

CASTp 3.0: Computed Atlas of Surface Topography of proteins

Wei Tian¹, Chang Chen^{1*}, Xue Lei^{1*}, Jieling Zhao² and Jie Liang^{1†}

December 13, 2018

Abstract

Geometric and topological properties of protein structures, including surface pockets, interior cavities and cross channels, are of fundamental importance for proteins to carry out their functions. Computed Atlas of Surface Topography of proteins (CASTp) is a webserver that provides online services for locating, delineating, and measuring these geometric and topological properties of protein structures. It has been widely used since its inception in 2003. In this article, we present the latest version of the webserver, CASTp 3.0. CASTp 3.0 continues to provide reliable and comprehensive identifications and quantifications of protein topography. In addition, it now provides: 1) imprints of the negative volumes of pockets, cavities and channels, 2) topographic features of biological assemblies in the Protein Data Bank, 3) improved visualization of protein structures and pockets, and 4) more intuitive structural and annotated information, including information of secondary structure, functional sites, variant sites, and other annotations of protein residues. The CASTp 3.0 webserver is freely accessible at <http://sts.bioe.uic.edu/castp/>

1 Introduction

Protein structures are complex and are sculpted with numerous surface pockets, internal cavities, and cross channels. These topographic features provide structural basis and microenvironments for proteins to carry out their functions such as ligand binding, DNA interaction, and enzymatic activity. Identification and quantification of these topographic features of proteins are therefore of fundamental importance for understanding the structure-function relationship of proteins [1], in engineering proteins for desired properties [2], and in developing therapeutics against protein targets [3].

2 The CASTp server

The CASTp server aims to provide comprehensive and detailed quantitative characterization of topographic features of proteins [4, 5]. Since its release 15 years ago, the CASTp server has $\sim 45,000$ visits and fulfills

*These authors contributed equally to this work.

†To whom correspondence should be addressed. Tel: +1 312 355 1789; Fax: +1 312 413 2 18; Email: jliang@uic.edu



Figure 1: The surface of the ATP binding pocket of CDK2 (PDB ID 2xmy), computed by CASTp 3.0.

~ 33,000 calculation requests annually. It has been proven to be a useful tool for a wide range of studies, including investigations of signaling receptors [1], discoveries of cancer therapeutics [6], understanding of mechanism of drug actions [7], studies of immune disorder diseases [8], analysis of protein-nanoparticle interactions [9], inference of protein functions [10], and development of high-throughput computational tools [11, 12].

To provide additional useful information and to deliver improved user experience, we introduce here an updated server called CASTp 3.0. All important features of the previous versions of the server are retained, including detecting and characterizing cavities, pockets, and channels of protein structures (Figure 1). In addition, we have substantially extended its functions by providing precomputed topographic features of biological assemblies in the PDB database, as well as imprints of negative volumes of these topographic features. Furthermore, the user interface has been redesigned so it is more intuitive and informative.

2.1 New features of CASTp 3.0

Precomputed results for biological assemblies. The atomic coordinates of a PDB entry in the Protein Data Bank describe an asymmetric unit, which is the minimum structure that can produce the unit cell of the crystal through duplication and crystal symmetry operations. For many PDB entries, the asymmetric unit differs from the biological assembly, which is the unit that has either been shown or is thought to be the functional form of the protein. This difference can be significant, and can make the computation results of a PDB entry biologically irrelevant. For example, PDB entry 2iiv of the outer membrane porin OmpG contains four units as deposited. Its direct computation leads to the detection of a giant artificial pocket (Figure 2A), which obscures the actual functional channel (Figure 2B). To uncover topographic features that are biologically most relevant, the new CASTp server precomputes topographic features for the biological assemblies of PDB entries. Users now are able to navigate effortlessly between results of the asymmetric unit and results of biological assemblies of a PDB entry.

Imprints of negative volumes of topographic features. In the previous versions of CASTp servers, topological features, such as cavities, pockets, and channels, were shown only through the representation of surface atoms participating in their formation. The new CASTp server has added solid imprints of the negative volumes as a visualization option, which can give users direct and intuitive understanding of these

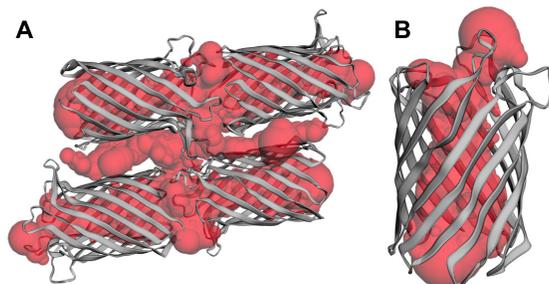


Figure 2: When the asymmetric unit differs from the biological assembly. **(A)** Asymmetric unit in the PDB entry of 2iwv and the artificial giant pocket. **(B)** Biological structure of 2iwv (biological assembly 1) and its channel detected by the CASTp server.

important structural features of proteins [13, 14].

Improved user interface. The visualization techniques employed in previous versions of CASTp servers over 12-year ago are out-of-date and are incompatible with the modern browsers. The new CASTp server now uses 3Dmol.js [15] for structural visualization, which allows users to view, to interact with protein structures, and to examine computation results in modern desktop or mobile web browsers such as Chrome and Firefox. Users can choose the representation style of the atoms that form each topographic feature. The imprint of these features can also be shown with user-selected colors (Figure 3A).

An intuitive sequence panel is also presented to users with secondary structures color-coded, where residues in user-selected pockets are highlighted. Residues annotated from UniProt [16] are also labeled. Both information on topographic feature-forming atoms and UniProt annotations are conveniently displayed when the cursor hovers over the relevant residues. The annotations are also summarized in the annotation panel (Figure 3B). In addition, both the sequence panel and the annotation panel are linked to the structure viewer, so users can conveniently click on one residue on the sequence map to have the structure viewer zoomed into that specific residue of interest.

Furthermore, a floating structure viewer has been added (Figure 3B). When a user inspects information in the sequence panel or the annotation panel, the structure viewer will automatically follow the webpage scrolling of the user, which saves the user from unnecessary and unproductive efforts in scrolling up and down.

3 Input and output

3.1 Input

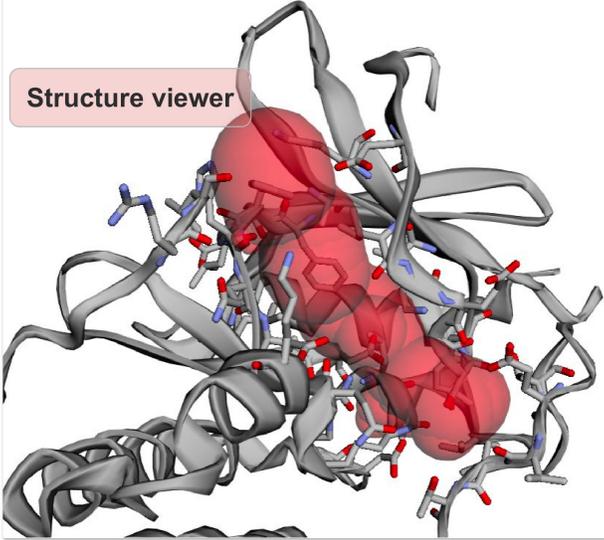
The CASTp server takes protein structures in the PDB format and a probe radius as input for topographic computation. Through the intuitive interface, users can either search for precomputed results using a four-letter PDB ID, or submit their own protein structures to request customized computation. For precomputed results, a default probe radius of 1.4 Å is used, which is the standard value for computing solvent accessible surface area. For customized computation request, users can specify any probe radius desired.

A **2XMY**
Discovery and Characterisation of 2-Anilino-4-(thiazol-5-yl) pyrimidine Transcriptional CDK Inhibitors as Anticancer Agents

Other assembly [2XMY1](#)

Basic info

Structure viewer



Biological assemblies

PocID	Area	Volume
2	461.221	262.111

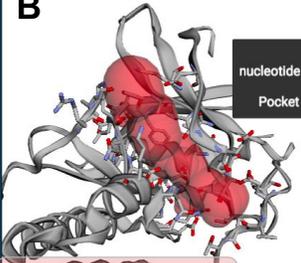
Pocket info

PocID	Chain	SeqID	AA	Atom
2	A	8	GLU	OE2
2	A	10	ILE	O
2	A	10	ILE	CB
2	A	10	ILE	CG1

Atom info

B

Floating viewer



Sequence panel

Chain A.E81
nucleotide phosphate-binding region: ATP
Pocket 2: E81.O
Pocket 4: E81.CA, E81.CB, E81.OE1

F E F L H Q D L K K F M D A S A L T G I P L P L I K S Y L
 V L H R D L K P Q N L L I N T E G A I K L A
 Y T H E V V T L W Y R A P E I L L G C K Y Y
 S L G C I F A E M V T R R A L F P G D S E I D Q L F R I F
 E V V W P G V T S M P D Y K P S F P K W A R Q D F S K V V
 R S L L S Q M L H Y D P N K R I S A K A A L A H P F F Q D
 V L K P V P H L R L

Features

Feature	Position(s)	Description	Reference
nucleotide phosphate-binding region	A: 10-18	ATP	Uniprot: P24941
nucleotide phosphate-binding region	A: 81-83	ATP	Uniprot: P24941
nucleotide phosphate-binding region	A: 129-132	ATP	Uniprot: P24941
active site	A: 127	Proton acceptor	Uniprot: P24941
metal ion-binding site	A: 132	Magnesium	Uniprot: P24941

Annotation panel

Figure 3: The main user interface of the CASTp server. (A) the pocket panel. (B) the sequence and annotation panels.

3.2 Output

The CASTp server identifies all surface pockets, interior cavities and cross channels in a protein structure, and provides detailed delineation of all atoms participating in their formation. It also measures their exact volumes and areas, as well as sizes of the mouth openings if exist. These metrics are calculated analytically, using both the solvent accessible surface model (Richards’ surface) [17] and the molecular surface model (Connolly’s surface) [18]. In addition, the CASTp server also provides imprints of topographic features. These results can be directly downloaded from CASTp server, which can be visualized using either the UCSF Chimera [19] or our PyMOL plugin, CASTpyMOL.

3.3 Case study

Figure 3 illustrates the ATP binding pocket on the cyclin-dependent kinase 2 (CDK2, PDB ID 2xmy) that is automatically identified by the CASTp server. CDK2 is an important enzyme that regulates cell cycle with other CDKs. Many functionally important residues are located in this pockets, including three residues Lys33, Asp86, and Asp145 in the ATP binding site, one residue Asp127 in the catalytic site, two residues Asn132 and Asp145 in Mg binding sites, and one residue Lys89 in the binding site that interacts with CDK7, another important cell-cycle-controlling enzyme. Due to the importance of this pocket, it has been widely studied as a target site for antitumor agents [6].

In the CASTp server, users can choose different representation of pockets. For example, the pocket can be displayed as surface (Figure 1). In Figure 3, the imprint of the pocket is shown in red, and the atoms forming the pocket are shown as sticks. Volume and area measures, and the list of all atoms participating in the pocket formation are also given (Figure 3A). Users can view the pocket from the sequence panel. Important annotations can also be found easily in either the sequence panel or the annotation panel (Figure 3B).

Additional examples illustrating how CASTp computations have been used can be found in the literature [11, 10, 20, 21, 22, 23, 24, 25].

4 Discussion

This paper describes significant updates to the CASTp server. We believe these updates with additional information and improved user experience will continue to facilitate studies of protein structures and functions.

5 Materials and methods

The CASTp server uses the alpha shape method [26] developed in computational geometry to identify topographic features, to measure area and volume, and to compute imprint [27, 28, 13, 14]. The secondary structures are calculated using DSSP [29]. Residue annotations of proteins are obtained from UniProt database [16] and mapped to PDB structures with residue-level information from the SIFTS database [30].

6 Acknowledgements

This work was supported by NIH grant R01CA204962.

6.0.1 Conflict of interest statement.

None declared.

References

- [1] Shigeo Toh, Duncan Holbrook-Smith, Peter J. Stogios, Olena Onopriyenko, Shelley Lumba, Yuichiro Tsuchiya, Alexei Savchenko, and Peter McCourt. Structure-function analysis identifies highly sensitive strigolactone receptors in *Striga*. *Science*, 350(6257):203–207, 2015. [PubMed:26450211] [doi:10.1126/science.aac9476].
- [2] Manfred T. Reetz. *Directed evolution of selective enzymes: Catalysts for organic chemistry and biotechnology*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, nov 2016. [doi:10.1002/9783527655465].
- [3] Jason M. Brouwer, Ping Lan, Angus D. Cowan, Jonathan P. Bernardini, Richard W. Birkinshaw, Mark F. van Delft, Brad E. Sleebs, Adeline Y. Robin, Ahmad Wardak, Iris K. Tan, Boris Reljic, Erinna F. Lee, W. Douglas Fairlie, Melissa J. Call, Brian J. Smith, Grant Dewson, Guillaume Lessene, Peter M. Colman, and Peter E. Czabotar. Conversion of Bim-BH3 from Activator to Inhibitor of Bak through Structure-Based Design. *Molecular Cell*, 68(4):659–672, 2017. [PubMed:29149594] [doi:10.1016/j.molcel.2017.11.001].
- [4] T. Andrew Binkowski, Shapor Naghibzadeh, and Jie Liang. CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Research*, 31(13):3352–3355, 2003. [PubMed:12824325] [doi:10.1093/nar/gkg512].
- [5] Joe Dundas, Zheng Ouyang, Jeffery Tseng, Andrew Binkowski, Yaron Turpaz, and Jie Liang. CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Research*, 34(WEB. SERV. ISS.), 2006. [PubMed:16844972] [doi:10.1093/nar/gkl282].
- [6] Shudong Wang, Gary Griffiths, Carol A. Midgley, Anna L. Barnett, Michael Cooper, Joanna Grabarek, Laura Ingram, Wayne Jackson, George Kontopidis, Steven J. McClue, Campbell McInnes, Janice McLachlan, Christopher Meades, Mokdad Mezna, Iain Stuart, Mark P. Thomas, Daniella I. Zhelleva, David P. Lane, Robert C. Jackson, David M. Glover, David G. Blake, and Peter M. Fischer. Discovery and characterization of 2-anilino-4-(thiazol-5-yl)pyrimidine transcriptional CDK inhibitors as anticancer agents. *Chemistry and Biology*, 17(10):1111–1121, 2010. [PubMed:21035734] [doi:10.1016/j.chembiol.2010.07.016].
- [7] Rameshwar U. Kadam and Ian A. Wilson. Structural basis of influenza virus fusion inhibition by the antiviral drug Arbidol. *Proceedings of the National Academy of Sciences*, 114(2):206–214, 2017. [PubMed:28003465] [doi:10.1073/pnas.1617020114].
- [8] Adam Shahine, Ildiko Van Rhijn, Tan-Yun Cheng, Sarah Iwany, Stephanie Gras, D. Branch Moody, and Jamie Rossjohn. A molecular basis of human T cell receptor autoreactivity toward self-phospholipids. *Science Immunology*, 2(16):eaao1384, 2017. [PubMed:29054999] [doi:10.1126/sciimmunol.aao1384].
- [9] Shivendu Ranjan, Nandita Dasgupta, Sudandiradoss Chinnappan, Chidambaram Ramalingam, and Ashutosh Kumar. A Novel Approach to Evaluate Titanium Dioxide Nanoparticle–Protein Interaction Through Docking: An Insight into Mechanism of Action. *Proceedings of the National Academy of Sciences India Section B - Biological Sciences*, 87(3):937–943, 2017. [doi:10.1007/s40011-015-0673-z].

- [10] Yan Yuan Tseng and Jie Liang. Estimation of amino acid residue substitution rates at local spatial regions and application in protein function inference: a Bayesian Monte Carlo approach. *Molecular Biology and Evolution*, page msj048, 2005. [PubMed:16251508] [doi:10.1093/molbev/msj048].
- [11] T. Andrew Binkowski, Andrzej Joachimiak, and Jie Liang. Protein surface analysis for function annotation in high-throughput structural genomics pipeline. *Protein Science*, 14(12):2972–2981, 2005. [PubMed:16322579] [doi:10.1110/ps.051759005].
- [12] Federico Melaccio, María del Carmen Marín, Alessio Valentini, Fabio Montisci, Silvia Rinaldi, Marco Cherubini, Xuchun Yang, Yoshitaka Kato, Michael Stenrup, Yoelvis Orozco-Gonzalez, Nicolas Ferré, Hoi Ling Luk, Hideki Kandori, and Massimo Olivucci. Toward Automatic Rhodopsin Modeling as a Tool for High-Throughput Computational Photobiology. *Journal of Chemical Theory and Computation*, 12(12):6020–6034, dec 2016. [PubMed:27779842] [doi:10.1021/acs.jctc.6b00367].
- [13] Jerry Osagie Ebalunode, Zheng Ouyang, Jie Liang, and Weifan Zheng. Novel approach to structure-based pharmacophore search using computational geometry and shape matching techniques. *Journal of chemical information and modeling*, 48(4):889–901, 2008. [PubMed:18396858] [doi:10.1021/ci700368p].
- [14] Wei Tian and Jie Liang. On quantification of geometry and topology of protein pockets and channels for assessing mutation effects. *Proceedings - IEEE International Conference on Biomedical and Health Informatics, BHI 2018*, pages 263–266.
- [15] Nicholas Rego and David Koes. 3Dmol.js: Molecular visualization with WebGL. *Bioinformatics*, 31(8):1322–1324, 2015. [PubMed:25505090] [doi:10.1093/bioinformatics/btu829].
- [16] The UniProt Consortium. UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, 45(D1):D158–D169, 2017. [PubMed:27899622] [doi:10.1093/nar/gkw1099].
- [17] B. Lee and F. M. Richards. The interpretation of protein structures: Estimation of static accessibility. *Journal of Molecular Biology*, 55(3), 1971. [PubMed:5551392] [doi:10.1016/0022-2836(71)90324-X].
- [18] M. Connolly. Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 221(4612):709–713, 1983. [PubMed:6879170] [doi:10.1126/science.6879170].
- [19] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J Comput Chem*, 25:1605–1612, 2004. [PubMed:15264254] [doi:10.1002/jcc.20084].
- [20] Prafulla Aryal, Hay Dvir, Senyon Choe, and Paul A. Slesinger. A discrete alcohol pocket involved in GIRK channel activation. *Nature Neuroscience*, 12(8):988–995, 2009. [PubMed:19561601] [doi:10.1038/nn.2358].
- [21] Jieling Zhao, Joe Dundas, Sema Kachalo, Zheng Ouyang, and Jie Liang. Accuracy of functional surfaces on comparatively modeled protein structures. *Journal of structural and functional genomics*, 12(2):97–107, 2011. [PubMed:21541664] [doi:10.1007/s10969-011-9109-z].
- [22] David Jimenez-Morales, Jie Liang, and Bob Eisenberg. Ionizable side chains at catalytic active sites of enzymes. *European Biophysics Journal*, 41(5):449–460, 2012. [PubMed:22484856] [doi:10.1007/s00249-012-0798-4].
- [23] Alisa Glukhova, Vania Hinkovska-Galcheva, Robert Kelly, Akira Abe, James A. Shayman, and John J.G. Tesmer. Structure and function of lysosomal phospholipase A2 and lecithin:cholesterol acyltransferase. *Nature Communications*, 6, 2015. [PubMed:25727495] [doi:10.1038/ncomms7250].

- [24] Reem Smoum, Saja Baraghithy, Mukesh Chourasia, Aviva Breuer, Naama Mussai, Malka Attar-Namdar, Natalya M. Kogan, Bitya Raphael, Daniele Bolognini, Maria G. Cascio, Pietro Marini, Roger G. Pertwee, Avital Shurki, Raphael Mechoulam, and Itai Bab. CB2 cannabinoid receptor agonist enantiomers HU-433 and HU-308: An inverse relationship between binding affinity and biological potency. *Proceedings of the National Academy of Sciences*, 112(28):8774–8779, 2015. [PubMed:26124120] [doi:10.1073/pnas.1503395112].
- [25] Anshika N. Singh, Meghna M. Baruah, and Neeti Sharma. Structure Based docking studies towards exploring potential anti-androgen activity of selected phytochemicals against Prostate Cancer. *Scientific Reports*, 7(1), 2017. [PubMed:28512306] [doi:10.1038/s41598-017-02023-5].
- [26] Herbert Edelsbrunner and Ernst P. Mücke. Three-dimensional alpha shapes. *ACM Transactions on Graphics*, 13(1):43–72, 1994. [PubMed:19775113] [doi:10.1145/174462.156635].
- [27] Jie Liang, Herbert Edelsbrunner, and Clare Woodward. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Science*, 7(9):1884–1897, 1998. [PubMed:9761470] [doi:10.1002/pro.5560070905].
- [28] Herbert Edelsbrunner, Michael Facello, and Jie Liang. On the definition and the construction of pockets in macromolecules. *Discrete Applied Mathematics*, 88:83–102, 1998. [PubMed:9390238] [doi:10.1016/S0166-218X(98)00067-5].
- [29] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, dec 1983. [PubMed:6667333] [doi:10.1002/bip.360221211].
- [30] Sameer Velankar, José M. Dana, Julius Jacobsen, Glen Van Ginkel, Paul J. Gane, Jie Luo, Thomas J. Oldfield, Claire O’Donovan, Maria Jesus Martin, and Gerard J. Kleywegt. SIFTS: Structure Integration with Function, Taxonomy and Sequences resource. *Nucleic Acids Research*, 41(D1), 2013. [PubMed:23203869] [doi:10.1093/nar/gks1258].