

# ChannelsDB 2.0: a comprehensive database of protein tunnels and pores in AlphaFold era

Anna Špačková <sup>1</sup>, Ondřej Vávra<sup>2,3</sup>, Tomáš Raček <sup>4,5</sup>, Václav Bazgier <sup>1</sup>, David Sehnal <sup>4,5</sup>, Jiří Damborský <sup>2,3</sup>, Radka Svobodová <sup>4,5,\*</sup>, David Bednář <sup>2,3,\*</sup> and Karel Berka <sup>1,\*</sup>

<sup>1</sup>Department of Physical Chemistry, Faculty of Science, Palacký University, tř. 17. listopadu 12, 771 46 Olomouc, Czech Republic

<sup>2</sup>Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic

<sup>3</sup>International Clinical Research Center, St. Anne's University Hospital Brno, Pekařská 53, 656 91 Brno, Czech Republic

<sup>4</sup>CEITEC – Central European Institute of Technology, Masaryk University Brno, Kamenice 5, 625 00 Brno, Czech Republic

<sup>5</sup>National Centre for Biomolecular Research, Faculty of Science, Masaryk University Brno, Kamenice 5, 625 00 Brno, Czech Republic

\*To whom correspondence should be addressed. Tel: +420 585 634 769; Fax: +420 585 634 761; Email: karel.berka@upol.cz

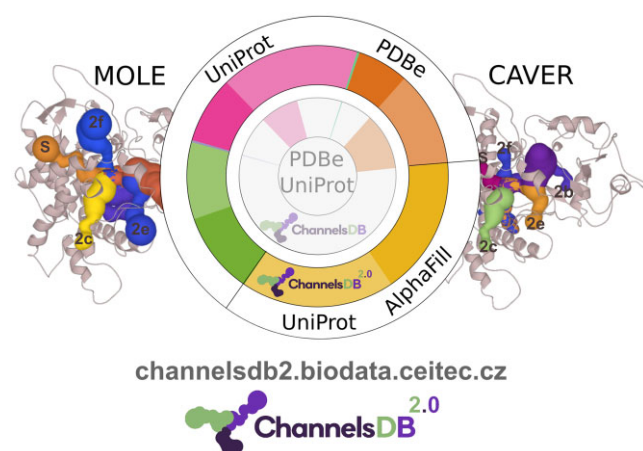
Correspondence may also be addressed to Radka Svobodová. Email: radka.svobodova@ceitec.muni.cz

Correspondence may also be addressed to David Bednář. Email: davidbednar1208@gmail.com

## Abstract

ChannelsDB 2.0 is an updated database providing structural information about the position, geometry and physicochemical properties of protein channels—tunnels and pores—within deposited biomacromolecular structures from PDB and AlphaFoldDB databases. The newly deposited information originated from several sources. Firstly, we included data calculated using a popular CAVER tool to complement the data obtained using original MOLE tool for detection and analysis of protein tunnels and pores. Secondly, we added tunnels starting from cofactors within the AlphaFill database to enlarge the scope of the database to protein models based on UniProt. This has enlarged available channel annotations ~4.6 times as of 1 September 2023. The database stores information about geometrical features, e.g. length and radius, and physico-chemical properties based on channel-lining amino acids. The stored data are interlinked with the available UniProt mutation annotation data. ChannelsDB 2.0 provides an excellent resource for deep analysis of the role of biomacromolecular tunnels and pores. The database is available free of charge: <https://channelsdb2.biodata.ceitec.cz>.

## Graphical abstract



## Introduction

There are over 200 000 structures present in the Protein Data Bank (PDB) (1,2). Still, while the growth of the number of experimental structures is substantial, it was dwarfed by the availability of high-quality structural models due to AlphaFold (3). Over 200 000 000 structural models were collected in AlphaFoldDB (4), thus enriching our view of the pro-

tein structural universe, including the channels within proteins important for the access routes to active sites of soluble proteins (tunnels) or transport through transmembrane proteins (pores).

To analyse the access routes important for ligand transport from/into protein binding sites, our teams have previously developed highly used tools MOLEonline (5,6), CAVER (7) and

Received: September 14, 2023. Revised: October 12, 2023. Editorial Decision: October 13, 2023. Accepted: October 23, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

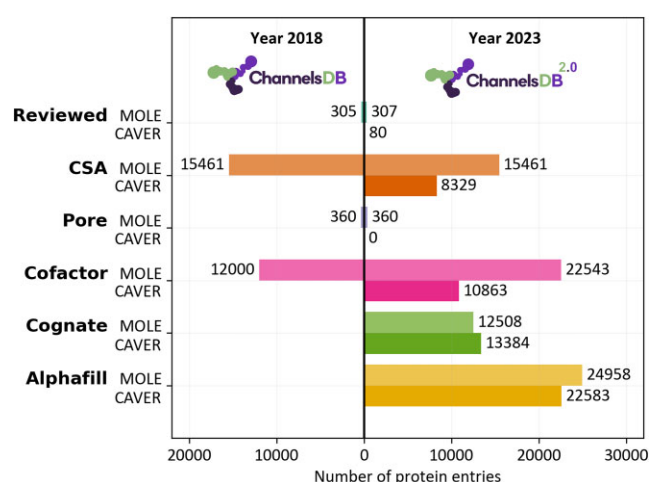
Caver Web (8). These tools have enabled the detection and analysis of channels, tunnels, and pores for over a decade, and their results have been collected in the publicly accessible database ChannelsDB (9). Since 2018, the database has grown from ~28 000 entries to ~40 000 experimental entries due to the assignments of channels for newer structures, the addition of results on cognate ligands based on recent studies (10,11), and complementing MOLE channels with CAVER ones to get consensus prediction. Moreover, the data in ChannelsDB were based only on a limited number of experimental structures. Identifying the start of the tunnel is nontrivial to fill the knowledge gap in model AlphaFold structures. Fortunately, the AlphaFill database (12) has been developed to fill the AlphaFold models with ligands and cofactors that can serve as a starting point for channel analysis. AlphaFill provides another ~25 000 model entries and is continuously growing.

ChannelsDB has been freely accessible since March 2017 at <https://channelsdb.ncbr.muni.cz/> and last year had close to 5000 users/year. This frequent usage of ChannelsDB is supported by its integration into a comprehensive PDBe-KB database (13). The usage of ChannelsDB is also reinforced by the fact that its source services (MOLE, MOLEonline, CAVER and Caver Web) are popular tools for detecting protein channels, tunnels and pores in protein structures.

## Database content

ChannelsDB 2.0 is a comprehensive database that combines data layers obtained from PDBe (2) and UniProt (14), along with predicted protein structures obtained based on deep learning and stored in a database AlphaFill (12). Experimental structures from the PDB database and model structures from AlphaFill database can be accessed through separate search windows, enabling the linkage to the corresponding stored tunnels, pores, and channels. UniProt provides essential details about the studied proteins, including residue annotations. ChannelsDB 2.0 utilises several diverse methods to obtain tunnels and pores (Figure 1):

- ~80 and ~300 proteins with **manually reviewed** tunnels from literature calculated by CAVER 3.02 (7) and MOLE 2.0 (15), respectively. See details of the calculation below.
- ~8300 and ~15 500 proteins with tunnels starting from catalytic sites annotated by **Catalytic Site Atlas** (16) calculated by CAVER and MOLE, respectively.
- ~350 proteins with **transmembrane pores** predicted by the pore mode of MOLEonline (5,6).
- ~10 800 and ~22 500 proteins with tunnels starting from **cofactors** calculated by CAVER and MOLE, respectively.
- ~13 400 and ~12 500 newly added proteins with tunnels containing **cognate ligands**, based on experimental evidence or functional annotations (10,11) calculated by CAVER and MOLE, respectively. More information is below in the next section.
- ~22 600 and ~25 000 newly added protein models with tunnels starting from cofactors transplanted into AlphaFold models via the **AlphaFill** (12) database, calculated by CAVER and MOLE, respectively. More information is below in the next sections.



**Figure 1.** The growth of the content of ChannelsDB from the original version released in 2018 (9) to the current day (this study). The colour coding indicates different data categories - light coloured are MOLE calculated entries, and CAVER is a darker colour. The updated database ChannelsDB 2.0 contains three times more data than the original release from 2018. The database has grown approximately ~4.6 times in size of channel annotations. Calculation and visualisation of tunnels and channels using MOLE 2.0 is newly complemented by analysis using CAVER 3.02, thus allowing utilisation of consensus tunnel annotation.

## Cognate dataset

The Cognate dataset utilises data collected from publications by Tyzack *et al.* and Vavra *et al.* (10,11) about enzyme-ligand complexes ranked by the similarity of the bound ligand with the cognate ligand, a putative natural substrate or product of the enzyme, from the KEGG database (17), calculated by the PARITY algorithm (11). We used all of the enzymes whose bound ligands had a similarity score above 0.6 and successfully calculated tunnels in their crystal structures, resulting in 13531 CAVER and 12508 MOLE entries.

## AlphaFill dataset

As of 21 August 2023, the AlphaFill database (12) contained 2325 transplanted compounds, with a cumulative total of 337 060 structures, when considering a 70% identity threshold. The same 58 distinct cofactors were selected for tunnel calculation as in the cofactor dataset, leading to 33 689 structures, out of which 16 257 (48%) or 24 958 (74%) structures contained at least one tunnel calculated with CAVER or MOLE, respectively.

## Tunnels calculated by MOLE 2.0

Except transmembrane pore dataset calculation, where MOLEonline (5,6) pore mode is used by first identifying the membrane boundaries and then calculating transmembrane pore from one boundary to another through the protein centre; standard MOLE 2.0 calculation parameters (15) were used, including BottleneckRadius 1.2 Å (the size of a water molecule), BottleneckTolerance 3 Å (allowing for the length of a bottleneck), MaxTunnelSimilarity 0.7 (to discard too similar tunnels), OriginRadius 4.0 Å (to optimise starting point around specified geometric centre of selection), and SurfaceCoverRadius 10 Å (to distinguish endpoints by distance of at least 10 Å). In the case of the Cognate dataset, the starting

point identified by the CAVER procedure below was used as an input for MOLE 2.0 calculation.

### Tunnels calculated by CAVER 3.02

In the first step of CAVER 3.02 calculations (7), the analysed structure is cleared off any molecules that are not cofactors based on the lists of cofactors from FPOCKET (18) and Co-Factor Database (19). The next essential step is calculating pockets by FPOCKET 2 (18) with parameters: -m 2.8 -n 10 -r 4.5 -s 2.5. Then, the starting points for tunnel calculation by CAVER 3.02 are defined similarly to the previously mentioned pipeline (10), with slight differences depending on the data source:

- (i) In the case of the Cognate dataset (see above), we select the pocket that contains the first location of the bound ligand molecule, find the nearest protein chain, and select the protein residue with the C $\alpha$  atom closest to the geometrical centre of the chain. Next, the geometrical centre of the pocket is calculated, and in the previously selected residue, the nearest side-chain atom to the pocket centre is found. After that, we save the coordinates of the point placed further into the space by 0.5 Å on the vector connecting the closest atom and the centre of the pocket. The coordinates of this point are saved as the starting point for the CAVER 3.02 tunnel calculation.
- (ii) For the data from ChannelsDB (reviewed, CSA and cofactor datasets), each starting point is found in the pocket containing the original MOLE starting point for tunnels in the database.
- (iii) The structures from AlphaFill are cleared from all molecules but the matched cofactor. In this case, we analyse the pockets that contain the reactive atoms from each cofactor position in the model structures.

The tunnels are then calculated with the same settings for all structures by CAVER 3.02 with a probe radius of 0.9 Å and remaining default settings.

The conversion algorithm processes protein tunnel data from the CAVER results for visualisation. The 'residues.txt' file is used to identify the residues around each tunnel, and these are stored as HetResidues in a JSON file for subsequent visualisation on the ChannelsDB web page. Next, the algorithm uses the 'summary.txt' file to obtain the starting points of the tunnels. It then divides each analysed tunnel into layers based on the surrounding residues. Calculating the five nearest atoms to the centre axis of the tunnel determines the residues for each layer, with a maximum of five residues per layer. The algorithm calculates the MinRadius for each layer, representing the minimum radius within that layer. It also calculates the start and end layer radius for the visualisation. Tunnel properties are determined by analysing the properties of lining amino acids and summing their relevant characteristics, normalised by the number of amino acids considered. Finally, all calculated data, including HetResidues, layer information, and tunnel properties, are written to a JSON file for display on the ChannelsDB 2.0 web page.

The ChannelsDB 2.0 homepage allows users to perform queries for protein structures from PDB or AlphaFill databases containing annotated channels. This flexible search functionality allows users to use any PDB-related metadata, such as PDB ID, protein name, protein family, cofactor, lig-

and, author or even the journal in which the research was published. The underlying search engine is powered by the PDB RESTful API, which efficiently and accurately retrieves relevant records. The ChannelsDB 2.0 website offers comprehensive documentation explaining the methodology used for channel annotation within the database and several examples of typical channel systems.

## Results and discussion

### Case study 1—tunnels in AlphaFill and X-ray structures of cytochrome P450 2D6

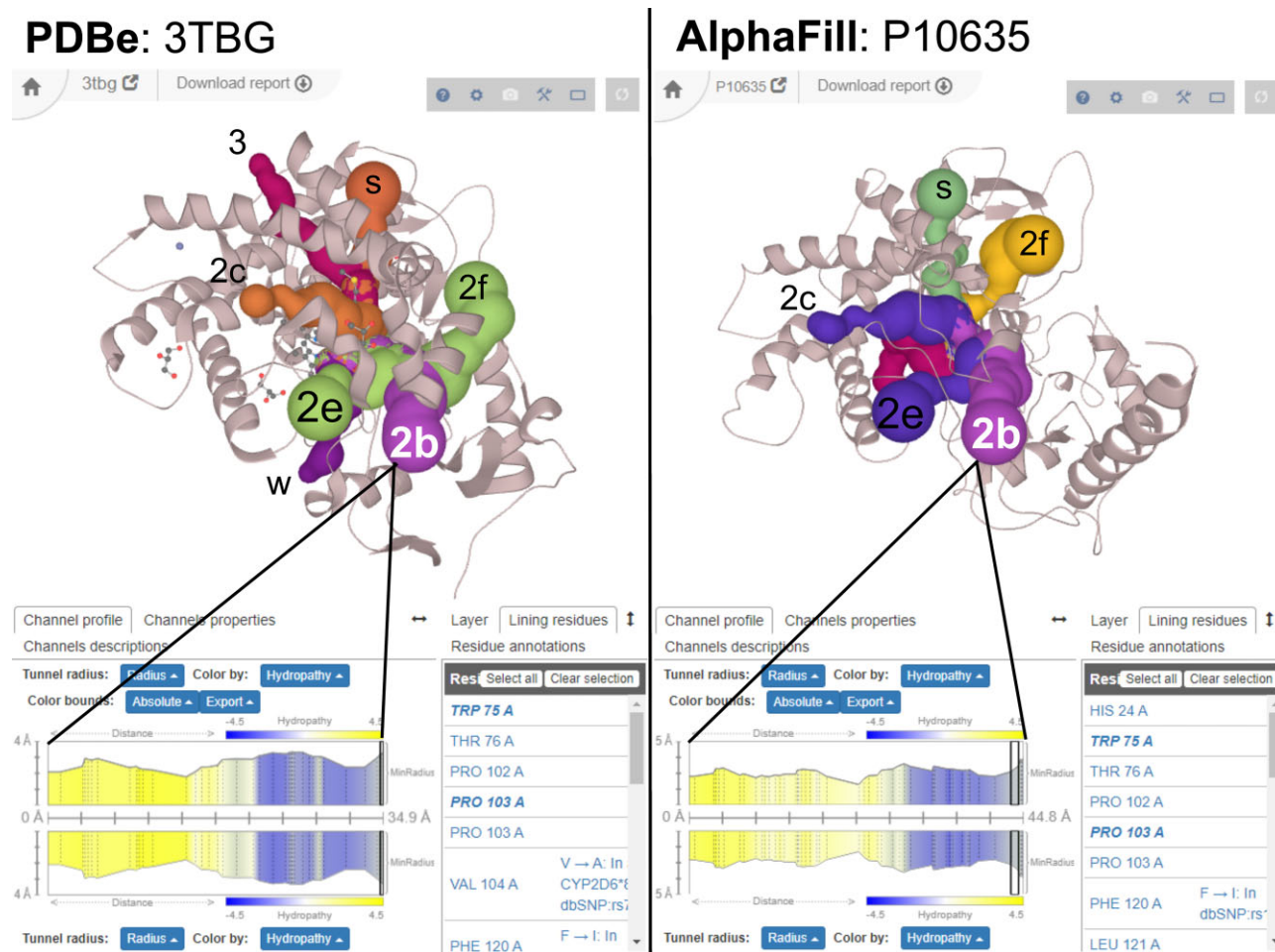
Cytochromes P450 (CYPs, EC 1.14.-.-) are a versatile and most typically monooxygenase enzyme family known in all kingdoms of life. In humans, these enzymes are involved in the metabolism of fatty acids, steroids or drug metabolism. One of the drug-metabolizing enzymes is CYP2D6, which is responsible for the metabolism and elimination of approximately 25% of commercially available drugs—especially those containing nitrogen (20). The active sites of oxidoreductases are typically buried (21); hence, CYPs have their heme-containing active sites accessible only via tunnels. Tunnels have nomenclature with respect to secondary structures (22). Membrane-bound channels are used for substrate entry, while the channels leading above the membrane are used for metabolite release (23). This fact is reflected by the properties of individual tunnels based on their lining residues (Figure 2).

CYP2D6 is thus a great, well-described system, which can be used to showcase the similarity of tunnels within AlphaFold models to the X-ray structure side-by-side (Figure 2). Since AlphaFold structures do not contain ligands nor cofactors, we have used transplanted cofactors from the AlphaFill database as starting points for tunnel calculation. As shown in Figure 2 from the membrane perspective, most of the named reviewed channels from the X-ray structure are also present in the AlphaFill model - especially the major access (2b, 2e, 2f) and egress (2c, S) channels. Narrower channels (3, W) are, however, not retained, and the imperfect transplantation of the heme cofactor opens an additional tunnel following the edges of the heme cavity. We can thus conclude that AlphaFold models can be, in principle, 'tunnelled' for ligand access/egress routes.

### Case study 2—tunnels calculated based on cognate ligands in haloalkane dehalogenase LinB

Haloalkane dehalogenases (EC 3.8.1.5) belong to the  $\alpha/\beta$ -hydrolase superfamily of enzymes. They perform the catalysis of the carbon-halogen bond cleavage in halogenated compounds. Due to their robust properties and wide substrate range, these enzymes have been used in various practical applications, including biodegradation, bioremediation, and cell imaging (24). The access tunnels in haloalkane dehalogenases have been well studied and are a target of engineering by mutagenesis to change their properties (25,26).

Haloalkane dehalogenase LinB from *Sphingobium japonicum* UT26 is involved in the biodegradation pathway to degrade the organochlorine pesticide lindane. LinB is one of the few dehalogenases capable of slowly processing the toxic environmental pollutant 1,2,3-trichloropropane (27). The crystal structure with bound product (2S)-2,3-dichloropropan-1-ol was obtained in a soaking experiment with 1,2,3-trichloropropane (PDB ID 2BFN) and was selected for this



**Figure 2.** Comparison of named reviewed cytochrome P450 2D6 tunnels from X-ray structure (left, PDB ID 3TBG) from Wang *et al.* (20) with automatically calculated tunnels from AlphaFill model (right, UniProt ID P10635). The structure is oriented from the membrane perspective (see the position of the transmembrane anchor in the AlphaFill model pointing towards the viewer), and thus, the major substrate entry tunnels (2b, 2e and 2f) are the most visible and retained in the AlphaFill model. Also, major product release tunnels (2c and S) pointing above the membrane (and sideways for the viewer) are retained. A narrower rare egress tunnel 3 and water channel (W) leading to the proximal side toward the cytosol are missing. As seen from the tunnel 2b profile at the bottom of the figure, the profiles and their properties are almost identical, including channel-lining residues (bottom right) and their UniProt annotation (e.g. Phe 120 in chain A). All ligands in both structures were hidden for clarity.

case study. This structure is a part of the new Cognate dataset. The bound ligand in the crystal structure was paired with the information about the cognate reaction and ligand of the enzyme. The location of the bound molecule was used to select the binding pocket automatically, define the starting point and analyse the tunnels. We have previously shown that the approach employing cognate ligands as a starting point has a higher success rate than selecting the pocket based on FPOCKET and Druggability scores (10). Both MOLE and CAVER found the main tunnels in the structure of the dehalogenase (Figure 3).

### Usage and integration with other databases and bioinformatics tools

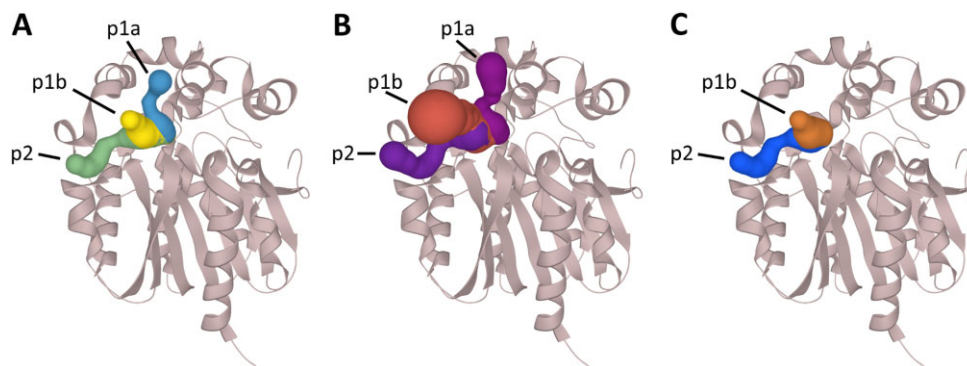
The ChannelsDB database was accessed nearly 5000 times last year. ChannelsDB data is a part of the PDBe-KB database (12,13), integrating a rich set of annotations of biomacromolecules. ChannelsDB was also included in the PDBsum database (27), summarising key information about individual protein structures. Visualisation of protein tunnels and pores, developed for ChannelsDB, was included in a frequently used molecular viewer, LiteMol (26). ChannelsDB is also referenced

in bio.tools (28)—a comprehensive database of bioinformatics services, and in FAIRsharing (29), a database of data and metadata standards interrelated to databases and data policies. The accessibility of ChannelsDB can also benefit from the inclusion of REST API service that can be found together with its documentation at the address <https://channelsdb2.biodata.ceitec.cz/api>.

### Limitations

Ligand transport through the protein is dynamic in nature, and a single snapshot, albeit experimental, might not be representative enough. Compared to a real system, some channels may be closed and some open. Hence, it is advisable to test multiple protein conformers or even run molecular dynamics to evaluate the influence of protein movement on the ligand transport process along the channel. Unfortunately, ChannelsDB 2.0 is based on static structures either from PDBe or AlphaFill databases. If multiple structures are available, we advise analysing the channels in all.

Another limitation is the automated database-wide configuration of both software tuned according to experience for their best performance for typical organic molecular ligands.



**Figure 3.** Visualisation of tunnels identified in the structure of haloalkane dehalogenase LinB (PDB ID: 2BFN). **(A)** Reviewed tunnels extracted from the publication by Brezovsky *et al.* (26), tunnels calculated based on starting points defined by an automatic pipeline using the information about bound and cognate ligands with **(B)** MOLE 2.0 and **(C)** CAVER 3.02. Even with the automatically generated starting points and default settings, both tools correctly identified the main tunnel p1 and the slot tunnel p2.

For larger systems interacting with larger biomacromolecules (e.g. ribozymes), it is necessary to change parameters accordingly, especially the probe size. Any parameter changes can significantly affect the accuracy and calculated tunnels in protein structures. We are also limited to subsets of cofactors and cognate ligands for tunnel calculation. Unoccupied or unannotated active sites are thus missed, but they can be tested for channel presence with both software packages: <https://mole.upol.cz/> or <https://loschmidt.chemi.muni.cz/caverweb/>.

Finally, it is important to note that ChannelsDB 2.0 is based on PDBe and AlphaFill structures availability, and PDBe and UniProt APIs. Hence, when any of these services have an outage, the visualisation of the results will fail. ChannelsDB currently does not use automatic updates for the entire database, so any recently added structure at PDBe or AlphaFill databases will likely not be added to ChannelsDB 2.0 yet. The current update cycle is set to 6-month period.

## Conclusions

ChannelsDB represents a unique resource for any analysis, which includes the transport of ligands and other small molecules within biomacromolecular structures. The updated ChannelsDB 2.0 represents a significant step forward in further channel analyses, which may facilitate future studies devoted to a deeper understanding of the biological roles and evolution of these functional features of biomacromolecular interactions with other molecules.

## Data availability

The updated ChannelsDB 2.0 website is available at <https://channelsdb2.biodata.ceitec.cz> with three times more protein entries (69776 on 1st September 2023) than the previous version, covered by two established software packages, CAVER <https://loschmidt.chemi.muni.cz/caverweb> and MOLE <https://mole.upol.cz/>.

## Acknowledgements

We would like to thank Prof. Dame Janet Thornton and colleagues from her group at EMBL-EBI for their long-standing support and fruitful discussions. Computational resources were provided by the e-INFRA CZ project (ID:90254), sup-

ported by the Ministry of Education, Youth and Sports of the Czech Republic. Biological Data Management and Analysis Core Facility of CEITEC Masaryk University, funded by ELIXIR CZ research infrastructure (MEYS Grant No: LM2023055), is gratefully acknowledged for supporting the research presented in this paper.

## Funding

ELIXIR CZ and RECETOX RI research infrastructure projects (MEYS) [LM2023055; LM2023069 to O.V., D.B., R.S., T.R., J.D.]; Palacky University Olomouc [IGA\_PrF\_2023\_018 to A.Š.]; Grant Agency of Czech Republic JuniorStar project [22-30571M to D.S.]; the project National Institute for Cancer Research (Programme EXCELES, ID Project No. LX22NPO5102)—Funded by the European Union—Next Generation EU (to O.V., D.B., J.D.). Funding for open access charge: CzechELib.

## Conflict of interest statement

None declared.

## References

- Burley, S.K., Bhikadiya, C., Bi, C., Bittrich, S., Chen, L., Crichlow, G.V., Christie, C.H., Dalenberg, K., Di Costanzo, L., Duarte, J.M., *et al.* (2021) RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.*, **49**, D437–D451.
- Armstrong, D.R., Berrisford, J.M., Conroy, M.J., Gutmanas, A., Anyango, S., Choudhary, P., Clark, A.R., Dana, J.M., Deshpande, M., Dunlop, R., *et al.* (2020) PDBe: improved findability of macromolecular structure data in the PDB. *Nucleic Acids Res.*, **48**, D335–D343.
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, A., Potapenko, A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
- Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., *et al.* (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, **50**, D439–D444.

5. Berka,K., Hanák,O., Sehnal,D., Banáš,P., Navrátilová,V., Jaiswal,D., Ionescu,C.-M., Svobodová Vařeková,R., Koča,J. and Otyepka,M. (2012) MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. *Nucleic Acids Res.*, **40**, W222–W227.
6. Pravda,L., Sehnal,D., Toušek,D., Navrátilová,V., Bazgier,V., Berka,K., Svobodová Vařeková,R., Koča,J. and Otyepka,M. (2018) MOLEonline: a web-based tool for analyzing channels, tunnels and pores (2018 update). *Nucleic Acids Res.*, **46**, W368–W373.
7. Chovancova,E., Pavelka,A., Benes,P., Strnad,O., Brezovsky,J., Kozlikova,B., Gora,A., Sustr,V., Klvana,M., Medek,P., *et al.* (2012) CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.*, **8**, e1002708.
8. Stourac,J., Vavra,O., Kokkonen,P., Filipovic,J., Pinto,G., Brezovsky,J., Damborsky,J. and 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.
9. Pravda,L., Sehnal,D., Svobodová Vařeková,R., Navrátilová,V., Toušek,D., Berka,K., Otyepka,M. and Koča,J. (2018) ChannelsDB: database of biomacromolecular tunnels and pores. *Nucleic Acids Res.*, **46**, D399–D405.
10. Vavra,O., Tyzack,J., Haddadi,F., Stourac,J., Damborsky,J., Mazurenko,S., Thornton,J. and Bednar,D. (2023) Large-scale annotation of biochemically relevant pockets and tunnels in cognate enzyme-ligand complexes bioinformatics. *bioRxiv* doi: <https://doi.org/10.1101/2023.03.29.534735>, 30 March 2023, preprint: not peer reviewed.
11. Tyzack,J.D., Fernando,L., Ribeiro,A.J.M., Borkakoti,N. and Thornton,J.M. (2018) Ranking enzyme structures in the PDB by bound ligand similarity to biological substrates. *Structure*, **26**, 565–571.
12. Hekkelman,M.L., De Vries,I., Joosten,R.P. and Perrakis,A. (2023) AlphaFill: enriching AlphaFold models with ligands and cofactors. *Nat. Methods*, **20**, 205–213.
13. PDBe-KB consortium, Varadi,M., Anyango,S., Armstrong,D., Berrisford,J., Choudhary,P., Deshpande,M., Nadzirin,N., Nair,S.S., Pravda,L., *et al.* (2022) PDBe-KB: collaboratively defining the biological context of structural data. *Nucleic Acids Res.*, **50**, D534–D542.
14. The UniProt Consortium, Bateman,A., Martin,M.-J., Orchard,S., Magrane,M., Ahmad,S., Alpi,E., Bowler-Barnett,E.H., Britto,R., Bye-A-Jee,H., *et al.* (2023) UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.*, **51**, D523–D531.
15. Sehnal,D., Svobodová Vařeková,R., Berka,K., Pravda,L., Navrátilová,V., Banáš,P., Ionescu,C.-M., Otyepka,M. and Koča,J. (2013) MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *J. Cheminformatics*, **5**, 39.
16. Furnham,N., Holliday,G.L., de Beer,T.a.P., Jacobsen,J.O.B., Pearson,W.R., Thornton,J.M. and Beer,T.A.P.D. (2014) The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res.*, **42**, D485–D489.
17. Kanehisa,M. (2000) KEGG: kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **28**, 27–30.
18. Le Guilloux,V., Schmidtke,P. and Tuffery,P. (2009) Fpocket: an open source platform for ligand pocket detection. *BMC Bioinf.*, **10**, 168.
19. Fischer,J.D., Holliday,G.L. and Thornton,J.M. (2010) The CoFactor database: organic cofactors in enzyme catalysis. *Bioinformatics*, **26**, 2496–2497.
20. Wang,A., Stout,C.D., Zhang,Q. and Johnson,E.F. (2015) Contributions of ionic interactions and protein dynamics to cytochrome P450 2D6 (CYP2D6) substrate and inhibitor binding. *J. Biol. Chem.*, **290**, 5092–5104.
21. Pravda,L., Berka,K., Svobodová Vařeková,R., Sehnal,D., Banáš,P., Laskowski,R.A., Koča,J. and Otyepka,M. (2014) Anatomy of enzyme channels. *BMC Bioinf.*, **15**, 379.
22. Cojocar,V., Winn,P.J. and Wade,R.C. (2007) The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta*, **1770**, 390–401.
23. Berka,K., Paloncýová,M., Anzenbacher,P. and Otyepka,M. (2013) Behavior of Human cytochromes P450 on lipid membranes. *J. Phys. Chem. B*, **117**, 11556–11564.
24. Koudelakova,T., Bidmanova,S., Dvorak,P., Pavelka,A., Chaloupkova,R., Prokop,Z. and Damborsky,J. (2013) Haloalkane dehalogenases: biotechnological applications. *Biotechnol. J.*, **8**, 32–45.
25. Kokkonen,P., Sykora,J., Prokop,Z., Ghose,A., Bednar,D., Amaro,M., Beerens,K., Bidmanova,S., Slanska,M., Brezovsky,J., *et al.* (2018) Molecular gating of an engineered enzyme captured in real time. *J. Am. Chem. Soc.*, **140**, 17999–18008.
26. Brezovsky,J., Babkova,P., Degtjarik,O., Fortova,A., Gora,A., Iermak,I., Rezacova,P., Dvorak,P., Smatanova,I.K., Prokop,Z., *et al.* (2016) Engineering a de novo transport tunnel. *ACS Catal*, **6**, 7597–7610.
27. Monincová,M., Prokop,Z., Vévodová,J., Nagata,Y. and Damborský,J. (2007) Weak activity of haloalkane dehalogenase LinB with 1,2,3-trichloropropane revealed by X-ray crystallography and microcalorimetry. *Appl. Environ. Microbiol.*, **73**, 2005–2008.
28. Ison,J., Ienasescu,H., Chmura,P., Rydza,E., Ménager,H., Kalaš,M., Schwämmle,V., Grüning,B., Beard,N., Lopez,R., *et al.* (2019) The bio.tools registry of software tools and data resources for the life sciences. *Genome Biol.*, **20**, 164.
29. the FAIRsharing Community, Sansone,S.-A., McQuilton,P., Rocca-Serra,P., Gonzalez-Beltran,A., Izzo,M., Lister,A.L. and Thurston,M. (2019) FAIRsharing as a community approach to standards, repositories and policies. *Nat. Biotechnol.*, **37**, 358–367.