ChannelsDB: database of biomacromolecular tunnels and pores

Lukáš Pravda[1,2], David Sehnal[1,2], Radka Svobodová Vařeková[1,2], Veronika Navrátilová[3], Dominik Toušek[1,3], Karel Berka[3], Michal Otyepka[3,*] and Jaroslav Koča[1,2,*]

[1]CEITEC - Central European Institute of Technology, Masaryk University Brno, Kamenice 5, 625 00 Brno-Bohunice, Czech Republic, [2]National Centre for Biomolecular Research, Faculty of Science, Kamenice 5, 625 00 Brno-Bohunice, Czech Republic and [3]Regional Centre of Advanced Technologies and Materials, Department of Physical Chemistry, Faculty of Science, Palacký University, tř. 17. listopadu 12, 771 46 Olomouc, Czech Republic

ABSTRACT

ChannelsDB (http://ncbr.muni.cz/ChannelsDB) is a database providing information about the positions, geometry and physicochemical properties of channels (pores and tunnels) found within biomacromolecular structures deposited in the Protein Data Bank. Channels were deposited from two sources; from literature using manual deposition and from a software tool automatically detecting tunnels leading to the enzymatic active sites and selected cofactors, and transmembrane pores. The database stores information about geometrical features (e.g. length and radius profile along a channel) and physicochemical properties involving polarity, hydrophobicity, hydropathy, charge and mutability. The stored data are interlinked with available UniProt annotation data mapping known mutation effects to channel-lining residues. All structures with channels are displayed in a clear interactive manner, further facilitating data manipulation and interpretation. As such, ChannelsDB provides an invaluable resource for research related to deciphering the biological function of biomacromolecular channels.

INTRODUCTION

Channels (tunnels and pores) are highly important structural features of biomacromolecules intimately connected with their biological function or structural stability.

Tunnels connect internal spaces of biomacromolecules with the exterior, enabling substrates to travel inwards to and product outwards from enzymes' active sites (1); make internal passages between two active sites, establishing substrate channeling in between them (2–4), e.g. within photosystem II (5); or facilitating the release of nascent synthe-sized proteins to leave the ribosomal proteosynthetic center via the ribosomal exit tunnel (6,7) to name just a few examples (see Figure 1A and B). It should be noted that tunnels have been identified in 64% of enzymes with known crystal structures (8) documenting that channels are common features of enzyme structures.

Pores span through the structure from one side to another. For this reason, they are especially useful for guiding transport through cellular biomembranes (9), e.g. the passage of ions through ion channels (10–12) and within other transporters (13). Pores filled with ions also stabilize the structure of G-DNA (14) (see Figure 1C and D). These examples document important roles of channels in the biological function of biomacromolecules, and this knowledge has also been exploited, e.g. in mutagenesis studies focused on rationally engineering the substrate specificity of haloalkane dehalogenases (15) or cytochrome P450 enzymes (16–19).

The importance of biomacromolecular channels has motivated the development of tools and databases that provide information about these structural features. In the last few decades, we have witnessed the intensive development of many software tools for the detection and characterization of tunnels and pores (20,21). The most popular is HOLE (22) for pores, Caver (23,24) for tunnels, and MOLE and MOLEonline (25–27) for both, however many others are available with various functionalities and performances (20,28). In parallel, several databases collecting information about channel proteins were created. Unfortunately, most of them focus mainly on the features and ontologies of proteins rather than the structural features of channels; TransportDB (29) annotates transport channel proteins in genomes. The Transporter Classification Database (TCDB) (30) classifies transporter proteins and provides structural, functional, mechanistic, evolutionary and disease/medical information about transporters from organisms of all types. The Orientations of Proteins

*To whom correspondence should be addressed. Tel: +420 549492685; Fax: +420 549491060; Email: jaroslav.koca@ceitec.muni.cz
Correspondence may also be addressed to Michal Otyepka. Tel: +420 585634756; Fax: +420 585634761; Email: michal.otyepka@upol.cz
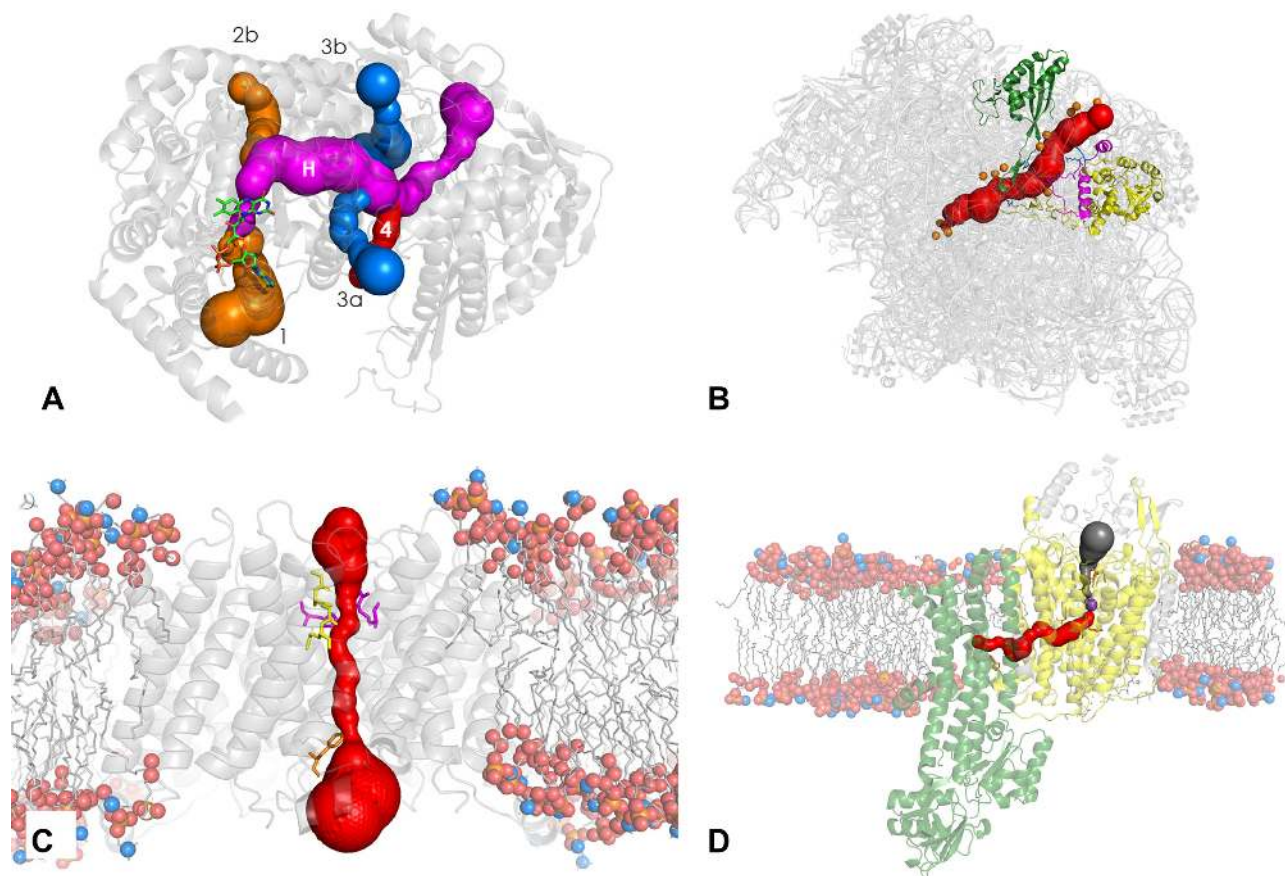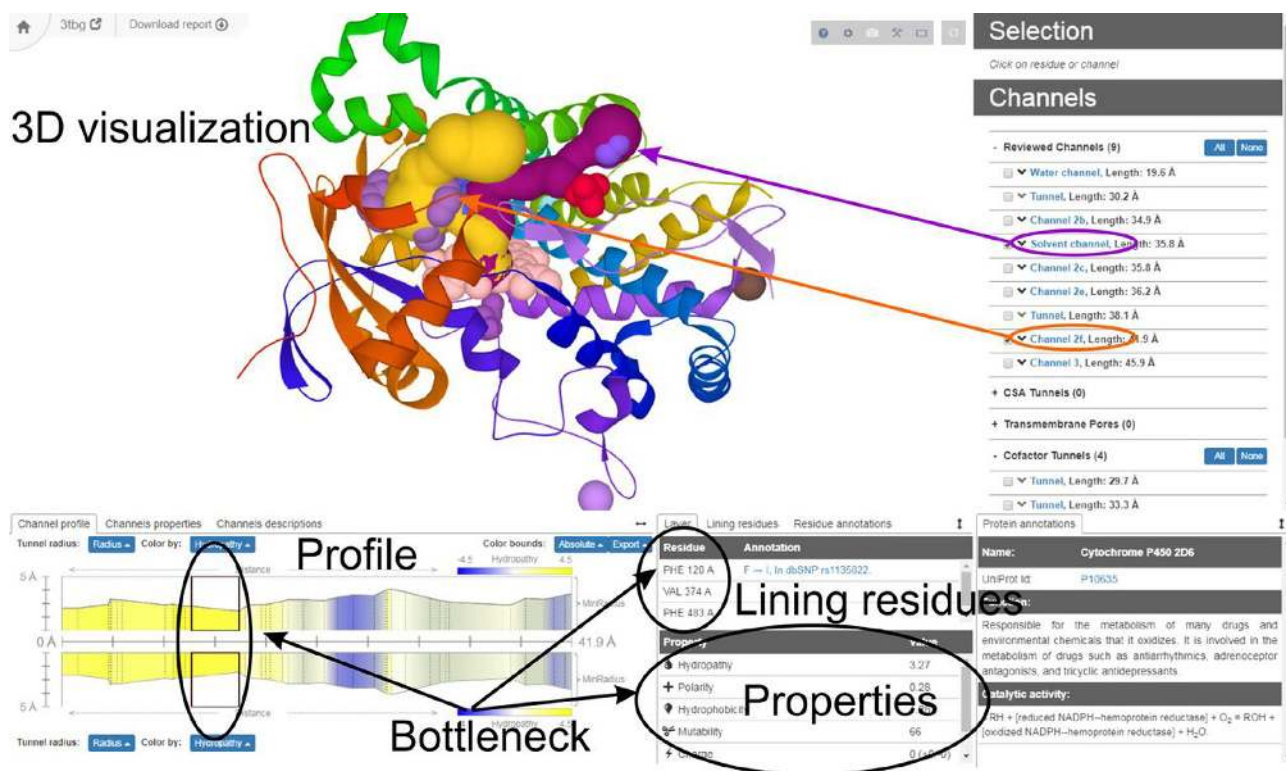
Figure 1. Visualization of selected channel systems. (**A**) Substrate channels 1 and 2b (orange), together with water access channels 3a and 3b (blue), are involved in the proline catabolism pathway catalyzed by proline utilization protein (PutA) in Gram-negative bacteria (PDB ID: 4NM9). PutA contains two active sites interconnected by the ∼75 Å long channel H (magenta), through a hydrolysis cavity. Finally, channel 4 connects the base of the other active site, suggesting a possible escape route for L-glutamate (4), for further details see the interactive view at http://ncbr.muni.cz/ChannelsDB/4nm9. (**B**) The ribosomal polypeptide exit tunnel directs a nascent protein from the peptidyl transferase center outside of the ribosome (here *Haloarcula marismortui* large ribosomal subunit PDB ID: 1JJ2). The exit tunnel is ∼100 Å long and its wall is made of a mosaic of negatively and positively charged residues to prevent the nascent protein from sticking inside the tunnel (6), see at http://ncbr.muni.cz/ChannelsDB/1jj2. (**C**) The ∼30 Å long aquaporin water channel (PDB ID: 1YMG) is a transmembrane pore crucial for maintaining water homeostasis. Residues important for water molecule permeation are highlighted with a stick model and include a selectivity filter (ar/R; magenta); canonical hydrogen bond acceptors important for proper water orientation (yellow); and a constriction region (orange) (61), see at http://ncbr.muni.cz/ChannelsDB/1ymg. (**D**) The potassium-importing KdpFABC membrane complex (PDB ID: 5MRW) is a potassium transporter with two domains coupled together (13). Cellular potassium import via the access channel (gray) is regulated by a charge transfer via the intermolecular channel (red) from the KdpA (yellow) to the KdpB (green), which eventually leads to the conformation change allowing potassium release to the cytosol. The potassium ion is shown as a sphere and is located at the bottom of the selectivity filter, which is reachable by a gray tunnel. The functionally important residues Glu370, Asp583 and Arg493 and Arg116 are displayed as magenta sticks; see at http://ncbr.muni.cz/ChannelsDB/5mrw. The position of the lipid bilayer in figures C and D was obtained from MemProtMD (62).

in Membranes (OPM) (31) and PDBTM (32) databases are specialized in the identification of the membrane-associated regions of known structures of proteins. The ABC protein mutations (ABCMdb) (33) and γ-aminobutyric acid transporter mutagenesis (GATMD) (34) databases list the effects of mutations of individual transporters. The voltage-gated potassium channel database (VKCDB) (35) contains sequence data for various voltage-gated K$^+$ channels adjusted with their electrophysiological parameters. Channel-pedia (36) synthesizes ion channel information from the literature and connects it with Hodgkin–Huxley models. The IUPHAR/BPS Guide to Pharmacology (17) annotates pharmacological targets including channel proteins and their ligands. SuperPain (37) contains data about pain-relieving compounds targeting ion channels with measured binding affinities connected with predicted ligand-binding

poses. And finally, the PDBsum (38) database contains information about tunnels and pores predicted for individual PDB entries. These databases are very useful to their respective research communities, as reflected in the wealth of the scientific literature in these fields. However, to date there is no available comprehensive general purpose database collecting and combining information about channels that were described in the literature, channels connecting active sites with the exterior, transmembrane pores etc., together with supportive information for rationalization of their function (e.g. physicochemical properties).

To fill this gap, we have developed the ChannelsDB database—a comprehensive and up-to-date resource of the channels found in entries from the Protein Data Bank (PDB) (39). The channels were detected using the software MOLE (26) based on information from the literature, po-

Figure 2. Visualization of cytochrome P450 2D6 tunnels on entity page. Page shows the 3D structure of cytochrome P450 2D6 (PDB ID: 3TBG) with two channels selected in the Channels tab—solvent (yellow) and 2f (purple). The ligands present within the structure are shown as vdW spheres, including one of the thioridazine molecules resting within channel 2f (magenta). The red rectangle on the interactive Channel profile tab is locked by the mouse on the channel's bottleneck, and its properties are shown in the Layer tab. The lining residues tab then lists all residues along the selected channel. Residue and Protein annotations tabs use data from UniProt to show known residue mutations for this UniProt ID, its function and catalytic activity.

sitions of buried catalytic sites or cofactors (frequent end points of tunnels) or positions of transmembrane proteins in a membrane. The stored information is accessible via an advanced and easy-to-use user interface, which provides interactive visualization using the state-of-the-art web-based molecular viewer LiteMol Viewer (https://litemol.org) (Sehnal, D., Deshpande, M., Svobodová Vařeková, R., Mir, S., Berka, K., Midlik, A., Pravda, L., Velankar, S. and Koča, J. (2017) LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. *Nat. Methods*, in press) and detailed information about the geometrical and physicochemical features of individual channels.

DATABASE CONTENT

ChannelsDB is built over the data layers obtained from the PDBe (40) and UniProt (41) databases. Individual PDB ID entries are used as a key for a biological assembly structure upon which the channels are deposited, whereas the UniProt ID maps unique information about the protein under study and residue annotation. The annotation of channels in ChannelsDB was obtained in several ways:

a) hundreds of entries were **reviewed manually** by considering the information available in the scientific literature and mapped over the channels precalculated by MOLE,

b) ~15 000 enzyme entries were detected by MOLE as tunnels connecting the buried catalytic sites annotated in **Catalytic Site Atlas** (CSA) (42) with the protein surface.

c) ~12 000 protein entries were identified by MOLE as tunnels connecting selected **cofactors** (typical reaction centers) with the protein surface,

d) hundreds of entries were predicted by MOLE to be **transmembrane pores**.

The ChannelsDB homepage allows querying for structures with annotated channels using any PDB-related metadata, such as PDB ID, protein name, protein family, cofactor or other ligand, author or even journal. The search engine is provided by the PDBe RESTful API. When the PDB ID entry has channels identified in its structure, the details in the record card together with the static picture of channel system are displayed and are accessible upon clicking a PDB ID / protein name. The ChannelsDB website also includes documentation explaining the methodology, a tutorial of the database usage and five examples of typical channel systems (aquaporin channel, cytochrome P450 active site access channels, substrate channeling system in PutA, ribosomal polypeptide exit tunnel and potassium importer complex).
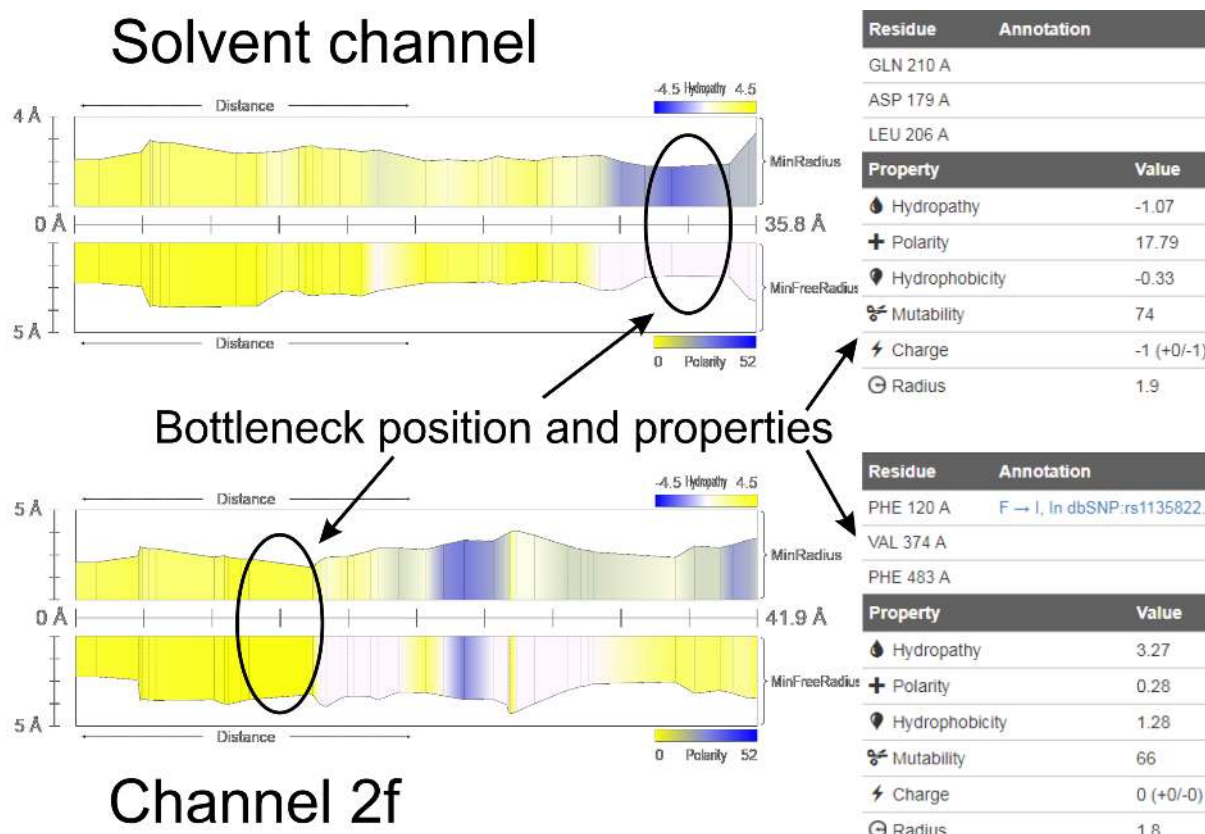
Figure 3. Comparison of tunnel bottlenecks′ properties. Comparison of two cytochrome P450 2D6 channels profiles shows different physicochemical properties of their bottlenecks. The solvent channel has a hydrophilic bottleneck located in the vicinity of amino acids Gln210, Asp179 and Leu206 ~30 Å from the starting point near the heme moiety within the active site. Channel 2f has a highly hydrophobic bottleneck in the vicinity of amino acids Phe120, Val374 and Phe483 closer to the active site. It should be noted that while Phe120Ile is a known tolerated mutation (63), the Phe120 contributes to the regiospecificity of the enzyme, as its mutation leads to the formation of a novel dextromethorphan metabolite (64). According to their properties, it can be hypothesized that the solvent channel is also an egress path for the generally more polar product of the monooxygenation reaction, whereas channel 2f might serve as a substrate access channel.

RESULTS

Description of prototypical database entity—protein with channels

The entity details page (Figure 2) allows users to interactively inspect the features of available channels listed in the *Channels* tab. The web page enables visualization and manipulation with the 3D structure using LiteMol Viewer. The inspected structure is downloaded as a biological assembly of a PDB entry according to the information provided by the PDBe. Channel 3D visualization is supplemented with the *Channel profile* tab - a plot of channel radius vs. distance from the starting point divided into interactive layers. Channel profile layers are colored according to a selected physicochemical property from the list of charge, hydropathy (43), polarity (44), hydrophobicity (45) or mutability (46), calculated by MOLE. The amino acids and values of these physicochemical properties for individual interactively selected layer are shown in the *Layer* tab. The *Channels properties* tab lists the properties of all detected channels. *Protein annotations* tabs show text annotations retrieved from UniProt and the literature using a publicly available APIs (47,48). Specifically, the protein name is retrieved together with its function. When the protein in question bears a catalytic activity, a list of known catalyzed reactions is also retrieved. Last but not least, *Residue annotations* tab lists sets of functional annotations for individual residues from both the UniProt and ChannelsDB databases. When these residues form a channel's walls, this information is highlighted. All the visuals are interactive—the *Selection* tab shows selected residues and channels. Annotations are directly bound to the source of information in the literature, and all the results are made available for download in several reporting formats (ZIP, PY, PDB and JSON) for further processing.

Case study—cytochrome P450 2D6 active site access channels

A well-known biological example of a protein family with tunnels is cytochrome P450 (16,49–58). These highly important metabolic enzymes have an active role in the biotransformation of both endobiotic and xenobiotic compounds. In humans, their broad substrate specificity affects the pharmacokinetic parameters of most marketed drugs and drug-drug interactions. Similarly to other oxidoreductases (1), their active site is deeply buried within the structure (49,52). Various substrates and products therefore have

to pass through the series of tunnels leading towards the active site heme cofactor. These tunnels have an already set nomenclature developed by Wade and coworkers (50,59) and it has been shown both theoretically (16,51,53–55) and experimentally (56–58) that they play a role in the substrate and product channeling to and from the active site and therefore in the substrate specificity of individual members of the cytochrome P450 family.

In our example (Figure 2; http://ncbr.muni.cz/ChannelsDB/3tbg) we show cytochrome P450 isoform 2D6 (CYP2D6) with two thioridazine molecules bound in the structure - one in the active site and the other one in the channel 2f. Channels from family #2 are thought to work as a substrate access route, and the X-ray structure of CYP2D6 (PDB ID: 3TBG) with one thioridazine bound in the 2f channel supports this idea (60). The active site of cytochrome P450 is thought to be hydrated via a solvent channel which may also function as a metabolite egress channel (51). The function of the access channels is also reflected in the different hydrophobicity and polarity of bottlenecks of these channels (Figure 3). The bottleneck of channel 2f is highly hydrophobic (hydropathy 3.27, polarity 0.28), whereas the bottleneck of the solvent channel is highly polar (hydropathy –1.07, polarity 17.79). One may anticipate that a nonpolar substrate would prefer channel 2f as an access route, whereas the more polar product of the monooxygenation reaction would select the solvent channel as an egress route. This example shows that ChannelsDB can be utilized not only for the visualization of structural features of channels, but also for the rationalization of their biological function. In addition, information acquired from ChannelsDB may be used to track the evolution of channels in organisms.

User feedback and integration with other databases and bioinformatics tools

To simplify contact with the user community, we also offer the possibility for anyone to contribute or point out not yet annotated systems with known channels. ChannelsDB contains a form page where the user can specify the system, PDB ID, DOI or Pubmed ID of a reference literature and even a list of residues to annotate. In the near future we will also provide an interactive annotation directly from updated MOLEonline web services.

Another significant asset of the ChannelsDB is that the data are provided over the API, and therefore everyone can take advantage of this resource and integrate it into an application freely or access its content programmatically.

DISCUSSION

ChannelsDB provides information about the presence and positions of the channels in biomacromolecular structures. It also contains information about local geometrical properties and residues lining the channel and their physicochemical properties. All the properties are transparently mapped onto the channels's profile via an easy-to-use and interactive user interface, aiding data interpretation. ChannelsDB also uses annotations from UniProt (e.g. a function of protein as well as individual residues) and maps them

on the channel profile. Thanks to this distinctive combination of calculated properties overlaid with a handful of residue annotations, ChannelsDB represents a unique resource for any analysis which includes the transport of ligands and other small molecules within biomacromolecular structures. We believe that ChannelsDB represents a significant step forward in channel analyses, which may facilitate future studies devoted to a deeper understanding of the biological roles and evolution of these structural features of biomacromolecules.

AVAILABILITY

The ChannelsDB database is available from http://ncbr.muni.cz/ChannelsDB.

REFERENCES

1. Davids,T., Schmidt,M., Böttcher,D. and Bornscheuer,U.T. (2013) Strategies for the discovery and engineering of enzymes for biocatalysis. *Curr. Opin. Chem. Biol.*, **17**, 215–220.
2. Thoden,J.B., Holden,H.M., Wesenberg,G., Raushel,F.M. and Rayment,I. (1997) Structure of carbamoyl phosphate synthetase: a journey of 96 Å from substrate to product. *Biochemistry*, **36**, 6305–6316.
3. Wheeldon,I., Minteer,S.D., Banta,S., Barton,S.C., Atanassov,P. and Sigman,M. (2016) Substrate channelling as an approach to cascade reactions. *Nat. Chem.*, **8**, 299–309.
4. Singh,H., Arentson,B.W., Becker,D.F. and Tanner,J.J. (2014) Structures of the PutA peripheral membrane flavoenzyme reveal a dynamic substrate-channeling tunnel and the quinone-binding site. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 3389–3394.
5. Guskov,A., Kern,J., Gabdulkhakov,A., Broser,M., Zouni,A. and Saenger,W. (2009) Cyanobacterial photosystem II at 2.9-Å resolution and the role of quinones, lipids, channels and chloride. *Nat. Struct. Mol. Biol.*, **16**, 334–342.
6. Voss,N.R., Gerstein,M., Steitz,T.A. and Moore,P.B. (2006) The geometry of the ribosomal polypeptide exit tunnel. *J. Mol. Biol.*, **360**, 893–906.
7. Nissen,P. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science*, **289**, 920–930.
8. 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 Bioinformatics*, **15**, 379.
9. Cellular gatekeepers (2016) *Nat. Struct. Mol. Biol.*, **23**, 463–463.
10. Guskov,A., Nordin,N., Reynaud,A., Engman,H., Lundbäck,A.-K., Jong,A.J.O., Cornvik,T., Phua,T. and Eshaghi,S. (2012) Structural insights into the mechanisms of Mg2+ uptake, transport, and gating by CorA. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 18459–18464.
11. Leontiadou,H., Mark,A.E. and Marrink,S.-J. (2007) Ion transport across transmembrane pores. *Biophys. J.*, **92**, 4209–4215.

12. Berneche,S. and Roux,B. (2001) Energetics of ion conduction through the K+ channel. *Nature*, **414**, 73.

13. Huang,C.-S., Pedersen,B.P. and Stokes,D.L. (2017) Crystal structure of the potassium-importing KdpFABC membrane complex. *Nature*, **546**, 681–685.

14. Neidle,S. (2009) The structures of quadruplex nucleic acids and their drug complexes. *Curr. Opin. Struct. Biol.*, **19**, 239–250.

15. Pavlova,M., Klvana,M., Prokop,Z., Chaloupkova,R., Banas,P., Otyepka,M., Wade,R.C., Tsuda,M., Nagata,Y. and Damborsky,J. (2009) Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat. Chem. Biol.*, **5**, 727–733.

16. Urban,P., Truan,G. and Pompon,D. (2015) Access channels to the buried active site control substrate specificity in CYP1A P450 enzymes. *Biochim. Biophys. Acta - Gen. Subj.*, **1850**, 696–707.

17. Ebert,M.C.C.J.C., Dürr,S.L., A. Houle,A., Lamoureux,G. and Pelletier,J.N. (2016) Evolution of P450 monooxygenases toward formation of transient channels and exclusion of nonproductive gases. *ACS Catal.*, **6**, 7426–7437.

18. Hendrychova,T., Berka,K., Navratilova,V., Anzenbacher,P. and Otyepka,M. (2012) Dynamics and hydration of the active sites of mammalian cytochromes P450 probed by molecular dynamics simulations. *Curr. Drug Metab.*, **13**, 177–189.

19. Misura,K.M.S., Morozov,A. V. and Baker,D. (2004) Analysis of anisotropic side-chain packing in proteins and application to high-resolution structure prediction. *J. Mol. Biol.*, **342**, 651–664.

20. Pravda,L., Berka,K., Sehnal,D., Otyepka,M., Svobodová Vařeková,R. and Koča,J. (2016) Detection of Channels. In: Koča,J, Svobodová Vařeková R, Pravda,L, Berka,K, Geidl,S, Sehnal,D and Otyepka,M (eds). *Structural Bioinformatics Tools for Drug Design*. Springer, Cham, pp. 59–69.

21. Simões,T., Lopes,D., Dias,S., Fernandes,F., Pereira,J., Jorge,J., Bajaj,C. and Gomes,A. (2017) Geometric detection algorithms for cavities on protein surfaces in molecular graphics: a survey. *Comput. Graph. Forum*, doi:10.1111/cgf.13158.

22. Smart,O.S., Neduvelil,J.G., Wang,X., Wallace,B.A. and Sansom,M.S.P. (1996) HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.*, **14**, 354–360.

23. Petřek,M., Otyepka,M., Banáš,P., Košinová,P., Koča,J. and Damborský,J. (2006) CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics*, **7**, 316.

24. 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.

25. Petřek,M., Košínová,P., Koča,J. and Otyepka,M. (2007) MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure*, **15**, 1357–1363.

26. 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. Cheminform.*, **5**, 39.

27. 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.

28. Krone,M., Kozlíková,B., Lindow,N., Baaden,M., Baum,D., Parulek,J., Hege,H.-C. and Viola,I. (2016) Visual analysis of biomolecular cavities: state of the art. *Comput. Graph. Forum*, **35**, 527–551.

29. Elbourne,L.D.H., Tetu,S.G., Hassan,K.A. and Paulsen,I.T. (2017) TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res.*, **45**, D320–D324.

30. Saier,M.H., Reddy,V.S., Tsu,B. V., Ahmed,M.S., Li,C. and Moreno-Hagelsieb,G. (2016) The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res.*, **44**, D372–D379.

31. Lomize,M.A., Pogozheva,I.D., Joo,H., Mosberg,H.I. and Lomize,A.L. (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.*, **40**, D370–D376.

32. Kozma,D., Simon,I. and Tusnády,G.E. (2013) PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res.*, **41**, D524–D529.

33. Tordai,H., Jakab,K., Gyimesi,G., András,K., Brózik,A., Sarkadi,B. and Hegedűs,T. (2017) ABCMdb reloaded: updates on mutations in ATP binding cassette proteins. *Database*, **2017**, doi:10.1093/database/bax023.

34. Anderson,C.M., Kidd,P.D. and Eskandari,S. (2010) GATMD: aminobutyric acid transporter mutagenesis database. *Database*, **2010**, baq028.

35. Gallin,W.J. and Boutet,P.A. (2011) VKCDB: voltage-gated K+ channel database updated and upgraded. *Nucleic Acids Res.*, **39**, D362–D366.

36. Ranjan,R., Khazen,G., Gambazzi,L., Ramaswamy,S., Hill,S.L., Schürmann,F. and Markram,H. (2011) Channelpedia: an integrative and interactive database for ion channels. *Front. Neuroinform.*, **5**, 36.

37. Gohlke,B.O., Preissner,R. and Preissner,S. (2014) SuperPain—a resource on pain-relieving compounds targeting ion channels. *Nucleic Acids Res.*, **42**, D1107–D1112.

38. de Beer,T.A.P., Berka,K., Thornton,J.M. and Laskowski,R.A. (2014) PDBsum additions. *Nucleic Acids Res.*, **42**, D292–D296.

39. Berman,H., Henrick,K. and Nakamura,H. (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.*, **10**, 980.

40. Velankar,S., van Ginkel,G., Alhroub,Y., Battle,G.M., Berrisford,J.M., Conroy,M.J., Dana,J.M., Gore,S.P., Gutmanas,A., Haslam,P. *et al.* (2016) PDBe: improved accessibility of macromolecular structure data from PDB and EMDB. *Nucleic Acids Res.*, **44**, D385–D395.

41. Wasmuth,E.V. and Lima,C.D. (2016) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **45**, 1–12.

42. Furnham,N., Holliday,G.L., De Beer,T.A.P., Jacobsen,J.O.B., Pearson,W.R. and Thornton,J.M. (2014) The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res.*, **42**, 1–5.

43. Kyte,J. and Doolittle,R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, **157**, 105–132.

44. Zimmerman,J.M., Eliezer,N. and Simha,R. (1968) The characterization of amino acid sequences in proteins by statistical methods. *J. Theor. Biol.*, **21**, 170–201.

45. Cid,H., Bunster,M., Canales,M. and Gazitúa,F. (1992) Hydrophobicity and structural classes in proteins. *Protein Eng. Des. Sel.*, **5**, 373–375.

46. Jones,D.T., Taylor,W.R. and Thornton,J.M. (1992) The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*, **8**, 275–282.

47. Nightingale,A., Antunes,R., Alpi,E., Bursteinas,B., Gonzales,L., Liu,W., Luo,J., Qi,G., Turner,E. and Martin,M. (2017) The Proteins API: accessing key integrated protein and genome information. *Nucleic Acids Res.*, **45**, W539–W544.

48. Velankar,S., Dana,J.M., Jacobsen,J., van Ginkel,G., Gane,P.J., Luo,J., Oldfield,T.J., O'Donovan,C., Martin,M.-J. and Kleywegt,G.J. (2013) SIFTS: structure integration with function, taxonomy and sequences resource. *Nucleic Acids Res.*, **41**, D483–D489.

49. Otyepka,M., Skopalík,J., Anzenbacherová,E. and Anzenbacher,P. (2007) What common structural features and variations of mammalian P450s are known to date? *Biochim. Biophys. Acta*, **1770**, 376–389.

50. Wade,R.C., Winn,P.J., Schlichting,I and Sudarko (2004) A survey of active site access channels in cytochromes P450. *J. Inorg. Biochem.*, **98**, 1175–1182.

51. Paloncýová,M., Navrátilová,V., Berka,K., Laio,A. and Otyepka,M. (2016) Role of enzyme flexibility in ligand access and egress to active site: bias-exchange metadynamics study of 1,3,7-trimethyluric acid in cytochrome P450 3A4. *J. Chem. Theory Comput.*, **21**, 2101-2109.

52. Denisov,I.G., Makris,T.M., Sligar,S.G. and Schlichting,I. (2005) Structure and chemistry of cytochrome P450. *Chem. Rev.*, **105**, 2253–2278.

53. Li,W., Shen,J., Liu,G., Tang,Y. and Hoshino,T. (2011) Exploring coumarin egress channels in human cytochrome p450 2a6 by random acceleration and steered molecular dynamics simulations. *Proteins Struct. Funct. Bioinforma.*, **79**, 271–281.

54. Fishelovitch,D., Shaik,S., Wolfson,H.J. and Nussinov,R. (2009) Theoretical characterization of substrate access/exit channels in the human cytochrome P450 3A4 enzyme: involvement of phenylalanine residues in the gating mechanism. *J. Phys. Chem. B*, **113**, 13018–13025.

55. Otyepka,M., Berka,K. and Anzenbacher,P. (2012) Is there a relationship between the substrate preferences and structural flexibility of cytochromes P450? *Curr. Drug Metab.*, **13**, 130–142.
56. Scott,E.E., He,Y.Q. and Halpert,J.R. (2002) Substrate routes to the buried active site may vary among cytochromes P450: mutagenesis of the F−G region in P450 2B1. *Chem. Res. Toxicol.*, **15**, 1407–1413.
57. Porubsky,P.R., Battaile,K.P. and Scott,E.E. (2010) Human cytochrome P450 2E1 structures with fatty acid analogs reveal a previously unobserved binding mode. *J. Biol. Chem.*, **285**, 22282–22290.
58. Porubsky,P.R., Meneely,K.M. and Scott,E.E. (2008) Structures of human cytochrome P-450 2E1. *J. Biol. Chem.*, **283**, 33698–33707.
59. Cojocaru,V., Winn,P.J. and Wade,R.C. (2007) The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta*, **1770**, 390–401.
60. Yuan,F., Wang,S. and Larson,R.G. (2015) Potentials of mean force and escape times of surfactants from micelles and hydrophobic surfaces using molecular dynamics simulations. *Langmuir*, **31**, 1336–1343.
61. Harries,W.E.C., Akhavan,D., Miercke,L.J.W., Khademi,S. and Stroud,R.M. (2004) The channel architecture of aquaporin 0 at a 2.2-A resolution. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 14045–14050.
62. Stansfeld,P.J., Goose,J.E., Caffrey,M., Carpenter,E.P., Parker,J.L., Newstead,S. and Sansom,M.S.P. (2015) MemProtMD: automated insertion of membrane protein structures into explicit lipid membranes. *Structure*, **23**, 1350–1361.
63. Solus,J.F., Arietta,B.J., Harris,J.R., Sexton,D.P., Steward,J.Q., McMunn,C., Ihrie,P., Mehall,J.M., Edwards,T.L. and Dawson,E.P. (2004) Genetic variation in eleven phase I drug metabolism genes in an ethnically diverse population. *Pharmacogenomics*, **5**, 895–931.
64. Flanagan,J.U., Maréchal,J.-D., Ward,R., Kemp,C.A., McLaughlin,L.A., Sutcliffe,M.J., Roberts,G.C.K., Paine,M.J.I. and Wolf,C.R. (2004) Phe120 contributes to the regiospecificity of cytochrome P450 2D6: mutation leads to the formation of a novel dextromethorphan metabolite. *Biochem. J.*, **380**, 353–360.