Database, 2023, 1–7
DOI: https://doi.org/10.1093/database/baad010
Original article

CaviDB: a database of cavities and their features in the structural and conformational space of proteins

Ana Julia Velez Rueda*, Franco Leonardo Bulgarelli, Nicolás Palopoli and Gustavo Parisi*

Departamento de Ciencia y Tecnologia, Universidad Nacional de Quilmes, Roque Saenz Pena 182, Bernal B1876BXD, Argentina

*Corresponding author: Tel: +54 11 43657100; Fax: +54 11 43657182; Email: anavelezrueda@gmail.com
Correspondence may also be addressed to Gustavo Parisi. Tel: +54 11 43657100; Fax: +54 11 43657182; Email: gusparisi@gmail.com

Citation details: Velez Rueda, A.J., Bulgarelli, F.L., Palopoli, N. *et al.* CaviDB: a database of cavities and their features in the structural and conformational space of proteins. *Database* (2023) Vol. 2023: article ID baad010; DOI: https://doi.org/10.1093/database/baad010

Abstract

Proteins are the structural, functional and evolutionary units of cells. On their surface, proteins are shaped into numerous depressions and protrusions that provide unique microenvironments for ligand binding and catalysis. The dynamics, size and chemical properties of these cavities are essential for a mechanistic understanding of protein function. Here, we present CaviDB, a novel database of cavities and their features in known protein structures. It integrates the results of commonly used cavity detection software with protein features derived from sequence, structural and functional analyses. Each protein in CaviDB is linked to its corresponding conformers, which also facilitates the study of conformational changes in cavities. Our initial release includes ∼927 773 distinct proteins, as well as the characterization of 36 136 869 cavities, of which 1 147 034 were predicted to be drug targets. The structural focus of CaviDB provides the ability to compare cavities and their properties from different conformational states of the protein. CaviDB not only aims to provide a comprehensive database that can be used for various aspects of drug design and discovery but also contributes to a better understanding of the fundamentals of protein structure–function relationships. With its unique approach, CaviDB represents an indispensable resource for the large community of bioinformaticians in particular and biologists in general.

Database URL: https://www.cavidb.org

Introduction

Proteins are the functional, structural and evolutionary units of cells. They consist of chains of amino acids that interact in complex and highly interconnected networks. On their surface, proteins are shaped into numerous cavities and protrusions that provide unique microenvironments for ligand binding or catalysis (1). The dynamic of these cavities are fundamental for understanding protein function, and their variations can explain changes in protein activity (2–5). Protein movements, even the smallest, can affect cavity architecture (6, 7). On different time scales, the movements are required not only to bind the substrate or determine its affinity constant but also to allow ligand transit from the surface to the active site (8).

The size and geometry of the cavities, as well as their accessibility, have proven useful in making predictions about protein–protein interactions, protein pharmacology and binding specificity (9–11). For example, physicochemical properties of the cavities such as their charge or hydrophobicity can also be used to predict the binding probability of specific ligands (12, 13). Residues are known to shift their pK_a values based on various structural and environmental features (14, 15), which favors various biological activities (16, 17). In addition, it has been shown that the shape and location of cavities in proximity to each other can determine their relative flexibility and influence their catalytic and binding promiscuity (4, 11, 18).

Functional cavities are generally located within protein domains, which are evolutionarily conserved protein regions with specific stability, function and dynamics. The biological activity of individual cavities is not always correlated with domain function, and the conservation of cavities may exceed that of a particular domain family. Therefore, knowledge of domain activity is not sufficient to fully understand protein function, and the integrative characterization of all domains and their cavities may be a better approach (19).

Here, we present CaviDB (https://www.cavidb.org/), an interactive online database that integrates the results of commonly used cavity detection software with protein features retrieved from sequence, structural and functional analyses. CaviDB implements established cavity detection methods (20, 21) that allow local structural characterization but is also useful to understand protein anatomy and function on a global scale (22). Our database allows users to explore protein dynamics through an easy-to-use interface that facilitates the comparison of the properties of protein conformers

Received 15 October 2022; Revised 4 January 2023

and their predicted cavities. CaviDB provides structural data on every known protein structure available in the Protein Data Bank (23) and on the protein structure predictions of entire proteomes from model organisms available in the AlphaFold database (24). Our goal is to provide a comprehensive resource for use in various biotechnological applications, such as drug development and discovery, but also for a better understanding of the fundamentals of the relationship between protein structure and function.

Materials and methods

Cavity prediction and categorization

CaviDB provides users with structural and sequential features to characterize protein cavities. Cavity predictions were performed using the widely used Fpocket software (25) with default settings for all entries in the Protein Data Bank (26, 27) and all the AlphaFold database entries (28). We retrieved and annotated all properties (Supplementary Table S1) associated with each cavity and all its lining residues. The cavity was considered to be druggable if it had an affectability value >0.5, as suggested in previous work (20).

Cavities features' calculation

To provide users with information on possible activated cavities, we estimated the pK_a values (at pH = 7) of the ionizable residues and their shifts (pK_a predicted – pK_a expected) using PROPKA (29). The net pK_a shift values per cavity were calculated as the sum of all absolute pK_a shifts of each ionizable residue belonging to a cavity.

Using PROPKA, we also retrieved data on inter-residue contacts per site to annotate the contacts of the cavities as side-chain hydrogen bonds, backbone hydrogen bonds and coulombic bonds. We created a network of cavities that have at least one contact between the same sites, which can be displayed as an interactive diagram. The binding energy heat maps show the contacts between cavities by calculating the sum of the absolute binding energies between the residues that make contact in the corresponding pair of cavities and rendering colored squares.

Different physicochemical properties per site were calculated using Classification of Intrinsically Disordered Ensemble Regions (30), modlAMP (31) and Biopython (32) and assigned to each cavity as the mean values of the properties of its residues.

Global protein features' calculation and annotation

Global protein features were calculated as described in the previous section. Each Protein Data Bank entry (PDB) chain or AlphaFold model was annotated via Structure Integration with Function, Taxonomy and Sequence (33) with identifiers of relevant biological databases such as CATH (34) and Pfam (35) to facilitate subsequent analysis by users.

Conformational comparisons

For the conformers' cavities comparisons, we used the PDBSWS—PDB/UniProt Mapping (36). This database maps PDB residues to residues in UniProtKB (Swiss-Prot and TrEMBL) entries (37), consequently allowing the precise comparison between cavities of different entries.

Web application overview

A responsive web interface was developed to display the data stored in a non-relational database, allowing easier navigation and visualization of the database contents on different devices. The web application was implemented in HTML, CSS, Ruby (on Rails) and JavaScript (using NodeJS).

The first step for running CaviDB is to provide a valid PDB or UniProt ID. The web server automatically loads all chains related to the search, as well as their general data, including their length, the number of predicted cavities and relevant cross-reference identifiers (Figure 1A and B). The search can be filtered using the AlphaFold selector if the user is only interested in these sorts of entries. The features obtained for each entry are organized into two main sections describing the general cavity descriptors, including an interactive display for visualizing the cavities, a network representation of the interactions and cavities including activated residues with pK_a shifts and the global protein descriptors (Figure 1D).

CaviDB allows users to explore the conformational diversity of proteins and its impact on cavity dynamics by providing a conformational comparator (Figure 1B) that displays a comparison page with the listed cavities for each chain and, when selected, their properties and residues.

Results and Test cases

Globular protein test case

Promiscuous proteins are a breaking point in the "structure–function" paradigm and the concept of biological specificity (38, 39). Promiscuous protein behavior presents both challenges and opportunities for drug discovery programs and has been explored as a strategy for drug repurposing (40–42).

Human serum albumin (HSA) is the major protein in plasma, binds multiple ligands (43) and has recently emerged as a very important drug carrier (44, 45). It has several high-affinity binding sites, but most drugs and ligands bind to the so-called sites I (from Met 1 to Asn 111) and II (from Gln 196 to Pro 303) (46). HSA has previously been described not only as a transport protein but also as a promiscuous enzyme possibly related to salicylic acid metabolism and side effects (18, 47–50).

It has been proposed that the basis for the great ability of albumins to catalyze various reactions lies in the existence of activated amino acids with abnormal pK_a in the hydrophobic cavity of the AII binding site, which creates a microenvironment favorable for catalysis (18, 51). As shown by the per-site solvent accessibility plot (ASA) generated by CaviDB for the 1AO6A:0 entry, there is a local minimum around Lys199 and Arg222 (see Figure 2B), a region described as important for catalysis (50, 52). These important catalytic residues are located in the AII binding site identified by CaviDB as the largest cavity (Cavity 1) in the entry's star plot with the highest relative length parameter (equal to 1), also showing a large number of contacts between cavities and the presence of activating residues. Residues Lys199 and Arg222 show essential pK_a shift in order to sustain the catalytic activity, showing abnormally acidic properties (Lys, 199, ∼7.51 and Arg, 222, ∼9.49 (18). Using the information deposited in the CoDNaS database (53), we found the pair of HSA conformers showing the maximum conformational diversity (pairs 3LU6_A and 1O9X_A with an Root-mean-square deviation

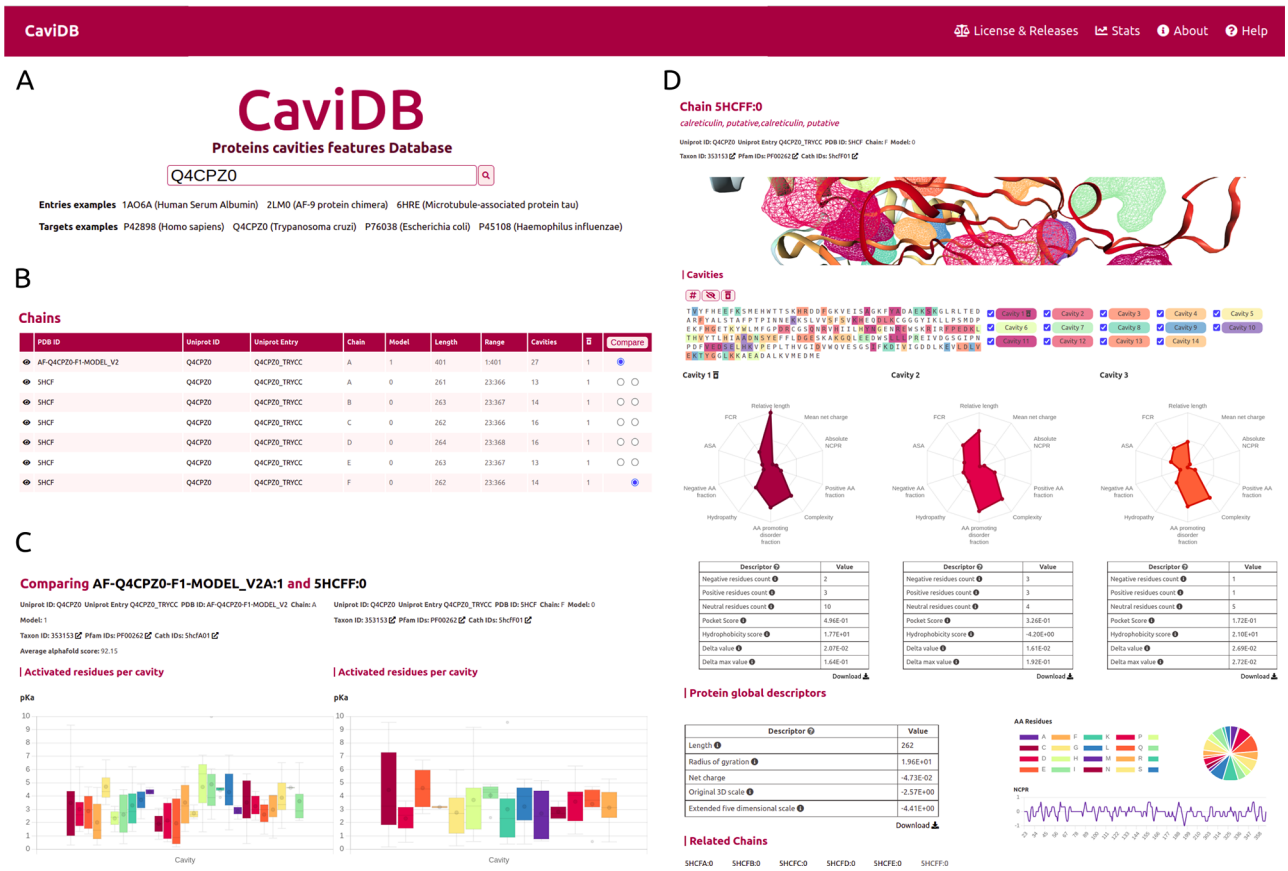

Figure 1. Overview of the CaviDB web application. (A) CaviDB search allows users to search for a specific PDB or UniProt identifier. A selector is also provided to focus the search on AlphaFold models. (B, C) Cavity dynamics can be explored using the comparison tool provided by CaviDB, where predicted cavities and their features can be selected and displayed for different protein conformations simultaneously. (D) Schematic example of chain feature display. The information of each entry is divided into two main sections, one containing the general cavity descriptors (top) and the other containing the global protein descriptors (bottom).

= 6.27 Å). Using this information and the comparison capability of conformers in CaviDB, it is also possible to compare the change in some cavity features. It is then possible to observe differences in the acid–base properties of Cavity 1, such as in the mean pK_a of Cavity 1 (Figure 2E) and changes in charge and hydrophobicity (Figure 2F).

A second cavity (Cavity 2) containing residues Arg410 and Tyr411, previously described as part of the catalytic active site, was also identified (47) (Figure 2). In addition, Cavity 2 contains tyrosine 411 and arginine 410 (belonging to Cavity 4), two residues that have been shown to be important for the esterase-like activity of the protein (52) and that interact with each other through coulombic forces (Figure 2C). In this way, CaviDB gathers important information that provides a mechanistic explanation for the promiscuous behavior of HSA as described previously (18, 54).

Using AlphaFold models for better predictions

The recent breakthrough of AlphaFold in predicting 3D models provides new opportunities for exploring protein–structure relationships. In CaviDB, we have included 1 029 746 AlphaFold models, but we plan to include all recently released models in future upgrades (https://alphafold.ebi.ac.uk/). Recently, AlphaFold models were found to correctly predict some of the native conformations of protein ensembles (55). In some cases, high-quality models could help to assess the functional implications of cavities. Pyridoxal 5′-phosphate (PLP) synthase (PLPS) is a biosynthetic pathway enzyme that produces PLP from glutamine, ribose 5-phosphate and glyceraldehyde 3-phosphate. The native state of PLP synthase consists of 12 synthase and 12 glutaminase subunits, and its chemical mechanism has already been described (56). The active site contains active Lys81 and Asp24 (57, 58). In some conformers of the enzyme, this active site is open, which is due to the presence of a disordered region over the binding site (residues 49–56) (58). When known PLPS conformers are tested for the presence of cavities in CaviDB (using UniProt ID Q5L3Y2 or PDBs 4wy0 and 4wxz), no cavities containing biologically active residues are found. This is likely due to the fact that the binding site is open in these experimental structures. However, when AlphaFold models of PLPs are considered, a new cavity is discovered that contains the biologically relevant residues (Figure 3.) In this sense, the use of high-quality AlphaFold models could help in the estimation of cavities and their potential biological role.

The advantages of CaviDB over existing services

CaviDB has a total count of 927 773 distinct proteins, with 740 140 conformers from the PDB and 1 029 746 from the AlphaFold database. It annotates proteins from 14 871 species

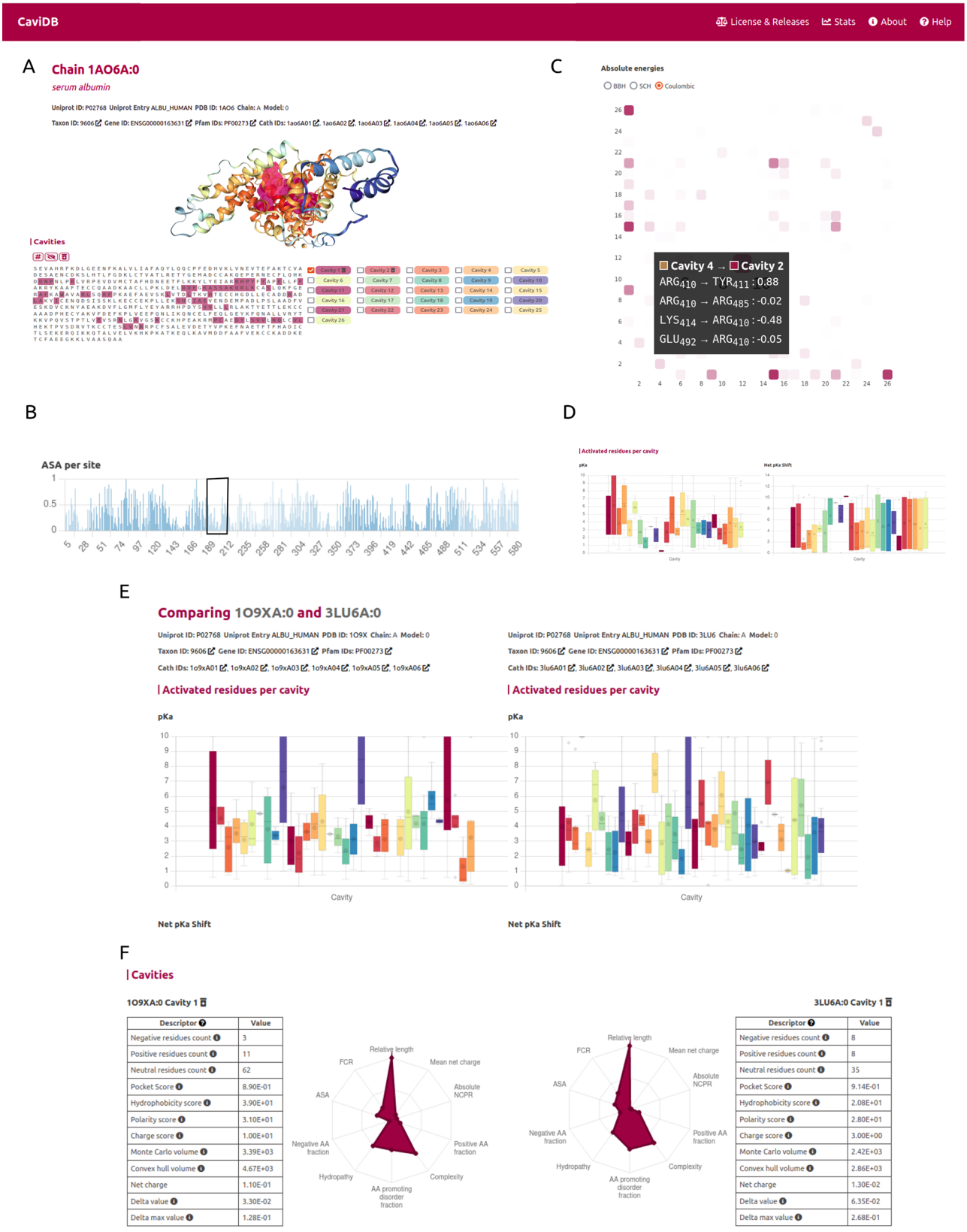

Figure 2. CaviDB display of HSA (PDB ID: 1AO6, chain: A, model: 0). (A) Display of protein and residues per cavity, with Cavities 1 and 2 highlighted in the sequence (magenta and pink, respectively). (B) Bar plot of normalized accessible surface area per site. A box highlights a local minimum of ASA in the vicinity of Lys199. (C) Interactive heat map display of all possible pairwise contacts established between residues in different cavities. A popup window provides details on the absolute energies of the selected interaction. Heat map colors correspond to the number of interactions per cavity, the larger the interaction number is, the darker the color. (D) Boxplot distribution of pK_a values (left) and pK_a shifts (right) per residue in each detected cavity. (E) Comparison of HSA conformers with high RMSD. The panel shows important changes in acid–base properties in Cavity 1 resulting from the conformational changes (mean $pK_a = 3.99$ in 3LU6_A vs. mean $pK_a = 5.16$ in 1O9X_A), along with changes in other physicochemical features such as the number of charged amino acids, hydrophobicity and charge of residues (F).

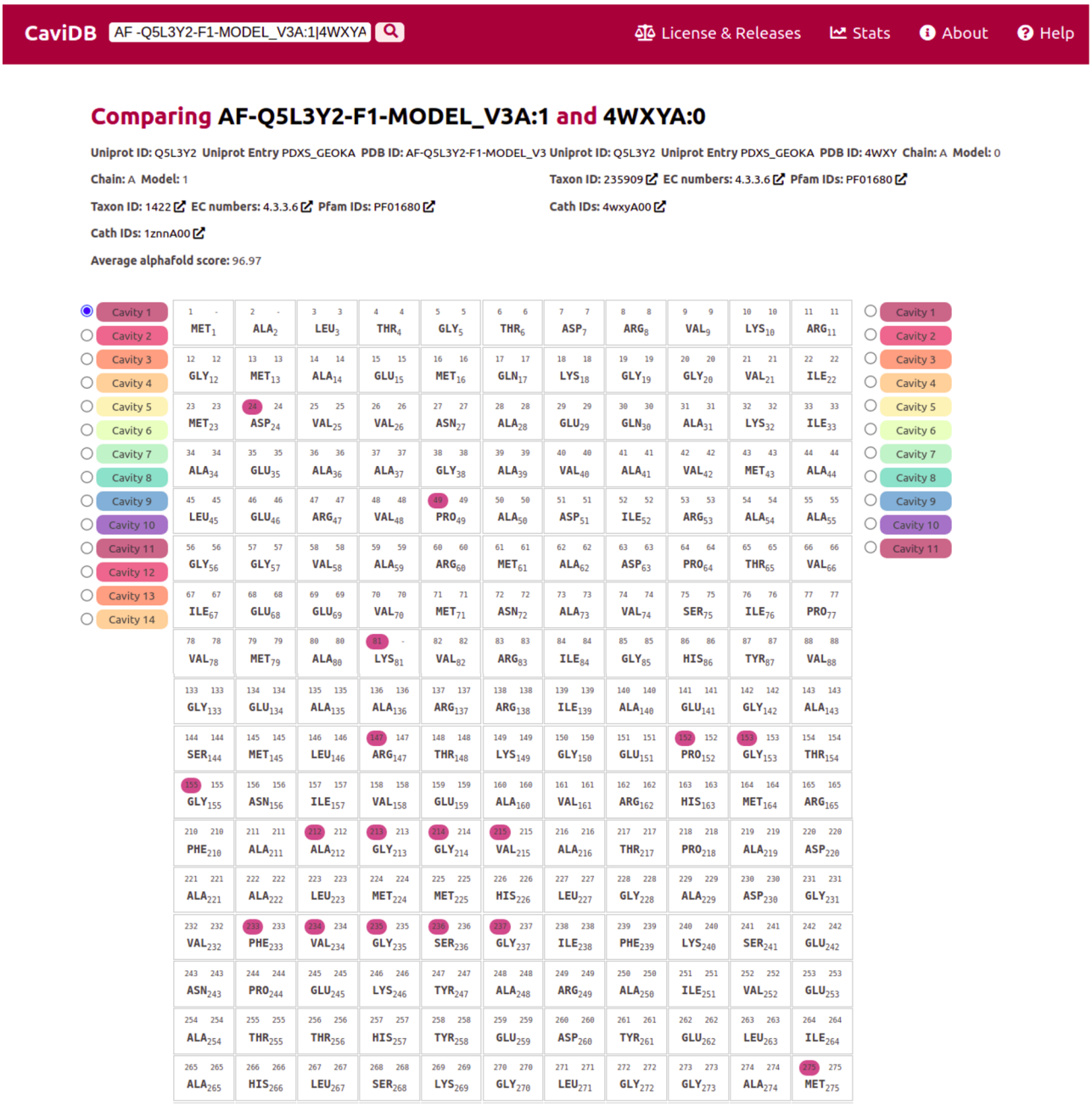

Figure 3. Comparison of the presence of cavities in PLP synthase conformers (UniProt ID Q5L3Y2). Using the expression AF-Q5L3Y2-F1-MODEL_V3A:1|4WXYA:0 to search in CaviDB allows comparing the presence of cavities in both selected conformers. It can be seen that the AlphaFold model contains a biologically relevant cavity (Cavity 1) that contains the key residues described in the bibliography (56). This cavity is absent in other conformers due to the presence of disordered regions.

representing 10 181 Pfam families. With the number of entries in our first release, we were able to characterize a total of 36 136 869 cavities, of which 1 147 034 are druggable. Since CaviDB provides gene IDs and Ensembl IDs, the data of each entry can be easily linked to metabolic pathways and evolutionary information in which each protein might be involved. Moreover, CaviDB is the first repository of information regarding protein cavities that explicitly considers the state-of-the-art AlphaFold models as targets for cavity discovery. Of AlphaFold models in CaviDB, 8042% are above a pLDDT score of 70, offering in this way a substantial amount of 3D models with a considerable level of predicted quality. Furthermore, this is also especially interesting for intrinsically disordered proteins or proteins with flexible regions, in which much of the structural information of biological relevance is not observable to experimental techniques such as X-ray crystallography. There are many tools focused on protein structural characterization and cavity prediction (59, 60), such as CavitySpace, a library focused on cavities in human proteins predicted by AlphaFold, or CavityPlus, a web server for cavity detection. In addition, the number of predicted 3D models is growing very rapidly, characterizing almost the entire known sequence space (https://alphafold.ebi.ac.uk/) (24) and providing unprecedented opportunities to

study the structure–function relationship of proteins. However, as we have shown, CaviDB is not only a tool for determining the properties of protein cavities and their dynamics in a large number of different species and proteins but also provides a simple and accessible way to analyze structural data.

Discussion

Identification of binding cavities is critical for understanding the relationship between protein structure and function and is a crucial step for drug design (13, 59, 61, 62). Since conformational diversity is a key concept for understanding protein biology, CaviDB provides not only a freely accessible, comprehensive database of features of proteins and their cavities but also a simple and user-friendly tool for analyzing the data with a dynamic perspective at multiple levels.

Supplementary material

Supplementary material is available at *Database* online.

Author contributions

A.J.V.R. designed the study and was responsible for the overall planning and management of the project. G.P. and A.J.V.R. were responsible for the theoretical validation. A.J.V.R. and F.L.B. performed software development and implementation. F.L.B. provided technical oversight. A.J.V.R., G.P., N.P. and F.L.B. wrote the manuscript.

Funding

A.J.V.R. is a postdoctoral fellow from National Scientific and Technical Research Council (CONICET). G.P. and N.P. are researchers from CONICET. Universidad Nacional de Quilmes (PUNQ 1004/11); National Agency for Scientific and Technological Promotion (ANPCyT) (PICT-2014-3430, PICT-2013-0232); AWS-CONICET INNOVA 2021 (project 2022011357003403). The funders had no role in the study design, data collection, analysis, decision to publish or preparation of the manuscript.

Conflict of 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

1. Guo,R., Cang,Z., Yao,J. *et al.* (2020) Structural cavities are critical to balancing stability and activity of a membrane-integral enzyme. *Proc. Natl. Acad. Sci. USA*, **117**, 22146–22156.
2. Hasenahuer,M.A., Barletta,G.P., Fernandez-Alberti,S. *et al.* (2017) Pockets as structural descriptors of EGFR kinase conformations. *PLoS One*, **12**, e0189147.
3. Rueda,A.J.V., Monzon,A.M., Ardanaz,S.M. *et al.* (2018) Large scale analysis of protein conformational transitions from aqueous to non-aqueous media. *BMC Bioinformatics*, **19**, 27.
4. Stank,A., Kokh,D.B., Fuller,J.C. *et al.* (2016) Protein binding pocket dynamics. *Acc. Chem. Res.*, **49**, 809–815.
5. Liang,J., Edelsbrunner,H. and Woodward,C. (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.*, **7**, 1884–1897.
6. Kamerlin,S.C.L. and Warshel,A. (2010) At the dawn of the twenty-first century: is dynamics the missing link for understanding enzyme catalysis? *Proteins*, **78**, 1339–1375.
7. Hammes-Schiffer,S. and Benkovic,S.J. (2006) Relating protein motion to catalysis. *Annu. Rev. Biochem.*, **75**, 519–541.
8. Nashine,V.C., Hammes-Schiffer,S. and Benkovic,S.J. (2010) Coupled motions in enzyme catalysis. *Curr. Opin. Chem. Biol.*, **14**, 644–651.
9. Laskowski,R.A., Luscombe,N.M., Swindells,M.B. *et al.* (1996) Protein clefts in molecular recognition and function. *Protein Sci.*, **5**, 2438–2452.
10. Chen,B.Y. and Honig,B. (2010) VASP: a volumetric analysis of surface properties yields insights into protein-ligand binding specificity. *PLoS Comput. Biol.*, **6**, e1000881.
11. Campbell,S.J., Gold,N.D., Jackson,R.M. *et al.* (2003) Ligand binding: functional site location, similarity and docking. *Curr. Opin. Struct. Biol.*, **13**, 389–395.
12. Andersson,C.D., Chen,B.Y. and Linusson,A. (2010) Mapping of ligand-binding cavities in proteins. *Proteins*, **78**, 1408–1422.
13. Weisel,M., Proschak,E., Kriegl,J.M. *et al.* (2009) Form follows function: shape analysis of protein cavities for receptor-based drug design. *Proteomics*, **9**, 451–459.
14. Grimsley,G.R., Scholtz,J.M. and Pace,C.N. (2009) A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci.*, **18**, 247–251.
15. Bartlett,G.J., Porter,C.T., Borkakoti,N. *et al.* (2002) Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.*, **324**, 105–121.
16. Harris,T.K. and Turner,G.J. (2002) Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. *IUBMB Life*, **53**, 85–98.
17. Gutteridge,A. and Thornton,J.M. (2005) Understanding nature's catalytic toolkit. *Trends Biochem. Sci.*, **30**, 622–629.
18. Velez Rueda,A.J. *et al.* (2022) Structural and evolutionary analysis unveil functional adaptations in the promiscuous behavior of serum albumins. *Biochimie*, **197**, 113–120.
19. Schmitt,S., Kuhn,D. and Klebe,G. (2002) A new method to detect related function among proteins independent of sequence and fold homology. *J. Mol. Biol.*, **323**, 387–406.
20. Schmidtke,P. and Barril,X. (2010) Understanding and predicting druggability. A high-throughput method for detection of drug binding sites. *J. Med. Chem.*, **53**, 5858–5867.
21. Zhang,Z., Li,Y., Lin,B. *et al.* (2011) Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics*, **27**, 2083–2088.
22. Faccio,G. (2018) From protein features to sensing surfaces. *Sensors*, **18**, 1204.
23. Touw,W.G., Baakman,C., Black,J. *et al.* (2015) A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.*, **43**, D364–D368.
24. Jumper,J., Evans,R., Pritzel,A. *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**, 583–589.
25. Le Guilloux,V., Schmidtke,P. and Tuffery,P. (2009) Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics*, **10**, 168.
26. Varadi,M., Berrisford,J. and Deshpande,M., PDBe-KB consortium. (2020) PDBe-KB: a community-driven resource for structural and functional annotations. *Nucleic Acids Res.*, **48**, D344–D353.
27. Berman,H.M., Westbrook,J., Feng,Z. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
28. Varadi,M., Anyango,S., Deshpande,M. *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.
29. Olsson,M.H.M., Søndergaard,C.R., Rostkowski,M. *et al.* (2011) PROPKA3: consistent treatment of internal and surface residues

in empirical pK predictions. *J. Chem. Theory Comput.*, **7**, 525–537.
30. Holehouse,A.S., Das,R.K., Ahad,J.N. *et al.* (2017) CIDER: resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. *Biophys. J.*, **112**, 16–21.
31. Müller,A.T., Gabernet,G., Hiss,J.A. *et al.* (2017) modlAMP: Python for antimicrobial peptides. *Bioinformatics*, **33**, 2753–2755.
32. Chapman,B. and Chang,J. (2000) Biopython: Python tools for computational biology. *ACM SIGBIO Newsl.*, **20**, 15–19.
33. Velankar,S., Dana,J.M., Jacobsen,J. *et al.* (2013) SIFTS: structure integration with function, taxonomy and sequences resource. *Nucleic Acids Res.*, **41**, D483–D489.
34. Sillitoe,I., Lewis,T. and Orengo,C. (2015) Using CATH-Gene3D to analyze the sequence, structure, and function of proteins. *Curr. Protoc. Bioinformatics*, **50**, 1–28.
35. Finn,R.D., Coggill,P., Eberhardt,R.Y. *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.
36. Martin,A.C.R. (2005) Mapping PDB chains to UniProtKB entries. *Bioinformatics*, **21**, 4297–4301.
37. Boutet,E., Lieberherr,D., Tognolli,M. *et al.* (2016) UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. *Methods Mol. Biol.*, **1374**, 23–54.
38. Khersonsky,O. and Tawfik,D.S. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.*, **79**, 471–505.
39. Atkins,W.M. (2015) Biological messiness vs. biological genius: mechanistic aspects and roles of protein promiscuity. *J. Steroid Biochem. Mol. Biol.*, **151**, 3–11.
40. Valdés-Jiménez,A., Jiménez-González,D., Kiper,A.K. *et al.* (2022) A new strategy for multitarget drug discovery/repositioning through the identification of similar 3D amino acid patterns among proteins structures: the case of tafluprost and its effects on cardiac ion channels. *Front. Pharmacol.*, **13**, 855792.
41. Gupta,M.N., Alam,A. and Hasnain,S.E. (2020) Protein promiscuity in drug discovery, drug-repurposing and antibiotic resistance. *Biochimie*, **175**, 50–57.
42. Fernández,A., Tawfik,D.S., Berkhout,B. *et al.* (2005) Protein promiscuity: drug resistance and native functions—HIV-1 case. *J. Biomol. Struct. Dyn.*, **22**, 615–624.
43. van der Vusse,G.J. (2009) Albumin as fatty acid transporter. *Drug Metab. Pharmacokinet.*, **24**, 300–307.
44. Di Masi,A., Gullotta,F., Bolli,A. *et al.* (2011) Ibuprofen binding to secondary sites allosterically modulates the spectroscopic and catalytic properties of human serum heme-albumin. *FEBS J.*, **278**, 654–662.
45. Yang,F., Zhang,Y. and Liang,H. (2014) Interactive association of drugs binding to human serum albumin. *Int. J. Mol. Sci.*, **15**, 3580–3595.
46. Kragh-Hansen,U., Chuang,V.T.G. and Otagiri,M. (2002) Practical aspects of the ligand-binding and enzymatic properties of human serum albumin. *Biol. Pharm. Bull.*, **25**, 695–704.
47. Watanabe,H., Tanase,S., Nakajou,K. *et al.* (2000) Role of arg-410 and tyr-411 in human serum albumin for ligand binding and esterase-like activity. *Biochem. J.*, **349**, 813–819.
48. Spanidis,Y., Priftis,A., Stagos,D. *et al.* (2017) Oxidation of human serum albumin exhibits inter-individual variability after an ultra-marathon mountain race. *Exp. Ther. Med.*, **13**, 2382–2390.
49. Sakurai,Y., Ma,S.-F., Watanabe,H. *et al.* (2004) Esterase-like activity of serum albumin: characterization of its structural chemistry using p-nitrophenyl esters as substrates. *Pharm. Res.*, **21**, 285–292.
50. Yang,F., Bian,C., Zhu,L. *et al.* (2007) Effect of human serum albumin on drug metabolism: structural evidence of esterase activity of human serum albumin. *J. Struct. Biol.*, **157**, 348–355.
51. Hollfelder,F., Kirby,A.J. and Tawfik,D.S. (1996) Off-the-shelf proteins that rival tailor-made antibodies as catalysts. *Nature*, **383**, 60–62.
52. Kragh-Hansen,U. (2013) Molecular and practical aspects of the enzymatic properties of human serum albumin and of albumin-ligand complexes. *Biochim. Biophys. Acta*, **1830**, 5535–5544.
53. Monzon,A.M., Rohr,C.O., Fornasari,M.S. *et al.* (2016) CoDNaS 2.0: a comprehensive database of protein conformational diversity in the native state. *Database (Oxford)*, **2016**.
54. Ardanaz,S.M., Velez Rueda,A.J., Parisi,G. *et al.* (2018) A mild procedure for enone preparation catalysed by bovine serum albumin in a green and easily available medium. *Catal. Lett.*, **148**, 1750–1757.
55. Saldaño,T., Escobedo,N., Marchetti,J. *et al.* (2022) Impact of protein conformational diversity on AlphaFold predictions. *Bioinformatics*, **38**, 2742–2748.
56. Smith,A.M., Brown,W.C., Harms,E. *et al.* (2015) Crystal structures capture three states in the catalytic cycle of a pyridoxal phosphate (PLP) synthase. *J. Biol. Chem.*, **290**, 5226–5239.
57. Strohmeier,M., Raschle,T., Mazurkiewicz,J. *et al.* (2006) Structure of a bacterial pyridoxal 5′-phosphate synthase complex. *Proc. Natl. Acad. Sci. USA*, **103**, 19284–19289.
58. Zhu,J., Burgner,J.W., Harms,E. *et al.* (2005) A new arrangement of (beta/alpha)8 barrels in the synthase subunit of PLP synthase. *J. Biol. Chem.*, **280**, 27914–27923.
59. Wang,S., Lin,H., Huang,Z. *et al.* (2022) Cavityspace: a database of potential ligand binding sites in the human proteome. *Biomolecules*, **12**, 967.
60. Konc,J., Lešnik,S., Škrlj,B. *et al.* (2021) ProBiS-dock database: a web server and interactive web repository of small ligand-protein binding sites for drug design. *J. Chem. Inf. Model*, **61**, 4097–4107.
61. Yan,X.-Y., Zhang,S.-W. and He,C.-R. (2019) Prediction of drug-target interaction by integrating diverse heterogeneous information source with multiple kernel learning and clustering methods. *Comput. Biol. Chem.*, **78**, 460–467.
62. Nayal,M. and Honig,B. (2006) On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. *Proteins*, **63**, 892–906.