Bioinformatics, 37(10), 2021, 1480–1481
doi: 10.1093/bioinformatics/btaa848
Advance Access Publication Date: 25 November 2020
Applications Note

Data and text mining

BEST: a Shiny/R web-based application to easily retrieve cross-related enzyme functional parameters and information from BRENDA

Juan Saez Hidalgo[1,2], Karen Y. Oróstica [1], Anamaria Sanchez–Daza [1,3,*] and Álvaro Olivera–Nappa [1]

[1]Centre for Biotechnology and Bioengineering – CeBiB, Department of Chemical Engineering, Biotechnology and Materials, University of Chile, 8370456 Santiago, Chile, [2]Department of Computer Science, University of Chile, 8370459 Santiago, Chile and [3]Faculty of Physical and Mathematical Sciences, University of Chile, Institute for Cell Dynamics and Biotechnology (ICDB), 8370450 Santiago, Chile

*To whom correspondence should be addressed.

Associate Editor: Wren Jonathan
Received on June 25, 2020; revised on September 3, 2020; editorial decision on September 16, 2020; accepted on September 17, 2020

Abstract

Motivation: BRENDA is the largest enzyme functional database, containing information of 84 000 experimentally characterized enzyme entries. This database is an invaluable resource for researchers in the biological field, which classifies enzyme-related information in categories that are very useful to obtain specific functional and protein engineering information for enzyme families. However, the BRENDA web interface, the most used by researchers with a non-informatic background, does not allow the user to cross-reference data from different categories or sub-categories in the database. Obtaining information in an easy and fast way, in a friendly web interface, without the necessity to have a deep informatics knowledge, will facilitate and improve research in the enzymology and protein engineering field.

Results: We developed the Brenda Easy Search Tool (BEST), an interactive Shiny/R application that enables querying the BRENDA database for complex cross-tabulated characteristics, and retrieving enzyme-related parameters and information readily and efficiently, which can be used for the study of enzyme function or as an input for other bioinformatics tools.

Availability and implementation: BEST and its tutorial are freely available from https://pesb2.cl/best/.

Contact: ana.sanchez@ing.uchile.cl

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Enzymes are a diverse and complex group of proteins, which are essential for most biological processes in living systems and are remarkably relevant in modern technology. In the past, most of these enzymatic properties used to be dispersed in thousands of different scientific publications until the appearance of the BRaunschweig ENzyme DAtabase (BRENDA) (Jeske *et al.*, 2019; Schomburg et al., 2017). BRENDA is currently the primary enzyme and enzyme–ligand information system, centralizing data for 84 000 enzyme entries with their functional parameters and other properties, including enzyme engineering results. BRENDA has three ways to retrieve enzyme-related data: SBML Output, BRENDA Textfile and the SOAP Web Service. The first two are text files that can be processed by the user. The third option is to use the web version of BRENDA, which includes the SOAP web service to query the database. However, this system can only be used by users with knowledge of programming. Still, information retrieval from BRENDA is not particularly intuitive or easy for users with a biological background, especially when information has to be cross-tabulated or the user needs to perform more complex searches. This usage difficulty arises mainly because the web search interface presents the available information classified by parameter, instead of presenting all the information regarding a particular enzyme family. Hence, the user must assemble the information manually to build the database for an enzyme of interest. Also, the visual interface is not customizable, making it impossible to retrieve the desired data, and the tables BRENDA generates may be challenging to download for users with no informatics background. The problematics mentioned above present an unnecessary bottleneck for research work and should be corrected.

In this work, we present the Brenda Easy Search Tool (BEST), an interactive Shiny/R web-based application that enables the user to readily utilize different filtering, visualization and data analysis methods on data from BRENDA. In particular, BEST enables a user to (i)perform multiple enzyme searches with the ability to download and visualize the data as tables or graphics, with low execution time, (ii)perform searches for multiple functional parameter criteria

on one or more enzyme classes, filtering by UniProtKB code or by phylogenetic, (iii)cross the available information on different functional parameters for one or more enzymes, personalizing the information shown in a table and (iv)generated tables can be easily modified by using standard software like Excel.

2 Materials and methods

BEST uses the Shiny/R package and Java built-in functions to connect with the SOAP Web Service from the BRENDA database. The use of BEST requires the user to have a BRENDA account. Through built-in functions in the Brenda SOAP, available on GitHub (https://github.com/StarBrand/BrendaSOAP), user can readily query the information in BRENDA. The tool retrieves information using the numerical classification scheme known as the Enzyme Commission number (EC number) or by the functional enzyme name.

3 Usage

BEST is designed to retrieve, visualize and analyse all enzyme-related information available in BRENDA. For this, BEST has three main modules: Protein Search, Data Visualization and Analysis.

3.1 Protein search

Protein Table The user has the option to search for enzyme data using EC Number or functional name. When the user selects an enzyme of interest, BEST generates a table referred as the Protein Table. This table contains the EC Number, systematic name, recommended name, UniProt identifier and source organism name for each protein retrieved. The user can optionally add comments from publications and their corresponding PubMedID reference (Fig. 1). Once the Protein Table is generated, the user can select all records or a subset of enzymes, in order to retrieve their functional parameters, PDB codes, UniProt identifiers or amino acid sequences. The user can download the generated table in different formats, such as .txt, .csv and others, allowing it to be easily processed in Excel or other common software.

Physicochemical parameters From the Protein Table, the user can select the physicochemical parameters to be searched for in Brenda for each protein selected. BEST allows retrieving KM values, turnover numbers, temperature and pH ranges, pI values, among other reported parameters. The result is a Parameter Table, which contains the EC number, organism, systematic name, recommended name and the retrieved parameters for each protein. The user can also optionally filter the results by ranges of parameter values through slider bars. In this way, the filtered functional parameters can be downloaded according to needs and interests.

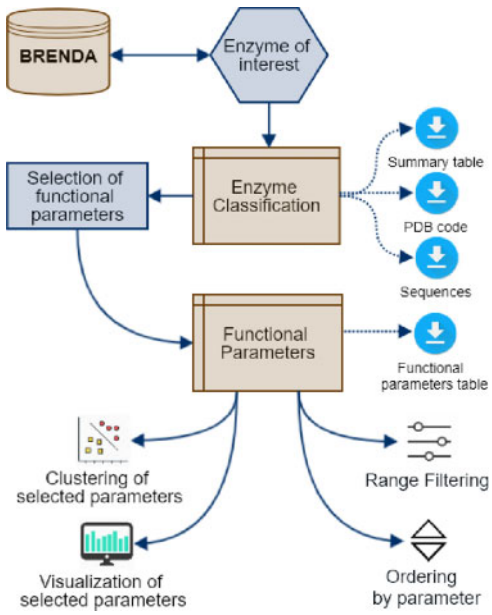

Fig. 1. Workflow of BEST. The user can choose an enzyme using the EC number. Resulting in an enzyme classification table where FASTA sequences could be downloaded, could get the PDB code and a Summary Table. Then, the user can select some functional parameters obtaining a Summary Table, which can be sorted, and filtered by a parameter range

Filtering data and additional information BEST allows the user to obtain the amino acid sequence in FASTA format and the PDB code for all the records found in the Protein Table or for a list of proteins of interest chosen by the user. These features are important since they facilitate data to be used as input for later analyses, such as multiple alignments, BLAST searches, structure modelling, among others.

3.2 Visualization and analysis

An essential feature of BEST is allowing the user to visualize and analyse the values of functional parameters obtained in the Parameter Table through interactive graphics, such as histograms, heatmaps and scatterplots. These features can help the user to decide which parameters are relevant to be considered in their work.

4 Application

As an application example of the operation of BEST, we selected endo-1,4-beta-xylanases, classified as EC 3.2.1.8. We focused our search on three parameters: 'Temperature Optimum', 'Specific activity' and 'Molecular Weight', for enzymes with a 'UNIPROT code', looking for psychrophilic xylanases. After had followed these simple search and filter steps, we easily visualized in the Parameter Table that in the BRENDA database there are no xylanases (with associated UniProt code) reported with an optimal temperature lower than 40 Celsius degrees and that their specific activity and molecular weight are reported at the same time. This type of information is very valuable for a researcher who is characterizing new xylanase that catalyses a reaction at low temperature because the already reported xylanases could be identified quickly. Allowing to focus the research on novel characteristics of the enzyme which is working on (for more details see Supplementary Material).

5 Conclusion

BEST is a tool that allows enzyme researchers without computer skills to obtain, cross-reference information, and visualize enzymatic information in an easy and fast way, being able to advance in their investigations without computer barriers limitations.

Acknowledgments

The authors would like to thank David Medina for technical support and Sebastián Saez for supporting graphic design.The authors would like to thank the researchers at the Centre for Biotechnology and Bioengineering (CeBiB) for beta testing BEST and helping us to improve the tool's usability and utility.

Funding

CeBiB (PIA FB0001). A.S.-D. thanks PAI Programme (I7818010006). K.Y.O. received PhD Scholarship 21182123, ANID, Chile.

Conflict of Interest

none declared.

References

Jeske,L. *et al.* (2019) BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.*, **47**, 542–549.

Schomburg,I. *et al.* (2017) The BRENDA enzyme information system-From a database to an expert system. *J. Biotechnol.*, **261**, 194–206.