

CBPDdb: a curated database of compounds derived from Coumarin–Benzothiazole–Pyrazole

Shailima Rampogu^{1,*}, Mohammed Rafi Shaik², Merajuddin Khan², Mujeeb Khan²,
Tae Hwan Oh³ and Baji Shaik^{3,*}

¹Cachet Big Data Lab, Hyderabad, Telangana 500045, India

²Department of Chemistry, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia

³School of Chemical Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea

*Corresponding author: Email: shailima.rampogu@gmail.com

Correspondence may also be addressed to Baji Shaik. Tel: +82538102775; Email: shaikbaji@yu.ac.kr

Citation details: Rampogu, S., Shaik, M.R., Khan, M. *et al.* CBPDdb: a curated database of compounds derived from Coumarin–Benzothiazole–Pyrazole. *Database* (2023) Vol. 2023: article ID baad062; DOI: <https://doi.org/10.1093/database/baad062>

Abstract

The present article describes the building of a small-molecule web server, CBPDdb, employing R-shiny. For the generation of the web server, three compounds were chosen, namely coumarin, benzothiazole and pyrazole, and their derivatives were curated from the literature. The two-dimensional (2D) structures were drawn using ChemDraw, and the .sdf file was created employing Discovery Studio Visualizer v2017. These compounds were read on the R-shiny app using ChemmineR, and the dataframe consisting of a total of 1146 compounds was generated and manipulated employing the dplyr package. The web server is provided with JSME 2D sketcher. The descriptors of the compounds are obtained using propOB with a filter. The users can download the filtered data in the .csv and .sdf formats, and the entire dataset of a compound can be downloaded in .sdf format. This web server facilitates the researchers to screen plausible inhibitors for different diseases. Additionally, the method used in building the web server can be adapted for developing other small-molecule databases (web servers) in RStudio.

Database URL: https://srampogu.shinyapps.io/CBPDdb_Revised/

Introduction

Computer-aided drug design (CADD) has been instrumental in retrieving plausible inhibitors for a given target for the past three decades (1). This method allows quick screening of compounds at a very low cost (2–4). This CADD is accomplished either by structure-based drug design (SBDD) (5) or by ligand-based drug design (LBDD) (4).

In the SBDD, the presence of the resolved three-dimensional (3D) structure and its inbound ligand (small molecule) plays an important role (5). The interactions between the target and the ligand are critical in understanding the probable binding mode (6, 7) and important residues that might bring out the biological activity. An approach that demonstrates the association between the structure of a compound and its physicochemical properties that determine the biological activity of a compound is called the LBDD (8). The selection of the potential inhibitors is done either by mapping the compounds to a pharmacophore model and molecular docking (9–12) or directly by molecular docking (13). This process can be termed as screening or virtual screening. In 1997, the term virtual screening was first used in the literature (14) and is ‘defined as a set of computational methods that analyses large databases or collections of compounds in order to identify potential hit candidates’ (15). Generally, the search for the

compounds is performed using the chemical libraries (16–19). Usually, the compounds are additionally filtered based on their drug-like properties in order to find favour during the development process.

A detailed account of different web servers embedded with small molecules is given in a study (14), while another web server provides information on different natural compounds with anticancer activity (20). However, a database with compound derivatives of coumarin, benzothiazole and pyrazole has not yet been built. Therefore, in the current study, we have built a web server of Coumarin–Benzothiazole–Pyrazole Derivatives Database (CBPDdb), with derivatives of coumarin, benzothiazole and pyrazole that have demonstrated biological activity towards various diseases.

Materials and methods

Collection of the compounds

In this study, three compounds, namely coumarin, benzothiazole and pyrazole, were selected to search for derivatives in the literature. These compounds were specifically chosen as there are an increasing number of experiments available on the biological activities of these derivatives. These compounds

Received 28 April 2023; Revised 1 August 2023; Accepted 26 August 2023

© The Author(s) 2023. Published by Oxford University Press.

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

have demonstrated varied biological activities and therapeutic applications. We aim to provide the researchers in the field of CADD with most of the compounds with biological activities that would help them discover novel compounds for different diseases.

Specifically, the compounds that have shown biological activity was selected. The derivatives were collected by giving 'compound names and their derivatives', 'compound name + synthesis', 'compound name + biological activity' as the key words in PubMed, NCBI (<https://pubmed.ncbi.nlm.nih.gov/>), Google Scholar and Google.

The polyphenolic compounds coumarin (2H-1-benzopyran-2-one) are a group of oxygenated, colourless, crystalline compounds. These compounds were initially isolated from *Dipteryx odorata* Willd. (Fabaceae) in 1820 by Vogel. This plant is commonly called *Coumarou* (21, 22). Structurally, this compound is made up of a fused benzene ring and α -pyrone ring (23).

Benzothiazole is a heterocyclic structure that is usually bioactive (24). These compounds have a heterocyclic nucleus called a thiazole that confers various biological properties (25). The π -excess aromatic heterocyclic compound pyrazole is a five-membered structure, which is a widely studied group in theazole family (26). The pyrazole template has gained popularity due to its potential therapeutic applications (26). In this compound, the fourth position is preferred for the electrophilic substitution reaction, while the third and the fifth positions are preferred by the nucleophilic reactions (26). To the pyrazole ring, several varied functional groups can be added, substituted, removed or fused to correspondingly synthesize the biologically potent compounds (27). These three compounds have various medicinal applications and hence are chosen to generate a web server with their derivatives (25, 28–33).

Building of the webserver

The two-dimensional (2D) structures were initially sketched employing ChemDraw and saved in .mol format. These structures were upgraded to Discovery Studio Visualizer to obtain their 3D forms and saved them in .sdf format. The therapeutic action of the compounds and the source of curation were prepared in a .csv file that was used to develop the server along with the .sdf files of the compounds. The overview of the web server is given in Figure 1.

To build the web server, the ChemmineR (34) was used that enables compound similarity search, clustering, visualization and function of compounds. Here, we have employed the DT (renderDataTable) to display the data of the compounds into a data table form.

Results

Collection of the compounds and building of CBPDdb

For building a web server that could help the computational chemist, computational biologist or CADD researchers, we have selected coumarins, benzothiazole and pyrazole as a first attempt. A total of 1146 compounds (coumarin, 140; benzothiazole, 451 and pyrazole, 555) were curated from various literature sources. Using the read.SDFset available with ChemmineR, the compounds were imported into the RStudio. The properties/descriptors for these compounds were generated employing propOB. This feature can be adapted post

instalment of ChemmineOB package and the OpenBabel software (35). The so-obtained results are transformed into a data table (DT1).

Furthermore, a different file was generated in .csv format that included the therapeutic action and source of data curation. This file was also read on RStudio using read.csv and a data table (DT2) was created. The two data tables (DT1 and DT2) were merged to join the descriptors with the therapeutic action using the merge function and dplyr. This final data table was displayed on the web server. This pattern was followed to generate the data table for the derivatives, which were displayed under three tabs.

How to use the database

The web server is divided into three major sections: (1) full dataset with filters, (2) full dataset graphical frequency analysis of descriptors and (3) extracting cansmi (smiles) column: filtered data.

Full dataset with filters

This section shows the full dataset of the compounds. The derivatives of the three compounds are included in a separate tab that can be downloaded in the .csv or .sdf formats. Each of the data tables is provided with a top filter that allows the users to choose their choice of descriptors. The filtered data can be downloaded as a .csv file and checked if the selected compounds are downloaded by counter-verifying the Chemical Name in both the files (Supplementary Figure 1). The DT is equipped with clickable links that correspondingly connects to the compound articles. The DT is provided with a search bar that allows the users to search a given input. For instance, if anticancer is given as an input, the results in the DT will display only those compounds with anticancer property.

Full dataset graphical frequency analysis of descriptors

The sidebar panel of the web server is equipped with a histogram plot that displays the frequency of the compounds. The users can select the descriptor from the sidebar panel and view the result as a histogram with the selection option for bins (Supplementary Figure 2).

Extracting the cansmi (smiles) column: filtered data

Section 3 is linked to Section 1, which specifically retrieves a single column upon selection. Once the data is filtered (Section 1), the cansmiName column is selected in Section 1. The selected column with the filtered data will be displayed in Section 3. Here, the display corresponds to the selected tab. The results (filtered data) can be downloaded in the .csv and .sdf formats. The .sdf files can be used to generate the 3D structures (Supplementary Figure 3).

Visualizing the 2D structures

The sidebar panel of the server is embedded with JSME Molecular Editor (Supplementary Figure 2) (36), which facilitates the visualization of the structure of the compounds. The 2D structures can be viewed by giving the SMILES (cansmi, which are the Canonical SMILES) as an input at the Molecular Editor by clicking the downward arrow, selecting the Paste Mol or SDF or SMILES and clicking Accept. The 2D structure

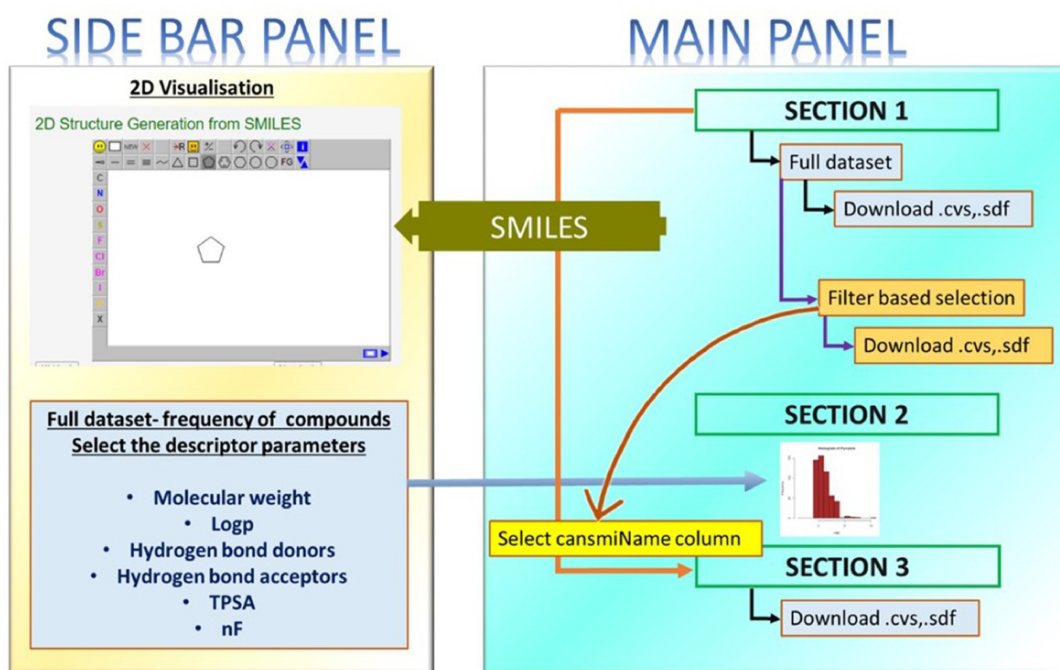


Figure 1. Overview of the web server.

appears on the editor (Supplementary Figure 4). The editor also has other parameters through which the compound's appearance can be changed. Additionally, the users can copy and save the compound in several formats. The modification of the molecules is supported by JSME by clicking the FG (36) (Supplementary Figure 5).

Discussion and conclusion

In order to discover new drugs with therapeutic ability, the CADD process plays a very effective role. In contrast, traditional drug discovery methods are time- and money-consuming processes (2). The term CADD includes saving the compounds, organizing and evaluating them and further modelling the compounds (2). The efficiency of CADD can be evidently seen during the recent pandemic times, when there was an urgency to identify the potential candidate compounds (37–39). Earlier, our group had computationally designed butein analogues that demonstrated anticancer activity (40). Furthermore, these compounds have shown in silico antibacterial activity (41). In another study, computational design of PARP inhibitors was performed against SARS-CoV-2 (42).

Virtual screening is an important step in retrieving the best molecule against a given target (43, 44). The screening process may proceed via SBDD and/or LBDD (43). In either methods, the main purpose is to discover a highly potent putative compound against a target (44, 45). The molecular docking is also included in the virtual screening step. Molecular docking primarily imparts knowledge on the binding mode of the ligand at the active site of the protein (46). The small molecules can be prepared using Gypsum-DL for structure-based virtual screening (47).

Accordingly, in the present study, we have built a web server called the CBPDdb, consisting of derivatives

of compounds from coumarin, benzothiazole and pyrazole curated from different literature sources. These compounds have displayed biological activities such as anticancer, antifungal, antiviral, etc. We believe that these compounds will be useful for the CADD researchers to work with the compounds for using them against several diseases. This web server is equipped with JSME, a 2D sketcher that enables the users to visualize the 2D structures of the compounds. Furthermore, the compounds can be selected based on filter parameters to facilitate the user's choice of compounds.

In the following versions, the web server will be regularly updated to increase the number of compounds with the coumarin, benzothiazole and pyrazole derivatives and other derivatives. Furthermore, the web server will be incorporated with different analysis methods and predictions relevant to medicinal chemistry and CADD.

In conclusion, we believe that this web server could help the computational chemist or computational biologist in their research progress. Furthermore, our attempt may also help the researchers design new small-molecule web servers.

Supplementary Material

Supplementary material is available at *Database* online.

Data availability

The data underlying this article are available in https://github.com/SRampogu/CBPDdb_revised.

Author Contribution

S.R., B.S. and T.H.O. conceived the idea of the project; S.R. built the web server, wrote the manuscript and curated the

compounds from literature; B.S., M.R.S., Me.K. and Muj.K. curated the compounds from literature; M.R.S., Me.K. and Muj.K. provided funding acquisition and B.S. and T.H.O. did sketching of the 2D structures.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

The authors extend their appreciation to the Deputyship for Research and Innovation, “Ministry of Education” in Saudi Arabia for funding this research (IFKSUOR3–103–3).

References

1. Sliwoski,G., Kothiwale,S., Meiler,J. *et al.* (2014) Computational methods in drug discovery. *Pharmacol. Rev.*, **66**, 334–395.
2. Ou-Yang,S., Lu,J., Kong,X. *et al.* (2012) Computational drug discovery. *Acta Pharmacol. Sin.*, **33**, 1131–1140.
3. Sliwoski,G., Kothiwale,S., Meiler,J. *et al.* (2013) Computational methods in drug discovery. *Pharmacol. Rev.*, **66**, 334–395.
4. Yu,W. and Mackerell,A.D. (2017) Computer-aided drug design methods. *Methods Mol. Biol.*, **1520**, 85–106.
5. Batool,M., Ahmad,B. and Choi,S. (2019) A structure-based drug discovery paradigm. *Int. J. Mol. Sci.*, **20**, 2783.
6. Ferreira,L.G., Dos Santos,R.N., Oliva,G. *et al.* (2015) Molecular docking and structure-based drug design strategies. *Molecules*, **20**, 13384–13421.
7. Anderson,A.C. (2012) Structure-based functional design of drugs: from target to lead compound. *Methods Mol. Biol.*, **823**, 359–366.
8. Shim,J. and MacKerell,A.D.J. Jr. (2011) Computational ligand-based rational design: role of conformational sampling and force fields in model development. *Medchemcomm*, **2**, 356–370.
9. Joshi,S.D., Dixit,S.R., Basha,J. *et al.* (2018) Pharmacophore mapping, molecular docking, chemical synthesis of some novel pyrrolyl benzamide derivatives and evaluation of their inhibitory activity against enoyl-ACP reductase (InhA) and Mycobacterium tuberculosis. *Bioorg. Chem.*, **81**, 440–453.
10. Simon,L., Imane,A., Srinivasan,K.K. *et al.* (2017) In silico drug-designing studies on flavanoids as anticancer agents: pharmacophore mapping, molecular docking, and monte carlo method-based QSAR modeling. *Interdiscip. Sci. Comput. Life Sci.*, **9**, 445–458.
11. Tian,X., Zhao,Q., Chen,X. *et al.* (2022) Discovery of novel and highly potent inhibitors of SARS CoV-2 papain-like protease through structure-based pharmacophore modeling, virtual screening, molecular docking, molecular dynamics simulations, and biological evaluation. *Front. Pharmacol.*, **13**, 817715.
12. Rampogu,S. and Lee,K.W. (2021) Pharmacophore modelling-based drug repurposing approaches for SARS-CoV-2 Therapeutics. *Front. Chem.*, **9**, 636362.
13. Rampogu,S., Parameswaran,S., Lemuel,M.R. *et al.* (2018) Exploring the therapeutic ability of fenugreek against type 2 diabetes and breast cancer employing molecular docking and molecular dynamics simulations. *Evidence-based Complement. Altern. Med.*, **2018**, 1943203.
14. Singh,N., Chaput,L. and Villoutreix,B.O. (2021) Virtual screening web servers: designing chemical probes and drug candidates in the cyberspace. *Brief. Bioinform.*, **22**, 1790–1818.
15. Wermuth,C.G., Villoutreix,B., Grisoni,S. *et al.* (2015) Chapter 4: strategies in the search for new lead compounds or original working hypotheses. In: Wermuth CG, Aldous D, Raboisson P *et al.* (eds.) *The Practice of Medicinal Chemistry*. 4th edn. Academic Press, San Diego, pp. 73–99.
16. Perola,E., Xu,K., Kollmeyer,T.M. *et al.* (2000) Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *J. Med. Chem.*, **43**, 401–408.
17. Irwin,J.J. and Shoichet,B.K. (2005) ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.*, **45**, 177–182.
18. Yao,H. (2022) Virtual screening of natural chemical databases to search for potential ACE2 inhibitors. *Molecules*, **27**, 1740.
19. Carracedo-Reboredo,P., Liñares-Blanco,J., Rodríguez-Fernández,N. *et al.* (2021) A review on machine learning approaches and trends in drug discovery. *Comput. Struct. Biotechnol. J.*, **19**, 4538–4558.
20. Mangal,M., Sagar,P., Singh,H. *et al.* (2013) NPACT: naturally occurring plant-based anti-cancer compound-activity-target database. *Nucleic Acids Res.*, **41**, D1124–9.
21. Küpeli Akkol,E., Genç,Y., Karpuz,B. *et al.* (2020) *Coumarins and Coumarin-Related Compounds in Pharmacotherapy of Cancer*. Vol. 12. Cancers (Basel), Basel, Switzerland, p. 1959.
22. Jain,P.K. and Joshi,H. (2012) Coumarin: chemical and pharmacological profile. *J. Appl. Pharm. Sci.*, **2**, 236–240.
23. Venugopala,K.N., Rashmi,V. and Odhav,B. (2013) Review on natural coumarin lead compounds for their pharmacological activity. *Biomed Res. Int.*, **2013**, 963248.
24. Ali,R. and Siddiqui,N. (2013) Biological aspects of emerging benzothiazoles: a short review. *J. Chem.*, **2013**, 345198.
25. Pathak,N., Rathi,E., Kumar,N. *et al.* (2020) A review on anticancer potentials of benzothiazole derivatives. *Mini Rev. Med. Chem.*, **20**, 12–23.
26. Karrouchi,K., Radi,S., Ramli,Y. *et al.* (2018) Synthesis and pharmacological activities of pyrazole derivatives: a review. *Molecules*, **23**, 134.
27. Costa,R.F., Turones,L.C., Cavalcante,K.V.N. *et al.* (2021) Heterocyclic compounds: pharmacology of pyrazole analogs from rational structural considerations. *Front. Pharmacol.*, **12**, 2021.
28. Burger,A. and Sawhney,S.N. (1968) Antimalarials. III. Benzothiazole amino alcohols. *J. Med. Chem.*, **11**, 270–273.
29. Ansari,A., Ali,A. and Asif,M. (2017) Review: biologically active pyrazole derivatives. *New J. Chem.*, **41**, 16–41.
30. Naim,M.J., Alam,O., Nawaz,F. *et al.* (2016) Current status of pyrazole and its biological activities. *J. Pharm. Bioallied Sci.*, **8**, 2–17.
31. Bairagi,S.H., Salaskar,P.P., Loke,S.D. *et al.* (2012) Medicinal significance of coumarins: a review. *Int. J. Pharm. Res.*, **4**, 16–19.
32. Poumale,H.M.P., Hamm,R., Zang,Y. *et al.* (2013) *Coumarins and Related Compounds from the Medicinal Plants of Africa. Medicinal Plant Research in Africa*. Elsevier, Amsterdam, Netherlands, pp. 261–300.
33. Gouda,M.A., Hussein,B.H.M., El-Demerdash,A. *et al.* (2020) A review: synthesis and medicinal importance of coumarins and their analogues (Part II). *Curr. Bioact. Compd.*, **16**, 993–1008.
34. Cao,Y., Charisi,A., Cheng,L.-C. *et al.* (2008) ChemmineR: a compound mining framework for R. *Bioinformatics*, **24**, 1733–1734.
35. Horan,K. and Girke,T. (2023) ChemmineOB: R interface to a subset of OpenBabel functionalities. *R package version 1.38.0*, <https://github.com/girke-lab/ChemmineOB>.
36. Bienfait,B. and Ertl,P. (2013) JSME: a free molecule editor in JavaScript. *J. Cheminform.*, **5**, 24.
37. Muratov,E.N., Amaro,R., Andrade,C.H. *et al.* (2021) A critical overview of computational approaches employed for COVID-19 drug discovery. *Chem. Soc. Rev.*, **50**, 9121–9151.
38. Gurung,A.B., Ali,M.A., Lee,J. *et al.* (2021) An updated review of computer-aided drug design and its application to COVID-19. *Biomed Res. Int.*, **2021**, 8853056.
39. Onawole,A.T., Sulaiman,K.O., Kolapo,T.U. *et al.* (2020) COVID-19: CADD to the rescue. *Virus Res.*, **285**, 198022.

40. Rampogu,S., Kim,S.M., Shaik,B. *et al.* (2021) Novel butein derivatives repress DDX3 expression by inhibiting PI3K/AKT signaling pathway in MCF-7 and MDA-MB-231 cell lines. *Front. Oncol.*, **11**, 712824.
41. Rampogu,S., Shaik,B., Kim,J.H. *et al.* (2023) Explicit molecular dynamics simulation studies to discover novel natural compound analogues as Mycobacterium tuberculosis inhibitors. *Heliyon*, **9**, e13324.
42. Rampogu,S., Jung,T.S., Ha,M.W. *et al.* (2023) Repurposing and computational design of PARP inhibitors as SARS-CoV-2 inhibitors. *Sci. Rep.*, **13**, 10583.
43. Li,Q. (2020) Chapter 4—virtual screening of small-molecule libraries. In: Trabocchi A, Lenci EBT-SMDD (eds.) *Small Molecule Drug Discovery*, Vol. 2020. Elsevier, Amsterdam, Netherlands, pp. 103–125.
44. Ekhteiari Salmas,R., Unlu,A., Bektaş,M. *et al.* (2017) Virtual screening of small molecules databases for discovery of novel PARP-1 inhibitors: combination of in silico and in vitro studies. *J. Biomol. Struct. Dyn.*, **35**, 1899–1915.
45. Cuccioloni,M., Bonfli,L., Cecarini,V. *et al.* (2020) Structure/activity virtual screening and in vitro testing of small molecule inhibitors of 8-hydroxy-5-deazaflavin:NADPH oxidoreductase from gut methanogenic bacteria. *Sci. Rep.*, **10**, 13150.
46. Morris,G.M. and Lim-Wilby,M. (2008) Molecular docking. *Methods Mol. Biol.*, **443**, 365–382.
47. Ropp,P.J., Spiegel,J.O., Walker,J.L. *et al.* (2019) Gypsum-DL: an open-source program for preparing small-molecule libraries for structure-based virtual screening. *J. Cheminform.*, **11**, 34.