalloprotein in collaboration with the experimental group of Prof. Dibyendu Mondal, Institute of Plant Genetics, Polish Academy of Sciences, Poznan. To effectively evaluate the binding of those molecules to conformational ensemble of Cyt c, I used HTMD guided by Markov models, which I became familiar with during the assessment of LinB86-DBE (un)binding processes.

To evaluate the molecular effects of hydrotropic solvents ATP and IL in dynamics and peroxidase activity of Cyt c, four molecular systems were investigated, one was the control with Cyt c in explicit water, the second system consisted of ATP added at 5 mM concentration and Cyt c, the third system contained IL at 300 mg/mL concentration and Cyt c and finally, the fourth system was simulated with Cyt c in presence of both ATP and IL at given concentrations. All simulations were performed using adaptive sampling methods in explicit water and at 300 K and 363.15 K

to assess the role of interacting molecules in room and near boiling temperatures. Hydrotrope molecules can solubilize and enhance protein solubility, specifically ATP at a concentration 2-10 mM can act as an effective hydrotropic solvent.[54] Additionally, IL is of particular interest due to their hydrotropic properties and marked effects on protein activity and stability.[55]

Simulations of the four designed molecular systems revealed significant role of ATP and IL in regulation of the dynamics of Cyt c, at room and higher temperature. ATP was found to provide thermal stability to Cyt c by interacting with charged amino acids specifically Lys and Arg by forming hydrogen bonds. Additionally, both the molecules (ATP and IL) could synergistically effect the dynamics of Ω 40-54 functional loop required for substrate access to the active site of Cyt c (role in apoptotic interactions)[56] in facilitating its reversible openings at room temperature. While, in higher temperature the hydrotrope solvents were stabilizing the Ω 70-85 functional region (role in peroxidase activity).[57] A representative view of the modulation and effects of hydrotropes is depicted in **Figure 5**. Overall, computational analysis revealed the backbone dynamics of Cyt c was nearly reduced to half in presence of ATP or IL or both the solvents. On the other hand, the interaction of the solvents with heme core revealed very limited contacts with indications the solvents could modulate thermal stability and activity without competing for accessibility of Cyt c active site for peroxidase activity. The work has been extensively discussed in **Publication-3**.

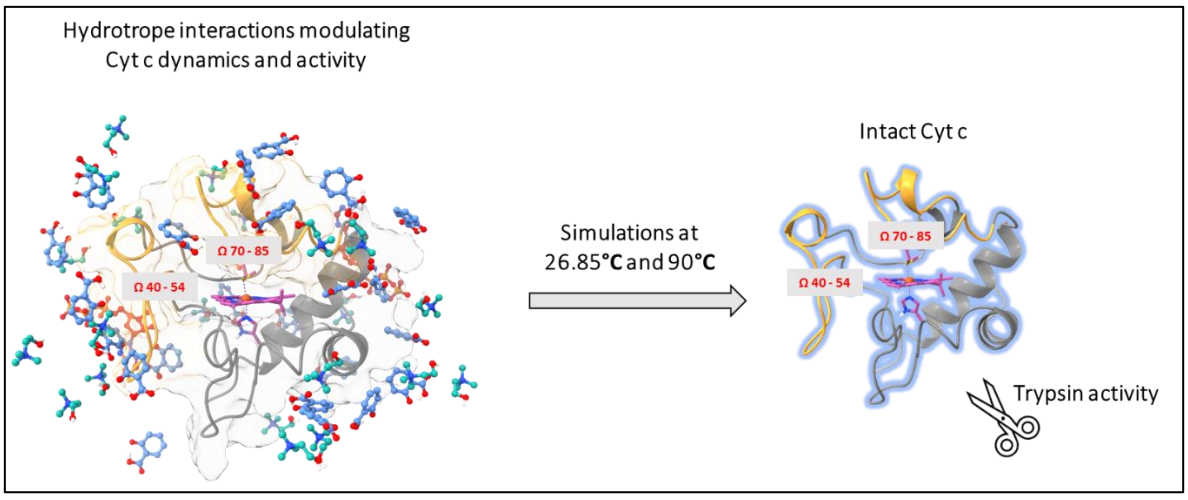


Figure 5. Representative view of hydrotropes interacting with Cyt c. *The hydrotropes viz. ATP and choline salicylate ionic liquid were synergistically stabilizing the Ω 40-54 and Ω 70-85 functional parts at 26.85°C and 90°C respectively. The hydrotropes were mostly interacting with charged amino acids Lys and Arg and were shown to keep the peroxidase activity of Cyt c intact.*

CONCLUSIONS AND FUTURE PERSPECTIVES

Using a more detailed exploration and critical assessment of ligand binding processes in proteins with buried active sites, the current thesis considers various considerations for the applicability of high-throughput molecular dynamics towards effective study of ligand transport and critical interactions that are essential for modulating protein dynamics and functions.

Here, the challenges of exploring and exploiting rare event sampling processes using adaptive sampling protocols were investigated. By employing knowledge-based seeding schemes and adaptive sampling simulations guided by Markov State Models (MSMs), the study revealed the role of applying more knowledge to seed conformations while sampling rare binding events in the context of potential variants of *haloalkane dehalogenase*. Such a seeding approach could also be beneficial for enhanced sampling MD methods relying on path sampling that need more appropriate collective variables to sample biological processes of protein-ligand association.

Additionally, the issues that may arise from large datasets of protein-ligand binding processes while systematically tackling and investigating the utilization of transport pathways by small molecules, were outlined, and the TransportTools package was introduced as their possible solution.

Finally, the impact of different solvents on Cytochrome c dynamics and peroxidase activity was explored. Notably, the study revealed the stabilizing effects of the hydrotropic solvents ATP and IL on distinct functional regions of the metalloprotein.

Overall, this thesis contributes to the field of molecular dynamics simulations of protein systems. It provides novel insights into the role of initial seed conformations, transport pathways, and solvent interactions in modulating ligand binding processes and protein behavior. The methodologies and tools developed herein offer alternative avenues for future research in understanding protein-ligand binding processes and revealing the molecular determinants governing the association and dissociation of ligands with potential hot spots for mutagenesis. Hence, the developed methodology and tools in the current thesis provide a vivid implication in drug design and *in silico* protein engineering applications.

REFERENCES

(1) Nooren, I. M. A.; Thornton, J. M. Diversity of Protein-Protein Interactions. *EMBO Journal*. 2003. https://doi.org/10.1093/emboj/cdg359.

(2) Wang, J.; Do, H. N.; Koirala, K.; Miao, Y. Predicting Biomolecular Binding Kinetics: A Review. *J Chem Theory Comput* **2023**, *19* (8), 2135–2148. https://doi.org/10.1021/acs.jctc.2c01085.

(3) Chu, K.; Vojtchovský, J.; McMahon, B. H.; Sweet, R. M.; Berendzen, J.; Schlichting, I. Structure of a Ligand-Binding Intermediate in Wild-Type Carbonmonoxy Myoglobin. *Nature* **2000**, *403* (6772). https://doi.org/10.1038/35002641.

(4) 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.

(5) Moreira, C.; Calixto, A. R.; Richard, J. P.; Kamerlin, S. C. L. The Role of Ligand-Gated Conformational Changes in Enzyme Catalysis. *Biochemical Society Transactions*. 2019. https://doi.org/10.1042/BST20190298.

(6) Ferruz, N.; De Fabritiis, G. Binding Kinetics in Drug Discovery. *Molecular Informatics*. 2016. https://doi.org/10.1002/minf.201501018.

(7) Marques, S. M.; Daniel, L.; Buryska, T.; Prokop, Z.; Brezovsky, J.; Damborsky, J. Enzyme Tunnels and Gates As Relevant Targets in Drug Design. *Medicinal Research Reviews*. 2017. https://doi.org/10.1002/med.21430.

(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). https://doi.org/10.1371/journal.pcbi.1005398.

(9) 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). https://doi.org/10.1186/s12859-014-0379-x.

(10) Kingsley, L. J.; Lill, M. A. Substrate Tunnels in Enzymes: Structure-Function Relationships and Computational Methodology. *Proteins* **2015**, *83*, 599–611. https://doi.org/10.1002/prot.24772.

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

(12) Kokkonen, P.; Bednar, D.; Pinto, G.; Prokop, Z.; Damborsky, J. Engineering Enzyme Access Tunnels. *Biotechnol Adv* **2019**, *37* (6), 1–13. https://doi.org/10.1016/j.biotechadv.2019.04.008.

(13) Schuetz, D. A.; de Witte, W. E. A.; Wong, Y. C.; Knasmueller, B.; Richter, L.; Kokh, D. B.; Sadiq, S. K.; Bosma, R.; Nederpelt, I.; Heitman, L. H.; Segala, E.; Amaral, M.; Guo, D.; Andres, D.; Georgi, V.; Stoddart, L. A.; Hill, S.; Cooke, R. M.; De Graaf, C.; Leurs, R.; Frech, M.; Wade, R. C.; de Lange, E. C. M.; IJzerman, A. P.; Müller-Fahrnow, A.; Ecker, G. F. Kinetics for Drug Discovery: An Industry-Driven Effort to Target Drug Residence Time. *Drug Discovery Today*. 2017. https://doi.org/10.1016/j.drudis.2017.02.002.

(14) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron*. 2018. https://doi.org/10.1016/j.neuron.2018.08.011.

(15) Jagger, B. R.; Ojha, A. A.; Amaro, R. E. Predicting Ligand Binding Kinetics Using a Markovian Milestoning with Voronoi Tessellations Multiscale Approach. *J Chem Theory Comput* **2020**, *16* (8). https://doi.org/10.1021/acs.jctc.0c00495.

(16) Huang, D.; Caflisch, A. The Free Energy Landscape of Small Molecule Unbinding. *PLoS Comput Biol* **2011**, *7* (2). https://doi.org/10.1371/journal.pcbi.1002002.

(17) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J Am Chem Soc* **2011**, *133* (24). https://doi.org/10.1021/ja202726y.

(18) Bernetti, M.; Cavalli, A.; Mollica, L. Protein-Ligand (Un)Binding Kinetics as a New Paradigm for Drug Discovery at the Crossroad between Experiments and Modelling. *MedChemComm*. 2017. https://doi.org/10.1039/c6md00581k.

(19) Ojha, A. A.; Thakur, S.; Ahn, S. H.; Amaro, R. E. DeepWEST: Deep Learning of Kinetic Models with the Weighted Ensemble Simulation Toolkit for Enhanced Sampling. *J Chem Theory Comput* **2023**, *19* (4). https://doi.org/10.1021/acs.jctc.2c00282.

(20) Donyapour, N.; Roussey, N. M.; Dickson, A. REVO: Resampling of Ensembles by Variation Optimization. *Journal of Chemical Physics* **2019**, *150* (24). https://doi.org/10.1063/1.5100521.

(21) Lotz, S. D.; Dickson, A. Wepy: A Flexible Software Framework for Simulating Rare Events with Weighted Ensemble Resampling. *ACS Omega* **2020**, *5* (49). https://doi.org/10.1021/acsomega.0c03892.

(22) Nunes-Alves, A.; Zuckerman, D. M.; Arantes, G. M. Escape of a Small Molecule from Inside T4 Lysozyme by Multiple Pathways. *Biophys J* **2018**, *114* (5). https://doi.org/10.1016/j.bpj.2018.01.014.

(23) Narayan, B.; Buchete, N. V.; Elber, R. Computer Simulations of the Dissociation Mechanism of Gleevec from Abl Kinase with Milestoning. *Journal of Physical Chemistry B* **2021**, *125* (22). https://doi.org/10.1021/acs.jpcb.1c00264.

(24) Votapka, L. W.; Amaro, R. E. Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning. *PLoS Comput Biol* **2015**, *11* (10). https://doi.org/10.1371/journal.pcbi.1004381.

(25) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J Chem Theory Comput* **2015**, *11* (8). https://doi.org/10.1021/acs.jctc.5b00436.

(26) Tiwary, P.; Parrinello, M. From Metadynamics to Dynamics. *Phys Rev Lett* **2013**, *111* (23). https://doi.org/10.1103/PhysRevLett.111.230602.

(27) Wang, Y.; Valsson, O.; Tiwary, P.; Parrinello, M.; Lindorff-Larsen, K. Frequency Adaptive Metadynamics for the Calculation of Rare-Event Kinetics. *Journal of Chemical Physics* **2018**, *149* (7). https://doi.org/10.1063/1.5024679.

(28) Plattner, N.; Noé, F. Protein Conformational Plasticity and Complex Ligand-Binding Kinetics Explored by Atomistic Simulations and Markov Models. *Nat Commun* **2015**, *6* (May). https://doi.org/10.1038/ncomms8653.

(29) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations. *Proc Natl Acad Sci U S A* **2011**, *108* (25), 10184–10189. https://doi.org/10.1073/pnas.1103547108.

(30) Nunes-Alves, A.; Kokh, D. B.; Wade, R. C. Ligand Unbinding Mechanisms and Kinetics for T4 Lysozyme Mutants from TRAMD Simulations. *Curr Res Struct Biol* **2021**, *3*. https://doi.org/10.1016/j.crstbi.2021.04.001.

(31) Schuetz, D. A.; Bernetti, M.; Bertazzo, M.; Musil, D.; Eggenweiler, H. M.; Recanatini, M.; Masetti, M.; Ecker, G. F.; Cavalli, A. Predicting Residence Time and Drug Unbinding Pathway through Scaled Molecular Dynamics. *J Chem Inf Model* **2019**, *59* (1). https://doi.org/10.1021/acs.jcim.8b00614.

(32) 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). https://doi.org/10.1021/acs.jctc.0c00395.

(33) Wehmeyer, C.; Scherer, M. K.; Hempel, T.; Husic, B. E.; Olsson, S.; Noé, F. Introduction to Markov State Modeling with the PyEMMA Software [Article v1.0]. *Living J Comput Mol Sci* **2019**, *1* (1). https://doi.org/10.33011/livecoms.1.1.5965.

(34) Doerr, S.; Harvey, M. J.; Noé, F.; de Fabritiis, G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J Chem Theory Comput* **2016**, *12* (4), 1845–1852. https://doi.org/10.1021/acs.jctc.6b00049.

(35) Wang, W.; Cao, S.; Zhu, L.; Huang, X. Constructing Markov State Models to Elucidate the Functional Conformational Changes of Complex Biomolecules. *Wiley Interdiscip Rev Comput Mol Sci* **2018**, *8* (1). https://doi.org/10.1002/wcms.1343.

(36) Bringas, M.; Lombardi, L. E.; Luque, F. J.; Estrin, D. A.; Capece, L. Ligand Binding Rate Constants in Heme Proteins Using Markov State Models and Molecular Dynamics Simulations. *ChemPhysChem* **2019**, *20* (19), 2451–2460. https://doi.org/10.1002/cphc.201900589.

(37) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. Automatic Discovery of Metastable States for the Construction of Markov Models of Macromolecular Conformational Dynamics. *Journal of Chemical Physics* **2007**, *126* (15), 1–17. https://doi.org/10.1063/1.2714538.

(38) Betz, R. M.; Dror, R. O. How Effectively Can Adaptive Sampling Methods Capture Spontaneous Ligand Binding? *J Chem Theory Comput* **2019**, *15* (3), 2053–2063. https://doi.org/10.1021/acs.jctc.8b00913.

(39) Janssen, D. B. Evolving Haloalkane Dehalogenases. *Curr Opin Chem Biol* **2004**, *8* (2), 150–159. https://doi.org/10.1016/j.cbpa.2004.02.012.

(40) Koudelakova, T.; Bidmanova, S.; Dvorak, P.; Pavelka, A.; Chaloupkova, R.; Prokop, Z.; Damborsky, J. Haloalkane Dehalogenases: Biotechnological Applications. *Biotechnology Journal*. 2013. https://doi.org/10.1002/biot.201100486.

(41) 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). https://doi.org/10.1021/acs.jcim.7b00070.

(42) Damborský, J.; Kutý, M.; Němec, M.; Koča, J. A Molecular Modeling Study of the Catalytic Mechanism of Haloalkane Dehalogenase: 1. Quantum Chemical Study of the First Reaction Step. *J Chem Inf Comput Sci* **1997**, *37* (3). https://doi.org/10.1021/ci960483j.

(43) Damborský, J.; Koča, J. Analysis of the Reaction Mechanism and Substrate Specificity of Haloalkane Dehalogenases by Sequential and Structural Comparisons. *Protein Eng* **1999**, *12* (11). https://doi.org/10.1093/protein/12.11.989.

(44) Hynková, K.; Nagata, Y.; Takagi, M.; Damborský, J. Identification of the Catalytic Triad in the Haloalkane Dehalogenase from Sphingomonas Paucimobilis UT26. *FEBS Lett* **1999**, *446* (1). https://doi.org/10.1016/S0014-5793(99)00199-4.

(45) Streltsov, V. A.; Prokop, Z.; Damborský, J.; Nagata, Y.; Oakley, A.; Wilce, M. C. J. Haloalkane Dehalogenase LinB from Sphingomonas Paucimobilis UT26: X-Ray Crystallographic Studies of Dehalogenation of Brominated Substrates. *Biochemistry* **2003**, *42* (34), 10104–10112. https://doi.org/10.1021/bi027280a.

(46) 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.

(47) Gora, A.; Brezovsky, J.; Damborsky, J. Gates of Enzymes. *Chem Rev* **2013**, *113* (8), 5871–5923. https://doi.org/10.1021/CR300384W.

(48) 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), 23–30. https://doi.org/10.1371/journal.pcbi.1002708.

(49) Vavra, O.; Filipovic, J.; Plhak, J.; Bednar, D.; Marques, S. M.; Brezovsky, J.; Stourac, J.; Matyska, L.; Damborsky, J. CaverDock: A Molecular Docking-Based Tool to Analyse Ligand Transport through Protein Tunnels and Channels. *Bioinformatics* **2019**, *35* (23), 4986–4993. https://doi.org/10.1093/BIOINFORMATICS/BTZ386.

(50) Doerr, S.; Harvey, M. J.; Noé, F.; De Fabritiis, G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J Chem Theory Comput* **2016**, *12* (4), 1845–1852. https://doi.org/10.1021/ACS.JCTC.6B00049.

(51) Matthews, C.; Weare, J.; Kravtsov, A.; Jennings, E. Umbrella Sampling: A Powerful Method to Sample Tails of Distributions. *Mon Not R Astron Soc* **2018**, *480* (3), 4069–4079. https://doi.org/10.1093/MNRAS/STY2140.

(52) 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). https://doi.org/10.1093/bioinformatics/btab872.

(53) Magdziarz, T.; Mitusińska, K.; Gołdowska, S.; Płuciennik, A.; Stolarczyk, M.; Lugowska, M.; Góra, A. AQUA-DUCT: A Ligands Tracking Tool. *Bioinformatics* **2017**, *33* (13), 2045–2046. https://doi.org/10.1093/bioinformatics/btx125.

(54) Patel, A.; Malinovska, L.; Saha, S.; Wang, J.; Alberti, S.; Krishnan, Y.; Hyman, A. A. ATP as a Biological Hydrotrope. *Science (1979)* **2017**, *356* (6339), 753–756. https://doi.org/10.1126/science.aaf6846.

(55) Sindhu, A.; Kumar, S.; Mondal, D.; Bahadur, I.; Venkatesu, P. Protein Packaging in Ionic Liquid Mixtures: An Ecofriendly Approach towards the Improved Stability of β-Lactoglobulin in Cholinium-Based Mixed Ionic Liquids. *Physical Chemistry Chemical Physics* **2020**, *22* (26), 14811–14821. https://doi.org/10.1039/d0cp02151b.

(56) Karsisiotis, A. I.; Deacon, O. M.; Wilson, M. T.; MacDonald, C.; Blumenschein, T. M. A.; Moore, G. R.; Worrall, J. A. R. Increased Dynamics in the 40-57 Î©-Loop of the G41S Variant of Human Cytochrome c Promote Its pro-Apoptotic Conformation. *Sci Rep* **2016**, *6*. https://doi.org/10.1038/srep30447.

(57) Amacher, J. F.; Zhong, F.; Lisi, G. P.; Zhu, M. Q.; Alden, S. L.; Hoke, K. R.; Madden, D. R.; Pletneva, E. V. A Compact Structure of Cytochrome c Trapped in a Lysine-Ligated State: Loop Refolding and Functional Implications of a Conformational Switch. *J Am Chem Soc* **2015**, *137* (26). https://doi.org/10.1021/jacs.5b01493.

PUBLICATIONS

Publication 1

Incorporating prior knowledge to seeds of adaptive sampling molecular dynamics simulations of ligand transport in enzymes with buried active sites

Impact Factor 2022: NA

MNiSW points 2023: NA

Citations: 0

Main text

https://doi.org/10.1101/2023.09.21.558608

Supplementary information

https://www.biorxiv.org/content/10.1101/2023.09.21.558608v1.supplementary-material

Incorporating prior knowledge to seeds of adaptive sampling molecular dynamics simulations of ligand transport in enzymes with buried active sites

Dheeraj Kumar Sarkar,[1,2] Bartlomiej Surpeta,[1,2] Jan Brezovsky*[1,2]

[1] Laboratory of Biomolecular Interactions and Transport, Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Uniwersytetu Poznanskiego 6, 61-614 Poznan, Poland.

[2] International Institute of Molecular and Cell Biology in Warsaw, Ks Trojdena 4, 02-109 Warsaw, Poland.

*Corresponding author: JB: janbre@amu.edu.pl; jbrezovsky@iimcb.gov.pl

Abstract

Given that most proteins have buried active sites, protein tunnels or channels play a crucial role in mitigating the transport of small molecules to the buried cavity for enzymatic catalysis. Tunnels can critically modulate the biological process of protein-ligand recognition. Various molecular dynamics methods have been developed for exploring and exploiting the protein-ligand conformational space to extract high-resolution details of the binding processes, one of the most recent represented by energetically unbiased high-throughput adaptive sampling simulations. The current study systematically contrasts the role of integrating prior knowledge while generating useful initial protein-ligand configurations, called seeds, for these simulations. Using a non-trivial system of haloalkane dehalogenase mutant with multiple transport tunnels leading to a deeply buried active site, these simulations were employed to derive kinetic models describing the process of association and dissociation of the substrate molecule. The more knowledge-based seed generation enabled high-throughput simulations that could more consistently capture the entire transport process, effectively explore the complex network of transport tunnels, and predict equilibrium dissociation constants, k_{off}/k_{on}, on the same order of magnitude as experimental measurements. Overall, the infusion of more knowledge into the initial seeds of adaptive sampling simulations could render analyses of transport mechanisms in enzymes more consistent even for very complex biomolecular systems, thereby promoting the rational design of enzymes with buried active sites and drug development efforts.

Keywords: *Protein-ligand, tunnels, ligand transport, seeding, adaptive sampling, kinetics*

1. Introduction

Given the fact that molecular recognition is critical for all biological processes, in this context, the intrinsically dynamic and volatile nature of protein-ligand (un)binding processes makes it a long-standing quest to capture the high-resolution sampling and resolve meaningful kinetics of ligand binding processes in structure-based drug design[1,2]. Additionally, a ligand can prefer multiple routes of entry to interact with the environment of active site[3–5]. These routes, often referred to as tunnels, are seen to have equivalent importance as the catalytic properties of enzymes[6]. While in the majority of enzymes, the active site is buried[7,8], the underlying molecular properties of the tunnels can control the entry and exit of ligands to a greater extent, specifically by gating residues[9]. In this context, the ligand binding processes via those transport pathways are a critical component in biocatalysis, also for identifying critical residues underlying the transport processes for mutagenesis and rational drug design[6,10]. Hence, protein tunnels are well-placed when considering improved catalysis and features like specificity and altered activity of small molecules. Because very often, tunnel lining residues or other gating residues can act as hot spots other than the active site residues[9,10].

The transport processes, like a migration of ligands from the active site to the bulk solvent, are often connected with the requirement of overcoming a high energy barrier, resulting in the rare nature of such an event[11]. Because molecular dynamics (MD) simulations can observe biologically relevant processes even at atomistic resolution, they are extensively used to study mechanisms, dynamics, and functions of biomolecular complexes[12,13]. Numerous computational approaches have been developed in recent years to sample such rare events of ligand transport processes involving the association and dissociation of ligands and receptors[14]. These approaches benefit from the improvement of computational hardware in terms of GPUs as well as the implementation of various path sampling methods and methods for sampling rare events[15,16]. Specifically enhanced sampling methods like milestoning[17], weighted ensemble[18], Gaussian accelerated MD[19,20], metadynamics[21,22], adaptive sampling MD (ASMD) based on Markov state models (MSM)[5,23,24], Random Acceleration MDs[5,25,26], gained popularity in studying such rare events. While most methods use additional potential or force to bias the simulations along a designed collective variable, the ASMD methods utilizing MSMs can avoid such perturbations[27–29]. Extensive ASMD simulations have been successfully used to study ligand binding processes[5,24,30–33]. The ASMD is an energetically unbiased protocol comprising iterative rounds of intelligently respawned equilibrium simulations of protein-ligand configurations. This is achieved by using a scoring function to select the least explored configurations from a preliminary MSM build on the so far generated simulations and employing those configurations to initiate subsequent batches of simulations (called epochs)[28,29].

Given the rising success of ASMD simulations in ligand transport studies, the impact of designing individual components in ASMD workflow on the efficacy of sampling relevant regions of protein-ligand configurational space is of interest[24,31,33–35]. Betz and Dror investigated the role of a scoring function for selecting the configuration for the successive iterations to partially overcome the exploration-exploitation tradeoff using the well-known test system of trypsin with benzamidine inhibitor and a more complex yet realistic system of membrane-bound adrenergic receptor β_2 with dihydroalprenolol inhibitor[33]. They compared three scoring functions based on simple counts, in which states are resampled with probability inversely proportional to their occurrence in simulation; the population scores, which prefer states with smaller populations in MSMs; and hub scores, which select states with lower connectivity in MSMs, the measure of connectivity of states in MSMs. On the membrane-bound system, the count score could not govern the ASMD toward investigating inhibitor migration through the protein, focusing entirely on the membrane region. In contrast, the other two scores successfully sampled the relevant configurations. Hence, the use of more information-rich functions markedly benefited the study of ligand transport in more complex settings.

Here, we investigated the role of employing relevant information as early as in preparing initial seeding structures for ASMD. We designed four schemes (Figure 1A) from random positioning of the ligand around the protein to more knowledge-based poses of the ligand bound in the active site or along the tunnels precomputed from apo simulation. We tested the capabilities of ASMD, initiated from these seeding schemes, in exploring and exploiting the transport tunnels in haloalkane dehalogenase mutant LinB86 (Figure 1B), in which an additional functional tunnel was introduced *de novo*[10]. By performing intensive ASMD simulations of LinB86 with one of its substrates, 1,2-dibromoethane (DBE), for each scheme, we were able to compare to what degree the initial seeding impacts ability of ASMD to i) capture entire process of substrate association and dissociation, ii) identify metastable states adopted by substrate consistently, iii) predict kinetic parameters of the process, and finally iv) describe complexity of transport via multiple transport tunnels.

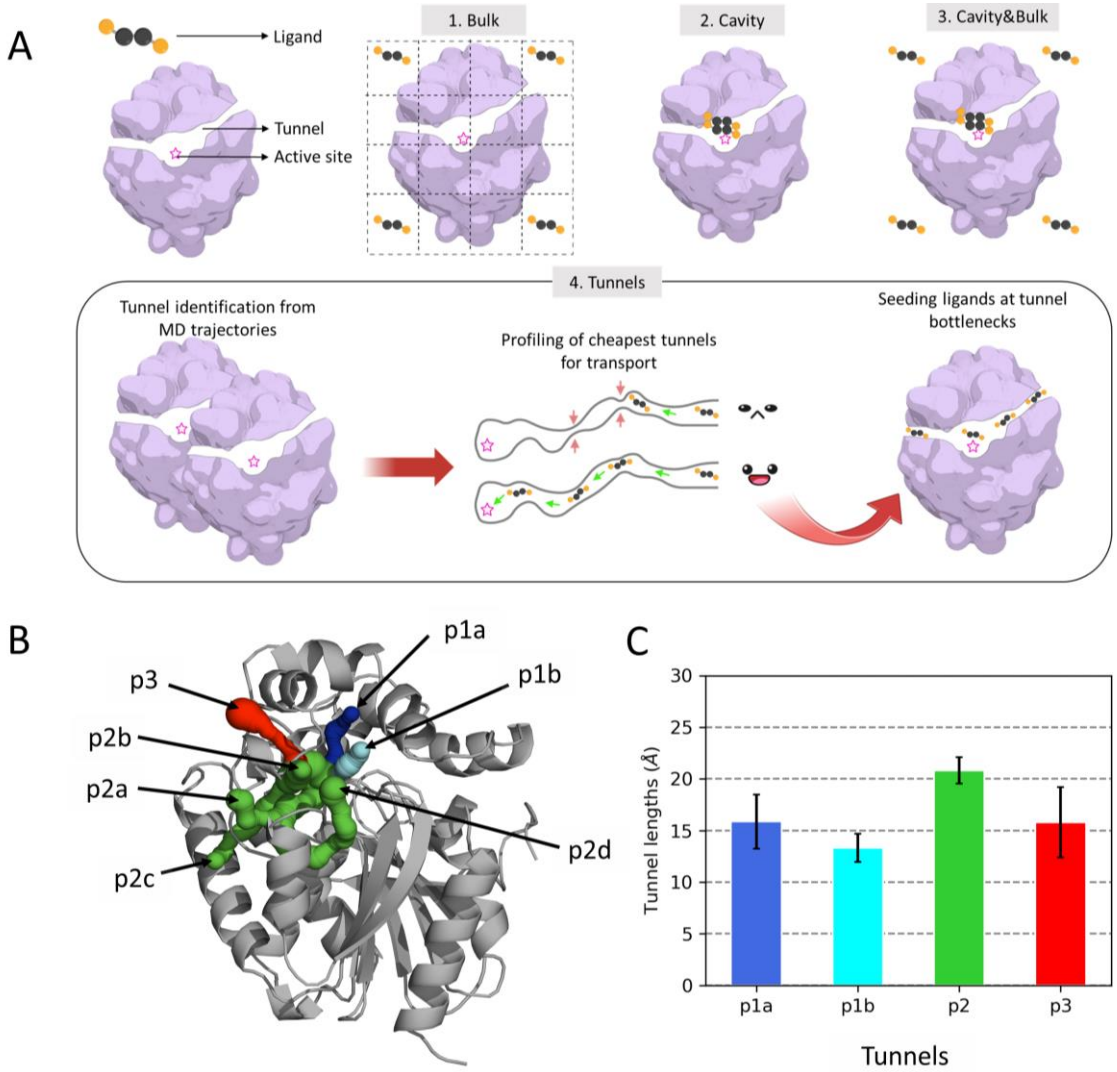


Figure 1. Overview of evaluated seeding schemes and model system used. A) Schematic representation of studied schemes and seeding of the substrate molecule from random to knowledge-based positions. B) Representative structure of tunnel network from 100 ns MD simulation of LinB86 (see Table S1 for other tunnel properties). The known tunnels are shown as sets of colored spheres: *p1a* (blue), *p1b* (cyan), four branches of *p2* (green), and *p3* (red). The protein structure is shown as a gray cartoon. C) Average lengths of ensembles of the known tunnels observed in MD simulation.

2. Materials and Methods

2.1. Seed generation for ASMD simulations

The input model was based on the crystallographic structure of the mutant of haloalkane dehalogenase enzyme LinB86 (PDB code: 5LKA). The protein structure was further protonated using H++ web server[36,37] pH 8.5. The protein was modeled using the AMBER ff14SB[36] force field and the substrate DBE with the General Amber Force Field - GAFF[38,39]. The partial atomic charges on the DBE molecule were derived using multi-conformational, multi-orientational restrained electrostatic potential fit[40]. Each DBE conformation was geometry optimized at MP2/6-31G(d) level of theory, and

their multi-orientational molecular electrostatic potential was calculated at HF/6-31G(d) level using Gaussian v09[41]. Finally, two-stage charge fitting was conducted for all conformations and orientations using *resp* and *antechamber* modules of AMBERTools18[42].

The substrate molecule was placed according to the four designed schemes to investigate the role of knowledge in ASMD seeding systematically (Figure 1A). DBE was placed at 30 different positions in each scheme (Figure S1) as follows. i) In the *Bulk* scheme, DBE was positioned on an equally spaced grid in the bulk solvent surrounding the protein using *"drawgridbox [selection], nx=5, ny=5, nz=5, padding=5, lw=1, r=0, g=0"* function of PyMOL[43]. ii) In the *Cavity* scheme, DBE was docked to the enzyme's active site using AutoDock Vina[44]. The 30 docked poses were derived by defining the grid box centered at COM of catalytic residues (N38, D108, W109, and H272) with a dimension of 22.5 Å and exhaustiveness of 1000. iii) In *the Cavity&Bulk* scheme, 15 DBE positions were taken from the *Cavity* scheme and 15 DBE positions from the *Bulk* scheme. Finally, iv) in the *Tunnels* scheme, putative transport tunnels in LinB86 were detected from 100 ns trajectory of ligand-free LinB86 simulation, and the most open tunnels were then explored for binding of DBE molecules along these tunnels. Finally, the composite tunnels, formed from parts of tunnels with conformations ensuring minimal energy costs for DBE migration, were generated (see Text S1 and Figures S2-S9 for details of this protocol).

The generated protein-ligand complexes were then solvated based on the 3D Reference Interaction Site Model theory[45] using the Placevent[46] algorithm. Such a system was then processed with the *tleap* module of AMBERTools18, placing the pre-solvated proteins in the octahedral box of OPC water molecules[47] to the distance of 10 Å and neutralizing them with counter ions (Na^+ and Cl^-) to the ionic strength of 0.1 M. Finally, the hydrogen mass repartitioning method was applied to produce topologies to enable 4 fs timestep[48].

2.1. Equilibration MD simulations of seeds

The systems were then minimized and equilibrated using PMEMD and PMEMD.CUDA modules[49] of AMBER18[42], respectively. All complexes were energy minimized in five consecutive stages, each composed of 100 steps of the steepest descent followed by 400 steps of the conjugate gradient method, with gradually decreasing restraints on the protein atoms (initially 500 to heavy atom, and later restraints of 500, 125, 25, and 0.001 $kcal.mol^{-1}.Å^{-2}$ applied only to the backbone atoms). Minimization was followed by 20 ps heating from 0 to 200 K in the NVT ensemble using the Langevin thermostat with a collision frequency of 2 ps^{-1} and coupling constant of 1 ps while keeping the protein heavy atoms restrained with a force constant of 5 $kcal{\cdot}mol^{-1}{\cdot}Å^{-2}$. Next, the temperature was raised to the target value of 310 K in 100 ps of NVT simulation and kept constant for 900 ps, employing the same parameters as previously described. This was followed by NPT simulation at 1 atm enforced by

the weak-coupling barostat with a coupling constant of 1 ps using positional restraints of 5 kcal·mol^{-1}·Å^{-2} on the backbone atoms for 1 ns, followed by 1 ns without any positional restraints. All MD simulation stages were run using a 4 fs timestep enabled by SHAKE[50] and hydrogen mass repartitioning algorithms, periodic boundary conditions, and particle mesh Ewald method[51]. The trajectories were generated by saving coordinates every 20 ps. The MD trajectories were analyzed using the *cpptraj* module of AMBERTools23[52,53]. The last snapshots from the unrestrained simulation were used as the initial input structures for ASMD.

2.2. High throughput ASMD to study substrate un(binding) processes

The ASMD was set up with 30 epochs, each consisting of 30 separate production simulations. To build an MSM model after each epoch, we used the distances between the Cα atoms of the protein and four heavy atoms of DBE and reduced the high dimensional space to three dimensions using time-dependent component analysis (TICA)[54] with a lag time of 2 ns. The ASMD simulations were performed using HTMD v1.13.10[27] and AMBER18[42] software packages. The equilibration phase in HTMD consisted of two 250 ps NVT and NPT simulations, during which the systems were heated from 0 to 310 K with a Langevin thermostat and harmonic positional restraints to the backbone atoms with a force constant of 5 kcal·mol^{-1}·Å^{-2}. Finally, a 50 ns unrestrained production MD was performed in the NVT ensemble using a weak-coupling thermostat and a saving frequency of 100 ps. Such ASMD runs were performed in three replicates for each investigated seeding scheme.

2.3. Final MSM construction and validation

All the MSM were built using the HTMD[27], which internally uses PyEMMA program[55], following the standard PyEMMA protocol. The high dimensional data from adaptive sampling projected with distance feature was reduced to three dimensions using TICA[54] with a lag time of 2 ns. Next, the reduced TICA coordinates were clustered into 1000 microstates using the MiniBatchKMeans[56] method. The metastable states were lumped using PCCA++ method[57], with the number of metastable states based on spectral analysis[55] and verified against plots of linear implied timescales (Figures S10-S12). The lag time of 20 ns was used during MSM construction. Finally, the Chapman-Kolmogorov[58] test was performed to confirm the Markovianity of the generated MSM models (Figures S13-S15).

2.4. MSM analysis and comparison

In order to quantify the ability of ASMD to sample the whole (un)binding process of DBE, the distances between the center of mass (COM) of the DBE molecule and COM of three catalytic residues (N38, D108, and W109, Figure S16A) were measured from the ~900 trajectories for each replicate using

cpptraj module of AMBERTools23[52,53]. Based on this distance, we can define the location of DBE in the active site (0-5 Å), tunnel (5-19 Å), and bulk (>19 Å). The cutoff of 19 Å for tunnels was derived from the average lengths of investigated tunnels measured by CAVER (Figure 1C and Table S1). Finally, the transition path theory approach implemented in PyEMMA was used to derive transition probability matrices and compute the mean first pass times of each association and dissociation process in MSMs. Here, the metastable states with the most prevalent bound conformation of DBE were used as sink states, while the metastable states featuring DBE mainly in the bulk solvent were considered as source states to perform the transition flux analysis and derive the transition probabilities and kinetics rates. Furthermore, the most frequently occurring bottleneck residues were shortlisted from the CAVER results as follows (Figure S16B-E): *p1a* (D147, F151, and V173), *p1b* (D147, W177, and L248), *p2* (L211 and L248), and *p3* (L143, F151, and I213) and the distance between their COM to COM of DBE was calculated to assess localization of DBE with respect to these tunnels.

Ensembles of 1000 representative structures of metastable states generated from individual MSMs were clustered to establish the correspondence of these metastable states across explored schemes. For this purpose, mean, 25th, 50th, and 75th percentiles were calculated for each set of characteristic distances to bottleneck residues and catalytic machinery described above (Figure S16). They were used cumulatively as a vector of 20 variables describing each metastable state. Principal component analysis (PCA) implemented in the Python scikit-learn library[56] was used to reduce the dimensionality of each vector. The set of the first three principal components for each metastable state was clustered with HDBSCAN[59] using *min_cluster_size* of 2, with the remaining parameters kept as default.

2.5. Analysis of substrate utilization of tunnels

Time-evolution of distances (Figure S16) for the entire set of trajectories was used to estimate the approximate position of the ligand in the context of the tunnel network. By tracking the change of the relative position, the movement through a particular tunnel was assigned where possible. Therefore, the approximate tunnels' utilization was estimated across investigated schemes by analyzing the transition between subsequent positions. The procedure was composed of three stages as follows.

i) Position assignment. First, the closest bottleneck at a particular frame to the DBE molecule was defined. Further, this information was used to define the approximate length of the closest tunnel, i.e., the distance between the COM of catalytic machinery and the COM of the particular bottleneck. These two distances were contrasted with the distance of the ligand to the catalytic machinery, which altogether resulted in the identification of the approximate ligand position. Importantly, at this point, additional parameters were introduced to classify the ligand position, namely *bt_cutoff_along=2.0 Å* defining the region around the bottleneck, distinguishing whether the ligand is in the bulk, bottleneck

region or tunnel, and *bt_cutoff_across=5.0 Å* that defines whether the ligand is not too far from the bottleneck horizontally in case it is within the bottleneck region. Considering these three distances and introduced cutoffs, the following scenarios and corresponding ligand states were considered:

- *Bulk (out_)*: Ligand is further from the active site than the sum of tunnel length and *bt_cutoff_along*.
- *Bottleneck (bt_)*: Ligand is within the bottleneck region, either further than the tunnel length or closer than the tunnel length but within *bt_cutoff_along* and *bt_cutoff_across.*
- *Unknown bottleneck (bt_unknown)*: Ligand is within the bottleneck region, either further than the tunnel length or closer than the tunnel length within *bt_cutoff_along* but exceeding the *bt_cutoff_across.*
- *Inside (in_)*: Ligand distance to catalytic machinery is shorter than the tunnel length decreased by the *bt_cutoff_along.*

ii) Transition detection and classification. Considering defined states for each frame, the transitions between bulk (out) and interior (in) and vice versa were identified. Transitions via bottleneck regions (in-bt-out or out-bt-in) were also considered. In the case that the mismatch between assigned tunnel in-out/out-in was detected, we applied an additional *dist_tolerance=1.0 Å* parameter, which defined the tolerance distance that is considered for swapping the classification of one of the sides of the transition, promoting the tunnel that was seen in the bottleneck region for scenarios where the intermediate state was seen. The transitions were tracked as follows:

- If the transition occurred from the bulk to inside or from the inside to bulk directly - the transition in-out/out-in was assigned by applying the *dist_tolerance* for cases where the mismatch between both sides occurred.
- If the ligand moved from the interior to the bottleneck region or from the bulk to the bottleneck region – the transition was not assigned, only the information regarding the temporary state.
- If the temporary state was a bottleneck and the closest tunnel changed, the transition was not assigned; only the bottleneck temporary state was updated.
- If the temporary state was a bottleneck and the ligand moved to the same general state but related to a different tunnel, the transition was not assigned; only the general state was updated.
- If the temporary state was a bottleneck and the ligand moved to the other general state (from in to out or from out to in), the transition was assigned, also collecting the information about the bottleneck used applying the *dist_tolerance* for cases where the mismatch between both sides occurred promoting the tunnel of the assigned transition bottleneck state.

iii) Characterization of tunnel utilization. Finally, all types of unique transitions were counted across all simulations from each scheme and averaged across three replicates performed for each scheme. Importantly, we applied the following classification to assign transitions to particular categories:

- Tunnel (*p1a, p1b, p2,* and *p3*) – all transitions that passed through the bottleneck of a particular tunnel or the direct transitions in-out or out-in related to the same tunnel on both sides;
- Mixed – all direct transitions in-out or out-in, where both sides of transitions differ even considering applied distance tolerance;
- Unknown – all transitions that crossed through the unknown bottleneck.

3. Results and Discussions

Overall, ~900 MD trajectories (450,000 frames) with aggregated simulation time of 45 µs were produced using ASMD of LinB86-DBE complexes generated according to all four studied schemes (Table S2). For each scheme, ASMD was performed in three replicates to evaluate the abilities to consistently describe the transport processes in its entirety, focusing on the convergence among the ASMD replicates, the degree of quantitative agreement with experimental data, and the ability to consider the transport via all known tunnels.

3.1 Capturing DBE association and dissociation processes in LinB86

In order to study the applicability of the studied schemes, we initially investigated how effectively each scheme could sample the endpoints of the processes, i.e., the bound and unbound states of DBE in the active site cavity of LinB86 and bulk solvent, respectively. Those states could be effectively defined by the distance of DBE COM from the COM of three catalytic residues located at the bottom of the cavity (Figure S16A), defining the bound states within 5 Å distance, while the unbound state samples primarily distances above 19 Å, which are further than the length of the longest tunnels present in LinB86 (Figure 1C). The DBE explored the unbound state in all schemes and replicas for a substantial fraction of cumulative ASMD trajectories (Figure 2). Even in the *Cavity* scheme initiated from the DBE molecule bound deep in the active site, the substrate reached the bulk solvent, generating a minimum of 12 % unbound states. Over 10,000 unbound states were generated after at most seven epochs of ASMD simulations (Figure 3). Foreseeably, the unbound states were most prevalent (> 31 %) in the simulation seeded with DBE placed in the bulk solvent around the enzyme (scheme *Bulk*).

Concerning the ability of ASMDs to reach the bound pose of DBE in the buried active site of LinB86, all schemes except for *Bulk* could consistently sample the bound states in all three replicates. In the case of the *Bulk* scheme, the DBE molecule was able to find a path to the active site in replicate 1, producing a total of 4 % of simulations in the bound state (Figure 2), with a significant ensemble of more than 1,000 bound configurations sampled already until the fifth epoch (Figure 3). However, no bound state was observed in the other two replicated ASMD from the *Bulk* scheme (Figures 2 and 3). This is no surprise since unbiased simulations of ligand associations are generally rather time-consuming, even for less complex systems[60–62].

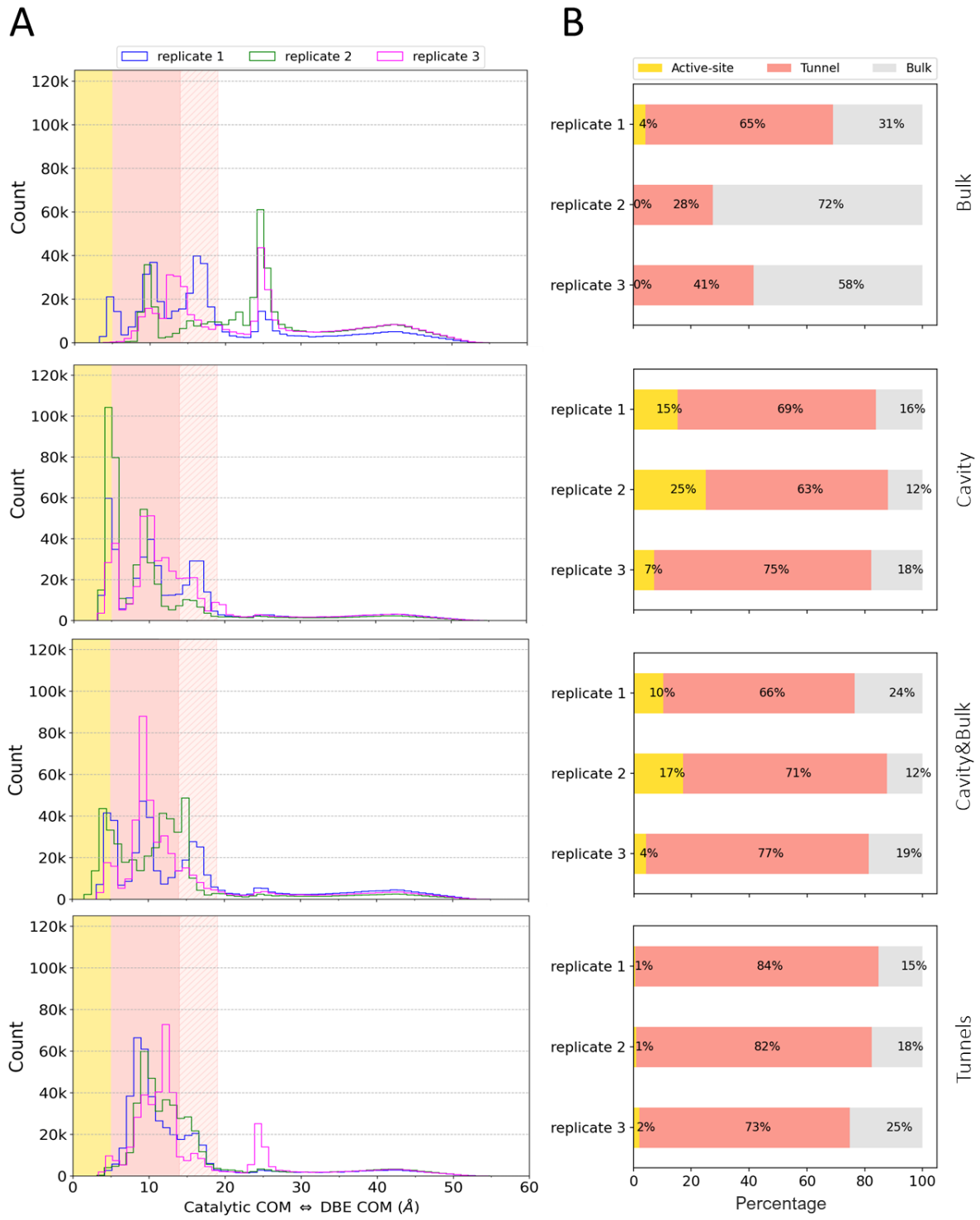


Figure 2. Substrate (un)binding to the active site of LinB86 captured by ASMD simulations with four seeding schemes. A) The distance distribution of DBE to catalytic residues for three replicated ASMD for each seeding scheme. The regions corresponding to DBE in the active site (0-5 Å), shortest (p1b, 5-14 Å) and longest (p2, 14-19 Å) tunnel lengths (Figure 1C), and bulk solvent (>19 Å) are highlighted as gold, pink, shaded pink, and white, respectively. The distances are between COM of DBE and COM of catalytic residues (N38, D108, and W109), measured in 45 μs simulations. B) The fraction of DBE seen in individual regions.

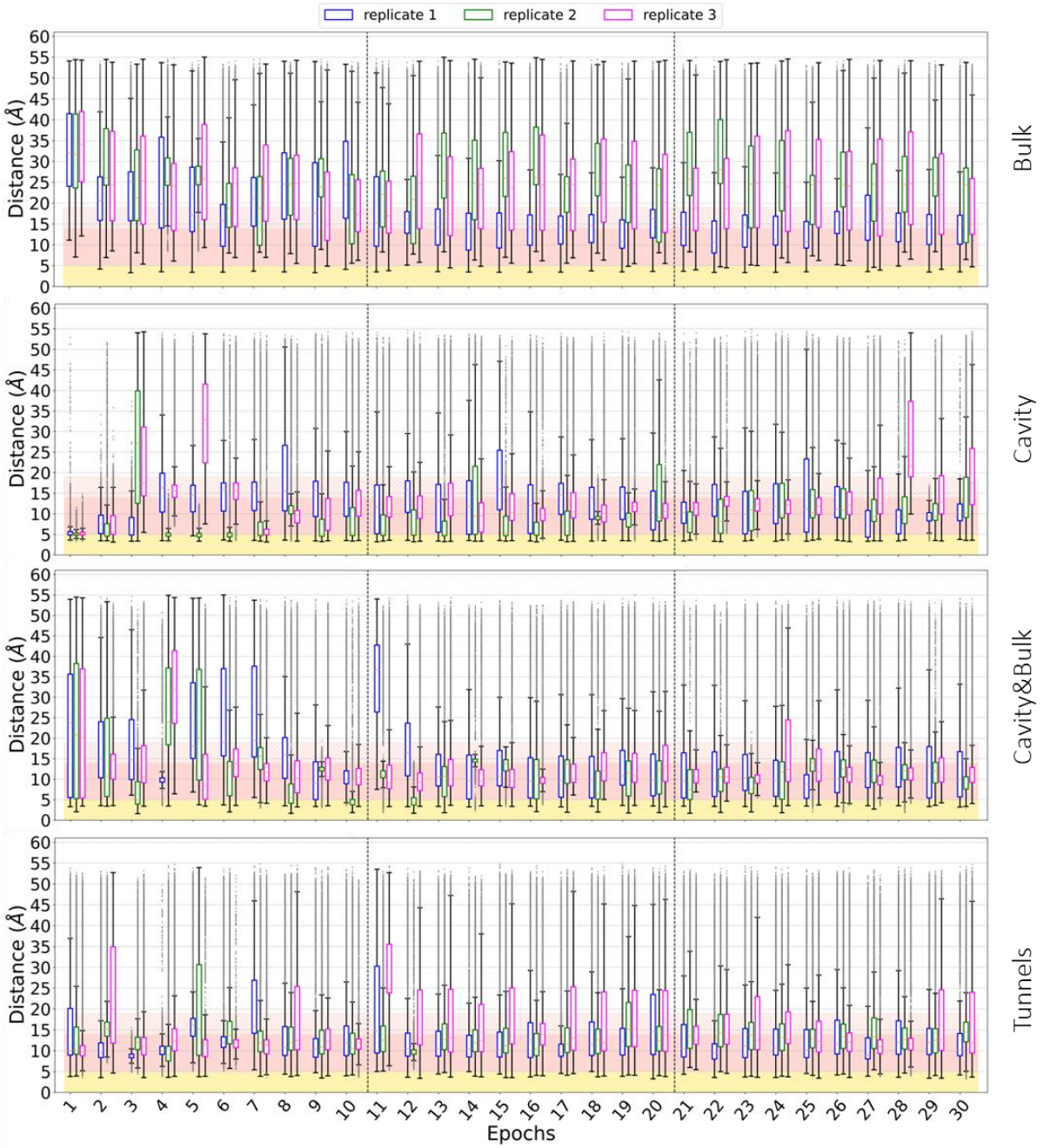


Figure 3. An epoch-wise sampling of the distance of DBE to catalytic residues of LinB86. The regions corresponding to DBE in the active site (0-5 Å), shortest (p1b, 5-14 Å) and longest (p2, 14-19 Å) tunnel lengths (Figure 1C), and bulk solvent (>19 Å) are highlighted as gold, pink, shaded pink, and white, respectively.

Among the remaining schemes, *Cavity* AMSDs exploited the bound states the most frequently, as expected from the initial seeding with docked poses of DBE (Figure 2). Such setup led to the accumulation of over 5,000 bound states already during the first epoch in all three replicates (Figure 3). Such behavior was also partially retained in the *Cavity&Bulk* scheme, where more than 2,000 bound states were systematically observed in the first epoch of ASMDs. Here, the additional seeds of DBE placed in the bulk solvent resulted in considerable sampling of more than 10,000 unbound states within the first three epochs of ASMDs, about twice faster than in the pure *Cavity* scheme (Figure 3). Finally, we have observed the DBE spending most of the time exploring the regions

corresponding to transport tunnels in ASMDs from the *Tunnels* scheme (Figure 2). Having sufficient coverage of bound and unbound states, we progressed to the creation of MSMs from the assembled trajectories and the calculation of kinetic parameters of descriptions of (un)binding processes. Due to the lack of bound states in the *Bulk* scheme, these AMSDs were not considered for constructing MSMs.

3.2. Identifying metastable states of DBE interacting with LinB86 and predicting kinetic parameters from MSMs

To further test the capabilities of the studied seeding schemes in the diversity and consistency of identified metastable states, we have generated MSMs from the individual ASMD replicates. These MSMs consisted of three to six metastable states for the *Cavity* (Figures S17-S19) and *Cavity&Bulk* (Figures S20-S22) schemes, whereas six to eight metastable states were identified in MSMs from the *Tunnel* schemes (Figures S23-S25). To understand the mutual correspondence among these states across all generated MSMs, we have generated 1,000 representative structures of each metastable state and measured the distances of DBE to the catalytic residues, as well as to the bottlenecks of the known transport tunnels in LinB86 (Figure S16). These distances represent fingerprints characterizing the metastable states (Figures S26-S28), clearly identifying not only unbound and bound states but also their alignment to individual transport tunnels.

Finally, those unified fingerprints enabled us to cluster the metastable states (Figure S29), forming the unified non-redundant ligand states (*ULS*) across all MSMs (Figure 4A). The only state consistently present in all replicates of each seeding scheme (Figure 4B) was *ULS1*, which corresponded to the DBE molecules in the bulk solvent. *ULS2-ULS5* all represented DBE molecules inside the catalytic cavity, with DBE bound closest to the catalytic residues in *ULS2*, which was found only in the MSMs of the *Cavity* scheme. In *ULS3* the substrate was placed closer to the cavity center, while in *ULS4* and *ULS5,* the substrate was located near the exit from the cavity in the direction of *p1* or *p3* tunnels. *ULS6-ULS9* featured the DBE molecule bound on the LinB86 surface at the entrances to *p3, p2a, p2c*, and *p2d* tunnels, with the *p3* tunnel entrance (*ULS6*) being the most prevalent across the MSMs (Figure 4B). Curiously, in replicate 2 from the *Cavity&Bulk* scheme, we have observed several metastable states forming *ULS10*, which were composed of the DBE molecules exploring the cryptic pocket located back-to-back with the canonical active site cavity of LinB86 with the entrance located on the opposite side of the enzyme structure with respect to the *p1* tunnel entrance.

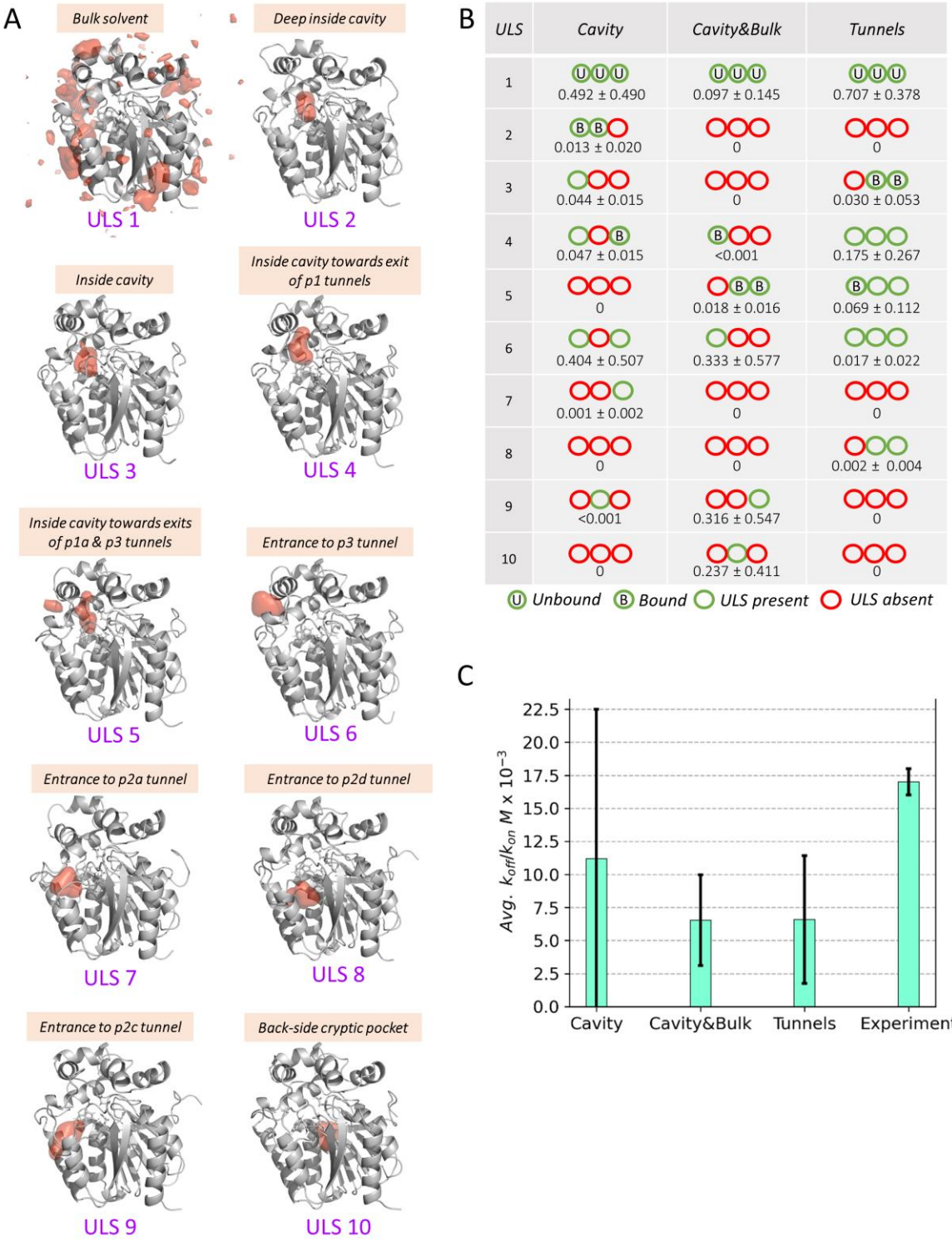


Figure 4. Inference into the (un)binding process of DBE to LinB86 from MSM analysis. A) Structurally unified ligand states (ULS) identified among all metastable states (Figures S17-S25) resolved by MSM analysis of three replicated ASMD simulations initiated from the studied seeding schemes. Protein structure is shown as a gray cartoon while the region occupied by DBE molecule in 20 % (1 % for bulk solvent state) of 1,000 structures representing given ULS is shown as red surface. B) The presence of ULS among metastable states in each ASMD replicates with their average probabilities. The unbound and bound metastable states used as source and sink states during the mean first pass time analyses are highlighted. C) Average equilibrium dissociation constants derived from MSMs as a ratio of dissociation and association rates (Figure S30). The experimental k_d was obtained from [66]. The data represents mean±stdev from the three replicates.

Some identified *ULSs* were also observed in the recent study of transient binding sites on the LinB wild-type conducted with seven halogenated compounds, including DBE molecules[63]. Out of nine sites, three could be matched to *ULSs* as follows: i) *site 5* corresponded to *ULS6,* the entrance to the *p3* tunnel, ii) *site 9* overlayed with *ULS9*, the entrance to *p2c* tunnel*,* and iii) *site 4* aligned to *ULS8*, the entrance to *p2d* tunnel. Such agreement suggests conservation of those interaction sites between LinB wild-type and LinB86 mutant despite the substitutions introduced into the *p1* and *p3* tunnels of the mutant. Considering the identification of *ULS* in replicated MSMs, the *Tunnels* scheme exhibited the best consistency since four *ULS* were found in all three replicates, while the other two *ULS* were found in two replicates. In contrast, only unbound *ULS1* was systematically found in the *Cavity* and *Cavity&Bulk* schemes. In fact, those two schemes frequently led to the formation of singleton *ULSs*, present in one MSM replicate only.

Finally, we have calculated the equilibrium dissociation constant (k_d) from the rates of DBE association and dissociation predicted from MSMs by the mean first pass time analyses. In contrast with the other schemes, we noted markedly faster DBE dissociation in the *Cavity* scheme (Figure S30), in line with the overrepresentation of bound states in the ASMDs. Interestingly, we found that the computed k_d values from all schemes were in good agreement with the experimentally determined one (Figure 4C). However, the computed values from the *Cavity* scheme were not well converged. The obtained less than an order of magnitude differences between simulations and experiments are still not common even in the case of much less complex biomolecular systems[64,65].

3.3. Exploration of different transport paths by substrate

Next, we investigated the utilization of individual transport pathways of LinB86 by the substrate DBE. Initially, we attempted to match the substrate migration traces to the tunnel ensembles using Transport Tools library[67]. However, we could observe only very few complete migration events of DBE molecule between the bulk solvent and the active site of LinB86 (Table S3), with replicate 2 of the *Cavity* scheme capturing 33 transport events of DBE via known tunnels. Since such data cannot provide sufficient inference, we have followed by considering a simplified transition of DBE molecule through the tunnel bottleneck only, which corresponds to the least favorable region along the migration path and hence controls the transport rates[9,60,66]. Considering the distances of DBE to the COM of bottleneck residues of each tunnel and the bottom of the active site cavity (Figure S16), we have traced the location of DBE in all simulations, focusing on the frames where DBE came close to any of the bottlenecks and whether if passed through them.

A thorough investigation of the transport in all schemes via particular tunnels revealed the following observations. We observed the highest total number of transitions for the *Tunnels* scheme, followed by *Cavity*, *Cavity&Bulk*, and finally, the lowest in the *Bulk* only (Table S4). The overall proportion of the particular tunnels being utilized to the total number of transitions was consistent across all schemes. The most frequently used tunnel was *p2*, followed by *p1b*, *p1a*, *p3* and finally mixed and unknown representing the smallest fraction of the data (Figure 5A). Interestingly, besides this trend consistency, schemes *Cavity&Bulk* and *Tunnels* displayed a higher percentage of *p2* tunnel as compared to the remaining schemes (Figure 5A), suggesting that the more complex seeding schemes enable relatively more efficient exploration of the longest and most complex branches of *p2* tunnel. In contrast, simplified schemes (*Bulk* and *Cavity*) tend to promote the sampling of more accessible primary conduits, *p1* tunnels, in agreement with their preferential utilization observed in TransportTools analyses of replicate 2 of the *Cavity* scheme (Table S3).

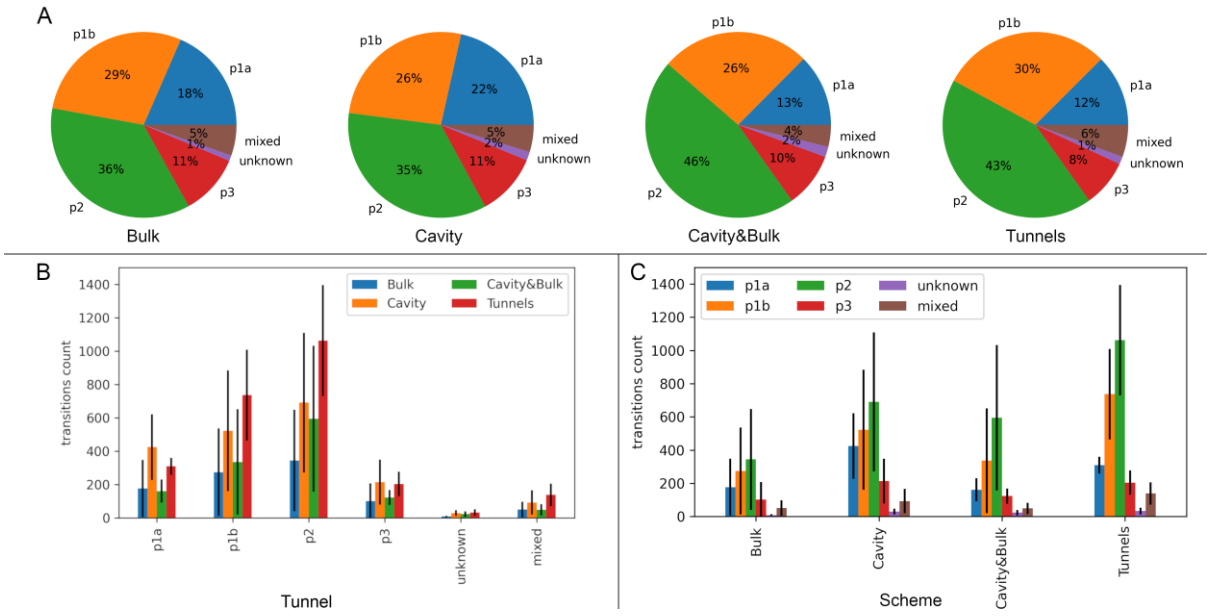


Figure 5. A statistical representation of tunnel utilization by the substrate for investigated seeding schemes. A) Relative utilization of particular tunnels for each scheme. B) Per tunnel average tunnel utilization. C) Per scheme average tunnel utilization. The data represents mean±stdev from the three replicates.

Besides this difference, the increased amount of transitions for particular schemes mainly came from the proportionally boosted sampling in each tunnel (Figure 5B). Importantly, considering the standard deviations calculated for combined statistics from three replicates for each scheme, it is clear that the *Tunnels* scheme presents the highest consistency from all tested schemes for all tunnels. This can be noticed when the transitions are considered for each run separately (Table S4). While the total sum of transitions for the *Bulk* scheme differs noticeably for particular replicates (1820, 49 and 1006 transitions for replica1, replica2 and replica3, respectively), the *Tunnels* scheme presents the lowest deviation when the three replicates are considered separately (replica1 – 2809, replica2 – 1680,

replica3 – 2968). *Cavity* and *Cavity&Bulk* schemes fall between these two extremes and present similar consistency to each other (Figure 5C).

4. Conclusions

This study aimed to test the effect of different seeding schemes on effective sampling in MSM-driven ASMD simulations, providing meaningful insights into kinetic rates and mechanisms of the transport of the substrate DBE in LinB86 from its deeply buried active site to the solvent environment via multiple transport tunnels. The four designed seeding schemes allowed the positioning of DBE by focusing on applying more knowledge to tackle the sampling of regions with higher energy barriers. The ensuing ASMD simulations could construct the kinetic models with different levels of detail based on the employed seeding scheme. All simulations could explore the entire transport process, visiting unbound and bound states except for the *Bulk* scheme that could not reach the bound state in two replicates of 45 μs ASMDs. Conversely, the *Tunnels* scheme was most consistent in sampling different metastable states of substrate in the transport-relevant regions. The application of more information-rich *Tunnels* and *Cavity&Bulk* schemes led to the enhanced exploration of auxiliary *p2* and *p3* tunnels. In contrast, primary *p1* tunnels were preferred in ASMDs initiated from the other two schemes. *Tunnels* and *Cavity&Bulk* schemes also provided the most converged k_d values from the rates of DBE association and dissociation, sufficiently close to the experimental measurements despite the complexity of the kinetic model. We expect that analogous methodology can also be beneficial for defining effective collective variables enhanced sampling methods like metadynamics,[21,22] and umbrella sampling[68]. Overall, the infusion of more knowledge into the initial seeds of ASMD simulations could render computational analyses of transport mechanisms in enzymes more consistent even for very complex biomolecular systems, having a clear potential to translate into faster rational protein design and drug development efforts.

Author contributions

Conceptualization: J.B.; Data curation: D.K.S, Formal analysis: D.K.S., B.S. J.B.; Funding acquisition: J.B., Investigation: D.K.S. (Simulations and MSMs), B.S. (Transitions and ULS clustering); Methodology: D.K.S. (Simulations and MSMs), B.S. (Transitions and ULS clustering); Project administration: J.B.; Resources: J.B.; Software: J.B.; Supervision: J.B.; Validation: J.B.; Writing – original draft: D.K.S. (Simulations and MSMs), B.S. (Transitions and ULS clustering); Writing – review & editing: J.B.

Acknowledgments

This work was supported by the National Science Centre, Poland (grant no. 2017/26/E/NZ1/00548). D.K.S. and B.S. were scholarship recipients provided by POWER projects POWR.03.02.00-00-I006/17

and POWR.03.02.00-00-I022/16, respectively. The simulations were performed in the Poznan supercomputing center.

References

(1) Ansari, N.; Rizzi, V.; Parrinello, M. Water Regulates the Residence Time of Benzamidine in Trypsin. *Nat Commun* **2022**, *13* (1). https://doi.org/10.1038/s41467-022-33104-3.

(2) Gallicchio, E.; Levy, R. M. Recent Theoretical and Computational Advances for Modeling Protein-Ligand Binding Affinities. *Adv Protein Chem Struct Biol* **2011**, *85*. https://doi.org/10.1016/B978-0-12-386485-7.00002-8.

(3) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and Mechanism of Drug Binding to G-Protein-Coupled Receptors. *Proc Natl Acad Sci U S A* **2011**, *108* (32). https://doi.org/10.1073/pnas.1104614108.

(4) Silva, D. A.; Bowman, G. R.; Sosa-Peinado, A.; Huang, X. A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the Lao Protein. *PLoS Comput Biol* **2011**, *7* (5). https://doi.org/10.1371/journal.pcbi.1002054.

(5) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations. *Proc Natl Acad Sci U S A* **2011**, *108* (25), 10184–10189. https://doi.org/10.1073/pnas.1103547108.

(6) Kokkonen, P.; Bednar, D.; Pinto, G.; Prokop, Z.; Damborsky, J. Engineering Enzyme Access Tunnels. *Biotechnol Adv* **2019**, *37* (6), 1–13. https://doi.org/10.1016/j.biotechadv.2019.04.008.

(7) 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). https://doi.org/10.1186/s12859-014-0379-x.

(8) Marques, S. M.; Daniel, L.; Buryska, T.; Prokop, Z.; Brezovsky, J.; Damborsky, J. Enzyme Tunnels and Gates As Relevant Targets in Drug Design. *Medicinal Research Reviews*. 2017. https://doi.org/10.1002/med.21430.

(9) Gora, A.; Brezovsky, J.; Damborsky, J. Gates of Enzymes. *Chem Rev* **2013**, *113* (8), 5871–5923. https://doi.org/10.1021/CR300384W.

(10) 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.

(11) Decherchi, S.; Cavalli, A. Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation. *Chemical Reviews*. 2020. https://doi.org/10.1021/acs.chemrev.0c00534.

(12) Zwier, M. C.; Chong, L. T. Reaching Biological Timescales with All-Atom Molecular Dynamics Simulations. *Current Opinion in Pharmacology*. 2010. https://doi.org/10.1016/j.coph.2010.09.008.

(13) Lee, C. T.; Amaro, R. E. Exascale Computing: A New Dawn for Computational Biology. *Comput Sci Eng* **2018**, *20* (5). https://doi.org/10.1109/MCSE.2018.05329812.

(14) Bruce, N. J.; Ganotra, G. K.; Kokh, D. B.; Sadiq, S. K.; Wade, R. C. New Approaches for Computing Ligand–Receptor Binding Kinetics. *Curr Opin Struct Biol* **2018**, *49* (Cmd), 1–10. https://doi.org/10.1016/j.sbi.2017.10.001.

(15) Chong, L. T.; Saglam, A. S.; Zuckerman, D. M. Path-Sampling Strategies for Simulating Rare Events in Biomolecular Systems. *Current Opinion in Structural Biology*. 2017. https://doi.org/10.1016/j.sbi.2016.11.019.

(16) Ahmad, K.; Rizzi, A.; Capelli, R.; Mandelli, D.; Lyu, W.; Carloni, P. Enhanced-Sampling Simulations for the Estimation of Ligand Binding Kinetics: Current Status and Perspective. *Frontiers in Molecular Biosciences*. 2022. https://doi.org/10.3389/fmolb.2022.899805.

(17) Votapka, L. W.; Stokely, A. M.; Ojha, A. A.; Amaro, R. E. SEEKR2: Versatile Multiscale Milestoning Utilizing the OpenMM Molecular Dynamics Engine. *J Chem Inf Model* **2022**, *62* (13). https://doi.org/10.1021/acs.jcim.2c00501.

(18) Ahn, S. H.; Jagger, B. R.; Amaro, R. E. Ranking of Ligand Binding Kinetics Using a Weighted Ensemble Approach and Comparison with a Multiscale Milestoning Approach. *J Chem Inf Model* **2020**, *60* (11). https://doi.org/10.1021/acs.jcim.9b00968.

(19) 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). https://doi.org/10.1021/acs.jctc.5b00436.

(20) Wang, J.; Arantes, P. R.; Bhattarai, A.; Hsu, R. V.; Pawnikar, S.; Huang, Y. ming M.; Palermo, G.; Miao, Y. Gaussian Accelerated Molecular Dynamics: Principles and Applications. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. 2021. https://doi.org/10.1002/wcms.1521.

(21) Bussi, G.; Laio, A. Using Metadynamics to Explore Complex Free-Energy Landscapes. *Nature Reviews Physics 2020 2:4* **2020**, *2* (4), 200–212. https://doi.org/10.1038/s42254-020-0153-0.

(22) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc Natl Acad Sci U S A* **2002**, *99* (20). https://doi.org/10.1073/pnas.202427399.

(23) Ribeiro, J. M. L.; Provasi, D.; Filizola, M. A Combination of Machine Learning and Infrequent Metadynamics to Efficiently Predict Kinetic Rates, Transition States, and Molecular Determinants of Drug Dissociation from G Protein-Coupled Receptors. *Journal of Chemical Physics* **2020**, *153* (12). https://doi.org/10.1063/5.0019100.

(24) Plattner, N.; Noé, F. Protein Conformational Plasticity and Complex Ligand-Binding Kinetics Explored by Atomistic Simulations and Markov Models. *Nat Commun* **2015**, *6* (May). https://doi.org/10.1038/ncomms8653.

(25) 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). https://doi.org/10.1021/acs.jctc.8b00230.

(26) Nunes-Alves, A.; Kokh, D. B.; Wade, R. C. Ligand Unbinding Mechanisms and Kinetics for T4 Lysozyme Mutants from TRAMD Simulations. *Curr Res Struct Biol* **2021**, *3*. https://doi.org/10.1016/j.crstbi.2021.04.001.

(27) Doerr, S.; Harvey, M. J.; Noé, F.; De Fabritiis, G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *J Chem Theory Comput* **2016**, *12* (4), 1845–1852. https://doi.org/10.1021/ACS.JCTC.6B00049.

(28) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything You Wanted to Know about Markov State Models but Were Afraid to Ask. *Methods*. 2010. https://doi.org/10.1016/j.ymeth.2010.06.002.

(29) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *Journal of the American Chemical Society*. 2018. https://doi.org/10.1021/jacs.7b12191.

(30) Lawrenz, M.; Shukla, D.; Pande, V. S. Cloud Computing Approaches for Prediction of Ligand Binding Poses and Pathways. *Sci Rep* **2015**, *5*, 1–5. https://doi.org/10.1038/srep07918.

(31) Doerr, S.; De Fabritiis, G. On-the-Fly Learning and Sampling of Ligand Binding by High-Throughput Molecular Simulations. *J Chem Theory Comput* **2014**, *10* (5), 2064–2069. https://doi.org/10.1021/CT400919U.

(32) Gu, S.; Silva, D. A.; Meng, L.; Yue, A.; Huang, X. Quantitatively Characterizing the Ligand Binding Mechanisms of Choline Binding Protein Using Markov State Model Analysis. *PLoS Comput Biol* **2014**, *10* (8). https://doi.org/10.1371/journal.pcbi.1003767.

(33) Betz, R. M.; Dror, R. O. How Effectively Can Adaptive Sampling Methods Capture Spontaneous Ligand Binding? *J Chem Theory Comput* **2019**, *15* (3), 2053–2063. https://doi.org/10.1021/acs.jctc.8b00913.

(34) Takahashi, R.; Gil, V. A.; Guallar, V. Monte Carlo Free Ligand Diffusion with Markov State Model Analysis and Absolute Binding Free Energy Calculations. *J Chem Theory Comput* **2014**, *10* (1), 282–288. https://doi.org/10.1021/ct400678g.

(35) Dandekar, B. R.; Ahalawat, N.; Sinha, S.; Mondal, J. Markov State Models Reconcile Conformational Plasticity of GTPase with Its Substrate Binding Event. *JACS Au* **2023**. https://doi.org/10.1021/jacsau.3c00151.

(36) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: A Server for Estimating PKas and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res* **2005**, *33* (SUPPL. 2). https://doi.org/10.1093/nar/gki464.

(37) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating PK Prediction and the Preparation of Biomolecular Structures for Atomistic Molecular Modeling and Simulations. *Nucleic Acids Res* **2012**, *40* (W1). https://doi.org/10.1093/nar/gks375.

(38) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J Mol Graph Model* **2006**, *25* (2), 247–260. https://doi.org/10.1016/j.jmgm.2005.12.005.

(39) He, X.; Man, V. H.; Yang, W.; Lee, T. S.; Wang, J. A Fast and High-Quality Charge Model for the next Generation General AMBER Force Field. *Journal of Chemical Physics* **2020**, *153* (11). https://doi.org/10.1063/5.0019056.

(40) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the Multimolecule and Multiconformational RESP Methodology to Biopolymers: Charge Derivation for DNA, RNA, and Proteins. *J Comput Chem* **1995**, *16* (11). https://doi.org/10.1002/jcc.540161106.

(41) M. J. Frisch, G.; Trucks, W.; Schlegel, H. B. ; Scuseria, G. E. ; Robb, M. A. ; Cheeseman, J. R.; Scalmani, G.; Barone, V. ; Mennucci, B. ; Petersson, G. A.; Nakatsuji, H.; Caricato, M. ; Li, X. ; Hratchian, H. P. ; Izmaylov, A. F. ; Bloino, J. ; Zheng, G. ; Sonnenberg, J. L. ; Frisch, M. J.; Trucks, G. W. ; Schlegel, H. B. ; Scuseria, G. E. ; Robb, M. A. ; Cheeseman, J. R.; Scalmani, G. ; Barone, V. ; Mennucci, B. ; Petersson, G. A.; Nakatsuji, H. ; Caricato, M. ; Li, X. ; Hratchian, H. P. ; Izmaylov, A. F. ; Bloino, J. ; Zheng, G. ; Sonnenberg, J. L. ; Hada, M.; Ehara, M.; Toyota, K. ; Fukuda, R. ; Hasegawa, J. ; Ishida, M. ; Nakajima, T. ; Honda, Y. ; Kitao, O. ; Nakai, H. ; Vreven, T. ; Montgomery, J. A., Jr.; Peralta, J. E. ; Ogliaro, F. ; Bearpark, M. ; Heyd, J. J. ; Brothers, E. ; Kudin, K. N. ; Staroverov, V. N.; Kobayashi, R. ; Normand, J.; Raghavachari, K. ; Rendell, A. ; Burant, J. C. ; Iyengar, S. S. ; Tomasi, J. ; Cossi, M. ; Rega, N. ; Millam, J. M. ; Klene, M. ; Knox, J. E. ; Cross, J. B. ; Bakken, V. ; Adamo, C. ; Jaramillo, J. ; Gomperts, R. ; Stratmann,

R. E. ; Yazyev, O. ; Austin, A. J. ; Cammi, R. ; Pomelli, C. ; Ochterski, J. W. ; Martin, R. L. ; Morokuma, K. ; Zakrzewski, V. G. ; Voth, G. A. ; Salvador, P. ; Dannenberg, J. J.; Dapprich, S. ; Daniels, A. D. ; Farkas, Ö. ; Foresman, J. B. ; Ortiz, J. V. ; Cioslowski, J. ;; Fox, D. J. Gaussian 09, Revision E. 01; Gaussian. *Gaussian, Inc.: Wallingford, CT* **2009**, *11* (1).

(42) Case, D. A.; Walker, R. C.; Cheatham, T. E.; Simmerling, C.; Roitberg, A.; Merz, K. M.; Luo, R.; Darden, T. Amber 18. *University of California, San Francisco. 2018* **2018**.

(43) Delano, W. L. The PyMOL Molecular Graphics System. *CCP4 Newsletter on protein crystallography* **2002**, *40* (1).

(44) 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* **2009**, *31* (2), NA-NA. https://doi.org/10.1002/jcc.21334.

(45) Kovalenko, A.; Luchko, T.; Gusarov, S.; Roe, D. R.; Simmerling, C.; Case, D. A.; Tuszynski, J. Three-Dimensional Molecular Theory of Solvation Coupled with Molecular Dynamics in Amber. *J Chem Theory Comput* **2010**, *6* (3). https://doi.org/10.1021/ct900460m.

(46) Sindhikara, D. J.; Yoshida, N.; Hirata, F. Placevent: An Algorithm for Prediction of Explicit Solvent Atom Distribution-Application to HIV-1 Protease and F-ATP Synthase. *J Comput Chem* **2012**, *33* (18). https://doi.org/10.1002/jcc.22984.

(47) Li, Z.; Song, L. F.; Li, P.; Merz, K. M. Systematic Parametrization of Divalent Metal Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. *J Chem Theory Comput* **2020**, *16* (7). https://doi.org/10.1021/acs.jctc.0c00194.

(48) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput* **2015**, *11* (4). https://doi.org/10.1021/ct5010406.

(49) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J Chem Theory Comput* **2012**, *8* (5). https://doi.org/10.1021/ct200909j.

(50) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J Comput Phys* **1977**, *23* (3). https://doi.org/10.1016/0021-9991(77)90098-5.

(51) Zwanzig, R. Nonlinear Generalized Langevin Equations. *J Stat Phys* **1973**, *9* (3). https://doi.org/10.1007/BF01008729.

(52) 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.

(53) Case, D. A.; Cheatham, T. E. I.; Simmerling, C.; Roitberg, A.; Merz, K. M.; Walker, R. C.; Luo, R.; Li, P.; Darden, T.; Sagui, C.; Pan, F.; Wang, J.; Roe, D. R.; Swails, J.; Götz, A. W.; Smith, J.; Cerutti, D.; Lee, T.; York, D.; Giese, T.; Luchko, T.; Forouzesh, N.; Man, V.; Cruzeiro, V. W. D.; Monard, G.; Miao, Y.; Wang, J.; Lin, C.; Cisneros, G. A.; Rahnamoun, A.; Shajan, A.; Manathunga, M.; Berryman, J. T.; Skrynnikov, N. R.; Mikhailovskii, O.; Xue, Y.; Izmailov, S. A.; Kasavajhala, K.; Belfon, K.; Shen, J.; Harris, J.; Onufriev, A.; Izadi, S.; Wu, X.; Gohlke, H.; Schott-Verdugo, S.; Qi, R.; Wei, H.; Wu, Y.; Zhao, S.; Zhu, Q.; King, E.; Giamba, G.; Liu, J.; Nguyen, H.; Brozell, S. R.; Kovalenko, A.; Gilson, M.; Ben-Shalom, I.; Kurtzman, T.; Pantano, S.; Machado, M.; Aktulga, H. M.; Kaymak, M. C.; O'Hearn, K. A.; Kollman, P. A. Amber 2023 Reference Manual. *University of California, San Francisco* **2023**.

(54) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of Slow Molecular Order Parameters for Markov Model Construction. *Journal of Chemical Physics* **2013**, *139* (1). https://doi.org/10.1063/1.4811489.

(55) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J. H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J Chem Theory Comput* **2015**, *11* (11), 5525–5542. https://doi.org/10.1021/acs.jctc.5b00743.

(56) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, *12*.

(57) Röblitz, S.; Weber, M. Fuzzy Spectral Clustering by PCCA+: Application to Markov State Models and Data Classification. *Adv Data Anal Classif* **2013**, *7* (2). https://doi.org/10.1007/s11634-013-0134-6.

(58) Prinz, J. H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schtte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *Journal of Chemical Physics* **2011**, *134* (17). https://doi.org/10.1063/1.3565032.

(59) McInnes, L.; Healy, J.; Astels, S. Hdbscan: Hierarchical Density Based Clustering. *The Journal of Open Source Software* **2017**, *2* (11). https://doi.org/10.21105/joss.00205.

(60) Bujotzek, A.; Weber, M. Efficient Simulation of Ligand-Receptor Binding Processes Using the Conformation Dynamics Approach. *J Bioinform Comput Biol* **2009**, *7* (5). https://doi.org/10.1142/S0219720009004369.

(61) Wolf, S. Predicting Protein–Ligand Binding and Unbinding Kinetics with Biased MD Simulations and Coarse-Graining of Dynamics: Current State and Challenges. *J Chem Inf Model* **2023**, *63* (10), 2902–2910. https://doi.org/10.1021/acs.jcim.3c00151.

(62) Lecina, D.; Gilabert, J. F.; Guallar, V. Adaptive Simulations, towards Interactive Protein-Ligand Modeling. *Sci Rep* **2017**, *7* (1), 1–11. https://doi.org/10.1038/s41598-017-08445-5.

(63) 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 February). https://doi.org/10.1371/journal.pone.0280776.

(64) Wang, J.; Do, H. N.; Koirala, K.; Miao, Y. Predicting Biomolecular Binding Kinetics: A Review. *J Chem Theory Comput* **2023**, *19* (8), 2135–2148. https://doi.org/10.1021/acs.jctc.2c01085.

(65) Sohraby, F.; Nunes-Alves, A. Advances in Computational Methods for Ligand Binding Kinetics. *Trends in Biochemical Sciences*. 2022. https://doi.org/10.1016/j.tibs.2022.11.003.

(66) Kokkonen, P.; Sykora, J.; Prokop, Z.; Ghose, A.; Bednar, D.; Amaro, M.; Beerens, K.; Bidmanova, S.; Slanska, M.; Brezovsky, J.; Damborsky, J.; Hof, M. Molecular Gating of an Engineered Enzyme Captured in Real Time. *J Am Chem Soc* **2018**, *140* (51). https://doi.org/10.1021/jacs.8b09848.

(67) 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). https://doi.org/10.1093/bioinformatics/btab872.

(68) Matthews, C.; Weare, J.; Kravtsov, A.; Jennings, E. Umbrella Sampling: A Powerful Method to Sample Tails of Distributions. *Mon Not R Astron Soc* **2018**, *480* (3), 4069–4079. https://doi.org/10.1093/MNRAS/STY2140.

Publication 2

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: 8

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

Bioinformatics, 38(6), 2022, 1752–1753
https://doi.org/10.1093/bioinformatics/btab872
Advance Access Publication Date: 31 December 2021
Applications Note

Structural bioinformatics

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

Jan Brezovsky [1,2,*], Aravind Selvaram Thirunavukarasu [1,2], Bartlomiej Surpeta[1,2], Carlos Eduardo Sequeiros-Borja[1,2], Nishita Mandal[1,2], Dheeraj Kumar Sarkar[1,2], Cedrix J. Dongmo Foumthuim [1,2] and Nikhil Agrawal[3]

[1]Laboratory of Biomolecular Interactions and Transport, Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland, [2]International Institute of Molecular and Cell Biology in Warsaw, 02-109 Warsaw, Poland and [3]Latvian Institute of Organic Synthesis, LV-1006 Riga, Latvia

*To whom correspondence should be addressed.
Associate Editor: Alfonso Valencia

Received on November 16, 2021; editorial decision on December 26, 2021; accepted on December 28, 2021

Abstract

Summary: Information regarding pathways through voids in biomolecules and their roles in ligand transport is critical to our understanding of the function of many biomolecules. Recently, the advent of high-throughput molecular dynamics simulations has enabled the study of these pathways, and of rare transport events. However, the scale and intricacy of the data produced requires dedicated tools in order to conduct analyses efficiently and without excessive demand on users. To fill this gap, we developed the TransportTools, which allows the investigation of pathways and their utilization across large, simulated datasets. TransportTools also facilitates the development of custom-made analyses.

Availability and implementation: TransportTools is implemented in Python3 and distributed as pip and conda packages. The source code is available at https://github.com/labbit-eu/transport_tools. Data are available in a repository and can be accessed via a link: https://doi.org/10.5281/zenodo.5642954.

Contact: janbre@amu.edu.pl or jbrezovsky@iimcb.gov.pl

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

At any moment, living systems contain thousands of small organic molecules that need to arrive at their sites of action to exert their function. The transport of these molecules around the cell (and beyond) is governed primarily by channels and tunnels (henceforth referred to as 'pathways') formed from the internal voids of biomolecules (Kingsley and Lill, 2015). These pathways enable the transport of ions and small molecules between different regions, connecting inner cavities with a surface, two different cavities with each other, or different cellular environments via transmembrane proteins. Operating as such, the investigation of these pathways is critical to drug discovery (Marques *et al.*, 2017) and protein engineering initiatives (Kokkonen *et al.*, 2019). Since pathways are often equipped with dynamic gates (Gora *et al.*, 2013), they are mostly transient and challenging to study.

One of the most common approaches used to characterize these rare events of ligand transmission via transiently open pathways is to run molecular dynamics (MD) simulations (Decherchi and Cavalli, 2020), analyzing the pathway dynamics using tools like CAVER (Jurcik *et al.*, 2018) or tracking ligand migration through the biomolecules with AQUA-DUCT (Magdziarz *et al.*, 2020); see Supplementary File S1 for an overview of the state-of-the-art tools to study ligand transport pathways. The intensive development seen in computing hardware and sampling algorithms over recent years has led to considerable growth in the size and complexity of datasets typically generated for a single protein system. It is not uncommon for such datasets to consist of thousands simulations. Such high-throughput approaches, however, impose a substantial burden on researchers in establishing the identity of the pathways observed across all simulations, determining which pathways are used by particular ligands, and developing means of specific quantitative analyses. To this end, we present TransportTools: a library designed to alleviate these difficulties by providing easy, efficient access to comprehensive details on transport processes—even for large-scale simulation sets—and offering an environment for the development of novel analyses and tools.

2 Features

TransportTools is available as a Python3 module distributed under the GNU General Public License v3.0, and available via pip and conda managers as the *transport_tools* package. In its standard workflow (Fig. 1), TransportTools utilizes outputs from CAVER and AQUA-DUCT analyses of MD simulation, integrating their complementary insights to investigate transport pathways and corresponding ligand migration events in soluble and membrane-embedded proteins. To achieve efficiency in such a high-throughput regimen, raw data on pathway ensembles and ligand-transport events is first coarse-grained, and positioned on a spherical grid. Next, TransportTools identifies relationships between pathway ensembles from individual simulations and joins them into superclusters, to which ligand-transport events are then assigned (see Supplementary File S2 for method details). Critical analysis parameters can be controlled via a configuration file. These parameters are thoroughly explained in the user guide, which also includes a detailed walk-through tutorial (Supplementary File S3). Aside from the ready-made workflow, the library offers many classes to process, manipulate and analyze pathways and events, simplifying the production of custom-made analyses and, hopefully, stimulating further development of new packages (Supplementary File S4).

Outputs

The main results generated by TransportTools are presented as a set of tables stored in text files. These contain data on the composition of pathway superclusters, on their geometrical properties and utilization by transport events, and on critical protein residues. Using generated scripts, the spatial representation of superclusters and assigned events can be visualized in PyMOL (PyMOL, Schrödinger, 2017). All results can be refined using various filters and split by individual simulation or by user-defined groups to facilitate their convenient comparison.

Performance and limitations

The performance of TransportTools was analyzed on three datasets of 50 simulations (each sampling 100 ns and consisting of 10 000 frames) of up to 500 residue-long enzymes with different accessibilities of their active sites, resulting in the detection of up to 5 000 000 transport pathways and 50 000 water-transport events, which were processed within 2–21 h on a standard workstation (Supplementary File S5). TransportTools inherits the limitations of the CAVER and AQUA-DUCT packages; their descriptions of pathway geometries and the definitions of their clusters (see Supplementary Section S2.2 of Supplementary File S3 for best practice guidelines). When MD trajectories are utilized directly, usage is restricted to file formats supported by either *MDtraj* or *pytraj* packages (McGibbon *et al.*, 2015; Roe and Cheatham, 2013).

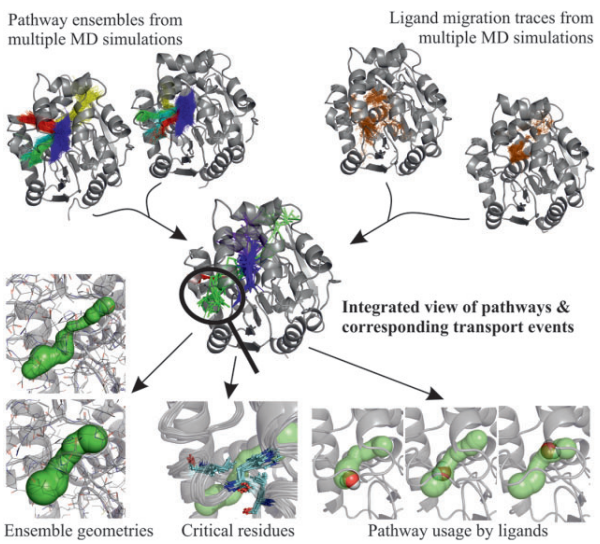


Fig. 1. Schematic of a standard TransportTools analysis workflow

Use cases

To illustrate the applicability of TransportTools, we applied it to the analysis of three representative examples of biological problems connected with ligand transport using an established model system—enzymes DhaA and LinB from the haloalkane dehalogenase family (Brezovsky *et al.*, 2016; Pavlova *et al.*, 2009). First, we analyzed 10 simulations of DhaA in an effort to discover rare transient tunnels and their usage by water molecules (Supplementary File S6). Next, we derived an understanding of the effect of mutations on the system by contrasting simulations of LinB wild-type, LinB32 mutant with a closed primary tunnel, and LinB86 mutant with a *de novo* created tunnel (Supplementary File S7). Finally, we studied the substrate molecule selectivity of the pathways leading to the active site of LinB86 in almost 600 simulations (Supplementary File S8).

3 Conclusions

The TransportTools library provides users with access to (i) efficient analyses of transport pathways across extensive MD simulations, including those originating from massively parallel calculations or very long simulations; (ii) integrated data regarding transport pathways and their actual utilization by small molecules; and (iii) rigorous comparisons of transport processes under different settings, e.g. by contrasting transport in an original system against the same system perturbed by mutations, different solvents or bound ligands.

Acknowledgement

Computations were performed at the Poznan Supercomputing and Networking Center.

Funding

This work was supported by the National Science Centre, Poland [2017/25/B/NZ1/01307 and 2017/26/E/NZ1/00548], POWER projects [POWR.03.02.00-00-I022/16 and POWR.03.02.00-00-I006/17] and a grant of the Dean of Faculty of Biology, Adam Mickiewicz University [GDWB-05/2020].

Conflict of Interest: none declared.

References

Brezovsky,J. *et al.* (2016) Engineering a de novo transport tunnel. *ACS Catal.*, **6**, 7597–7610.
Decherchi,S. and Cavalli,A. (2020) Thermodynamics and kinetics of drug–target binding by molecular simulation. *Chem. Rev.*, **120**, 12788–12833.
Gora,A. *et al.* (2013) Gates of enzymes. *Chem. Rev.*, **113**, 5871–5923.
Jurcik,A. *et al.* (2018) CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics*, **34**, 3586–3588.
Kingsley,L.J. and Lill,M.A. (2015) Substrate tunnels in enzymes: structure–function relationships and computational methodology. *Proteins*, **83**, 599–611.
Kokkonen,P. *et al.* (2019) Engineering enzyme access tunnels. *Biotechnol. Adv.*, **37**, 107386.
Magdziarz,T. *et al.* (2020) AQUA-DUCT 1.0: structural and functional analysis of macromolecules from an intramolecular voids perspective. *Bioinformatics*, **36**, 2599–2601.
Marques,S.M. *et al.* (2017) Enzyme tunnels and gates as relevant targets in drug design. *Med. Res. Rev.*, **37**, 1095–1139.
McGibbon,R.T. *et al.* (2015) MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.*, **109**, 1528–1532.
Pavlova,M. *et al.* (2009) Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat. Chem. Biol.*, **5**, 727–733.
Roe,D.R. and Cheatham,T.E. (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.*, **9**, 3084–3095.
Schrödinger, LLC. (2017) *The PyMOL Molecular Graphics System, Version 2.0*. Schrödinger, LLC.

Publication 3

Nano-structured hydrotrope-caged cytochrome c with boosted stability in harsh environments: a molecular insight

Impact Factor 2022: 9.8

MNiSW points 2023: 200

Citations: 0

Main text

https://doi.org/10.1039/D3GC01704D (featured as 2023 HOT *Green Chemistry* article)

Supplementary information

https://www.rsc.org/suppdata/d3/gc/d3gc01704d/d3gc01704d1.pdf

Repository

https://doi.org/10.5281/zenodo.7875755

Green Chemistry

PAPER

View Article Online
View Journal | View Issue

Cite this: *Green Chem.*, 2023, **25**, 6666

Nano-structured hydrotrope-caged cytochrome c with boosted stability in harsh environments: a molecular insight†

Pranav Bharadwaj,‡[a] Dheeraj Kumar Sarkar, ‡[b,c] Meena Bisht,§[d] Sachin M. Shet,[a] Nataraj Sanna Kotrappanavar, [a] Veeresh Lokesh,[d] Gregory Franklin, *[d] Jan Brezovsky *[b,c] and Dibyendu Mondal *[a,d]

Green and nano-structured catalytic media are vital for biocatalysis to attenuate the denaturation tendency of biocatalysts under severe reaction conditions. Hydrotropes with multi-faceted physiochemical properties represent promising systems for sustainable protein packaging. Herein, the ability of adenosine-5′-triphosphate (ATP) and cholinium salicylate ([Cho][Sal]) ionic liquid (IL) to form nano-structures and to nano-confine Cytochrome c (Cyt c) enhanced the stability and activity under multiple stresses. Experimental and computational analyses were undertaken to explain the nano-structured phenomenon of ATP and IL, structural organizations of nano-confined Cyt c, and site-specific interactions that stabilize the protein structure. Both ATP and IL form nano-structures in aqueous media and could cage Cyt c *via* multiple nonspecific soft interactions. Remarkably, the engineered molecular nano-cages of ATP (5–10 mM), IL (300 mg mL^{-1}), and ATP + IL surrounding Cyt c resulted in 9-to-72-fold higher peroxidase activity than native Cyt c with exceptionally high thermal tolerance (110 °C). The polar interactions with the cardiolipin binding site of Cyt c, mediated by hydrotropes, were well correlated with the increased peroxidase activity. Furthermore, higher activity trends were observed in the presence of urea, GuHCl, and trypsin without any protein degradation. Specific binding of hydrotropes in highly mobile regions of Cyt c (Ω 40–54 residues) and enhanced H-bonding with Lys and Arg offered excellent stability under extreme conditions. Additionally, ATP effectively counteracted reactive oxygen species (ROS)-induced denaturation of Cyt c, which was enhanced by the [Sal] counterpart of IL. Overall, this study explored the robustness of nano-structured hydrotropes to have a higher potential for protein packaging with improved stability and activity under extreme conditions. Thus, the present work highlights a novel strategy for real-time industrial biocatalysis to protect mitochondrial cells from ROS-instigated apoptosis.

Received 19th May 2023,
Accepted 7th August 2023

DOI: 10.1039/d3gc01704d

rsc.li/greenchem

Introduction

Adenosine-5′-triphosphate (ATP) has shown multi-faceted functions, including its participation in the electron transport chain, acting as the energy currency of cells, holding the key for caspase-9 activated apoptosis, and taking part in transcription by revamping chromatin complexes.[1–3] The role of physiological concentrations of ATP in preventing thermally induced protein aggregation in mammalian cells was previously demonstrated by Nguyen and Benusaude,[4] and the answer regarding the regulation of a high concentration of ATP (2–10 mM) in the cell was unveiled by Patel and co-workers, whereby ATP is claimed as a biological hydrotrope.[5] The term 'hydrotrope' was coined by Neuberg for amphiphilic molecules having a characteristic range of minimum hydrotrope concentration, which augment the solubility of partially soluble organic or hydrophobic substances in water.[6,7] This opened a

[a]*Centre for Nano and Material Sciences, Jain (Deemed-to-be University), Jain Global Campus, Kanakapura, Bangalore, Karnataka, 562112, India. E-mail: dmtapu@gmail.com, dmon@igr.poznan.pl*
[b]*Laboratory of Biomolecular Interactions and Transport, Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Uniwersytetu Poznanskiego 6, 61-614 Poznan, Poland. E-mail: janbre@amu.edu.pl, jbrezovsky@iimcb.gov.pl*
[c]*International Institute of Molecular and Cell Biology in Warsaw, Ks Trojdena 4, 02-109 Warsaw, Poland*
[d]*Institute of Plant Genetics (IPG) of the Polish Academy of Sciences, Strzeszyńska 34, 60-479 Poznań, Poland. E-mail: fgre@igr.poznan.pl*
†Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d3gc01704d
‡These authors have contributed equally to this work.
§Current address: Department of Chemistry, Sri Venkateswara college, University of Delhi, Dhaula Kuan, New Delhi 110021, India.

new channel to explore ATP in the medical field to treat Parkinson's and Alzheimer-like diseases by preventing phase separation of biological fluids, which causes amyloid fibrillation.[8,9] The solubilizing tendency of ATP was attributed to its nonspecific aggregation, charge reinforcement, and self-assembly through hydrophobic interactions (Fig. 1a).[10,11] Thus, the hydrotropic mechanism of ATP follows both the Neuberg and Hofmeister effects.[12] Self-aggregated ATP is found to provide thermal stability to lysozyme, malate dehydrogenase, and ubiquitin,[13] and so it is important to understand how the self-aggregation property of ATP is useful in protein packaging under multiple stressors, which has not been previously studied.

Unlike ATP, various amphiphilic nano-structured solvents have recently emerged with promising applications in energy-efficient bioprocesses.[14] Due to their highly manipulative nature, ionic liquids (ILs) have gained interest for their ability to increase enzyme activity, solubility and thermal stability.[15] Since the initial demonstration of ILs in biocatalysis, numerous research groups have focused on utilizing imidazolium, pyridium, tetraalkyl-ammonium, and tetraalkyl-phosphonium based ILs for catalysis involving various classes of enzymes.[16,17] Further, with extended studies over ILs, bio-based ILs of the cholinium family have shown hydrotropic properties[18] with greater hydrophobicity and ability to enhance the nano-structuring of amphiphilic solutes in aqueous media.[19] Although earlier reports showed protein packaging with long-term stability and activity using cholinium ILs,[15] but there is a lack of understanding of the underlying mechanisms by which hydrotropic ILs (potential amphiphilic molecules) can regulate protein stability and activity.[20] In this regard, an IL with a salicylate counterpart caught our attention as it contains carboxylic and hydroxyl groups, along with a hydrophobic aromatic ring featuring the salicylate ion and showing multiple interaction sites in the aqueous medium (Fig. 1b).[21,22] Additionally, salicylic acid is an important phytohormone that alters the mitochondrial processes by amplifying the concentration of reactive oxygen species (ROS) and induces cells to undergo programmed cell death.[23,24] During apoptosis, ATP plays a crucial role in caspase-9 activation, a process stimulated by cytochrome c (Cyt c) release from mitochondria.[25] Since both ATP and salicylate are coupled to Cyt c, the latter was chosen as a model protein for this study (Fig. 1c), and ATP and choline salicylate ([Cho][Sal]) IL as the nano-structured hydrotropes.

Cyt c is a mitochondrial-based 13 kDa monomeric protein comprising of approximately 104 residues and is widely distributed across eukaryotes, bacteria, and archaea.[26] Despite its significant role in electron transfer, this protein is of great interest in many technological applications like biosensors, synthetic receptors, and molecular imprinting among others.[26] Cyt c also shows peroxidase-like activity and serves as a biocatalyst in several catalytic transformations.[26] The protein is highly vulnerable to denaturation under prevailing stress conditions such as extreme temperatures, oxidative stress, and proteolytic and chemical denaturants.[27] Recently, a surface modification strategy using quantum dots, metal–organic frame-works,[28] and DNA was reported to enhance the stability of Cyt c.[27,29] However, the stability of the protein when confined within a hydrotrope system has not yet been reported. Considering the hydrotropism and probability of nano-structuring,[26] the present study aimed to cage Cyt c in ATP and [Cho][Sal] IL-based hydrotropes, and thereby improve the activity and stability of the protein in gentle and harsh environments (such as high temperature, chemical denaturants, protease digestion, and oxidative stress). Therefore, detailed experimental and *in silico* approaches were undertaken to study the structure–function relationship of nano-confined Cyt c in ATP, [Cho][Sal] and ATP + IL at a molecular level to develop a novel and sustainable solvent manipulation strategy for dynamic protein packaging and robust biocatalysis.

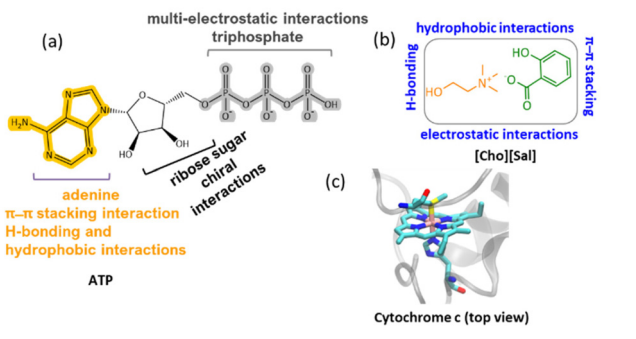


Fig. 1 Chemical structures and structural features of (a) ATP and (b) [Cho][Sal] IL, with possible interaction sites. (c) The modeled active site structure of Cyt c, based on the crystallographic structure of bovine heart Cyt c (PDB code: 2B4Z) at 1.5 Å resolution.

Results and discussion

ATP and [Cho][Sal] IL as nano-structured hydrotropes

To understand the extent to which the IL affects the self-aggregation of ATP, which is vital for its role as a biological hydrotrope, we performed molecular dynamics (MD) simulations of 5 mM ATP, 300 mg mL^{-1} [Cho][Sal], and their mixture. The pair-wise interactions of ATP and [Cho][Sal] molecules in water and the mixture were investigated using radial distribution functions (RDF). As expected, ATP exhibited a self-aggregating tendency at 4-to-5 Å and up to 15 Å (Fig. 2a). The self-aggregating propensity of ATP molecules from a chemical perspective in the presence of an IL can occur in three possible ways, first the anion–π interactions between the aromatic ring and negatively charged oxygen moiety of phosphate groups, secondly through a non-bonding aromatic ring interaction (H-bonding, π–π stacking interactions, and cation–π interactions) and third, H-bond formation between the sugar and triphosphate group.[11] It was evident from our analysis that two ATP molecules preferentially formed stack-like configurations in water (Fig. S1a†). In the presence of [Cho][Sal], the IL-mediated H-bonding (Fig. S1b†) markedly decreased the occurrence of

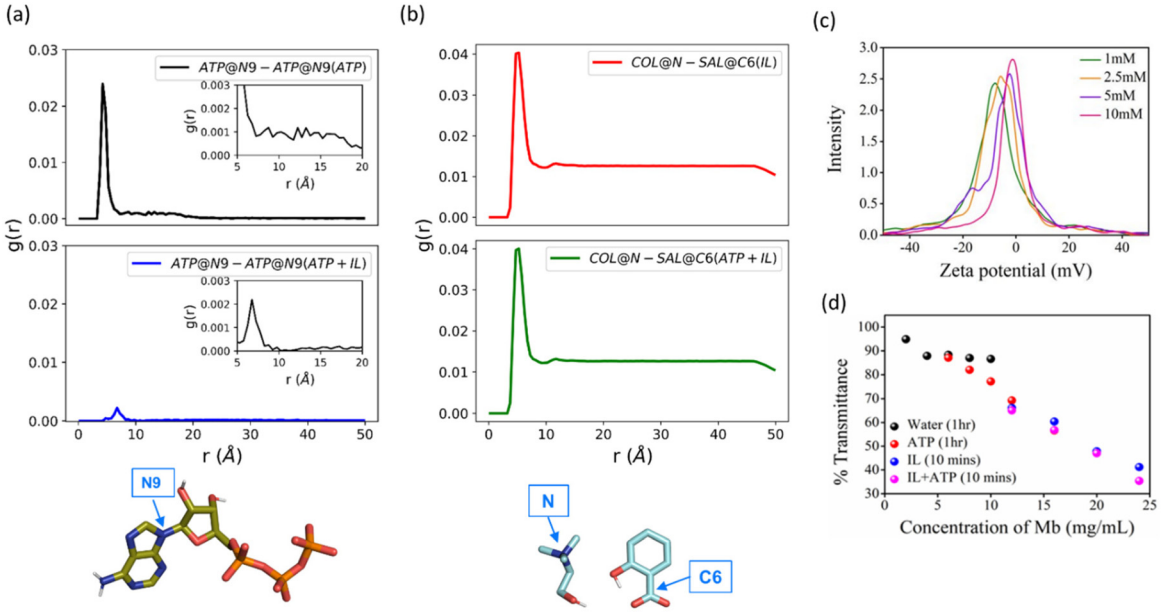


Fig. 2 Mutual effects of ATP and [Cho][Sal] on their respective nano-structure and interactions. (a) RDF of 5 mM ATP (black) and 5 mM ATP in presence of [Cho][Sal] (blue). The N9 atom of ATP used for RDF calculation is shown by the blue arrow. (b) RDF of 300 mg mL^{-1} [Cho][Sal] (red) and [Cho][Sal] with added ATP (green). The N and C6 atoms used for RDF calculation are shown by the blue arrows. (c) Concentration-dependent zeta potential analysis of ATP. (d) Solubility data of equine heart myoglobin in an aqueous medium at pH 7 in the presence and absence of 5 mM ATP, 100 mg mL^{-1} [Cho][Sal], and a combination of both ATP + IL.

the stack-like configurations of ATP (Fig. 2a). Notably, we found that the IL at a molecular concentration of 300 mg mL^{-1} forms dense H-bonding network mediated by cholinium and salicylate (Fig. S1c†). Unsurprisingly, we have not observed any perturbation of the [Cho][Sal] nano-structure upon the addition of ATP, given its low concentration (Fig. 2b). In addition to RDFs, the self-aggregating behavior of ATP and ATP + IL was further verified by a concentration-dependent ζ-potential study (Fig. 2c and Fig. S2†). At 1 mM concentration, which is below the critical hydrotropic concentration (2 mM), a single broad distribution peak was observed with maxima of −8.12 mV, indicating the existence of a single non-aggregated structure (Fig. 2c). At 2.5 mM, the distribution curve shifted to a lower negative value with peak maxima of −5.96 mV. This lowering in the ζ-potential indicates the beginning of aggregated structures similar to the self-aggregation of nano-particles as a result of the decrease in the electrophoretic mobility.[30] This is further evidenced by the shift of maxima from −5.96 to −2.35 mV for 5 mM ATP. Finally, at 10 mM concentration, a broad single peak with maxima at −1.42 mV was obtained, which indicates the presence of a self-assembled, nano-structured ATP state. Conversely, 200 mg mL^{-1} of [Cho][Sal] shows a very broad Gaussian peak with a maximum at −9.12 mV (Fig. S3†), which shifts to a lower value of −4.02 mV in the presence of 5 mM ATP, justifying the hypothesis that the IL enhanced the self-assembly of amphiphiles.[21] The hydrodynamic radius (R_h) of 5 mM ATP was found to be 0.83 nm (slightly higher than R_h of monomeric ATP, which is 0.57–0.65 nm),[10] suggesting the existence of nano-structured/oligomeric forms and supporting our earlier evidence. In the presence of the IL, the R_h of ATP increased to 1.12 nm (Fig. S4†).

To demonstrate the hydrotropic property of ATP and [Cho][Sal], solubility studies of myoglobin (Mb) were carried out in an aqueous medium at pH 7 in the presence and absence of 5 mM ATP, 100 mg mL^{-1} IL and a combination of both (Fig. 2d). After 1 h, the maximum solubility of Mb was found to be 6 mg mL^{-1}. In contrast, ATP-induced solubility of Mb increased to 16 mg mL^{-1} during the same duration, which directly shows the hydrotropic nature of ATP. In the case of [Cho][Sal], the same solubility was achieved within 5 min. Thus, the concentration of Mb was further increased and examined until 24 mg mL^{-1}. Interestingly, for ATP + IL, 24 mg mL^{-1} Mb showed higher solubility than the IL-only medium. These observations of enhanced solubility in the presence of two hydrotropes inspired us towards nano-confinement of Cyt c using ATP and [Cho][Sal]-based nano-structured hydrotropes to enable improved biological activity and stability of the protein.

Structural features of nano-structured hydrotrope-caged Cyt c

To understand the structural features of caged Cyt c, four molecular systems were prepared, namely keeping Cyt c in water, ATP, IL, and both ATP + IL at 26.85 °C (Fig. 3a–d). All molecular systems were solvated in a cubic box with a side length of 10 nm prepared using a PACKMOL package.[31] The RDFs of hydrotropes in all systems with and without the presence of the protein were in agreement, indicating that the presence of Cyt c had a negligible effect on the adopted hydrotrope nano-structures (Fig. 2a, b & Fig. S5†). A lower dielectric con-

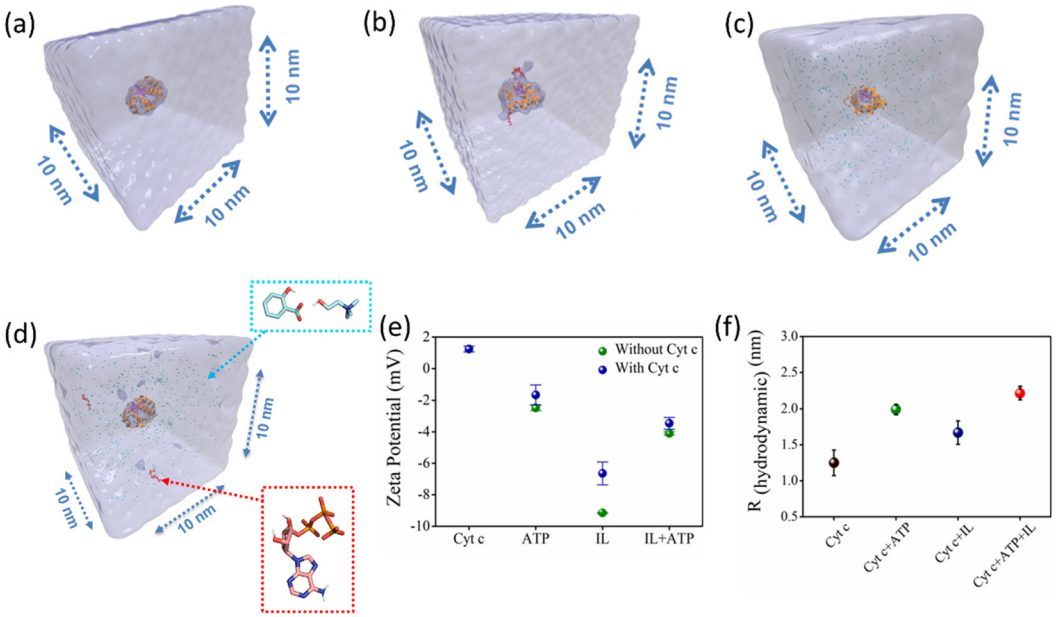


Fig. 3 (a–d) Representative figures of a solvent of cubic box length 10 nm for Cyt c in H_2O (a), ATP + H_2O (b), [Cho][Sal] + H_2O (c), and ATP + IL + H_2O molecular systems (d). Graph (e) shows zeta (ζ)-potential studies of nano-structured hydrotropes and Cyt c with and without nano-structured hydrotropes. Plot (f) depicts the hydrodynamic radius (R_h) of native Cyt c and Cyt c caged in different nano-structured hydrotropes.

stant of the protein or more negative zeta potential of ATP and IL can be related to the colloidal stability of caged Cyt C. Accordingly, ζ-potential analysis was employed to envisage the electrostatic interaction between Cyt c and nano-structured hydrotropes. We found the ζ-potential of native Cyt c solution (10 mM) to be +1.24 mV (Fig. 3e), and the ζ-potential of ATP (5 mM), IL (200 mg mL^{-1}), and ATP + IL was −2.48 mV, −9.16 mV, and −4.09 mV, respectively. After Cyt c incubation with ATP, IL, and IL + ATP, the ζ-potential values decreased to −1.66 mV, −6.64 mV, and −3.46 mV, respectively (Fig. 3e and Fig. S6†). These changes in the ζ-potential demonstrate the possible electrostatic interaction between the protein and the nano-structured hydrotropes. Further R_h of native Cyt c was found to be 1.25 nm, which agrees with a previous report,[32] and in the presence of ATP, the value increased to 1.99 nm showing self-assembly of ATP around the protein (Fig. 3f and Fig. S7†). In the presence of IL + ATP, R_h of Cyt c is further increased to 2.22 nm. Thus, the hydrotropic mechanism of ATP occurs initially by charge reinforcement, followed by the stacking phenomenon. Along with ζ-potential and DLS, structures of nano-confined Cyt c were further characterized by UV-vis spectra (Fig. S8†). In the presence of 5 mM ATP, a hyperchromic shift (a blue shift of 2 nm) in the Soret band (409 nm) was observed, suggesting that a non-polar environment dominates around the heme. In the case of IL (300 mg mL^{-1} [Cho][Sal]), absorbance in the Soret band enhanced considerably together with a blue shift, while the Q-band (520–550 nm) is broadened due to enhanced absorbance. This suggests changes in the tertiary conformations.[33] Overall, it is evident that the self-aggregation propensities of the hydrotropes persisted even in the presence of Cyt c, and the structure of caged Cyt c was partially affected (without any unfolding, as discussed below) due to multiple interactions with nano-structured hydrotropes.

Nano-structured hydrotrope-caged Cyt c with boosted peroxidase activity: molecular insights on the improved catalytic activity and enhanced stability

In the presence of H_2O_2, Cyt c undergoes covalent modifications, which facilitate peroxidase-like activity.[34] The activity of Cyt c in the presence of 5–10 mM ATP showed a similar trend with an >8-fold higher activity than native Cyt c (Fig. 4a). Since a typical nano-structure is formed around Cyt c in the range of hydrotropic concentration (5–10 mM), similar relative activity is achieved rather than a maxima–minima trend. ATP acts as a hydrotrope initially by unwinding the protein and then stabilizing the extended chain through electrostatic interactions,[11] which may be a reason for the high peroxidase activity. In the case of the [Cho][Sal] IL, nearly the same relative activities were seen from the 200-to-500 mg mL^{-1} range (Fig. 4b). A clear trend was observed with the highest relative activity of 67-fold at the concentration of the IL of 300 mg mL^{-1}. The activity data agree with the UV-vis absorption spectra that inferred the transition of Cyt c towards penta-coordinated high spin complex. When the Met80 ligand is displaced, Cyt c naturally exhibits pronounced peroxidase-like activity.[34] Interestingly, a replica of the IL optimization trend pattern was observed when using the ATP + IL mixture when [ATP] was kept constant (5 mM) and IL was varied (Fig. 4c). At 300 mg mL^{-1} IL and 5 mM ATP concentrations, the activity of Cyt c was enhanced 72-fold more than native Cyt c. This activity enhancement is indicating that the structure of Cyt c is

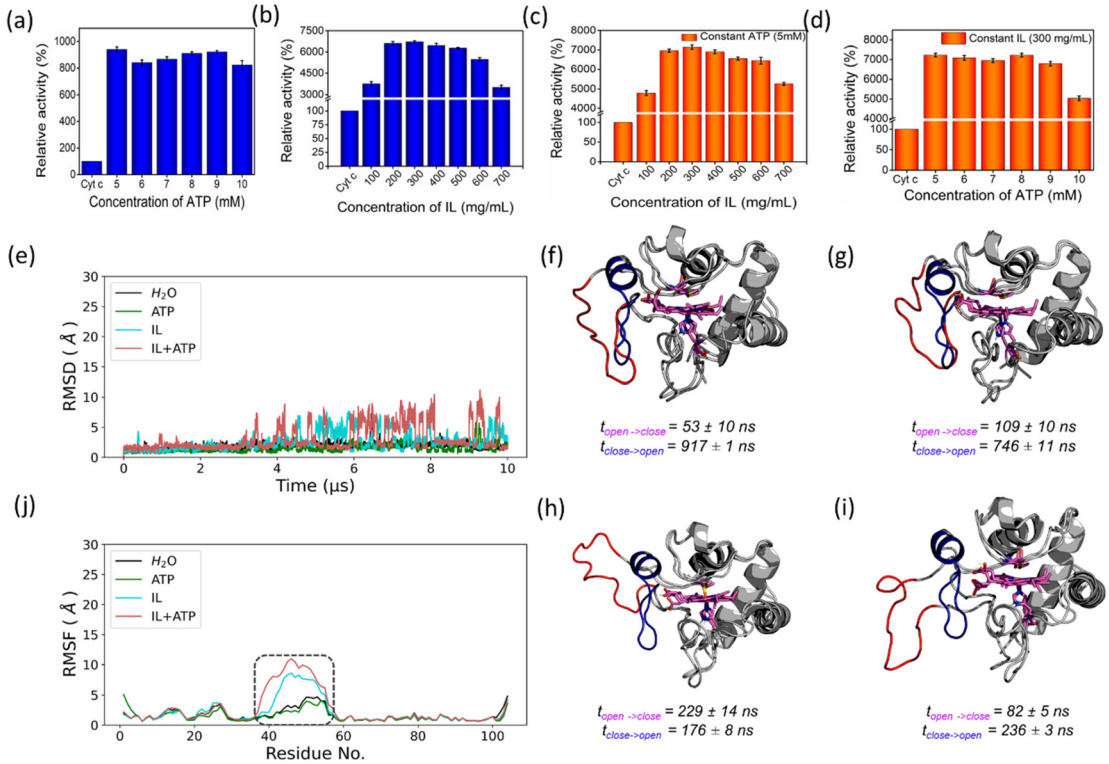


Fig. 4 Peroxidase-like activity of Cyt linked to its dynamics. Optimization results of peroxidase-like activity of Cyt c with different hydrotropic concentrations of (a) ATP, (b) [Cho][Sal] IL, (c) varying [IL] by keeping [ATP] constant (5 mM), and (d) varying [ATP] by keeping IL constant (300 mg mL^{-1}). Structural characterization of Cyt c in four solvent systems at 26.85 °C. (e, j) Stability and dynamics of Cyt c structure simulated in four solvent systems shown as backbone RMSD and RMSF, respectively. The peaks in the RMSD plot correspond to the reversible openings of Ω 40–54 residue region highlighted in the RMSF plot by a dashed rectangle. Figures (f–i) show the metastable states representing open (red) and closed (blue) structures of Cyt c in H_2O, ATP, IL, and ATP + IL, respectively. The timescales of opening and closing processes were calculated as mean first passage times between the two states presented as mean ± standard deviation.

still intact and combined hydrotropic systems boost the peroxidase activity further. Some ILs enhance the self-assembly of co-solutes in the media due to the virtue of their solvophobic effects.[19] Accordingly, the concentration of ATP was varied with IL being constant (300 mg mL^{-1}). A similar trend was observed with a >69-fold increase in activity (Fig. 4d). These trends demonstrate that both IL and ATP further enhance the conformational changes of Cyt c, followed by stabilization of the extended protein structure by the self-assembling of ATP to surround it. Similar inference can be observed from Fig. S5a (blue line),† in which ATP tends to self-aggregate, even in the presence of IL and Cyt c. To understand the behavior of nano-structured hydrotrope-caged Cyt c, adaptive sampling MD simulations with a total of 10 μs simulation time for each system were performed. In all systems, we observed reversible structural rearrangements of the protein backbone (Fig. 4e), which were more pronounced in mixtures containing IL and ATP + IL. Next, we traced these changes to the reversible opening of the loop in the Ω 40–54 residue region that was relatively more mobile than other parts of the protein for all four systems (Fig. 4j). This region corresponds to the mAb 1D3 binding site and is reported to be more dynamic,[35] while having specificity towards Cytochrome c oxidase.[36] To delineate the details of this process and its modulation by the action of nano-structured hydrotropes, we have constructed Bayesian Markov state models from the simulations. This inference revealed the presence of metastable states with an open conformation of the Ω 40–54 loop of Cyt c in all systems (Fig. 4f–i), which were characterized by their increased RMSD and elevated radius of gyration (Fig. S9†). The interactions of ATP and IL with Cyt c resulted in marked effects on the stability and dynamics of this region. In the presence of ATP, the closing process was approximately 2 times slower, whereas the opening became approximately 20% faster (Fig. 4g). Similarly, the presence of the IL resulted in a much faster opening process and the closing speed was slowed even more prominently than that observed for ATP (Fig. 4h). Finally, the dynamics of Cyt c in the mixture of ATP and IL exhibited ATP-like closing and IL-like opening times, enhancing the dynamics of this functionally important region (Fig. 4i). To understand if the hydrotropes molecules interact directly with the heme group of Cyt c, a structural analysis of was carried out by computational study. The simulations show infrequent interactions of hydrotropes with the heme group (Fig. S10†). Such interactions are unlikely to contribute to markedly improved activity observed with all three systems (ATP, IL, and

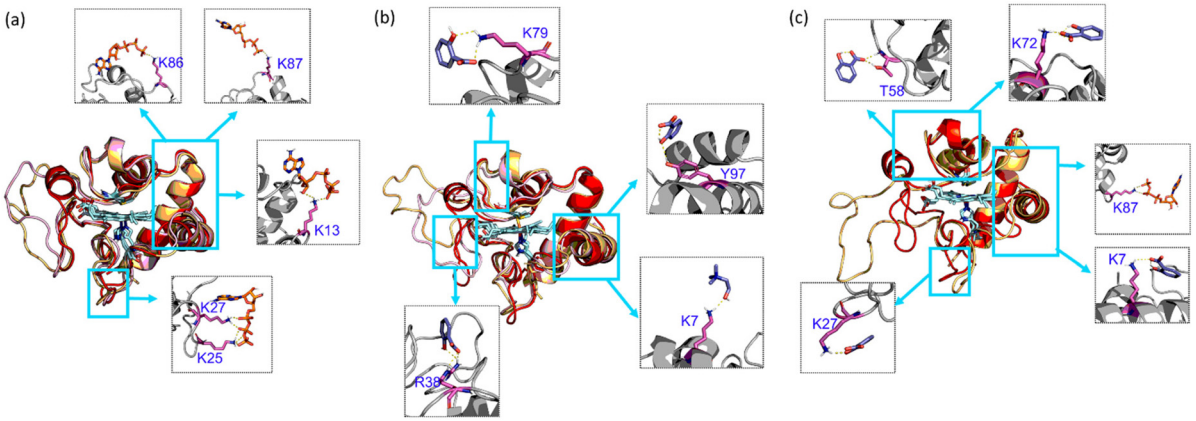


Fig. 5 Interaction of hydrotrope molecules in metastable states in the open (orange), closed (red), and transient (pink) Cyt c conformations from 10 μs of adaptive simulation data with (a) ATP, (b) [Cho][Sal] IL, and (c) ATP + IL. The top residues that have hydrogen bond interactions between solvent molecules and Cyt c are shown in gray squares (dashed), and the protein regions are highlighted in cyan (solid). All simulations were performed at 26.85 °C. In all cases, the concentration of ATP and IL was 5 mM and 300 mg mL^{-1}, respectively.

ATP + IL), in particular, considering their absence with ATP. Since both ATP and IL were significantly affecting the dynamics of Cyt c, we identified the residues exhibiting the most frequent H-bonding with hydrotropes in the simulations (Fig. 5). In general, ATP interacted preferentially with Lys residues (Fig. S11a and S12a†), which is in accordance with an earlier report demonstrating that triphosphate group of ATP can deliver weak, nonspecific interactions with conserved Lys or Arg residues.[10] Additionally, the IL had a wider range of residues as relevant binding partners, although Lys was the most populated in terms of H-bonding (Fig. S11b and S12b†). This trend was also maintained in the mixture of ATP and IL (Fig. S11c and S12c, d†). We observed marked overlap in particular interaction hot spots for IL and ATP on the surface of Cyt c (Fig. S11d and e†). ATP most often formed interactions with K13, K25, K27, K72, K86, and K87 (Fig. 5a and S11d†). Notably, K72 is involved in site A of Cyt c, which is crucial for interacting with cardiolipin,[37] and was reported to have a significant role in the peroxidase activity of Cyt c.[38] Also, IL demonstrated general preference to Lys, with somewhat higher affinity towards K7, K79, R38 (Fig. 5b and S11e†). Interestingly we also found considerable affinity towards Y97, which has been ascribed a significant role in keeping the terminal helix intact and hence providing structural stability to the protein.[39]

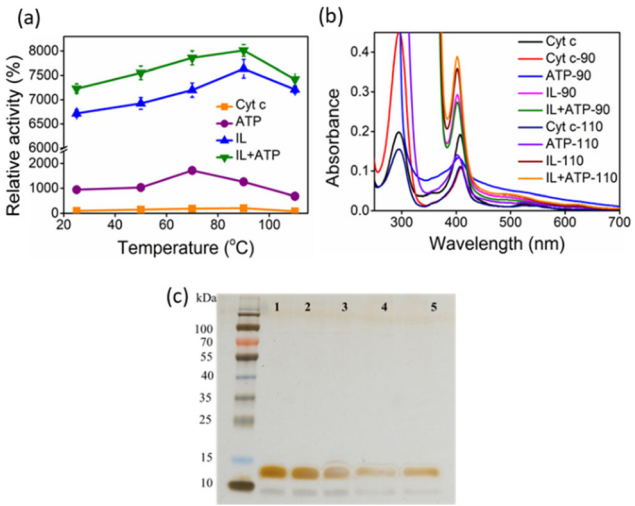


Fig. 6 (a) Peroxidase activity of Cyt c with and without nano-structured hydrotropes at different temperatures. (b) UV-vis spectra of Cyt c at higher temperatures (90 °C and 110 °C) with and without nano-structured hydrotropes. (c) SDS-PAGE image of native Cyt c (lane 1), Cyt c incubated at 100 °C (lane 2), Cyt c + ATP incubated at 100 °C (lane 3), Cyt c + [Cho][Sal] incubated at 100 °C (lane 4), and Cyt C + ATP + [Cho][Sal] incubated at 100 °C (lane 5).

Nano-structured hydrotrope-caged Cyt c with enhanced thermal stability for high-temperature bio-catalysis

The activity of Cyt c increases with temperature up to 70 °C, above which the protein starts to denature gradually (Fig. 6a). Above 80 °C, the activity drops due to significant denaturation of the protein.[27] A similar trend is observed in the presence of ATP, but the activity extends to approximately a 20-fold increase than for native Cyt c at 70 °C. In the case of IL and ATP + IL systems, the denaturation is prevented even up to extremes of 110 °C. At 90 °C, both Cyt c + IL and Cyt c + IL + ATP showed an accelerated activity value corresponding to 76–80-fold higher than native Cyt c. The stability aspects of Cyt c, when coupled to hydrotropes at elevated temperatures, were supported by UV-vis spectra (Fig. 6b). Enhancement in the Soret band intensity was observed for both IL and IL + ATP at 90 °C compared to the results obtained at 26.85 °C, which implies that the secondary conformation, formed due to weak interactions between the hydrotrope systems and Cyt c, was preserved. ATP + IL showed an intensified Soret peak compared to the other conditions at 110 °C, suggesting the advantage of engineering Cyt c with two different hydrotropes when aiming for high-temperature bio-catalysis. Additionally, as

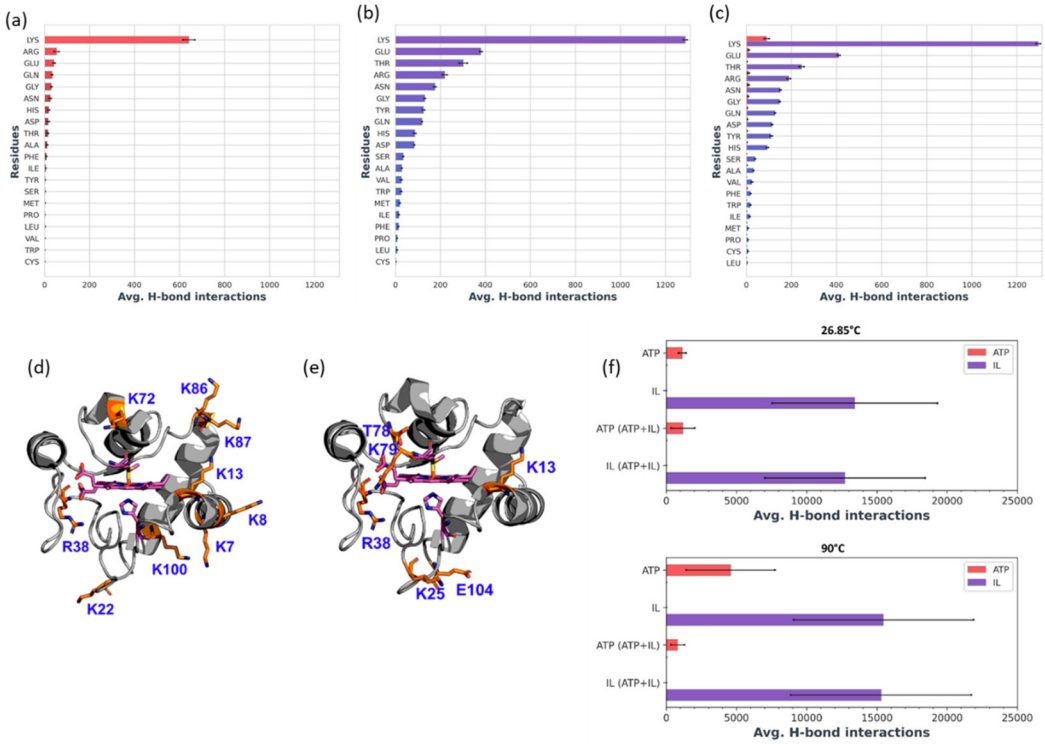


Fig. 7 (a–c) H-Bonding interaction counts of individual residue types obtained from adaptive simulation data of Cyt c in the presence of ATP (a), IL (b), ATP and IL in the ATP + IL system (c) are depicted with ATP in red and IL in blue. All simulations were performed at 90 °C. (d and e) The position of residues having a higher occurrence of H-bonding with ATP (d) and IL (e) also in their mixtures, respectively. The concentration of ATP and IL was 5 mM and 300 mg mL^{-1}, respectively. (f) Overall counts of H-bonding interaction between hydrotropes and Cyt c obtained from adaptive simulations, showing average across 100 performed simulations and corresponding standard error of mean.

evident from non-reducing SDS-PAGE (Fig. 6c), improved activity trends at high temperatures were achieved without structural degradation of the protein.

A systematic computational study was undertaken at 90 °C to gain structural insights into the higher stability of nano-structured-caged Cyt c at elevated temperatures. As expected, the Cyt c conformational ensemble was significantly altered at the higher temperature, with backbone RMSD being approximately 3–4 times higher than that at 26.85 °C for all four solvents (Fig. S13a†). No significant effects were observed between the mobility of the Ω 40–54 residue region in the four systems than that observed for 26.85 °C (Fig. 4e and S13b†). However, the thermal stability of the Ω 70–85 residues region was significantly enhanced with nano-structured hydrotropes compared to water (Fig. S13b†).

From Bayesian Markov state models constructed based on the high-temperature simulations (Fig. S13c–f†), we could observe that Cyt c in water exhibited severe conformational instability, underlined by a very high RMSD of the open state, as well as the presence of multiple secondary open-like states (Fig. S14†). In contrast, the presence of ATP or/and IL approximately halved the maximal range of displacements of the Cyt c backbone, providing heightened Cyt c resistance. Hence, it is evident that ATP and IL could thermally stabilize the functional component of Cyt c *via* multiple polar interactions.[26] Affinity of hydrotrope molecules towards specific amino acids at the higher temperature (Fig. 7a–c and Fig. S15†) was found to be similar to that observed at 26.85 °C (Fig. S11 & S12†). Overall, Lys and Arg were found to have frequent H-bonding at 90 °C to ATP (Fig. 7a, c and d). Similarly to the lower temperature environment, the IL had a broader range of specificity, and more affinity with Lys, Glu and Thr in terms of H-bonding (Fig. 7b, c and e). The highest affinity was seen towards K7, K8, K13, K22, R38, K72, K86, K87 and K100 with ATP (Fig. 7d), and K13, K25, R38, T78, K79, and E104 with IL (Fig. 7e), all of which are crucial in maintaining thermal stability.[40] It is important to highlight that the synergistic effects of ATP and IL were observed for the peroxidase activity of Cyt C (Fig. 4a–d) but not for stabilization, where IL and the mixture of IL with ATP grant similar enhancements to the Cyt C resistance towards higher temperature (Fig. 6a). Indeed, molecules of IL form numerous H-bonds with the protein at 26.85 °C as well as at 90 °C indicating their key role in the observed stabilisation (Fig. 7f). Moreover, the overall number of formed H-bonds is similar in IL and ATP + IL systems corresponding well with the observed enhancements in the Cyt C stability in the presence of IL as well as its mixture with ATP.

Improved stability of nano-structured hydrotrope-caged Cyt c when exposed to chemical denaturants

Above a particular concentration and on more prolonged exposure to some chaotropic chemicals like H_2O_2, guanidine

hydrochloride (GuHCl), and urea cause amide bond disruption or overexpose the metallic core of the enzyme, causing a loss of structural integrity and function. To investigate the stability of Cyt c against chemical denaturants, the protein was incubated with the three nano-structured hydrotropic systems for 10 min, followed by the addition of a chemical denaturant. After the interaction with 6 M GuHCl for 15 min, there was a complete disruption in the heme pocket of native Cyt c and the activity severely declined, with the remaining activity approximately ~20%. Conversely, hydrotrope-caged Cyt c maintained initial activity of 8.5-fold, 63-fold, and 66-fold in the presence of ATP, IL, and ATP + IL, respectively (Fig. 8a). While urea is known to be a hydrotrope, it is well recognized as a chemical denaturant at concentrations above 6 M. In the pre-denaturation concentrations of urea (1–6 M), a ~30 fold increase in peroxidase-like activity was obtained for Cyt c.[41] After 30 min incubation with 8 M urea, its activity was recorded at approximately 70% of the initial activity, while IL and ATP + IL hydrotrope-stabilized Cyt c showed similar activity (Fig. 8b).

Native Cyt c showed a denaturing pattern in the UV-vis spectra with GuHCl (Fig. 8c), where the Q band structure was entirely lost and the Soret band was severely blue-shifted. In the case of ATP, changes in the 280 nm shoulder peak suggest partial dissociation of aromatic residues like Tyr and Trp. In contrast, the structures remained almost the same in the cases of IL and ATP + IL, except that the absorbance of the IL-based system trended lower. In the case of ATP, the activity of Cyt c almost doubled (~24-fold higher than native Cyt c) in the presence of urea compared to the Cyt c + ATP system (Fig. 7b). This strange observation suggested that ATP cannot prevent further unfolding of Cyt c in the presence of urea, a behavior confirmed by the UV-vis spectra (Fig. S16†). Furthermore, SDS-PAGE analysis demonstrated that Cyt c did not degrade upon treatment with denaturing agents in the presence or absence of ATP and [Cho][Sal] (Fig. 8d), thus, confirming that the variation in the activity was due to the changes in the structural conformation of the protein.

Nano-structured hydrotrope-caged Cyt c is resistant to protease digestion

Trypsin, being a proteolytic enzyme, hydrolyses the peptide bond at basic amino acids such as Arg and Lys residues (Fig. 9a), thereby denaturing the target protein.[42] Native and hydrotrope-stabilized Cyt c systems were incubated with 6 μM trypsin at 37 °C for 24 h. After the interaction, only 25% of activity was retained for native Cyt c, whereas the hydrotropic systems showed excellent performance and maintained Cyt c resistant to the action of trypsin (Fig. 9b). In the presence of ATP, Cyt c lost only ~2-fold activity compared to activity at the 0 hour (~700% retained), whereas IL and ATP + IL systems retained ~61- and ~67-fold relative activity, respectively. This excellent performance against protease digestion can be deduced by close examination of the interactions of ATP and IL with Cyt c. H-bonding analysis revealed that ATP has a higher tendency to interact with Lys, Thr, and Gln at both 26.85 °C and 90 °C temperatures, thus suggesting the efficacy of ATP to impart stability to Cyt c (Fig. 7a–c and Fig. S13†). Additionally, IL also greatly affected the dynamic nature by

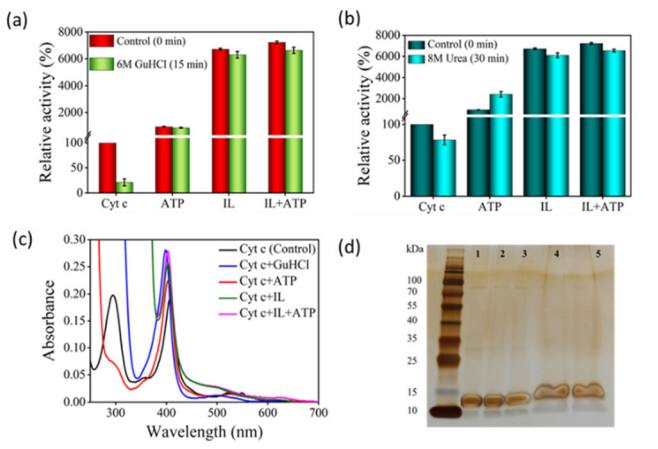


Fig. 8 Effect of chemical denaturants, 6 M GuHCl (a) and 8 M urea (b) on the peroxidase activity of Cyt c and nano-structured hydrotrope-caged Cyt c. (c) UV-vis spectra of Cyt c after incubation with 6 M GuHCl. (d) SDS-PAGE image of native Cyt c (lane 1), Cyt c incubated with 8 M urea (lane 2), Cyt c + ATP incubated with 8 M urea (lane 3), Cyt c + [Cho][Sal] incubated with 8 M urea (lane 4), and Cyt c + ATP + [Cho][Sal] incubated with 8 M urea (lane 5). In all cases, the concentration of ATP and IL was 5 mM and 300 mg mL^{-1}, respectively.

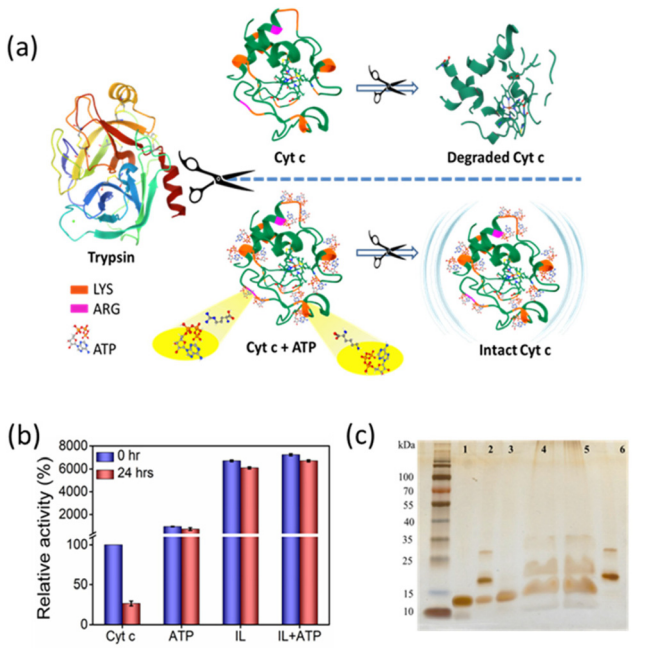


Fig. 9 (a) Plausible mechanism of Cyt c digestion in the presence and absence of nano-structured hydrotropes. (b) Effect of protease digestion (6 μM trypsin) on peroxidase activity of Cyt c. (c) SDS-PAGE of native Cyt c (lane 1), Cyt c incubated with trypsin (lane 2), Cyt c + ATP incubated with trypsin (lane 3), Cyt c + [Cho][Sal] incubated with trypsin (lane 4), Cyt c + ATP + [Cho][Sal] incubated with trypsin (lane 5), only trypsin (lane 6). In all cases, the concentration of ATP and IL was 5 mM and 300 mg mL^{-1}, respectively.

interacting with a wide range of residues. The affinity of ATP and IL towards Lys may be helpful to protect Cyt c by hindering the degradation mechanism of trypsin. The electrostatic and H-bond interactions that occur when Cyt c is nano-structured with ATP completely safeguard the protein from trypsin degradation (Fig. 9a). This was further shown by non-reducing SDS-PAGE analysis of the Cyt c before and after digestion with trypsin (Fig. 9c). Most of the native Cyt c in the presence of trypsin was found degraded (lane 2, Fig. 9c). However, the protein was intact when it was digested in the presence of ATP, IL and a combination of both. From UV-vis spectra (Fig. S17†), the Q band is lost in almost all the cases (except with ATP) after incubation with trypsin. However, only marginal changes were observed in the Soret band of Cyt c + ATP, and Cyt c + ATP + IL systems, indicating the stability induced by the nano-structured hydrotropes. In addition, the hydrotropic systems enhance the activity of trypsin (Fig. S18†) with 1.4 to 2 folds higher activity than control. These results give strong evidence that the higher activity of Cyt c in the presence of trypsin is due to the stabilisation provided by nanostructured hydrotropic systems and not by the inactivation of trypsin.

Nano-structured hydrotrope-caged Cyt c reaction to oxidative stress

Unlike other stress conditions where the peptide back-bone is attacked, H_2O_2 initially targets the catalytic site, followed by a breakdown of peptide bonds. Thus, even though H_2O_2 is essential for the peroxidase activity of Cyt c, upon prolonged exposure it causes severe impairment to the protein structure (Fig. 10a). In the peroxidase activity studies mentioned above, 1 mM of H_2O_2 was added at the end, whereas for this study, Cyt c was incubated with H_2O_2 for 30 min, and the substrate was added afterwards. The activity of bare Cyt c following the induction of oxidative stress was found to be reduced to approximately 60% of its original activity (Fig. 10b).

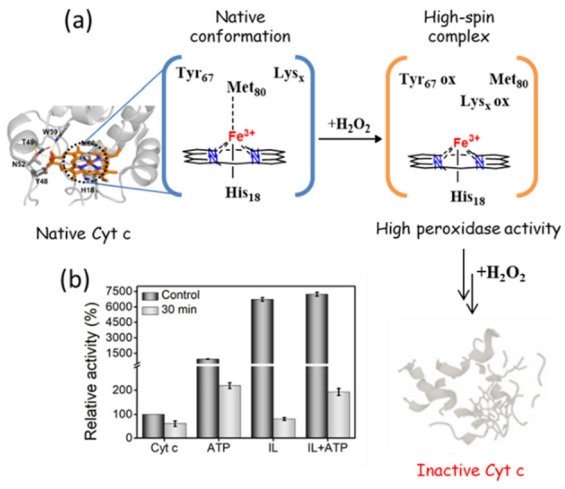


Fig. 10 (a) Schematic mechanism showing H_2O_2 induced structural degradation of Cyt c. (b) Effect of H_2O_2 on the peroxidase activity of Cyt c in the presence of different hydrotropes.

Conversely, in the presence of ATP, ~219% of activity was retained. This observation can be correlated with the earlier findings in which ATP was found to prevent the oxidation of Cyt c from the Cyt c oxidase enzyme in the mitochondria by hampering the electron flow rate.[43,44] Alternatively, there was a severe drop in the activity of Cyt c from 6700% to just 80% in the presence of IL. These results are not astonishing since the salicylate counterpart is known to enhance the formation of ROS.[24] However, the presence of ATP was found to prevent the oxidation and Cyt c from IL-induced oxidative stress and demonstrated 192% activity. These findings demonstrate the stabilizing effect of nano-structured ATP surrounding Cyt c.

Conclusions

In summary, this study systematically demonstrated the robustness of ATP and IL-based mixed nano-structured hydrotropes and their utility in the improvement of protein packaging in extreme conditions. From the Mb solubility data, it was shown that the hydrotropic nature of ATP accelerated with the addition of a co-hydrotrope like [Cho][Sal] and *vice versa*. Through RDF, DLS, ζ-potential, and UV-vis data, we provided convincing data that ATP forms oligomeric nano-structures at concentrations above 2.5 mM around Cyt c, which can be augmented further with the addition of [Cho][Sal]. A 9-fold increase in peroxidase activity with ATP, 67-fold with IL, and 72-fold with IL + ATP suggested a 'partial reversible unfolding-electrostatic stabilization' mechanism. Reversible binding of ATP and IL with the Ω 40–54 residue loop region of Cyt c showed higher stability at 26.85 °C. Moreover, specific binding of nano-structured hydrotropes with Cyt c to Lys and Arg residues through H-bonding and polar interactions with Ω 70–85 region, presented an exceptionally high thermal tolerance with 80-fold activity even at 90 °C. Because of such binding specificity, the structure and activity of nano-confined Cyt c were found to be retained against protease digestion, which explicitly cleaves at Lys and Arg residues. Furthermore, ATP and IL-based nano-structured hydrotropes showed efficacy in retaining the functional integrity of Cyt c even in the presence of chemical denaturants like urea and GuHCl, which suggested the suitability of these solvent manipulation strategies for industrial bio-catalysis. Through oxidative stress studies, a stark observation was made where [Cho][Sal] was found to enhance ROS formation, which could be efficiently subdued by adding ATP. These results have the potential for counteracting oxidative stress inside the living cells. Thus, the novel strategy of protein confinement in nano-structured hydrotropes can find significant usage in protein packaging under biotic and abiotic stresses, high-temperature bio-catalysis, and cell protection against ROS.

Experimental

Materials and characterization techniques, IL synthesis, Mb solubility study, SDS-PAGE analysis, peroxidase activity of Cyt c

at room temperature and under harsh conditions, system setup and Molecular Dynamics simulation details, and Markov state models are provided in ESI.†

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

D. M. acknowledges SERB, India (EEQ/2021/000059), for the research grant. D. K. S. was a recipient of a scholarship provided by POWER project POWR.03.02.00-00-I006/17. The computations were performed at the Poznan Supercomputing and Networking Center. D. M. and G. F. acknowledge the NANOPLANT project, which received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 856961.

References

1 Y. Eguchi, A. Srinivasan, K. J. Tomaselli, S. Shimizu and Y. Tsujimoto, *Cancer Res.*, 1999, **59**, 2174–2181.
2 D. V. Fyodorov and J. T. Kadonaga, *Nature*, 2002, **418**, 896–900.
3 M. Hochstrasser and J. Wang, *Nat. Struct. Biol.*, 2001, **8**, 294–296.
4 S.-J. Lee, *Antioxid. Redox Signaling*, 2003, **5**, 337–348.
5 A. Patel, L. Malinovska, S. Saha, J. Wang, S. Alberti, Y. Krishnan and A. A. Hyman, *Science*, 2017, **356**, 753–756.
6 C. Neuberg, *Biochem. Z.*, 1916, **76**, 107–108.
7 K. Werner, H. Krister and Z. Thomas, *Curr. Opin. Colloid Interface Sci.*, 2016, **22**, 99–107.
8 D. Mei, K. Jian, L. Liangzhong, L. Yifan, W. Lu and S. Jianxing, *Biochem. Biophys. Res. Commun.*, 2020, **522**, 247–253.
9 V. G. Jack and G. Thomas, *Exp. Eye Res.*, 2020, **190**, 107862.
10 M. Nishizawa, E. Walinda, D. Morimoto, B. Kohn, U. Scheler, M. Shirakawa and K. Sugase, *J. Am. Chem. Soc.*, 2021, **143**, 11982–11993.
11 S. Sarkar and J. Mondal, *J. Phys. Chem. B*, 2021, **125**, 7717–7731.
12 M. Johannes, D. Tuan-Minh, T. Didier, H. Max, K. Ali, H. Dominik and K. Werner, *Cell Rep. Phys. Sci.*, 2021, **2**, 100343.
13 X. Ou, Y. Lao, J. Xu, Y. Wutthinitikornkit, R. Shi, X. Chen and J. Li, *JACS Au*, 2021, **1**, 1766–1777.
14 H. J. Jiang, R. Atkin and G. G. Warr, *Curr. Opin. Green Sustainable Chem.*, 2018, **12**, 27–32.
15 A. Sindhu, S. Kumar, D. Mondal, I. Bahadur and P. Venkatesu, *Phys. Chem. Chem. Phys.*, 2020, **22**, 14811–14821.
16 T. Itoh, *Chem. Rev.*, 2017, **117**, 10567–10607.
17 R. A. Sheldon, *Green Chem.*, 2021, **23**, 8406–8427.
18 T. E. Sintra, D. O. Abranches, J. Benfica, B. P. Soares, S. P. M. Ventura and J. A. P. Coutinho, *Eur. J. Pharm. Biopharm.*, 2021, **164**, 86–92.
19 T. L. Greaves and C. J. Drummond, *Chem. Soc. Rev.*, 2013, **42**, 1096–1120.
20 L. Bui-Le, C. J. Clarke, A. Bröhl, A. P. S. Brogan, J. A. J. Arpino, K. M. Polizzi and J. P. Hallett, *Commun. Chem.*, 2020, **3**, 55.
21 R. Hayes, G. G. Warr and R. Atkin, *Chem. Rev.*, 2015, **115**, 6357–6426.
22 A. Sanchez-Fernandez, A. E. Leung, E. G. Kelley and A. J. Jackson, *J. Colloid Interface Sci.*, 2021, **581**, 292–298.
23 Z. Xie and Z. Chen, *Plant Physiol.*, 1999, **120**, 217–226.
24 P. Poór, *Biomolecules*, 2020, **10**, 341.
25 (*a*) I. Martínez-Reyes and N. S. Chandel, *Nat. Commun.*, 2020, **11**, 102; (*b*) Z. Duan, Y. Ouyang, Y. Fu, F. Huang, F. Xia and I. Willner, *Angew. Chem., Int. Ed.*, 2023, **62**, e2023014.
26 (*a*) D. Alvarez-Paggi, L. Hannibal, M. A. Castro, S. Oviedo-Rouco, V. Demicheli, V. Tórtora, F. Tomasina, R. Radi and D. H. Murgida, *Chem. Rev.*, 2017, **117**, 13382–13460; (*b*) N. Kapil, A. Singh and D. Das, *Angew. Chem., Int. Ed.*, 2015, **54**, 6492–6495.
27 S. M. Shet, P. Bharadwaj, M. Bisht, M. M. Pereira, S. K. Thayallath, V. Lokesh, G. Franklin, N. S. Kotrappanavar and D. Mondal, *Int. J. Biol. Macromol.*, 2022, **215**, 184–191.
28 Y. Chen, F. Jiménez-Ángeles, B. Qiao, M. D. Krzyaniak, F. Sha, S. Kato, X. Gong, C. T. Buru, Z. Chen, X. Zhang, N. C. Gianneschi, M. R. Wasielewski, M. Olvera de la Cruz and O. K. Farha, *J. Am. Chem. Soc.*, 2020, **142**, 18576–18582.
29 S. M. Shet, S. K. Thayallath, M. Bisht, M. M. Pereira, J. A. P. Coutinho, N. Sanna Kotrappanavar and D. Mondal, *ACS Sustainable Chem. Eng.*, 2021, **9**, 8327–8335.
30 M. Aurora, C. Ileana, T. Gheorghe, B. Liviu-Dorel, H. Ossi and T.-C. Maria, *Colloids Surf., A*, 2009, **338**, 93–101.
31 A.-R. Allouche, *J. Comput. Chem.*, 2011, **32**, 174–182.
32 I. Gomes, N. C. Santos, L. M. A. Oliveira, A. Quintas, P. Eaton, E. Pereira and R. Franco, *J. Phys. Chem. C*, 2008, **112**, 16340–16347.
33 U. K. Singh, M. Kumari, S. H. Khan, H. B. Bohidar and R. Patel, *ACS Sustainable Chem. Eng.*, 2018, **6**, 803–815.
34 V. Yin, G. S. Shaw and L. Konermann, *J. Am. Chem. Soc.*, 2017, **139**, 15701–15709.
35 V. E. Kagan, V. A. Tyurin, J. Jiang, Y. Y. Tyurina, V. B. Ritov, A. A. Amoscato, A. N. Osipov, N. A. Belikova, A. A. Kapralov, V. Kini, I. I. Vlasova, Q. Zhao, M. Zou, P. Di, D. A. Svistunenko, I. V. Kurnikov and G. G. Borisenko, *Nat. Chem. Biol.*, 2005, **1**, 223–232.
36 W. B. Gordon, V. L. Gordon and D. B. Gary, *J. Mol. Biol.*, 1990, **214**, 585–595.
37 J. D. Lambeth, *J. Biol. Chem.*, 1981, **256**, 4757–4762.
38 L. J. McClelland, T.-C. Mou, M. E. Jeakins-Cooley, S. R. Sprang and B. E. Bowler, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 6648–6653.

39 (*a*) B. San Francisco, E. C. Bretsnyder and R. G. Kranz, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, E788–E797; (*b*) A. D. Savino, J. M. Foerster, T. L. Haye, A. Blok, M. Timmer, G. M. Ullmann and M. Ubbink, *Angew. Chem., Int. Ed.*, 2020, **59**, 23239–23243.

40 W. Sato, S. Hitaoka, K. Inoue, M. Imai, T. Saio, T. Uchida, K. Shinzawa-Itoh, S. Yoshikawa, K. Yoshizawa and K. Ishimori, *J. Biol. Chem.*, 2016, **291**, 15320–15331.

41 T. Nataša, V. Rastislav, L. Veronika, M. Andrej and S. Erik, *Biochim. Biophys. Acta, Proteins Proteomics*, 2018, **1866**, 1073–1083.

42 M. Krieger, L. M. Kay and R. M. Stroud, *J. Mol. Biol.*, 1974, **83**, 209–230.

43 L. Smith, H. C. Davies and M. E. Nava, *Biochemistry*, 1980, **19**, 1613–1617.

44 D. B. Craig and C. J. A. Wallace, *Protein Sci.*, 1993, **2**, 966–976.

CO-AUTHORS STATEMENTS

Incorporating prior knowledge to seeds of adaptive sampling molecular dynamics simulations of ligand transport in enzymes with buried active sites

1. Sarkar, D. K.;
2. Surpeta, B.;
3. Brezovsky, J.

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

Poznan, September 24, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

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

Sarkar DK, Surpeta B, Brezovsky J, (2023). Incorporating prior knowledge to seeds of adaptive sampling molecular dynamics simulations of ligand transport in enzymes with buried active sites. *bioRxiv* 2023.09.21.558608. DOI: 10.1101/2023.09.21.558608.

I declare the following contribution to this publication:

performed all computational experiments, i.e., simulations and MSM generation; analyzed, interpreted, and visualized the simulations and MSM data; wrote the draft of manuscript except for clustering and transition analyses; and co-wrote the manuscript.

Dheeraj kumar Sarkar

.....
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. Poznan, September 25, 2023
Faculty of Biology, Adam Mickiewicz University, 61-614 Poznań, Poland
Email: bartlomiej.surpeta@amu.edu.pl

CO-AUTHOR CONTRIBUTION STATEMENT

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

Sarkar DK, **Surpeta B**, Brezovsky J, (2023). Incorporating prior knowledge to seeds of adaptive sampling molecular dynamics simulations of ligand transport in enzymes with buried active sites. *bioRxiv* 2023.09.21.558608. DOI: 10.1101/2023.09.21.558608.

I declare the following contribution to this publication:

analyses and investigation of ligand transitions and metastable state clustering; wrote the draft of ligand transitions and metastable state clustering; and co-wrote the manuscript.

Bartłomiej Surpeta
.....
Signature

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

Poznan, September 24, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

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

Sarkar DK, Surpeta B, **Brezovsky J**, (2023). Incorporating prior knowledge to seeds of adaptive sampling molecular dynamics simulations of ligand transport in enzymes with buried active sites. *bioRxiv* 2023.09.21.558608. DOI: 10.1101/2023.09.21.558608.

I declare the following contribution to this publication:

Conceptualization and design of the study; interpretation of computational results; supervision and fund acquisition; co-wrote, reviewed and edited the manuscript.

.....
Signature

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

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

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

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

Poznan, September 22, 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

.....
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

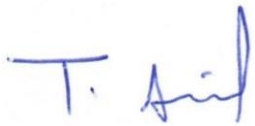
CO-AUTHOR CONTRIBUTION STATEMENT

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.

.....
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

Nano-structured hydrotrope-caged cytochrome c with boosted stability in harsh environments: a molecular insight

1. Sarkar, D.K.;
2. Bharadwaj, P.;
3. Franklin, G.;
4. Brezovsky, J.;
5. Mondal, D.

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

Poznan, September 22, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

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

Bharadwaj P, **Sarkar DK**, Bisht M, Shet SM, Nataraj SK, Lokesh V, Franklin G, Brezovsky J, Mondal D, (2023). Nano-structured Hydrotrope-Caged Cytochrome c with Boosted Stability in Harsh Environments: A Molecular Insight. *Green Chemistry* 2023, 25 (17), 6666–6676. DOI: 10.1039/d3gc01704d.

I declare the following contribution to this publication:

performed all computational experiments, i.e., preparation of structures for modeling with different solvent composition, running high throughput molecular dynamics simulations; analyzed, interpreted and visualized the data; and co-wrote the manuscript.

Dheeraj kumar Sarkar

.....
Signature

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

Centre for Nano and Material Sciences

4th Floor, Jain Global Campus, 45th km, NH-209, Kanakapura Taluk, Ramanagara District, INDIA-562112

Pranav Bharadwaj, M.Sc.
Centre for Nano and Material Sciences,
Jain University, Bangalore 562112, India.
Email: bpranav17@gmail.com

Poznan, September 3, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

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

Bharadwaj P, Sarkar DK, Bisht M, Shet SM, Nataraj SK, Lokesh V, Franklin G, Brezovsky J, Mondal D, (2023). Nano-structured Hydrotrope-Caged Cytochrome c with Boosted Stability in Harsh Environments: A Molecular Insight. *Green Chemistry* 2023, 25 (17), 6666–6676. DOI: 10.1039/d3gc01704d.

I declare the following contribution to this publication:

experimental data generation, in particular performed hydrotropic studies, activity assays, and in-situ characterizations; formally validated and investigated the data and results; co-wrote the manuscript.

.....
Signature

Gregory Franklin, Ph.D.
Institute of Plant Genetics (IPG),
Polish Academy of Sciences,
Strzeszyńska 34, 60-479 Poznań, Poland.
Email: fgre@igr.poznan.pl

Poznan, September 3, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

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

Bharadwaj P, Sarkar DK, Bisht M, Shet SM, Nataraj SK, Lokesh V, **Franklin G**, Brezovsky J, Mondal D, (2023). Nano-structured Hydrotrope-Caged Cytochrome c with Boosted Stability in Harsh Environments: A Molecular Insight. *Green Chemistry* 2023, 25 (17), 6666–6676. DOI: 10.1039/d3gc01704d.

I declare the following contribution to this publication:

contributed to formal analysis of the research; project supervised, reviewed and edited the manuscript; acquisition of project funding.

.....
Signature

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

Poznan, September 20, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

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

Bharadwaj P, Sarkar DK, Bisht M, Shet SM, Nataraj SK, Lokesh V, Franklin G, **Brezovsky J**, Mondal D, (2023). Nano-structured Hydrotrope-Caged Cytochrome c with Boosted Stability in Harsh Environments: A Molecular Insight. *Green Chemistry* 2023, 25 (17), 6666–6676. DOI: 10.1039/d3gc01704d.

I declare the following contribution to this publication:

contributed to conceptualization, design, and interpretation of computational results; supervision of computational work; co-wrote the manuscript.

.....
Signature

Dibyendu Mondal, Ph.D.
Institute of Plant Genetics (IPG),
Polish Academy of Sciences,
Strzeszyńska 34, 60-479 Poznań, Poland
Email: dmon@igr.poznan.pl

Poznan, September 3, 2023

CO-AUTHOR CONTRIBUTION STATEMENT

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

Bharadwaj P, Sarkar DK, Bisht M, Shet SM, Nataraj SK, Lokesh V, Franklin G, Brezovsky J, **Mondal D**, (2023). Nano-structured Hydrotrope-Caged Cytochrome c with Boosted Stability in Harsh Environments: A Molecular Insight. *Green Chemistry* 2023, 25 (17), 6666–6676. DOI: 10.1039/d3gc01704d.

I declare the following contribution to this publication:

conceptualization of the research project; project coordination and supervision; data analysis and interpretation; co-wrote the manuscript; acquired funding for the project.

……
Signature