Biotechnology Advances 60 (2022) 108009

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Fast approximative methods for study of ligand transport and rational design of improved enzymes for biotechnologies

Ondrej Vavra [a,b], Jiri Damborsky [a,b,c,*], David Bednar [a,b,*]

[a] Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic
[b] International Clinical Research Center, St. Anne's University Hospital Brno, Pekařská 53, 656 91 Brno, Czech Republic
[c] Enantis, INBIT, Kamenice 34, 625 00 Brno, Czech Republic

ARTICLE INFO

Keywords:
ART-RRT
Binding
Biotechnology
CaverDock
Catalysis
Cytochrome P450 CYP153A
Fe/α-ketoglutarate-dependent hydroxylase
GPathFinder
Channel
Docking
Ligand
MoMA-LigPath
Monoamine oxidase
Nanomotors
Protein engineering
SLITHER
Threonine deaminase
Transport
Tunnel

ABSTRACT

Acceleration of chemical reactions by the enzymes optimized using protein engineering represents one of the key pillars of the contribution of biotechnology towards sustainability. Tunnels and channels of enzymes with buried active sites enable the exchange of ligands, ions, and water molecules between the outer environment and active site pockets. The efficient exchange of ligands is a fundamental process of biocatalysis. Therefore, enzymes have evolved a wide range of mechanisms for repetitive conformational changes that enable periodic opening and closing. Protein-ligand interactions are traditionally studied by molecular docking, whereas molecular dynamics is the method of choice for studying conformational changes and ligand transport. However, computational demands make molecular dynamics impractical for screening purposes. Thus, several approximative methods have been recently developed to study interactions between a protein and ligand during the ligand transport process. Apart from identifying the best binding modes, these methods also provide information on the energetics of the transport and identify problematic regions limiting the ligand passage. These methods use approximations to simulate binding or unbinding events rapidly (calculation times from minutes to hours) and provide energy profiles that can be used to rank ligands or pathways. Here we provide a critical comparison of available methods, showcase their results on sample systems, discuss their practical applications in molecular biotechnologies and outline possible future developments.

1. Introduction

Enzymes are essential biomolecules catalysing chemical reactions and key components of molecular biotechnologies. A substrate molecule undergoes a chemical reaction and is converted into a new product molecule. To initiate the catalytic cycle, the substrate needs to pass into a microenvironment of the active site. The active site is often buried inside the protein, shielding it from the outer environment. Access pathways in enzymes have the shape of a tunnel connecting the outer environment with the active site and possessing one opening (Fig. 1). In contrast, proteins may contain also a channel enabling a ligand to pass through the entire macromolecule with openings on both sides, often connecting different cellular environments or several enzymes in a biochemical pathway (Gora et al., 2013). The shape and physicochemical properties of these access pathways influence the selectivity of the enzyme for particular ligands. Furthermore, residues lining these pathways can be targeted by protein engineering to design enzyme variants with modified substrate specificity, activity or stability (Brezovsky et al., 2016; Kaushik et al., 2018; Marques et al., 2016).

The classical approach for studying binding and unbinding processes is to perform enzyme kinetics experiments to measure the rates or residence times of the ligands (Schuetz et al., 2017). *In silico* methods can be used as a complementary approach for experimental studies. Molecular docking (Morris et al., 2009; Trott and Olson, 2010; Verdonk et al., 2003) aims to identify optimal binding modes of a ligand. These methods search for local minima of the binding free energy by

* Corresponding authors at: Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, Kamenice 5/A13, 625 00 Brno, Czech Republic.
E-mail addresses: jiri@chemi.muni.cz (J. Damborsky), 222755@mail.muni.cz (D. Bednar).

https://doi.org/10.1016/j.biotechadv.2022.108009
Received 30 April 2022; Received in revised form 12 June 2022; Accepted 16 June 2022
Available online 20 June 2022

perturbating the ligand conformation in the binding site and evaluating the binding energy by a scoring function. The use of these methods is essential for virtual screenings and drug design (Kitchen et al., 2004). Molecular docking provides information about the best binding mode, but it does not consider the transport processes. Therefore, molecular docking usually does not provide information on transport kinetics or energy barriers related to ligand passage through access pathways.

Molecular dynamics (MD) simulations are state-of-the-art methods to analyse the motion of protein systems in time and their interactions with ligands (Gelpi et al., 2015; Haliloglu et al., 1997). MD simulations can be used to study changes in the protein conformation or the binding and unbinding of ligands (Kokkonen et al., 2018; Marques et al., 2017). Unfortunately, ligand binding and unbinding are often beyond the time limits of classical MD simulations. Therefore, many enhanced sampling methods have been developed to sample a larger conformational space (Rydzewski and Nowak, 2017). These methods either implement an external force or apply different strategies to sample rare events during the simulation.

The setup, execution, and assessment of MD simulations require a broad knowledge of molecular modelling. To facilitate specific modelling scenarios, tools employing MD, sometimes in the form of user-friendly webservers, have been developed (Bruce et al., 2019; Kingsley and Lill, 2014; Stank et al., 2017; Yang et al., 2019). Even with the steady improvement of computational power in high-performance computing, methods based on MD are not feasible for large-scale screening studies. Therefore, prediction tools applying approximations to describe the process of ligand binding or unbinding have been developed. This Review provides a detailed and critical comparison of currently available methods, highlights the advantages and limitations of individual algorithms and their applicability for mechanistic studies and discusses possible future developments in this research field.

2. Approximative methods for simulation of ligand transport and their validation

Five different methods suitable for the study of ligand binding and unbinding are described in this section. Provided is a detailed description of each method and the extent of their validation by their respective authors. Validation of a theoretical method against experimental data is a critical development step that must be undertaken before applying theoretical methods to any new system. Thorough and quantitative validation with multiple protein targets indicates a trustable and generally applicable tool.

2.1. SLITHER

SLITHER (Lee et al., 2009) was developed to predict the passage of ligands through channels in proteins. This method applies iterative docking in a predefined grid box set up around the channel of interest. The docking can be carried out by either the Lamarckian genetic algorithm from AutoDock4 (Morris et al., 2009) or the evolutionary Gaussian algorithm implemented in MEDock (Chang et al., 2005). After the first round of docking has finished and the first energy minimum has been found, the method searches for additional minima close to the first docked conformation. The newly found conformations may overlap with the previously found ones or additional minima may be found in the grid box. All found conformations are clustered, and the tool reports the largest cluster. Owing to the original design of the method for studying channels, the tool cannot be used for the exploration of narrow and curved tunnels. The developers tested SLITHER with the human glucose transporter Glut1 and α-D-glucopyranose in the original publication (Lee et al., 2009). They were able to explore the potential conformations of the ligand in the channel and identify a steric bottleneck linked to a barrier in the energy profile.

2.2. MoMA-LigPath

This method (Devaurs et al., 2013) is not based on a docking algorithm but uses principles from robot motion planning. The underlying algorithm in MoMA-LigPath is the Manhattan-like variation of the rapidly-exploring random tree (ML-RRT) (Cortes et al., 2008). This method considers both ligand and protein sidechains as movable objects. The algorithm translates the conformation parameters of movable parts into so-called active and passive variables. The active variables are essential for the unbinding of a ligand, whereas residues defined as passive variables move when they block the movement of other movable atoms. During the simulation, the algorithm tries to find the possible unbinding pathways and analyses atoms that must be moved to allow

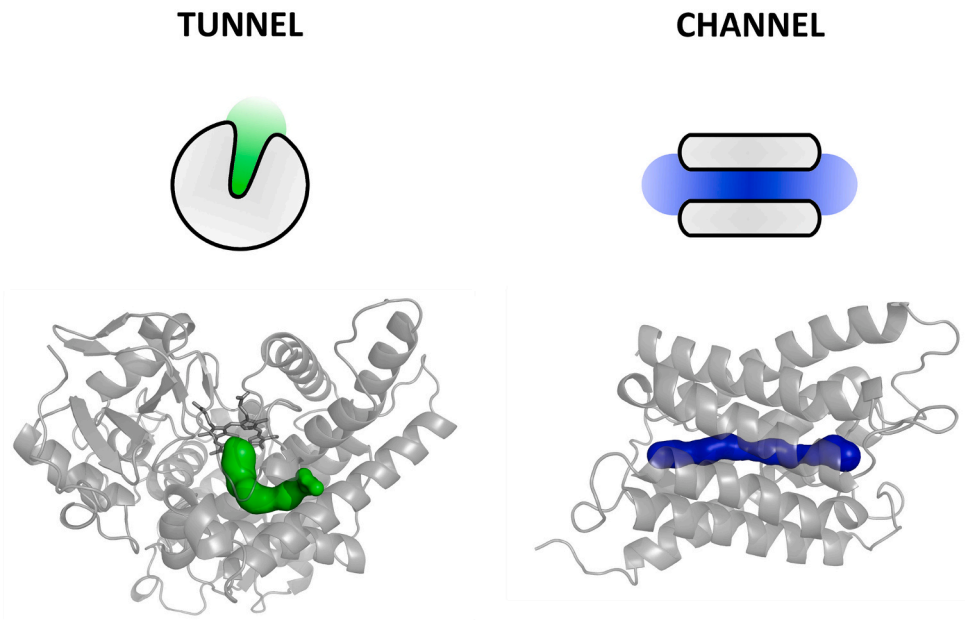

Fig. 1. Visualisation of an enzyme tunnel with one opening and a protein channel with two openings. Structure of cytochrome P450 2E1 (PDB ID 3LC4) with the buried active site and access tunnel (left) and aquaporin Z (PDB ID 1RC2) with a channel passing through the protein structure (right).

the unbinding. The algorithm changes the conformation of these atoms to help the ligand leave the protein and generate the ligand trajectory. Although the resulting trajectory is continuous, it only gives information about the possibility of a ligand fitting and passing through a certain pathway based on its size and geometry. This method does not consider the impact of non-covalent interactions and does not provide any information about the binding energy. The method was tested on a hexameric insulin complex with phenol in the original publication (Devaurs et al., 2013). The authors identified two unbinding pathways. When comparing the results with the random accelerated MD simulation, MoMA-LigPath could not find the third pathway that appeared in the simulation. The authors concluded that it would be necessary to consider the protein backbone flexibility to model the unbinding through this third pathway.

2.3. ART-RRT

ART-RRT (Nguyen et al., 2018a) applies the rapidly-exploring random tree (RRT) algorithm (LaValle, 1998) for motion planning to sample the conformational space and move the ligand through the pathway. The ligand atoms in the system are defined as active variables. In ART-RRT, the RRT part of the method samples only the active variables. In each step, a new ligand conformation further into the tunnel is obtained. The protein atoms are translated into passive variables sampled through the as-rigid-as-possible (ARAP) technique (Alexa et al., 2000). ARAP extends the implementation of RRT to reduce the dimensionality of the system and decrease the computational time. Each generated conformation of the system is minimised through the FIRE (Bitzek et al., 2006) method. This method uses minimisation constrained to the hyperplane orthogonal to the current expansion direction to relax the system but not revisit previous conformations. Lastly, ART-RRT performs a transition test similarly to Metropolis Monte Carlo methods (Binder, 1986) to accept or reject generated states. The ART-RRT method is part of the software package SAMSON (INRIA, 2017). During the calculation, the user can use the integrated GROMACS package (Abraham et al., 2015) to evaluate the system's energy in each step and score the predicted pathways. The method was subsequently enhanced (Nguyen et al., 2018b) by implementing a clash remover and adding a way to optimise the predicted pathways using the nudged elastic band method (Jónsson et al., 1998). The authors benchmarked the method with three protein-ligand complexes with available data based on biased MD simulations of ligand unbinding. They successfully reproduced pathways from previously published studies in all three cases. Furthermore, the authors compared their method with MoMA-LigPath and stated that their ART-RRT was up to 26-times faster.

2.4. CaverDock

CaverDock is a tool for studying ligand binding and unbinding through protein tunnels and channels (Filipovic et al., 2019; Vavra et al., 2019). The calculations are based on a predefined tunnel geometry calculated by CAVER 3.02 (Chovancova et al., 2012). The tunnel geometry can be acquired from a single crystal structure or models and also from ensembles of structures from MD simulations. If the user analyses multiple structures they must set up separate CaverDock calculations for each of the snapshots. The identified tunnel is discretised into a set of discs, which are used to guide the ligand through the protein during the simulation. It implements the docking engine from AutoDock Vina (Trott and Olson, 2010) for ligand minimisation and energy evaluation. The calculation comprises two parts. Firstly, the ligand is docked to each disc in the tunnel without any constraints to construct the lower-bound trajectory. The lower-bound represents the trajectory with the lowest energy but does not guarantee continuous movement. Secondly, the ligand is docked again, but any rapid changes in the conformation are disallowed between consecutive discs by another constraint. Furthermore, a specialised backtracking algorithm is applied to generate a smoothed, continuous upper-bound trajectory. We demonstrated the applicability of CaverDock with a large variety of proteins, tunnel geometries, and ligands. We showcased the tool's potential to reproduce docking results and complexes of bound ligands in crystal structures (Filipovic et al., 2019), find the tunnel preference, ligand specificity, analyse proteins with engineered tunnels, and make critical comparisons with all other published tools (Vavra et al., 2019). CaverDock showed high robustness and successfully completed jobs with >70 protein structures and 100 ligands.

2.5. GPathFinder

GPathFinder (Sánchez-Aparicio et al., 2019) was developed to extend the modular multi-objective package for molecular modelling Gaudi-MM (Rodríguez-Guerra Pedregal et al., 2017). It implements the NSGA-II multi-objective genetic algorithm (Deb et al., 2002). The GPathFinder method was developed as a new module for the Gaudi-MM tool. The workflow of the processes is separated into three modules: pathway generation, evaluation, and refinement. The algorithm produces possible trajectories connecting the predefined initial and final points in space during the calculation of pathways. The ligand trajectories can follow a straight line or be U, L, or S-shaped. In each step, the ligand is moved by a given distance from the starting point of the found pathway to its centroid (default 0.8 Å). The flexibility of the ligand is based on the free rotation of dihedral angles on single bonds. The protein offers global and local flexibility. For global flexibility, GPathFinder implements the normal mode analysis from the ProDy package, which models changes in the protein backbone caused by the position of the ligand. The local flexibility of sidechains is based on the Dunbrack or Dynameomics rotamer libraries (Scouras and Daggett, 2011; Shapovalov and Dunbrack, 2011). In the next step, the quality of generated frames is evaluated. The tool uses a scoring module, which scans the conformation of the protein-ligand complex for clashes. Afterwards, the ligand is minimised by the algorithm from AutoDock Vina (Trott and Olson, 2010). The scoring by Vina is also used to obtain the energy profile for the ligand passage. The resulting ligand trajectory is then refined by RRT-Connect (Kuffner and LaValle, 2000), which generates ligand positions in places where the ligand movement was not contiguous, such as around the narrowest parts (*i.e.* bottlenecks) in the pathways. The authors tested the method with a benchmarking set of 20 protein-ligand complexes. The data for testing were collected from previous computational and experimental analyses. GPathFinder found all the previously annotated pathways in this test and also new ones. The authors claimed that the novel pathways were biologically relevant based on the energy profiles and geometry.

3. Critical comparison of available approximative methods

This section critically compares four freely available tools for approximative simulations of ligand binding and unbinding: SLITHER, MoMA-LigPath, CaverDock and GPathFinder. Special attention is devoted to energy calculations and algorithms implemented to treat protein flexibility, which is currently the most challenging problem of approximative methods. Information about the input, output and other parameters of the compared tools is summarised in Table 1. Listed is also the envisaged interpretation of output data, which could assist with the identification and prioritization of transport pathways for specific ligands, study of ligand (substrate, product, inhibitor) passage through protein tunnels and channels during (un)binding processes, and identification of critical (bottleneck, gating) residues for site-directed mutagenesis of focused directed evolution experiments.

3.1. SLITHER

SLITHER is available as a webserver. The user must provide the protein and ligand structure in PDB or PDBQ files. The protein structure

Table 1
Comparison of requirements and parameters of approximative tools suitable for the analysis of ligand transport.

Software	SLITHER	MoMA-LigPath	CaverDock	GPathFinder
Input preparation	Receptor, ligand, tunnel aligned with the *y*-axis, grid box surrounding the tunnel	Ligand docked in the active site	Receptor and ligand, tunnel geometry, tunnel discretisation, grid box surrounding the tunnel, Vina configuration file	Complex with the ligand in the binding site, the configuration file
Input files	Protein PDB or PDBQ, ligand PDB or PDBQ	Complex PDB	Protein PDBQT, ligand PDBQT, discretised tunnel	Protein MOL2, ligand MOL2
Trajectory	Unbinding	Unbinding	Unbinding and binding	Unbinding and binding
Settings	SLITHER settings, docking algorithm, docking settings	Standard settings, advanced RRT algorithm settings	Vina docking parameters	Pathway settings, evaluation method
Additional options	Flexible receptor, relaxed receptor	Flexible receptor	Flexible receptor	Minimisation of receptor snapshots
Run time	Minutes	Minutes	Minutes–hours	Minutes–hours
Output data	Ligand trajectory in a single PDB file, energy information in REMARK	Ligand trajectory in multiple PDB files, missing information about energies	Ligand trajectory in single PDBQT file, energy information in REMARK, bottleneck residues, interacting residues	Ligand trajectory and receptor snapshots in multiple PDB files, score file with energies
Mechanistic interpretation	Identification and prioritization of transport pathways; analysis of ligand (un)binding process	Identification of potential transport pathways	Identification and prioritization of transport pathways; analysis of ligand (un)binding processes; identification of hot spots for mutagenesis	Identification and prioritization of transport pathways; analysis of ligand (un)binding processes

Individual tools provide different output data, which can be used for mechanistic enzymology, drug design, and protein design studies.

must be oriented so that the studied tunnel or channel is parallel to the *y*-axis of the coordinate system. The user can select either the AutoDock 4 or MEDock docking algorithm (Chang et al., 2005; Morris et al., 2009), the number of iterations for the docking scheme, and either set the grid box for the calculation manually or let SLITHER calculate it automatically. Furthermore, the user can select the flexible-receptor mode to specify residues with flexible sidechains during the computation or, through the relaxed-receptor mode, can upload a set of protein conformations. When the calculation has finished, the user can download the ligand trajectory file, which contains binding energy values for each ligand position.

3.2. MoMA-LigPath

MoMA-LigPath is available as a web server, and the source code is also available upon request. The user must prepare a PDB complex with the protein and bound ligand to run the calculations. Next, the number of calculated pathways must be specified and the simulation is started. In this tool, the user does not select any type of flexibility. The algorithm introduces flexibility to residues clashing with the ligand on the fly. Later, the user is presented with an archive containing PDB files with snapshots of the protein and ligand for each generated pathway. Unfortunately, this robotics-based algorithm does not apply any scoring of the ligand position. The user must re-score the complexes to obtain any information on energetics.

3.3. ART-RRT

ART-RRT tool is part of the commercial software platform SAMSON (INRIA, 2017). Since the software is not freely available to the scientific community, it is not included in this practical comparison of methods.

3.4. CaverDock

CaverDock is available as downloadable binaries for Linux, a singularity container image, and part of the webserver named Caver Web. A singularity container image is compatible with any operating system and does not require further compilation or installations. The developers of the tool also provide a Python API to improve the usability of CaverDock for screening calculations. To increase the method's reach in the community, CaverDock has also been implemented in the user-friendly Caver Web webserver (Stourac et al., 2019). Caver Web uses an interactive workflow to calculate and analyse tunnels in a protein of interest. Using the CaverDock modules, the user can prepare ligands, calculate identified tunnels, and evaluate the binding/unbinding in a single graphical user interface. When using the standalone version, the user must prepare a PDB file of the protein and a PDB or MOL2 file of the ligand. The standalone CaverDock software requires the installation of MGLtools to prepare PDBQT files. Using the scripts from MGLtools (Morris et al., 2009) the user can prepare PDBQT files similarly to AutoDock Vina docking. The user can set the residue sidechain flexibility during the preparation of the receptor or run CaverDock on individual protein snapshots to introduce protein dynamics. In the next step, the user needs to calculate tunnels in the protein using CAVER 3.02 (Chovancova et al., 2012) and discretise them to a set of discs using the provided script in the package. The user can set the distance between consecutive disks (resolution of the trajectory) and the direction of the ligand's pathway through the tunnel (in or out). Together with the grid box, a configuration file can be generated by another script in the package. The user can optionally use command line parameters to generate a file containing a list of bottleneck residues that have an unfavourable impact on the transport of the molecule and also an output file with a list of favourable interactions from each disk. These extra files can be used to simplify subsequent analyses of protein-ligand interactions. The results are provided in a PDBQT file containing the trajectory and energy values for each ligand snapshot. A provided script can convert it into a text file with a space-delimited file that can be used to generate plots of the energy profile of the (un)binding processes.

3.5. GPathFinder

GPathFinder is available for installation through the Conda package management system. This tool also requires the installation of the Chimera visualisation tool (Pettersen et al., 2004) because it uses some of its features. First, the user must prepare the protein-ligand complex, extract the ligand structure, and finally convert both into MOL2 files. In the next step, the user prepares the configuration file. Sample configuration files with different settings for the calculations are provided in the online documentation for the tool. The user can specify the binding or unbinding of the ligand, minimisation of the protein structures, the number of normal modes used for generating different protein conformations, the ligand step size, and whether to track clashes between the ligand and protein. The user can also select an algorithm to score the ligand snapshots in each trajectory part. The tool can automatically analyse a set number of possible pathways or use specific coordinates in a predefined direction specified by the user.

To test the individual tools practically, we analysed a test case of haloalkane dehalogenase DhaA (PDB ID 1BN7) with 1-chlorobutane. The purpose of this test was to critically compare individual tools and gain practical experience in setting the calculation parameters, the time

demands for job completion and interpretation of calculated results. We prepared the input files required for each software and ran the calculations to study the unbinding of the ligand from the active site of the enzyme. All four tools were used to analyse the unbinding process through the main P1 tunnel (Marques et al., 2019). The extracted ligand trajectories are shown in Fig. 2. Apparent differences in the step size and resolution of trajectories are seen in this visualisation. SLITHER generated the lowest number of ligand conformations. MoMA-LigPath was able to produce the most continuous trajectory. The implementation of the robotic algorithm in the tool allows the protein residues to move out of the way of the passing ligand molecule. Therefore, the ligand did not need to make larger conformational changes, which may not represent reality sufficiently. When we compared the resolution of generated trajectories for CaverDock and GPathFinder with default settings, CaverDock produced a smoother trajectory. However, both tools enable the step size to be decreased to obtain a better resolution with a higher computational time cost. The resulting ligand trajectories were not as smooth as in the case of the result from MoMA-LigPath. This is because the docking algorithm must fit the molecule in a certain position in the tunnel rather than moving the protein residues away from the ligand.

Apart from MoMa-LigPath, the tools also provided the binding energy based on the scoring implemented in the docking algorithms. SLITHER described only part of the studied pathway in the plot of energy profiles (Fig. 3). For the comparison, we utilised the scoring function from AutoDock Vina in the GPathFinder calculation, the same as implemented in CaverDock. We expected that the binding energy of the unbinding produced by GPathFinder would be very similar to the one obtained from CaverDock. However, higher energy barriers were observed for GPathFinder, which may have been due to the snapshots produced by normal mode analysis. While the crystal structure was sufficiently open for the ligand to pass through during CaverDock calculation, the pathway might have become narrower in the snapshots generated by GPathFinder.

4. Brief overview of applications of individual tools

This section provides an overview of the wide range of applications of individual tools to different problems in the fields of enzymology, protein engineering, metabolic engineering, and drug design. The list of individual studies is comprehensive, and readers are advised to study individual published articles for further details. Six selected case studies are discussed in depth in the following section, illustrating how approximative methods can facilitate the design of improved enzymes, provide mechanistic insights into enzyme enantioselectivity and assist in the development of novel drugs with high affinities and selectivities.

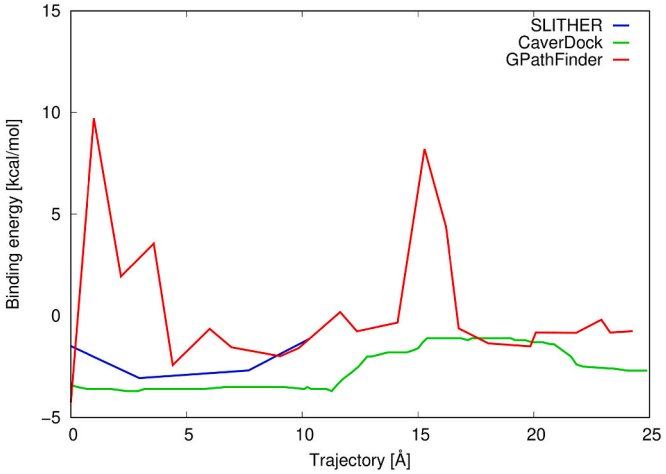

Fig. 3. Comparison of energy profiles calculated by individual software tools for the model enzyme haloalkane dehalogenase and substrate 1-chlorobutane. The energy profiles are provided for SLITHER (blue), CaverDock (green), and GPathFinder (red). MoMa-LigPath does not generate energy profiles, making it impossible to prioritize pathways and critically judge their biological relevance. The trajectory was measured as the distance between the central atom in 1-chlorobutane between consecutive snapshots. For the tested case, GPathFinder suggested a much more rugged energy profile compared to SLITHER and CaverDock. In general, it is recommended to compare different systems (wild type *versus* mutants; substrates *versus* products; various inhibitors) using one computational method rather than comparing the results from calculations using different methods. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.1. SLITHER

SLITHER has been used to find binding modes of three RNA aptamers in the light chain A of zinc-endopeptidase (Chang et al., 2010). The tool was used to identify the residues that bound to the aptamers. In another study, the authors used this method to study the potential egress pathway of small products in prolyl oligopeptidase (Kaushik and Sowdhamini, 2011). They used different snapshots from MD simulation to study how the energy profile of product passage changed based on the size of the β-propeller pore. SLITHER has also been applied to study the passage of nucleotides through an OmpF porin channel (Hadi-Alijanvand et al., 2014). It was used to generate potential binding modes and analyse interactions with the channel. Another study conducted screening of cytochrome EgP450 CYP71A with various herbicides (Phongdara et al., 2012) using SLITHER. The EgP450 enzyme was predicted to bind phenylurea type of herbicides with the highest affinity. The authors confirmed the computational results experimentally using

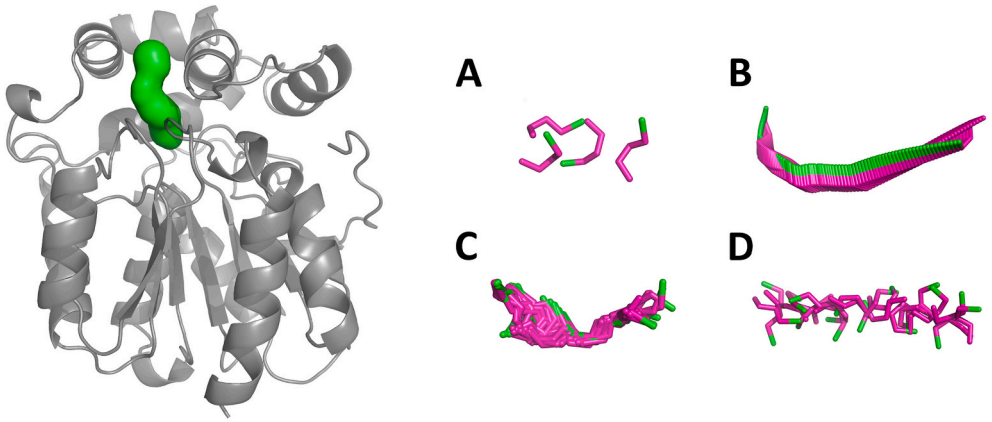

Fig. 2. Comparison of trajectories calculated by individual software tools obtained for the model enzyme haloalkane dehalogenase and substrate 1-chlorobutane. Structure of enzyme haloalkane dehalogenase DhaA (PDB ID 1BN7) with visualised P1 tunnel (green) and extracted ligand trajectories from the test run: (A) SLITHER, (B) MoMA-LigPath, (C) CaverDock, and (D) GPathFinder. Hydrogen atoms have been removed from the visualisation of ligand trajectories for clarity. Note the striking differences in the visualisation of substrate binding trajectories for methods A-D. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

herbicide tolerance tests.

4.2. MoMA-LigPath

This method has been used in a series of studies focusing on important cancer signalling targets, namely the PI3Kγ, AKT1/PKBα, Bcl-2, NF-κB, Stat3, and mTOR proteins (Jamal et al., 2014; Rehan, 2017, 2015; Rehan et al., 2014). These studies followed the same two-step scenario. Firstly, the inhibitor of interest was docked into the binding site and MoMA-LigPath was used to produce snapshots of the unbinding. Secondly, these snapshots were analysed in Ligplot+ (Laskowski and Swindells, 2011) to find residues that might interact with the inhibitors. A similar strategy was applied in studies of anti-diabetic drug sotagliflozin for the inhibition of sodium-glucose cotransporter-2 and acetylcholinesterase (Shakil, 2017), fluoxetine as an inhibitor of glutathione reductase (Dalmizrak et al., 2019), and the anti-cancer molecule niraparib as an inhibitor of acetylcholinesterase (Shakil, 2020). In these cases, the authors also used MoMA-LigPath to produce snapshots and searched for residues interacting with the inhibitor.

4.3. ART-RRT

ART-RRT has been applied in a complex study of aquaporin-3 (Yadav et al., 2020). The authors carried out molecular docking of >3 million small molecules. The best 20 hits were then explored by MD simulations, and the unbinding pathway of the best binder [1-(4-methoxyphenoxy)-3-((4-methoxyphenyl) amino) propan-2-ol] was analysed by the tool. The authors carried out 10 runs of the calculation with ART-RRT and observed that the molecule was unbound from the binding site through the extracellular part of the protein in all trajectories.

4.4. CaverDock

Apart from the various benchmarks used in the original publications (Filipovic et al., 2019; Vavra et al., 2019), we have continued using CaverDock in several other projects. We studied the tunnel preference of inhibitors in two important pharmacological targets: cytochrome P450 17A1 and leukotriene A4 hydrolase/aminopeptidase (Pinto et al., 2019). Moreover, we screened >100 ligands and >50 ligands, respectively, and defined the preferred tunnels based on transport barriers. In another study (Pinto et al., 2021), trajectories of >4000 drugs were screened in the tunnel of the S1 domain of the trimeric SARS-CoV-2 spike glycoprotein. We combined the calculation of energies with partial least squares analysis and identified the top binders as potential inhibitors with clearly defined quantitative structure-activity relationships. CaverDock has also been used in studies to identify the tunnel lining residues that interact with the ligand in acyltransferase (Contente et al., 2020) and rutinosidase (Brodsky et al., 2020).

4.5. GPathFinder

This method has been applied as a part of a screening pipeline to study potential inhibitors from *Uncaria tomentosa* on the SARS-CoV-2 main protease $3CL^{pro}$ (Yepes-Pérez et al., 2020). After virtual screening by classical molecular docking and following transport analysis with GPathFinder, the authors selected the top binding molecules: ciophylline, cadambine, and proanthocyanidin B2. The produced ligand-binding pathways showed that all three molecules had a good binding affinity and could access the S1 cleavage site of the protease. The obtained complexes were further studied by MD simulations and MM-GBSA (Genheden and Ryde, 2015) calculations.

5. Ligand design and protein design case studies

We made a comprehensive overview of articles published in established scientific journals, where protein tunnels were studied, and approximative methods were employed to simulate ligand (un)binding. Available methods have their strengths and limitations. Recently, we experienced rapid growth in their popularity thanks to the development of improved algorithms and user-friendly software tools. We have selected six case studies based on their quality, importance to the enzymology field and the journal impact. These studies consist of combined experimental and computational analyses focusing on using approximative methods for simulation of binding and/or unbinding, helping to design improved catalysts or explain mechanisms.

5.1. Developing enzymes with modified substrate specificities

5.1.1. A highly efficient variant of CYP153A$_{M.aq}$ with improved substrate anchoring

5.1.1.1. System description. This study focused on cytochrome P450 CYP153A ω-hydroxylase from *Marinobacter aquaeolei* (EC 1.14.15.3), which catalyses the following reaction: octane + 2 reduced rubredoxin + O_2 + $2H^+$ = 1-octanol + 2 oxidized rubredoxin + H_2O. Rapp et al. combined mutagenesis strategies to create a highly efficient three-point mutant variant of CYP153A$_{M.aq}$ (Rapp et al., 2021). Apart from experimental measurements, the influence of the mutated residues on the activity and specificity of the enzyme was investigated by multiple modelling approaches.

5.1.1.2. Protocol and findings. The simulations were conducted for the transport of octanoic acid into the active site through three distinct tunnels. It was concluded that the tunnel containing the mutated residues at the opening was most preferred by the substrate molecule. The simulated trajectory played a major role in explaining the molecular basis of the increased activity of the engineered enzymes. The observations supported the hypothesis that the increased activity towards octanoic acid was due to better stabilization in the access tunnel and active site (Fig. 4). While the mutation Q129R served as an improved anchor for keeping the substrate in the reactive position, the other two mutated residues, V141L and M228T, affected the flexibility of the BC-loop at the tunnel entrance and changed the shape of the tunnel.

5.1.1.3. Critical assessment. This study illustrates the application of approximative methods for an explanation of experimental observations. It is typical for site-directed mutagenesis or directed evolution to find exciting variants in protein engineering projects (Brezovsky et al., 2016; Kokkonen et al., 2018; Rapp et al., 2021). Protein engineers are always interested in a deeper understanding of how particular mutations contributed to improved catalysis. Approximative methods, mainly those available as web applications (Table 1), are straightforward to use by non-experts without the need for tedious and time-consuming installations. Calculated trajectories and energetic profiles provide intuitive information about (un)binding processes, which are essential parts and often rate-limiting steps of the catalytic cycle.

5.1.2. Broadening substrate specificity of threonine deaminase through mutagenesis of gating residues

5.1.2.1. System description. Song et al. have attempted to engineer an access tunnel in threonine deaminase from *Corynebacterium glutamicum* (EC 4.3.1.19) (Song et al., 2020). The enzyme catalyses the following deamination reaction: L-threonine = 2-oxobutanoate + NH_4^+. The authors combined computational analyses to identify the effect of the gate in the main tunnel, select residues for mutagenesis, and develop an improved enzyme variant. The novel range of substrate specificity was confirmed experimentally.

5.1.2.2. Protocol and findings. The study began with molecular docking of a selected set of natural and unnatural substrates to analyse their fit

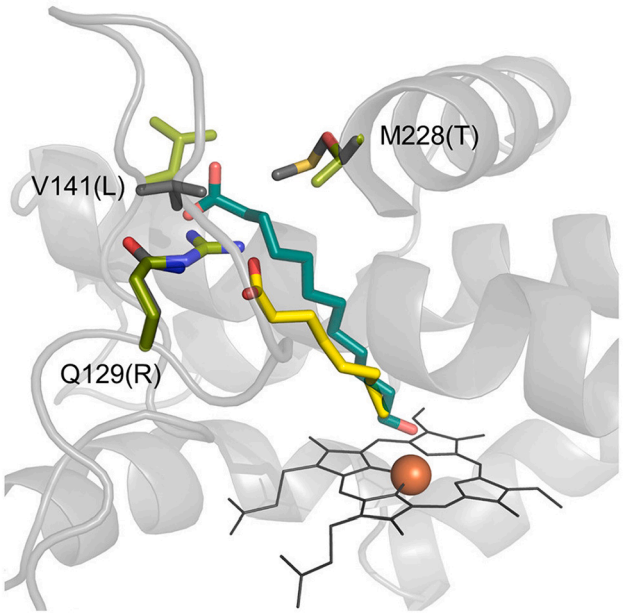

Fig. 4. Visualisation of a substrate molecule bound to the active site of $CYP153A_{M.aq}$. The crystal structure of the wild type (PDB ID 5FYG) is shown in grey with co-crystalized octanoic acid (dark green), the position of docked octanoic acid (yellow) and mutated residues (green). The heme is depicted as black lines and the iron atom as an orange sphere. The Q129R mutation stabilized the substrate, whereas the other mutations V141L and M228T reduced the flexibility of the loop at the tunnel entrance. The combined triple-point mutant exhibited a 151-fold increase in catalytic efficiency and improved substrate binding. The figure was reproduced from Rapp et al. (Rapp et al., 2021) https://pubs.acs.org/doi/10.1021/acscatal.0c05193, further permissions for readers related to the material excerpted should be directed to the ACS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

into the active site of the enzyme. From the analysis of the active site cavity and access tunnel by MD simulations, two flexible regions were identified that created a gate with open or closed conformations. Based on this knowledge, 47 nonconserved residues were selected from the gate, hinge, and anchoring regions. Several mutant libraries were designed and characterised, and the variant $CgTD_{Mu7}$ (V111A/V119N/K123S/V137I/K260S/R261T) was identified as the most efficient towards bulkier substrates. The variant carried the opened gate, which enabled bulky substrates to enter the protein. The introduced changes also improved substrate coupling, deprotonation, and imine hydrolysis. However, catalysis did not occur on the largest molecules in the set because the mutations in the best variant were only in the gate region and did not alter the shape and size of the active site. Therefore, these substrates were not able to bind properly. The improved variant could be used for the efficient production of natural and unnatural α-keto acids.

5.1.2.3. Critical assessment. This study demonstrates the power of the engineering strategy targeting the access tunnels rather than an active site. Mutagenesis focused on gating residues is particularly effective since these residues often play a crucial role in the physical steps of the catalytic cycle (Marques et al., 2017). Mutagenesis of access pathways is also generally safer than substitutions in the active site. Approximative methods could have been used to compare the transport of the ligand through the wild-type and engineered mutant. Although docking to the active site was analysed, the critical tunnel residues were identified only using MD simulations. Using an approximative method for (un)binding could have provided information about bottleneck residues in the tunnel and shortened the time needed for running MD simulations. Moreover, the speed of approximative methods (Table 1) allows for the testing of many potential designs *in silico* before constructing them in the laboratory.

5.2. Developing enzymes with novel catalytic functions

5.2.1. Re-programming of Fe/α-ketoglutarate-dependent hydroxylase into halogenase

5.2.1.1. System description. Papadopoulou et al. successfully converted L-proline *cis*-4-hydroxylase into a functional *cis*-3-halogenase (Papadopoulou et al., 2021). The original enzyme originates from *Sinorhizobium meliloti* (EC 1.14.11.56) and catalyses the reaction: 2-oxoglutarate + L-proline + O_2 = *cis*-4-hydroxy-L-proline + CO_2 + succinate. Mutagenesis was first carried out to create a variant with a novel function (Fig. 5). By combining experimental and computational approaches, it was possible to optimise the catalyst towards increased activity and a better ratio of created products, as well as explain the impact of the residue substitutions in the engineered enzyme.

5.2.1.2. Protocol and finding. Starting with the substitution of the active site residues in the coordination sphere of the Fe(II), D180G, the hydroxylase gained halogenating function with low reaction selectivity. Using molecular docking with L-proline, two possible binding poses were found: (i) hydroxylation binding mode, where the C4 of the substrate was close to the ferryl intermediate, and (ii) halogenation binding mode with the C3 atom being closer. Based on these observations, the study continued with site-directed mutagenesis of residues close to the docked substrate. This led to the best variant SmP4H-7 (V57L/S107T/D113E/T115P/R274H) with an 18.7-fold increase in chlorination. The docking was repeated with the model of the best variant and a slight decrease in the distance between the C3 and Fe atoms was found. The authors hypothesized that the increased chlorination activity might be due to changes in the hydrogen bond network. However, they missed the main reason behind the increase in activity, *i.e.* most of the mutated residues were close to the surface of the protein.

5.2.1.3. Critical assessment. The authors modelled the binding trajectory of the substrate. The approximative method enabled the identification of bottleneck residues, which either corresponded with the mutated residues (V57L/D113E/R274H) or were close to them. Furthermore, the studied tunnel in the best variant had a wider bottleneck radius, suggesting that the mutations facilitated the substrate transport. Papadopoulou et al. nicely combined theoretical and experimental approaches to achieve the functional switch in their target enzyme. Changes in protein function represent one of the most challenging tasks in the protein engineering field. It is advisable to combine approximative methods with other approaches, like structural analysis, molecular docking, molecular dynamics, quantum-chemical calculations, site-directed mutagenesis, and focused directed evolution, to collect complementary evidence for protein design as an interpretation of effective mutations (Kokkonen et al., 2018; Rapp et al., 2021).

5.3. Developing structural models for understanding stereochemistry

5.3.1. Studying stereoselective inhibitors of monoamine oxidase isoforms

5.3.1.1. System description. In this study, the authors analysed the inhibitory effect of selective 1-propargyl-4-styrylpiperidine-like analogues on human monoamine oxidase (MAO) isoforms A and B (EC 1.4.3.4). Both isoforms catalyse the same reaction: R-NH_2 + H_2O + O_2 = R-CHO + NH_3 + H_2O_2. Knez et al. studied both isoforms and the differences in binding and interaction of different inhibitor isomers (Knez et al., 2020).

5.3.1.2. Protocol and findings. In the first step, the authors compared

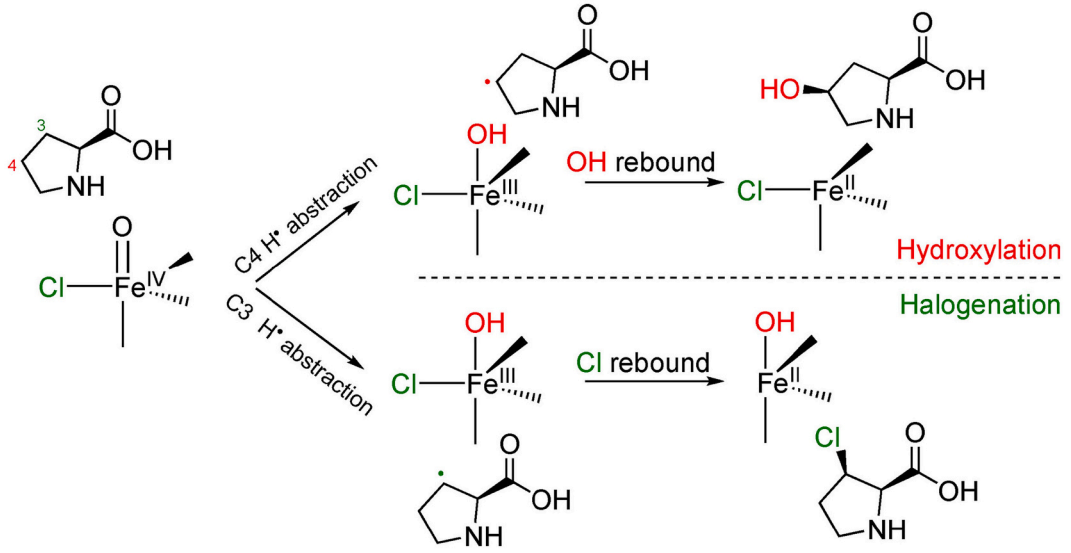

Fig. 5. Reaction scheme of the newly gained halogenation function of the Fe/α-ketoglutarate-dependent hydroxylase. After changing the active site residue chelating the ferryl, the enzyme acquired a new function. Depending on the orientation of the L-proline substrate upon binding, the carbon atom closest to the ferryl intermediate governed the direction of the catalytic reaction. If C4 was closer, hydroxylation of the substrate occurred, whereas if C3 was closer, the substrate underwent halogenation. The figure was reproduced from Papadopoulou et al. (Papadopoulou et al., 2021) © 2021 The Authors and ChemCatChem published by Wiley-VCH GmbH.

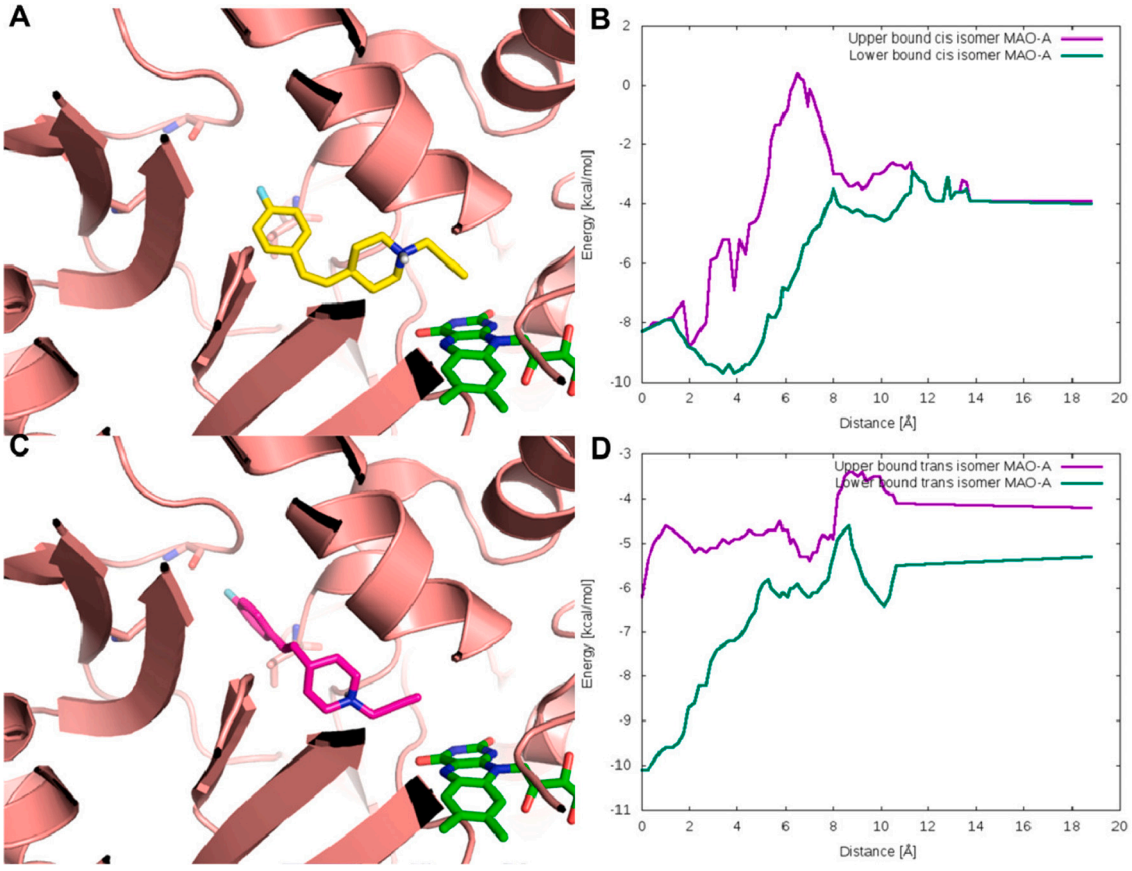

Fig. 6. Visualisation of the binding of selective isomers in human MAO-A. (A) *Cis* isomer (yellow sticks) with propargyl group facing the FAD cofactor (green sticks). (B) Energy profile for binding of the *cis* isomer. (C) *Trans* isomer (violet sticks) with propargyl group facing the FAD cofactor (green sticks). (D) Energy profile for binding of the *trans* isomer. In MAO-A, the *cis* isomer had lower binding energy in the visualised correct orientation. Analysis of the binding modes of the *trans* isomer showed the position with the lowest binding energy in the wrong orientation, suggesting that this isomer binds weakly because it is not able to form a covalent bond with the FAD cofactor in MAO-A. The figure was reproduced from Knez et al. (Knez et al., 2020) https://pubs.acs.org/doi/abs/10.1021/acs.jmedchem.9b01886, further permissions for readers related to the material excerpted should be directed to the ACS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the shape and size of the active sites of both isoforms. In both cases, the active sites are hydrophobic cavities, but the isoforms differ in the shape determined by gating residues. In MAO-A, the active site cavity is monopartite and spherical, whereas in MAO-B, it is flat and elongated. In an *in-house* screening campaign, the authors identified an inhibitor with a stilbene-like motif. The *trans* isomer of this compound selectively inhibited MAO-B, while the *cis* isomer inhibited MAO-A. The binding trajectories of both isomers were simulated in both MAO isoforms. The method was able to describe the stereoselectivity through binding energies of the optimal binding modes and differences in the passage through the tunnels. This structural and energetic analysis provided the following insights into the mechanism of inhibition. In agreement with experimental data, the tunnel diameter and energy barriers of the trajectory showed that the *cis* isomer inhibited MAO-A due to better binding energy. In MAO-A, the trans isomer had the lowest binding energy in an incorrect orientation, which prevented the formation of a covalent bond with the FAD cofactor (Fig. 6). In MAO-B, the *cis* isomer was too large to enter the narrower access tunnel and only the *trans* isomer was able to enter the active site and inhibit MAO-B with good binding affinity. Furthermore, the modelled conformation of the *trans* ligand MAO-B was in agreement with the position in an available crystalized complex. Based on information from the analysis, the authors studied the structure-activity relationships of both isomers and designed a new library of piperidine inhibitors with substituents on the phenyl ring. Using experimental analyses, they found even more potent inhibitor analogues for both MAO isoforms. The effects of the newly designed compounds were successfully tested *in vitro*, *ex vivo,* and behavioural *in vivo* experiments.

5.3.1.3. Critical assessment. This study provides an excellent example of applying approximative methods for understanding the differential binding of two stereoisomers. Enantiodiscrimination of substrates is one of the hallmarks of enzymatic catalysis. Differential binding of two enantiomers to the active site is typically considered. At the same time, the process of substrate entry and product egress is often neglected, even though these physical steps of the catalytic cycle can contribute to enantioselectivity. The study by Knez et al. could be easily extended by analysing products, providing a complete picture of the catalytic cycle. Moreover, approximative methods are sufficiently fast (Table 1) to allow virtual screening (Pinto et al., 2019), which can provide valuable data for quantitative structure-activity relationships (Pinto et al., 2021).

5.4. Studying the impact of tunnel anatomy on catalytic activity

5.4.1. Influence of enzyme conformation on the performance of lipase-powered nanomotors

5.4.1.1. System description. Wang et al. studied the effect of different types of immobilisation on the conformation of lipase on mesoporous silica nanoparticles to create improved chemically powered nanomotors (Wang et al., 2020). The selected lipase (EC 3.1.1.3) originates from *Candida rugosa* and catalyses the reaction: triacylglycerol + H_2O = diacylglycerol + fatty acid + H^+. The authors investigated the types of interactions for immobilisation and how the conformation of the immobilised lipase affects the dynamics of the protein and access of the substrate into the active site.

5.4.1.2. Protocol and finding. The studied lipase has a mobile domain (lid) located over the active site which governs the access of substrates. The authors analysed how different types of immobilisation (i) ionic adsorption, (ii) covalent bonding, and (iii) hydrophobic interaction, affected the orientation of the lipase on the nanoparticle (Fig. 7). Different immobilisations altered the range of motion of the mobile domain, and therefore the open and closed conformation of the enzyme. Experimental analyses showed that the most active conformation of the lipase was achieved with hydrophobic immobilisation. The open and closed conformations of the enzyme were studied to identify further details about this mechanism. MD simulations were performed with both types of conformation and revealed the presence of two distinct tunnels, T1 and T2, whose geometry was affected by the conformation of the mobile domain. The binding of the substrate triacetin was simulated through the two tunnels in both enzyme conformations. The energy profiles showed that the first tunnel was preferred in both cases. Furthermore, the profiles for both tunnels showed higher energy barriers due to bottleneck residues V86 and F87, which formed part of the lid domain and hindered the substrate access in the closed conformation.

5.4.1.3. Critical assessment. This paper describes a powerful application of approximative methods to identify preferred (un)binding routes for specific ligands. It is widespread for geometric algorithms to identify more than one access tunnel (Gora et al., 2013). Often, it is not straightforward to propose which tunnels are biochemically relevant and used for the ligands passage (Brezovsky et al., 2016). Tracking the water molecules in crystal structures can provide helpful clues. The

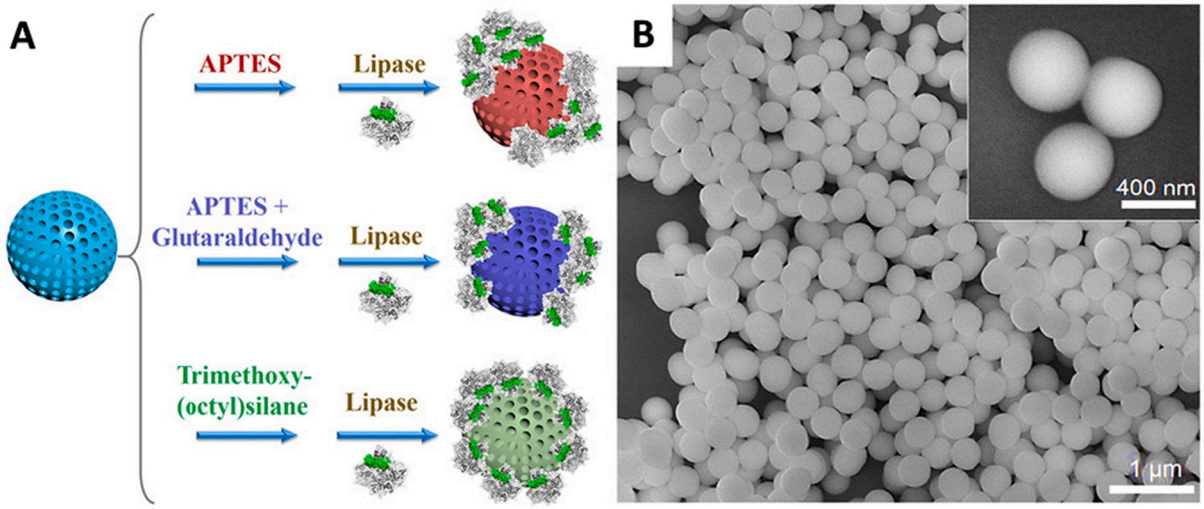

Fig. 7. Lipase-powered nanomotors. (A) Schematic of the synthesis of nanoparticles and subsequent immobilisation of lipase using three different techniques: (i) ionic adsorption; 3-aminopropyltriethoxysilane (APTES), (ii) covalent bonding; APTES + glutaraldehyde (APTES/GLY), (iii) hydrophobic interaction; trimethoxy-(octyl)silane (OTES). (B) SEM images showing the nanoparticles. The figure was reproduced from Wang et al.(Wang et al., 2020) © 2020 Wiley-VCH GmbH.

situation can become even more complex when MD simulations are used to explore protein conformational space, leading to some tunnels being temporarily open/closed (Kingsley and Lill, 2014). We advise calculation of tunnels for multiple snapshots, identification of the gating residues and tunnel bottlenecks (narrowest points) and selection of representative structures to calculate trajectories and corresponding energy profiles. Further analysis can follow the same protocol used in this study, *i.e.*, comparing energy profiles to decide on preferred routes for specific ligands (Wang et al., 2020).

5.5. Developing inhibitors for enzymes in a metabolic pathway

5.5.1. Analysis of the effect of antifungal essential oil from Zingiber officinale on the aflatoxin B_1 production cascade in Aspergillus flavus

5.5.1.1. System description. This study examined the impact of verbenol, the main constituent of an antifungal essential oil, on an enzymatic pathway producing aflatoxin B_1 (Singh et al., 2021). Based on earlier studies with *Aspergillus flavus*, the authors selected three enzymes: *Nor-1* norsolorinic acid ketoreductase (EC 1.1.1.349), which catalyses the reaction (1′*S*)-averantin + $NADP^+$ = H^+ + NADPH + norsolorinic acid; *Omt-1* sterigmatocystin 8-*O*-methyltransferase (EC 2.1.1.110), which catalyses the reaction S-adenosyl-L-methionine + sterigmatocystin = 8-*O*-methylsterigmatocystin + H^+ + S-adenosyl-L-homocysteine; and *Vbs* versicolorin B synthase, which catalyses the reaction (2*S*—3*S*)-versiconal hemiacetal = H_2O + versicolorin B. The effect of the essential oil and verbenol on fungal cells was studied experimentally together with a detailed computational study of the verbenol binding and mode of action.

5.5.1.2. Protocol and findings. The authors selected the essential oil from *Zingiber officinale* based on previous reports of antifungal activity. They characterised the content of the oil with GC–MS and found out that the major component in the oil was verbenol. The antifungal effect of the oil was validated on *Aspergillus flavus* cells. Treating the cells with the oil caused damage to their membrane integrity mainly due to the blocking of the carbohydrate catabolism pathway and reduction in progesterol production. This further hindered the function of mitochondria cells. The study continued by identifying the molecular effect of verbenol on the production of aflatoxin B_1. The authors selected three enzymes from the toxin production pathway and carried out a series of computational analyses. Starting with docking, verbenol was found to successfully bind to the active site of all three enzymes. The binding modes were validated with MD simulations. Analysis of protein dynamics showed that in the case of *Nor-1* and *Omt-1*, the verbenol molecule stabilised the complex through increased compactness and lower fluctuations in motions. These effects were not seen in the case of *Vbs*, which remained more relaxed. Furthermore, simulations showed that verbenol bound deeply in the interior of *Nor-1* and *Omt-1*, whereas in *Vbs,* the molecule travelled closer to the surface. Without clear evidence of the proposed mode of action of verbenol, the authors simulated the binding of verbenol into all three enzymes with an approximative method. The analysis of binding energies showed that verbenol binds best in the active site and that it targets this place rather than blocking the access tunnel.

5.5.1.3. Critical assessment. This is a well-designed and carefully conducted study, which provided valuable mechanistic information. A potential extension of the study would be the analysis of multiple snapshots from MD simulations. Fast approximative methods are well suited for this purpose. Such calculations are significantly more informative than those conducted with individual crystal structures since they consider protein structure fluctuations (Brezovsky et al., 2016; Bruce et al., 2019; Kokkonen et al., 2018; Rydzewski and Nowak, 2017; Wang et al., 2020), which may assist or hinder a passage of ligands *via* tunnels. Multiple calculations allow for statistical analysis and lead to more robust data.

6. Conclusions

Investigating ligand access pathways in enzymes is an essential component of biocatalysis, molecular enzymology, protein engineering, and drug design. The use of approximative methods for the simulation of ligand binding and unbinding allows users from a broad scientific community to study the mechanisms of these processes in molecular detail. State-of-the-art methods based on MD simulations, such as steered MD, random accelerated MD, and umbrella sampling (Kosztin et al., 1999; Lüdemann et al., 2000; Torrie and Valleau, 1977), use a high level of computational theory to analyse protein-ligand transport. These methods require extensive expertise in molecular modelling. In contrast, the methods and software tools summarised and critically compared in this review are faster, easier to use and can be applied to a broad range of applications.

The tools covered by this Review use several approximations that decrease the level of detail of the studied systems. Probably the most significant limitation is the inability to fully capture protein dynamics, although some tools provide solutions by introducing flexibility to backbone atoms. Current implementations address small backbone motions but will not work for large-scale conformational changes during substrate binding or product release (Gora et al., 2013). One solution to this problem is to run an MD simulation with the free protein, select several representative snapshots, and perform simulations with ligands using one of these approximative methods on such an ensemble of structures. However, the user must consider a reasonable balance between speed and a sufficiently realistic approximation of protein motions. Hybrid approaches combining several different computational techniques are an attractive prospect for future research, *e.g.* combination of robotic algorithms (efficiently exploring a conformational space) with molecular docking and force-field calculations (providing information on energy barriers).

The computational speed of approximative methods makes them considerably promising for large-scale screenings of ligands or enzyme variants. The increasing number of solved experimental structures, mainly due to advances in protein crystallography, NMR spectroscopy, and CryoEM (Callaway, 2015), as well as recent breakthroughs in the accurate structure prediction using AlphaFold2 (Jumper et al., 2021), provide a plethora of highly attractive protein targets. Large-scale whole proteome screening campaigns can be used to analyse crosstalk among individual proteins and small molecules inside living cells. We expect that further increases in the computational power of classical computers and quantum computers in the future will stimulate further exploration of these methods in the scientific and medical community.

What are the open challenges that future tools need to solve? Based on the literature search summarised in this article, there is an interest in using these tools for analysing ligand transport in target proteins and identification of critical interacting residues for experimental mutagenesis. We envisage the further development of specialised databases containing information about protein-ligand pairs, access pathways, protein-ligand interactions, and transport events. Development of such databases is possible owing to the speed and robustness of available algorithms which can be integrated into complex workflows. Automated workflows are necessary to keep the established databases updated. Large databases of protein structures, both experimental and those predicted by AlphFold2, provide an excellent starting point for comprehensive whole proteome analyses. There is still much to learn about substrate binding and molecular recognition, positioning of the transition state, and product release, as well as the critical role of biologically relevant conformational changes, involvement of water molecules, and evolution of structural features assisting ligand transport. The software tools discussed in this Review provide a unique opportunity for engineering and the discovery of significantly improved functionalities.

Funding information

The authors thank the Czech Ministry of Education (grant INBIO - CZ.02.1.01/0.0/0.0/16_026/0008451, RECETOX RI - LM2018121, ELIXIR - LM2018131, e-INFRA - LM2018140), the European Commission (grant nos. TEAMING 857560 and SINFONIA 814418), the Technology Agency of the Czech Republic (grant Permed TN01000013), and the Grant Agency of the Czech Republic (grant 20-15915Y) for financial support. OV is the recipient of a Ph.D. Talent award provided by Brno City Municipality.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E., 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

Alexa, M., Cohen-Or, D., Levin, D., 2000. As-rigid-as-possible shape interpolation. In: SIGGRAPH 2000 Conference Proceedings, Computer Graphics, pp. 157–164.

Binder, K., 1986. Introduction: Theory and "Technical" aspects of Monte Carlo simulations. In: Binder, K. (Ed.), Monte Carlo Methods in Statistical Physics. Topics in Current Physics, vol 7. Springer, pp. 1–45. https://doi.org/10.1007/978-3-642-82803-4_1.

Bitzek, E., Koskinen, P., Gähler, F., Moseler, M., Gumbsch, P., 2006. Structural relaxation made simple. Phys. Rev. Lett. 97, 170201 https://doi.org/10.1103/PhysRevLett.97.170201.

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., 2016. Engineering a de novo transport tunnel. ACS Catal. 6, 7597–7610. https://doi.org/10.1021/acscatal.6b02081.

Brodsky, K., Kutý, M., Pelantová, H., Cvačka, J., Rebroš, M., Kotik, M., Kutá Smatanová, I., Křen, V., Bojarová, P., 2020. Dual substrate specificity of the rutinosidase from aspergillus Niger and the role of its substrate tunnel. Int. J. Mol. Sci. 21, 5671. https://doi.org/10.3390/ijms21165671.

Bruce, N.J., Ganotra, G.K., Richter, S., Wade, R.C., 2019. KBbox: a toolbox of computational methods for studying the kinetics of molecular binding. J. Chem. Inf. Model. 59, 3630–3634. https://doi.org/10.1021/acs.jcim.9b00485.

Callaway, E., 2015. The revolution will not be crystallized: a new method sweeps through structural biology. Nature 525, 172–174. https://doi.org/10.1038/525172a.

Chang, D.T.-H., Oyang, Y.-J., Lin, J.-H., 2005. MEDock: a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm. Nucleic Acids Res. 33, W233–W238.

Chang, T.W., Blank, M., Janardhanan, P., Singh, B.R., Mello, C., Blind, M., Cai, S., 2010. In vitro selection of RNA aptamers that inhibit the activity of type A botulinum neurotoxin. Biochem. Biophys. Res. Commun. 396, 854–860. https://doi.org/10.1016/j.bbrc.2010.05.006.

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., 2012. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. PLoS Comput. Biol. 8, e1002708 https://doi.org/10.1371/journal.pcbi.1002708.

Contente, M.L., Roura Padrosa, D., Molinari, F., Paradisi, F., 2020. A strategic Ser/Cys exchange in the catalytic triad unlocks an acyltransferase-mediated synthesis of thioesters and tertiary amides. Nat. Catal. 3, 1020–1026. https://doi.org/10.1038/s41929-020-00539-0.

Cortes, J., Jaillet, L., Simeon, T., 2008. Disassembly path planning for complex articulated objects. IEEE Trans. Robot. 24, 475–481. https://doi.org/10.1109/TRO.2008.915464.

Dalmizrak, O., Teralı, K., Asuquo, E.B., Ogus, I.H., Ozer, N., 2019. The relevance of glutathione reductase inhibition by fluoxetine to human health and disease: insights derived from a combined kinetic and docking study. Protein J. 38, 515–524. https://doi.org/10.1007/s10930-019-09834-7.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. 6, 182–197. https://doi.org/10.1109/4235.996017.

Devaurs, D., Bouard, L., Vaisset, M., Zanon, C., Al-Bluwi, I., Iehl, R., Simeon, T., Cortes, J., Siméon, T., Cortés, J., 2013. MoMA-LigPath: a web server to simulate protein-ligand unbinding. Nucleic Acids Res. 41, W297–W302. https://doi.org/10.1093/nar/gkt380.

Filipovic, J., Vavra, O., Plhak, J., Bednar, D., Marques, S.M., Brezovsky, J., Matyska, L., Damborsky, J., 2019. CaverDock: a novel method for the fast analysis of ligand transport. IEEE/ACM Trans. Comput. Biol. Bioinforma. 1–1 https://doi.org/10.1109/TCBB.2019.2907492.

Gelpi, J., Hospital, A., Goñi, R., Orozco, M., 2015. Molecular dynamics simulations: advances and applications. Adv. Appl. Bioinforma. Chem. 37 https://doi.org/10.2147/AABC.S70333.

Genheden, S., Ryde, U., 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin. Drug Discovery 10, 449–461. https://doi.org/10.1517/17460441.2015.1032936.

Gora, A., Brezovsky, J., Damborsky, J., 2013. Gates of enzymes. Chem. Rev. 113, 5871–5923. https://doi.org/10.1021/cr300384w.

Hadi-Alijanvand, S., Mobasheri, H., Hadi-Alijanvand, H., 2014. Application of OmpF nanochannel forming protein in polynucleotide sequence recognition. J. Mol. Recognit. 27, 575–587. https://doi.org/10.1002/jmr.2381.

Haliloglu, T., Bahar, I., Erman, B., 1997. Gaussian dynamics of folded proteins. Phys. Rev. Lett. 79, 3090–3093. https://doi.org/10.1103/PhysRevLett.79.3090.

INRIA, 2017. SAMSON: Software for Adaptive Modeling and Simulation of Nanosystems.

Jamal, M.S., Parveen, S., Beg, M.A., Suhail, M., Chaudhary, A.G.A., Damanhouri, G.A., Abuzenadah, A.M., Rehan, M., 2014. Anticancer compound plumbagin and its molecular targets: a structural insight into the inhibitory mechanisms using computational approaches. PLoS One 9, e87309. https://doi.org/10.1371/journal.pone.0087309.

Jónsson, H., Mills, G., Jacobsen, K.W., 1998. Nudged elastic band method for finding minimum energy paths of transitions. In: Classical and Quantum Dynamics in Condensed Phase Simulations. World Scientific, pp. 385–404. https://doi.org/10.1142/9789812839664_0016.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2.

Kaushik, S., Sowdhamini, R., 2011. Structural analysis of prolyl oligopeptidases using molecular docking and dynamics: insights into conformational changes and ligand binding. PLoS One 6, e26251. https://doi.org/10.1371/journal.pone.0026251.

Kaushik, S., Marques, S.M., Khirsariya, P., Paruch, K., Libichova, L., Brezovsky, J., Prokop, Z., Chaloupkova, R., Damborsky, J., 2018. Impact of the access tunnel engineering on catalysis is strictly ligand-specific. FEBS J. 285, 1456–1476. https://doi.org/10.1111/febs.14418.

Kingsley, L.J., Lill, M.A., 2014. Including ligand-induced protein flexibility into protein tunnel prediction. J. Comput. Chem. 35, 1748–1756. https://doi.org/10.1002/jcc.23680.

Kitchen, D.B., Decornez, H., Furr, J.R., Bajorath, J., 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat. Rev. Drug Discov. 3, 935–949. https://doi.org/10.1038/nrd1549.

Knez, D., Colettis, N., Iacovino, L.G., Sova, M., Pišlar, A., Konc, J., Lešnik, S., Higgs, J., Kamecki, F., Mangialavori, I., Dolšak, A., Žakelj, S., Trontelj, J., Kos, J., Binda, C., Marder, M., Gobec, S., 2020. Stereoselective activity of 1-propargyl-4-styrylpiperidine-like analogues that can discriminate between monoamine oxidase isoforms a and B. J. Med. Chem. 63, 1361–1387. https://doi.org/10.1021/acs.jmedchem.9b01886.

Kokkonen, P., Sykora, J., Prokop, Z., Ghose, A., Bednar, D., Amaro, M., Beerens, K., Bidmanova, S., Slanska, M., Brezovsky, J., Damborsky, J., Hof, M., 2018. Molecular gating of an engineered enzyme captured in real time. J. Am. Chem. Soc. 140, 17999–18008. https://doi.org/10.1021/jacs.8b09848.

Kosztin, D., Izrailev, S., Schulten, K., 1999. Unbinding of retinoic acid from its receptor studied by steered molecular dynamics. Biophys. J. 76, 188–197. https://doi.org/10.1016/S0006-3495(99)77188-2.

Kuffner, J.J., LaValle, S.M., 2000. RRT-connect: An efficient approach to single-query path planning. In: Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065). IEEE, pp. 995–1001. https://doi.org/10.1109/ROBOT.2000.844730.

Laskowski, R.A., Swindells, M.B., 2011. LigPlot+: multiple ligand–protein interaction diagrams for drug discovery. J. Chem. Inf. Model. 51, 2778–2786. https://doi.org/10.1021/ci200227u.

LaValle, S.M., 1998. Rapidly-Exploring Random Trees: A New Tool for Path Planning. Iowa State Univ.

Lee, P.-H., Kuo, K.-L.K.-L., Chu, P.-Y.P.-Y., Liu, E.M., Lin, J.-H.J.-H., 2009. SLITHER: a web server for generating contiguous conformations of substrate molecules entering into deep active sites of proteins or migrating through channels in membrane transporters. Nucleic Acids Res. 37, W559–W564. https://doi.org/10.1093/nar/gkp359.

Lüdemann, S.K., Lounnas, V., Wade, R.C., 2000. How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms 1 1Edited by J. Thornton. J. Mol. Biol. 303, 797–811. https://doi.org/10.1006/jmbi.2000.4154.

Marques, S.M., Daniel, L., Buryska, T., Prokop, Z., Brezovsky, J., Damborsky, J., 2016. Enzyme tunnels and gates as relevant targets in drug design. Med. Res. Rev. 37 (5), 1095–1139. https://doi.org/10.1002/med.21430.

Marques, S.M., Dunajova, Z., Prokop, Z., Chaloupkova, R., Brezovsky, J., Damborsky, J., 2017. Catalytic cycle of haloalkane dehalogenases toward unnatural substrates explored by computational modeling. J. Chem. Inf. Model. 57, 1970–1989. https://doi.org/10.1021/acs.jcim.7b00070.

Marques, S.M., Bednar, D., Damborsky, J., 2019. Computational study of protein-ligand unbinding for enzyme engineering. Front. Chem. 6 https://doi.org/10.3389/fchem.2018.00650.

Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., 2009. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J. Comput. Chem. 30, 2785–2791. https://doi.org/10.1002/jcc.21256.

Nguyen, M.K., Jaillet, L., Redon, S., 2018a. ART-RRT: as-rigid-as-possible exploration of ligand unbinding pathways. J. Comput. Chem. 39, 665–678. https://doi.org/10.1002/jcc.25132.

Nguyen, M.K., Jaillet, L., Redon, S., 2018b. Generating conformational transition paths with low potential-energy barriers for proteins. J. Comput. Aided Mol. Des. 32, 853–867. https://doi.org/10.1007/s10822-018-0137-7.

Papadopoulou, A., Meierhofer, J., Meyer, F., Hayashi, T., Schneider, S., Sager, E., Buller, R., 2021. Re-programming and optimization of a L -proline cis −4-hydroxylase for the cis −3-halogenation of its native substrate. ChemCatChem 13, 3914–3919. https://doi.org/10.1002/cctc.202100591.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084.

Phongdara, A., Nakkaew, A., Nualkaew, S., 2012. Isolation of the detoxification enzyme EgP450 from an oil palm EST library. Pharm. Biol. 50, 120–127. https://doi.org/10.3109/13880209.2011.631019.

Pinto, G.P., Vavra, O., Filipovic, J., Stourac, J., Bednar, D., Damborsky, J., 2019. Fast screening of inhibitor binding/unbinding using novel software tool CaverDock. Front. Chem. 7 https://doi.org/10.3389/fchem.2019.00709.

Pinto, G.P., Vavra, O., Marques, S.M., Filipovic, J., Bednar, D., Damborsky, J., 2021. Screening of world approved drugs against highly dynamical spike glycoprotein of SARS-CoV-2 using CaverDock and machine learning. Comput. Struct. Biotechnol. J. 19, 3187–3197. https://doi.org/10.1016/j.csbj.2021.05.043.

Rapp, L.R., Marques, S.M., Zukic, E., Rowlinson, B., Sharma, M., Grogan, G., Damborsky, J., Hauer, B., 2021. Substrate anchoring and flexibility reduction in CYP153A M.aq leads to highly improved efficiency toward octanoic acid. ACS Catal. 11, 3182–3189. https://doi.org/10.1021/acscatal.0c05193.

Rehan, M., 2015. A structural insight into the inhibitory mechanism of an orally active PI3K/mTOR dual inhibitor, PKI-179 using computational approaches. J. Mol. Graph. Model. 62, 226–234. https://doi.org/10.1016/j.jmgm.2015.10.005.

Rehan, M., 2017. An anti-Cancer drug candidate OSI-027 and its analog as inhibitors of mTOR: computational insights into the inhibitory mechanisms. J. Cell. Biochem. 118, 4558–4567. https://doi.org/10.1002/jcb.26117.

Rehan, M., Beg, M.A., Parveen, S., Damanhouri, G.A., Zaher, G.F., 2014. Computational insights into the inhibitory mechanism of human AKT1 by an orally active inhibitor, MK-2206. PLoS One 9, e109705. https://doi.org/10.1371/journal.pone.0109705.

Rodríguez-Guerra Pedregal, J., Sciortino, G., Guasp, J., Municoy, M., Maréchal, J.-D., 2017. GaudiMM: A modular multi-objective platform for molecular modeling. J. Comput. Chem. 38, 2118–2126. https://doi.org/10.1002/jcc.24847.

Rydzewski, J., Nowak, W., 2017. Ligand diffusion in proteins via enhanced sampling in molecular dynamics. Phys Life Rev 22–23, 58–74. https://doi.org/10.1016/j.plrev.2017.03.003.

Sánchez-Aparicio, J.E., Sciortino, G., Herrmannsdoerfer, D.V., Chueca, P.O., Pedregal, J. R.G., Maréchal, J.D., 2019. Gpathfinder: identification of ligand-binding pathways by a multi-objective genetic algorithm. Int. J. Mol. Sci. 20 https://doi.org/10.3390/ijms20133155.

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., 2017. Kinetics for drug discovery: an industry-driven effort to target drug residence time. Drug Discov. Today 22, 896–911. https://doi.org/10.1016/j.drudis.2017.02.002.

Scouras, A.D., Daggett, V., 2011. The dynameomics rotamer library: amino acid side chain conformations and dynamics from comprehensive molecular dynamics simulations in water. Protein Sci. 20, 341–352. https://doi.org/10.1002/pro.565.

Shakil, S., 2017. Molecular interaction of anti-diabetic drugs with acetylcholinesterase and sodium glucose co-transporter 2. J. Cell. Biochem. 118, 3855–3865. https://doi.org/10.1002/jcb.26036.

Shakil, S., 2020. Molecular interaction of anti-cancer ligands with human brain acetylcholinesterase. J. Biomol. Struct. Dyn. 1–10 https://doi.org/10.1080/07391102.2020.1837678.

Shapovalov, M.V., Dunbrack, R.L., 2011. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure 19, 844–858. https://doi.org/10.1016/j.str.2011.03.019.

Singh, P.P., Jaiswal, A.K., Kumar, A., Gupta, V., Prakash, B., 2021. Untangling the multi-regime molecular mechanism of verbenol-chemotype *Zingiber officinale* essential oil against Aspergillus flavus and aflatoxin B1. Sci. Rep. 11, 6832. https://doi.org/10.1038/s41598-021-86253-8.

Song, W., Xu, X., Gao, C., Zhang, Y., Wu, J., Liu, J., Chen, X., Luo, Q., Liu, L., 2020. Open gate of *Corynebacterium glutamicum* threonine deaminase for efficient synthesis of bulky α-keto acids. ACS Catal. 10, 9994–10004. https://doi.org/10.1021/acscatal.0c01672.

Stank, A., Kokh, D.B., Horn, M., Sizikova, E., Neil, R., Panecka, J., Richter, S., Wade, R.C., 2017. TRAPP webserver: predicting protein binding site flexibility and detecting transient binding pockets. Nucleic Acids Res. 45, W325–W330. https://doi.org/10.1093/nar/gkx277.

Stourac, J., Vavra, O., Kokkonen, P., Filipovic, J., Pinto, G., Brezovsky, J., Damborsky, J., Bednar, D., 2019. Caver web 1.0: identification of tunnels and channels in proteins and analysis of ligand transport. Nucleic Acids Res. 47, W414–W422. https://doi.org/10.1093/nar/gkz378.

Torrie, G.M., Valleau, J.P., 1977. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. J. Comput. Phys. 23, 187–199. https://doi.org/10.1016/0021-9991(77)90121-8.

Trott, O., Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461. https://doi.org/10.1002/jcc.21334.

Vavra, O., Filipovic, J., Plhak, J., Bednar, D., Marques, S.M., Brezovsky, J., Stourac, J., Matyska, L., Damborsky, J., 2019. CaverDock: a molecular docking-based tool to analyse ligand transport through protein tunnels and channels. Bioinformatics 35, 4986–4993. https://doi.org/10.1093/bioinformatics/btz386.

Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W., Taylor, R.D., 2003. Improved protein-ligand docking using GOLD. Proteins Struct. Funct. Bioinforma. 52, 609–623. https://doi.org/10.1002/prot.10465.

Wang, L., Marciello, M., Estévez-Gay, M., Soto Rodriguez, P.E.D., Luengo Morato, Y., Iglesias-Fernández, J., Huang, X., Osuna, S., Filice, M., Sánchez, S., 2020. Enzyme conformation influences the performance of lipase-powered nanomotors. Angew. Chem. Int. Ed. 59, 21080–21087. https://doi.org/10.1002/anie.202008339.

Yadav, D.K., Kumar, S., Choi, E.-H., Chaudhary, S., Kim, M.-H., 2020. Computational modeling on aquaporin-3 as skin Cancer target: a virtual screening study. Front. Chem. 8 https://doi.org/10.3389/fchem.2020.00250.

Yang, J.-F., Wang, F., Chen, Y.-Z., Hao, G.-F., Yang, G.-F., 2019. LARMD: integration of bioinformatic resources to profile ligand-driven protein dynamics with a case on the activation of estrogen receptor. Brief. Bioinform. 21 (6), 2206–2218. https://doi.org/10.1093/bib/bbz141.

Yepes-Pérez, A.F., Herrera-Calderon, O., Sánchez-Aparicio, J.-E., Tiessler-Sala, L., Maréchal, J.-D., Cardona-G, W., 2020. Investigating potential inhibitory effect of Uncaria tomentosa (Cat's claw) against the Main protease 3CLpro of SARS-CoV-2 by molecular modeling. Evidence-based complement. Altern. Med. 2020, 1–14. https://doi.org/10.1155/2020/4932572.