BRENDA in 2017: new perspectives and new tools in BRENDA

Sandra Placzek, Ida Schomburg, Antje Chang, Lisa Jeske, Marcus Ulbrich, Jana Tillack and Dietmar Schomburg[*]

Technische Universität Braunschweig, Braunschweig Integrated Centre of Systems Biology (BRICS), Rebenring 56, 38106 Braunschweig, Germany

ABSTRACT

The BRENDA enzyme database (www.brenda-enzymes.org) has developed into the main enzyme and enzyme-ligand information system in its 30 years of existence. The information is manually extracted from primary literature and extended by text mining procedures, integration of external data and prediction algorithms. Approximately 3 million data from 83 000 enzymes and 137 000 literature references constitute the manually annotated core. Text mining procedures extend these data with information on occurrence, enzyme-disease relationships and kinetic data. Prediction algorithms contribute locations and genome annotations. External data and links complete the data with sequences and 3D structures. A total of 206 000 enzyme ligands provide functional and structural data. BRENDA offers a complex query tool engine allowing the users an efficient access to the data via different search methods and explorers. The new design of the BRENDA entry page and the enzyme summary pages improves the user access and the performance. New interactive and intuitive BRENDA pathway maps give an overview on biochemical processes and facilitate the visualization of enzyme, ligand and organism information in the biochemical context. SCOPe and CATH, databases for protein structure classification, are included. New online and video tutorials provide online training for the users. BRENDA is freely available for academic users.

THE BRENDA PROFILE

Founded in 1987 the BRENDA enzyme database (http://www.brenda-enzymes.org/) has developed into the main enzyme functional database system. BRENDA contains manually annotated literature-based data on a wide range of aspects of enzyme function, their metabolic role, involvement in disease processes, genomic and protein sequences, and enzyme structures (1). The characterized enzymes cover all taxonomic groups ranging from eukaryotes, archaea, bacteria to viruses. The enzyme information is categorized according to the Enzyme Commission (EC) nomenclature, the enzyme classification system of the IUBMB (International Union of Biochemistry and Molecular Biology) (2). The enzyme classes are defined according to the type of catalyzed reactions. The data for each entry are linked to the source organism, the protein sequence ID of UniProt (where available, (3)) and to the literature citations. Currently BRENDA contains manually curated data for 82 568 enzymes and 7.2 million enzyme sequences from UniProt. Among the currently stored 6953 EC classes there are 814 which have been deleted or transferred to other classes. These are not removed from the database but remain in the database with a comment on their fate. Four hundred and seventy EC classes are indicated with a 'B' in the four-digit number. These classes are not yet approved by the IUBMB and are marked as 'preliminary BRENDA-supplied EC number'. For most of these entries literature information regarding their specificity for substrates or cofactors is missing, or the reaction products have not been analyzed. These are kept under supervision and completed with data where possible. Since our last publication in 2015 the number of enzyme classes has increased by 478.

Manually annotated data

The manual annotation procedure covers ∼60 criteria in various fields such as kinetic and stability data, catalyzed reactions, procedures for purification, crystallization and cloning. Each value is connected to an organism name and a literature reference. Where available the UniProt protein sequence ID for the enzyme protein is connected and often supplemented with the strain denomination for enzymes from microbial origin. Organs, tissues and cell cultures, based on the terms of the BRENDA Tissue Ontology (BTO) are specified for eukaryotic enzymes (4). Table 1 shows the data content of a selection of data fields. This information originates from ∼146 000 annotated references

---

[*]To whom correspondence should be addressed. Tel: +49 531 391 55202; Fax: +49 531 391 55199; Email: d.schomburg@tu-bs.de

Table 1. Number of entries in selected data fields

Enzyme information	Entries
Substrates and products	407 446
Inhibitors	196 548
Cofactors	14 382
Metals and ions	36 711
Activating compounds	26 761
K_M-values	135 603
K_i-values	38 378
k_{cat}-values	62 445
Specific activity	45 773
IC_{50}- values	49 842
Localization and source/tissue	96 889
Enzyme names and synonyms	102 394
Citations (manually annotated)	146 221
Isolation and preparation/crystallization	88 849
Enzyme structure	158 397
Mutant enzymes	76 451
Enzyme stability	47 281
Enzyme application	15 080

The numbers refer to the combination of enzyme protein, source organism and literature reference. The term enzyme protein refers either to a protein sequence or to a protein isolated from a given organism without its sequence having been determined.

and describes enzymes from ∼11 300 different organisms. Since the last publication in 2015 ∼11 000 references have been manually annotated.

Text mining

Since it is impossible to perform a manual annotation for the complete literature for each enzyme class, text mining procedures have been developed to include more information and give a complete overview on the available literature. AMENDA (**A**utomatic **M**ining of **EN**zyme **DA**ta) and FRENDA (**F**ull **R**eference **EN**zyme **DA**ta) are additional databases (5) on enzyme occurrence created by text mining on PubMed abstracts and titles (6) based on enzyme names and synonyms from BRENDA, the BTO, Gene Ontology (7) and the NCBI Taxonomy (8). In the past 2 years the BTO has been supplemented with ∼200 new terms. These procedures add a large amount of information and a large number of information and citations as can be seen in Table 2. A verification system is integrated to make sure that incorrect connections are removed. The applied dictionaries are revised regularly in order to remove meaningless and misleading terms and to add new ones. In a similar text mining procedure data for diseases that are connected to enzyme malfunction are retrieved and can be found in the connected DRENDA database (**D**isease **R**elated **EN**zyme information **DA**tabase, (9)). The references are classified by support vector machines to indicate the described topic, such as *causal interaction* or *diagnostic usage*. Apart from the manually annotated kinetic data (∼332 000) a text mining process designated as KENDA (**K**inetic **EN**zyme **DA**ta, (10)) adds ∼8500 kinetic values from literature abstracts. An overview on the text mining-based additional data is shown in Table 2.

Table 2. BRENDA data retrieved by text mining

FRENDA	
reference—organism	6 661 235
AMENDA	
reference—organism	3 187 682
reference—organism—source tissue	691 887
reference—organism—subcellular localization	107 112
DRENDA	
diseases connected to enzymes	146 938
references for enzyme-connected diseases	1 227 800
KENDA	
literature abstracts with enzyme kinetic data	11 886

The BRENDA ligands

Great emphasis is laid on the presentation of the substrate-specificity and the enzyme-catalyzed reactions. Substrates, products, cofactors as well as inhibitors, or activating compounds are displayed with their chemical nomenclature but also graphically in structure diagrams. Traditionally chemical compounds are described in manifold ways, with various trivial names, systematic names or abbreviations. In order to bring together all denominations a special ligand section of BRENDA displays these compounds with their structures, a list of synonyms and the enzyme-specific functions. These can be substrates, products, inhibitors, cofactors, activating compounds or the kinetic values. Currently 207 720 compound names are stored, an increase of 6% relative to our last publication. About 81% of these are furnished with a structure yielding 122 600 different molecules. On average each compound has 1.4 synonyms. However, notably, the frequently applied chloromethylketone inhibitors are attributed with up to 70 different names.

Macromolecular or generic structures are generally not shown. However, some of these are displayed when the reaction takes place on an attached molecule and the body of the macromolecule remains unchanged. In these cases the macromolecule is displayed in an abbreviated form with emphasis on the part where the reaction takes place.

Integration of external data

BRENDA gets major additional value from the imported and connected data. Thus most of the cited reference are linked to their abstracts in PubMed. Enzyme sequences are from UniProt and enzymes structures can be viewed at the PDB (11). The Genome Browser provides the link between enzymes and their genomic context.

RETRIEVING ENZYME INFORMATION ON THE BRENDA WEBSITE

Text- and structure-based searches

The BRENDA enzyme portal offers a wide variety of ways to access enzyme information. If the user intuitively enters a search term on the entry page, the software will at first search enzyme and ligand names and display the amount of results in a table with links to enzyme and ligand information.

The enzyme information table provides a quick overview on the respective enzymes with icons to link to word maps,

protein sequences, the catalyzed reactions and the enzyme summary page with the detailed information for the enzymes.

The ligand information table gives a first overview where the molecule is involved in enzyme function. The user gets links to the nomenclature and structure and to the download of the molfile. Further links lead to the involved enzymes. Aside from searching by name ligands can be searched with their structure or a part of it by using a small drawing tool.

Further text-based queries can be performed separately for each of the information fields or by combining a variety of criteria (advanced search). Finally a full-text search which includes all parts of the database can be performed.

Nomenclature standards, ontologies and dictionaries

Large parts of BRENDA are based on established classifications, dictionaries and ontologies. The enzymes are stored according to the IUBMB system of EC classes. The hierarchical structure is displayed in the EC explorer. Organism names are linked to the taxonomic tree (Tax Tree) of the NCBI. A set of ontologies provides further classification. Among these is the BTO, a valuable tool for localizing enzymes in tissues and cell cultures. The CAVEman human anatomy is based on the Terminologia Anatomica and is the international standard on human anatomic terminology (12). SCOPe and CATH cover the structural classification of enzyme proteins (13,14). The subcellular localization of enzymatic activity is linked to the cellular component branch of the Gene Ontology (15).

Enzyme prediction

For the prediction of enzymatic activity in protein sequences BRENDA provides the EnzymeDetector, developed in 2011 and further improved in 2014 (16). It provides a fast and comprehensive overview of the available functional predictions for enzymes encoded in bacterial and archaeal genomes, thus giving enzyme information beyond the published literature.

Visualization

For a quick overview on the relevance and key aspects each enzyme class is attributed with a word map. They were introduced in 2014 and highlight the perceptions which scientific authors have when they publish facts on enzymes. The words differ in size depending on the frequency of their occurrence in the literature. Different colors indicate various topics. The majority is linked to corresponding BRENDA entries.

The Genome Explorer, introduced in 2006, closes the gap between genomic and enzymatic data and affords the alignment of genomes at a given enzyme-coding gene and its orthologs, thus allowing the user a visual comparison of the genomic environment of the gene in different organisms.

Interactive pathway charts are newly introduced (see detailed description below in BRENDA pathway maps).

NEW DEVELOPMENTS AND MAJOR IMPROVEMENTS

Entry page

An external study on the user experience of the BRENDA website has recently been compiled. The study was based on traffic analysis, user surveys and various further methods for measuring the user experience.

In order to meet the study's recommendations the BRENDA website has been revised. A new and modern easy-to-use interface has been implemented for a better user experience.

Six tiles represent the six main query categories. The text-based queries cover the full-text search, the advanced search, and the possibility to explore enzymes and diseases. The ligand substructure search constitutes the structure-based queries. The enzyme classification, a taxonomy tree, the protein folding and several ontologies can be explored in the third query category. The next category covers visualizations like word maps, the Genome Explorer, functional enzyme parameter statistics and metabolic pathways. Prediction tools like transmembrane helices and the EnzymeDetector amount to a further category. A sixth category extends BRENDA's query methods by supporting tools like the BTO and biochemical reactions (BKM-react, (17)).

A new footer focuses on the essential supporting information and completes the better user experience.

The enzyme summary page

The Enzyme Summary Page in BRENDA gives access to a wide range of information about each enzyme. The various data fields are classified in six main categories—enzyme nomenclature, enzyme–ligand interactions, functional parameters, organism related information, enzyme structure information, molecular properties as well as references and links to other databases. This summary page is most frequently called by the user and thus it is essential that it provides a comprehensive and quick access to the information stored for each specific enzyme. As this information is growing quickly with each update this page had become incomprehensible for many enzymes. Therefore, it has now been extensively optimized.

At the top of the page the user finds several search boxes for filtering the displayed amount of information. The corresponding data can be restricted to one or more organisms by using a selection box. AMENDA or FRENDA text mining results can be included with a check box.

An overview on the scientific context of the enzymes is provided by 3584 word maps which visualize enzyme-specific terms from PubMed titles and abstracts. If there are less than five references or less than five specific terms for the enzyme no word map is created because a statistically relevant term distribution is not expected.

The summary page also offers a variety of additional information and is linked to a large amount of specific BRENDA tools. Structure diagrams are available for reactions and ligands. Organisms expressing the enzyme are linked to the BRENDA TaxTree Explorer. More details about the sequence is shown by clicking at the UniProt ID. All references are linked to a literature page comprising the

△ top   print   hide 32 entries   ⊞   ⊟

INHIBITORS ▲▼	ORGANISM ▲▼	UNIPROT ▲▼	COMMENTARY ▲▼ ✕	LITERATURE ▲▼	IMAGE ▲▼
2-oxoglutarate	Arabidopsis thaliana	-	complete inhibition at 0.5 mM	705148	↻
ADP	⊞ 2 entries				
arginine ⊟	Chlamydomonas reinhardtii	-	pH-optimum for inhibition is 7.5	642353	↻
		-	allosteric enzyme which is inhibited by arginine	642354	↻
	Escherichia coli	-	enzyme has no allosteric properties and its activity is influenced neither by arginine nor by any of the intermediates of the arginine biosynthetic pathway	642354	↻
	Neurospora crassa	-	feedback inhibition is markedly dependent on pH, above pH 9 no inhibition	642357	↻
	Pseudomonas aeruginosa	-	feedback inhibition is markedly dependent on pH, above pH 9 no inhibition	642356	↻
		-	1 mM, 96% inhibition	642361	↻
		-	-	675366	↻
	Saccharomyces cerevisiae	-	feedback regulation of enzyme and N-acetylglutamate synthase is mutually interdependent, enzymes form a complex	486081	↻
	Synechococcus elongatus	-	feedback inhibition	659433	↻
	Thermotoga maritima	Q9X2A4	-	675366	↻
ATP	Escherichia coli	-	-	659704	↻
L-arginine	⊞ 11 entries				
L-arginine methyl ester	Thermotoga maritima	-	1 mM, 80% inhibition	659050	↻
L-canavanine	Chlamydomonas reinhardtii	-	-	642353	↻
L-citrulline	⊞ 2 entries				
MgCl2	⊞ 2 entries				
additional information	Thermotoga maritima	-	not inhibitory: D-arginine, agmatine, citrulline, L-canavanine, L-lysine, L-ornithine, guanidinium ions, urea at 1 mM, 37°C	659050	-

Figure 1. Inhibitor table of acetylglutamate kinase: rows with redundant entries are merged to a single line. The inhibitor arginine is listed with different organisms and commentaries in 10 lines of information. The user can click at the (+)-icon to show the specific data of an inhibitor or click at the (−)-icon to hide it.

extracted data from the respective reference as well as title, authors and abstract.

Additionally to a further optimization of the loading times, the usability has been improved. A new search function is added to display only the entries extracted from a particular reference. The navigation panel on the left side is useful for jumping directly to a desired data field. Each data field is represented by a table and is assigned to one of the six main categories mentioned above. A faster overview is provided by displaying the associated data fields when howering the mouseover a main category.

By default, the tables are sorted numerically or alphabetically by the first column, with the exception of the tables containing kinetic values. These are sorted alphabetically by their respective substrate. However, all columns of a table can be sorted individually by the user.

The enzyme summary page provides the option to hide data fields and display only specific tables of interest. This function is very useful to display large sets of data in a clearly arranged manner. A variety of tools including the options 'hide' and 'show' and the number of table entries can be found above all tables.

The tables containing the largest amount of entries, for example the diseases, references and amino acid sequences, are hidden by default and can be brought out by click. This reduces the loading time and prevents unnecessary scrolling. For clarity reasons the display of entries is lim-

ited to 20 records. The user can either navigate through the entries or display the complete table in a new browser tab.

In the enzyme summary page the commentary column might also take up much space. A white (x)-icon in the table header of the commentary column, which can be used to hide it, helps to reduce details and restrict the information to the essential parts. The original settings can be restored by the option 'show all columns' above the table.

Many records in the tables differ only in a few cells. For this purpose, the rows with redundant contents are merged to a single line. This row contains a (+)-icon and the count of summarized entries. For example the inhibitor arginine for acetylglutamate kinase has been recorded for seven organisms with eight different commentaries (Figure 1). This gives rise to 10 lines of information to this inhibitor. The user has the choice of either displaying all lines with a click at the (+)-icon or just a preview of these data with a mouseover. The (−)-icon behind the term arginine can be used to return to the previous view with the summarized rows. (+)- and (−)-icons can also be seen above the table to display all summarized entries or to hide all redundant entries again.

Additionally, the table tools contain a function to print the table separately in an individually chosen format. In the upper right corner of the page a further print option is added. Apart from the ability to print all tables of the page, it is possible to print only the tables which are visible and are not hidden by the user (see Figure 2).

Figure 2. Several print options on the enzyme summary page: (**A**) It is possible to print all tables of the page or only those, which are currently visible. (**B**) The red bordered function can be used to print a table separately.

A link to the abbreviated version of the enzyme summary page is located below these two print options. This version can be used to get a quick overview about the available data of an enzyme and to take a look at a single table directly. The abbreviated version of the enzyme summary page is divided into four sections—the catalyzed reaction, associated synonyms, a condensed view of the EC tree and an outline of the accessible information of the enzyme together with the number of entries of the appropriate data field. A single table can be displayed by choosing one of the data fields on the left side.

BRENDA pathway maps

The newly integrated BRENDA pathway maps offer an approach toward interactive and intuitive biochemical networks on metabolome level with diverse functionalities. The maps give an overview on metabolic pathways and biochemical processes. They represent an alternate access to enzyme and ligand data and allow the user to visualize the enzyme and ligand information.

The maps have been generated manually in a three-step process. Firstly, one large map is manually drawn with the help of Cytoscape, an open source software platform for visualizing complex networks (18). This manual step in the drawing procedures has been chosen in order to create maps in a design which is familiar to biochemists and similar to
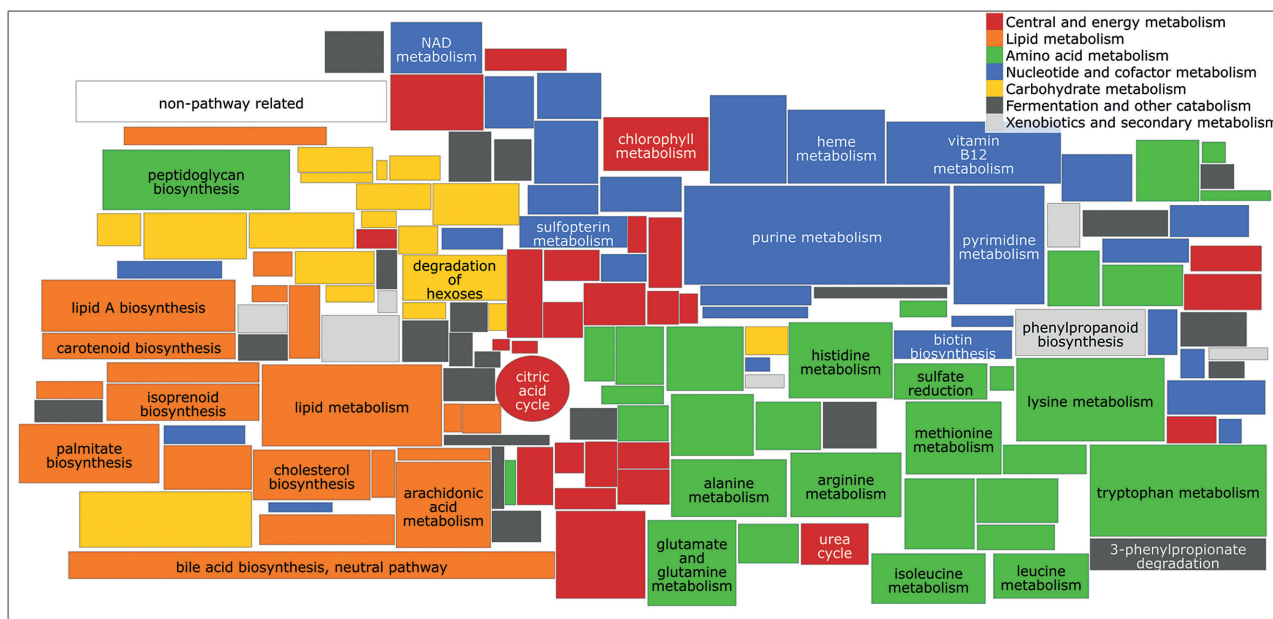
Figure 3. Overview map with pathway icons colored differently with respect to their metabolic roles. The pathways are categorized as central and energy metabolism (red), lipid metabolism (orange), amino acid metabolism (green), nucleotide and cofactor metabolism (blue), carbohydrate metabolism (yellow), fermentation and other catabolism (dark gray) and xenobiotics and secondary metabolism (light gray). This color scheme will be available in Release 2017.1.

maps in student textbooks. Map extensions and modifications are easily performed with Cytoscape. The current map contains 9774 nodes for ligands and enzymes and 10 198 edges which combine enzymes and reactants to reactions. The reactions are organized in 154 pathways of different sizes.

The contents are stored in a relational MySQL database and two main scripts are used for data handling. The complete information in the manually created Cytoscape file is exported as an xgmml-file and then imported into the maps database. An automatic script generates the final BRENDA pathway maps. The map information includes identifiers from BRENDA such as EC numbers and ligand IDs and thus easily combines the maps with biochemical information from the BRENDA database.

The complete set of BRENDA maps consists of an overview map including all pathways schematically with links to the single pathway map. Each of the 154 individual pathway maps can be displayed separately. Currently, the maps cover 1622 enzymes and 1649 metabolites. The pathway icons on the overview map are colored differently with respect to their metabolic roles and categorized as central and energy metabolism, lipid metabolism, amino acid metabolism, nucleotide and cofactor metabolism, carbohydrate metabolism, fermentation and other catabolism, and xenobiotics and secondary metabolism, respectively (see Figure 3).

In the overview map the central metabolic pathways (e.g. citric acid cycle, glycolysis) are placed in the middle. The maps for the nucleotide and amino acids metabolism are arranged mainly in the right section, whereas the maps for carbohydrate and lipid metabolism are found in the left section.

The BRENDA pathway maps offer a variety of search and visualization features. The multicolored overview map described above directly leads to the individual pathway maps. The names of the large pathways can be seen immediately. The notations for small pathways are shown on the tooltip with a mouseover function. However, pathways can also be selected from the alphabetical list of available maps in the left menu section of the page.

The query section offers the possibility to search for individual enzymes via the name or the EC number or parts thereof. Entering a search term will result in a partly colored overview map which highlights the pathways where the enzyme is involved. The metabolites can be searched in the ligand search box. After entering a partial or complete ligand name the respective pathways will be highlighted in the overview map. For the convenience of the users the colors can be changed individually. It is also possible to restrict the search for an enzyme to a particular organism or taxonomic group and in addition to a ligand. Clicking on 'Reset Map' always returns to the entry page.

The organism search can be performed on two sets of database information. The default mode reverts to the manually verified BRENDA data. It is possible to expand the displayed data by including the BRENDA text mining data from the AMENDA and FRENDA data subsets.

Frequently, the user looks for an overview on the metabolic capacity of an organism or of a taxonomic range of organisms. This can be obtained by a combination of an organism search and the visualization of taxonomic information. The taxonomic range is indicated by a coloring scheme. The further up in the taxonomic range the lighter the color becomes. Figure 4 shows the metabolic repertoire for *Zymomonas mobilis* including all taxonomically related organisms up to the genus Bacterium. Thereby, all taxo-
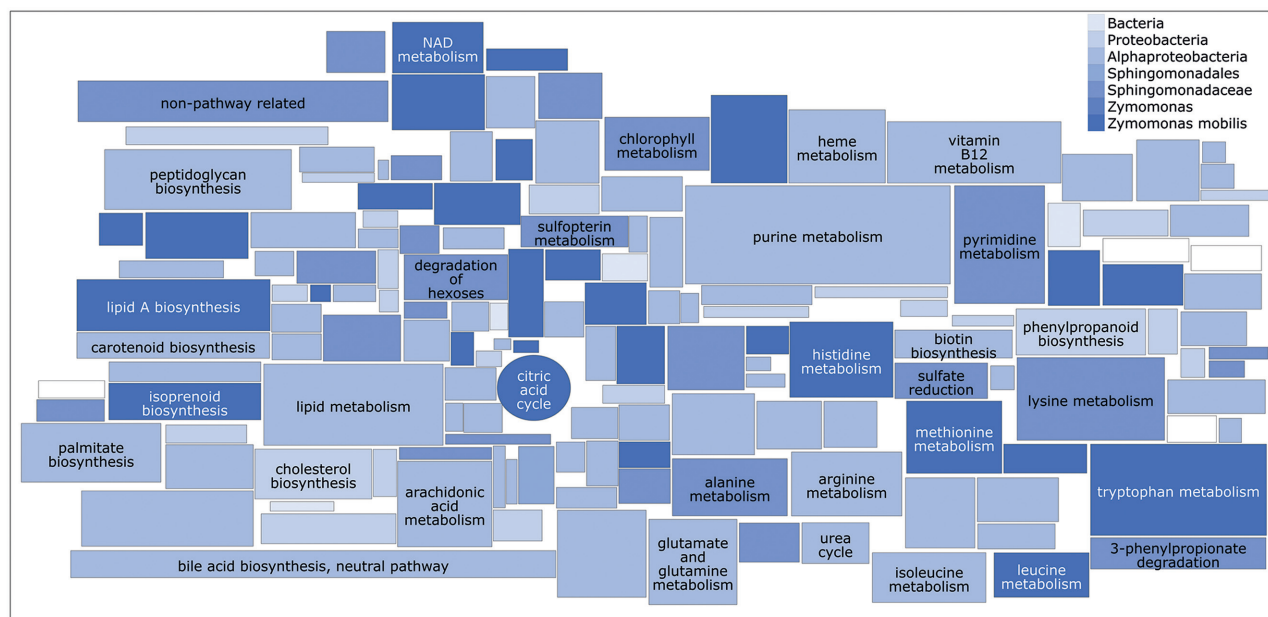
Figure 4. Taxonomic overview map showing the metabolic repertoire for *Zymomonasmobilis* including all taxonomically related organisms up to the genus bacterium. The taxonomic range is indicated by a coloring scheme. The further up in the taxonomic range the lighter the color becomes.

nomic information about enzymes starting from the organism of interest and ending with the bacterial level will be highlighted with decreasing color intensity. The user can select a taxonomic level by switching intermediate levels on or off. Finally, this result map also contains the coverage of highlighted nodes related to the number of all enzyme nodes in a pathway via a mouseover.

In the standard view of the individual pathway maps the metabolites are blue whereas the enzymes are red (see Figure 5). Clicking on a metabolite or an enzyme will lead to the BRENDA information of the particular entry. Co-metabolites can be switched on or off, depending on the user's preferences. All queries for the overview map can also be performed on individual pathway maps. Then either the enzymes are highlighted after an enzyme or an organism search or the metabolites are highlighted following a ligand search. Again, the taxonomic range can be displayed in these maps and may give detailed information on alternate substrate specificities for a metabolic step within a taxonomic range.

The BRENDA website provides access to the metabolic maps from several points. The direct approach is from the 'Metabolic Pathways' icon on the entry page. A more specific and direct link is available from the result page of a BRENDA enzyme search indicated by a symbol in the result table. Thirdly the newly designed enzyme summary page (see section 'The enzyme summary page' above) lists all pathways for an individual EC class. This listing also includes the respective pathways in the KEGG and MetaCyc databases (19,20). The ligand summary page which lists all enzyme-related functions of metabolites also provides links to the respective pathways.

In vitro reactions integrated into BKM-react

The module BKM-react presents a means to search the combined enzyme-catalyzed reactions of the three main metabolic databases BRENDA, KEGG and MetaCyc. The three databases differ in the annotation of the reactions and the amount of available data. The reactions are integrated into BKM-react and matched to a non-redundant database with 64 812 unique reactions currently.

The number of reactions with naturally occurring substrates increased by about 31% in the past five years. Since 2015, the *in vitro* reactions with artificial or modified natural substrates from the BRENDA database are also integrated. However these are clearly distinguished from those labeled as 'natural' in the column 'remark'. This information is not shown in the default view but can be viewed by clicking the respective box.

The distribution of unique reactions between the three databases is illustrated in Figure 6. Four thousand seven hundred and eighty four reactions occur in all three databases. A total of 3275 KEGG reactions, 7260 MetaCyc reactions and 46 526 BRENDA reactions can be originally found only in the respective database.

EnzymeDetector

For analyzing enzymatic functions of microbial organisms an accurate annotation of enzyme predictions is essential. Therefore it is not advisable to rely solely on one source of annotations, but to rather integrate and compare data of several databases and miscellaneous methods. The EnzymeDetector provides an extensive collection of genome-wide enzyme function predictions of more than 5000 archaeal and bacterial genomes including their plasmids by aggregating the information of automated annotation databases and manually revised or experimentally de-
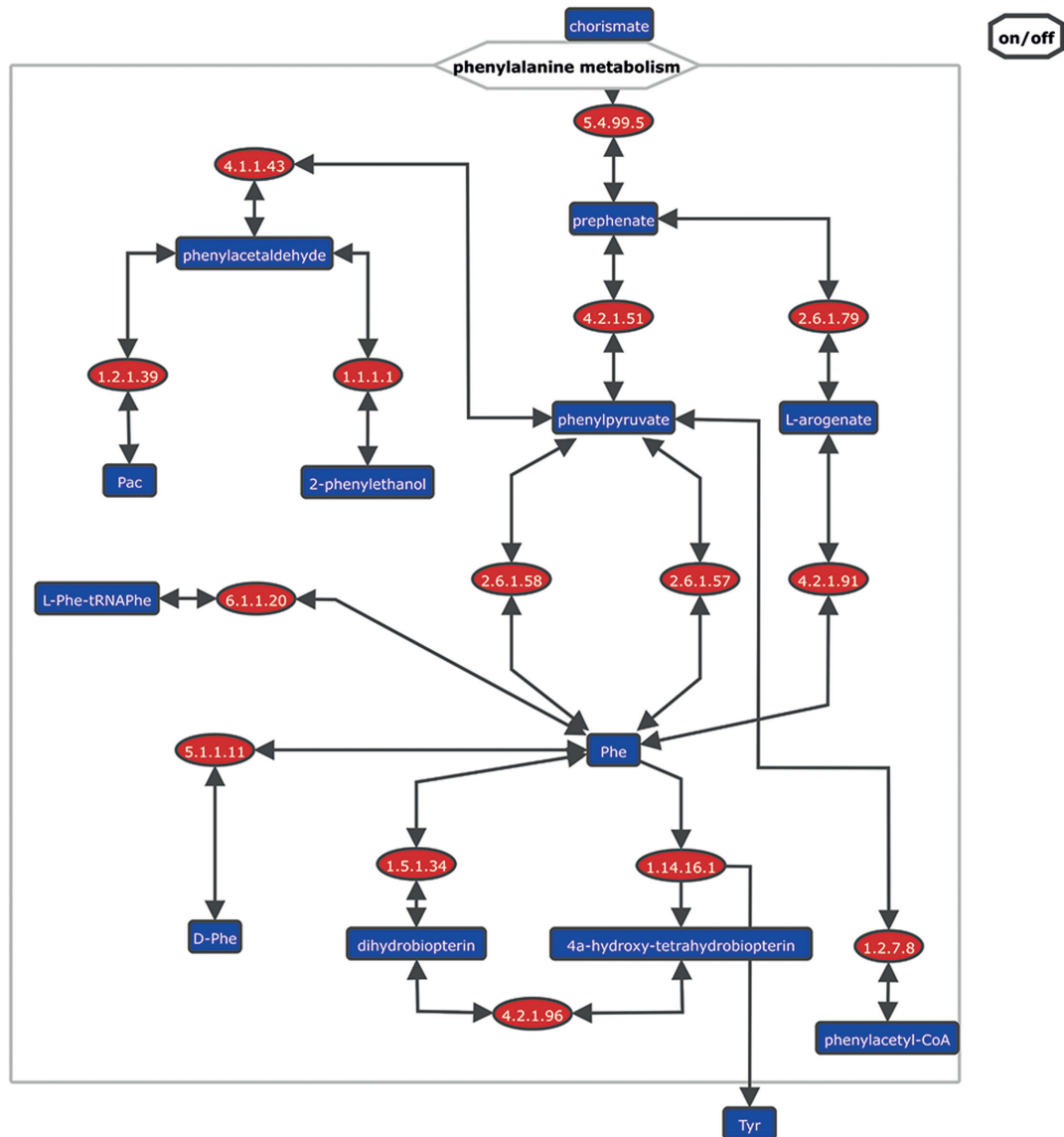
Figure 5. Pathway map of phenylalanine metabolism with metabolites (blue rectangles), enzymes (red ellipses) and hidden co-metabolites. Co-metabolites can be switched on or off using the 'on/off'-button in the right upper corner.

scribed functions. A fast up-to-date overview on sequence annotations of nearly 16 million genes improved by manually created organism-specific enzyme information from BRENDA, sequence pattern searches with BrEPS (21) and a filtered sequence based similarity analysis by BLAST (22) is provided by the web service at http://edbs.tu-bs.de and linked to BRENDA. Additional databases like KEGG orthology data and PATRIC (23) are integrated as annotation sources as well as sequence comparisons with precompiled function specific Pfam (24) Hidden Markov Models (HMMs) (25) to support the decision for the most probable enzymatic function of a gene of interest out of more than 4000 unique assigned enzyme classes.

BRENDA training, tutorials and videos

BRENDA provides online training materials for new and advanced users, including handouts, practice exercises and video tutorials (http://www.brenda-enzymes.org/tutorial.php). They give a short introduction to the database and detailed information on both how to run quick and advanced enzyme searches and how to retrieve the search results. The users learn about the enzyme classification, ligand and sequence searches, protein structure search, the Genome Explorer, the BTO, the TaxTree, the BRENDA pathway maps and the word maps. The tutorials and videos are constantly updated.
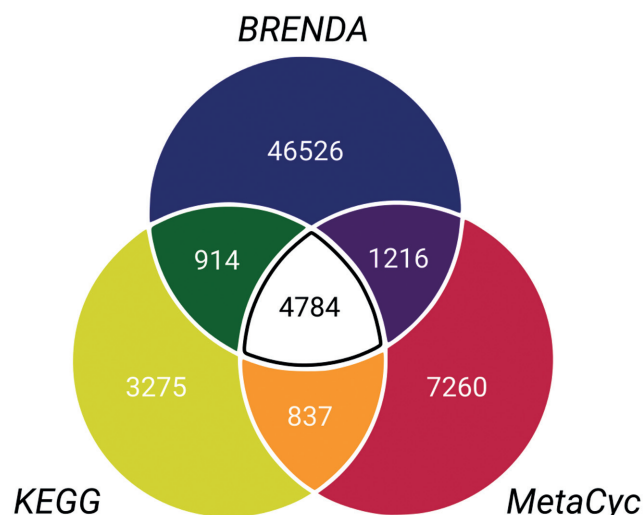
Figure 6. Distribution of unique reactions between BRENDA, KEGG and MetaCyc in BKM-react.

ACCESSIBILITY

All described databases and features are accessible via the main BRENDA website: http://www.brenda-enzymes.org/. The EnzymeDetector can also be accessed via http://edbs.tu-bs.de/ and a direct link to BKM-react is provided via http://bkm-react.tu-bs.de/. Commercial users need a license.

FUNDING

REFERENCES

1. Chang,A., Schomburg,I., Placzek,S., Jeske,L., Ulbrich,M., Xiao,M., Sensen,C.W. and Schomburg,D. (2015) BRENDA in 2015: exciting developments in its 25th year of existence. *Nucleic Acids Res.*, **43**, D439–D446.
2. McDonald,A.G., Boyce,S. and Tipton,K.F. (2009) ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res.*, **37**, D593–D597.
3. The UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
4. Gremse,M., Chang,A., Schomburg,I., Grote,A., Scheer,M., Ebeling,C. and Schomburg,D. (2011) The BRENDA tissue ontology (BTO): the first all-integrating ontology of all organisms for enzyme source. *Nucleic Acids Res.*, **39**, D507–D513.
5. Barthelmes,J., Ebeling,C., Chang,A., Schomburg,I. and Schomburg,D. (2007) BRENDA, AMENDA and FRENDA: the enzyme information system in 2007. *Nucleic Acids Res.*, **35**, D511–D514.
6. NCBI Resource Coordinators. (2014) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **42**, D7–D17.
7. The Gene Ontology Consortium. (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056.
8. Federhen,S. (2012) The NCBI Taxonomy database. *Nucleic Acids Res.*, **40**, D136–D143.
9. Söhngen,C., Chang,A. and Schomburg,D. (2011) Development of a classification scheme for disease-related enzyme information. *BMC Bioinformatics*, **12**, 329.
10. Schomburg,I., Chang,A., Placzek,S., Söhngen,C., Rother,M., Lang,M., Munaretto,C., Ulas,S., Stelzer,M., Grote,A. *et al.* (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res.*, **41**, D764–D772.
11. Rose,P.W., Prlić,A., Bi,C., Bluhm,W.F., Christie,C.H., Dutta,S., Green,R.K., Goodsell,D.S., Westbrook,J.D., Woo,J. *et al.* (2015) The RCSB protein data bank: views of structural biology for basic and applied research and education. *Nucleic Acids Res.*, **41**, D345–D356.
12. Turinsky,A.L., Fanea,E., Trinh,Q., Wat,S., Hallgrímsson,B., Dong,X., Shu,X., Stromer,J.N., Hill,J.W., Edwards,C. *et al.* (2008) CAVEman: standardized anatomical context for biomedical data mapping. *Anat. Sci. Educ.*, **1**, 10–18.
13. Fox,N.K., Brenner,S.E. and Chandonia,J.M. (2014) SCOPe: structural classification of proteins–extended, integrating SCOP and ASTRAL data and classification of new structures. *Nucleic Acids Res.*, **42**, D304–D309.
14. Sillitoe,I., Lewis,T.E., Cuff,A., Das,S., Ashford,P., Dawson,N.L., Furnham,N., Laskowski,R.A., Lee,D., Lees,J.G. *et al.* (2015) CATH: comprehensive structural and functional annotations for genome sequences. *Nucleic Acids Res.*, **43**, D376–D381.
15. Roncaglia,P., Martone,M.E., Hill,D.P., Berardini,T.Z., Foulger,R.E., Imam,F.T., Drabkin,H., Mungall,C.J. and Lomax,J. (2013) The gene ontology (GO) cellular component ontology: integration with SAO (Subcellular Anatomy Ontology) and other recent developments. *J. Biomed. Semantics*, **4**, 20.
16. Quester,S. and Schomburg,D. (2011) EnzymeDetector: an integrated enzyme function prediction tool and database. *BMC Bioinformatics*, **12**, 376.
17. Lang,M., Stelzer,M. and Schomburg,D. (2011) BKM-react, an integrated biochemical reaction database. *BMC Biochem.*, **12**, 42.
18. Shannon,P., Markiel,A., Ozier,O., Baliga,N., Wang,J., Ramage,D., Amina,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **11**, 2498–2504.
19. Kanehisa,M., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, **44**, D457–D462
20. Caspi,R., Altman,T., Billington,R., Dreher,K., Foerster,H., Fulcher,C.A., Holland,T.A., Keseler,I.M., Kothari,A., Kubo,A. *et al.* (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.*, **42**, D459–D471
21. Bannert,C., Welfle,A., aus dem Spring,C. and Schomburg,D. (2010) BrEPS: a flexible and automatic protocol to compute enzyme-specific sequence profiles for functional annotation. *BMC Bioinformatics*, **11**, 589.
22. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **21**, 403–410.
23. Gillespie,J.J., Wattam,A.R., Cammer,S.A., Gabbard,J.L., Shukla,M.P., Dalay,O., Driscoll,T., Hix,D., Mane,S.P., Mao,C. *et al.* (2011) PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infect. Immun.*, **79**, 4286–4298.
24. Finn,R.D., Coggill,P., Eberhardt,R.Y., Eddy,S.R., Mistry,J., Mitchell,A.L., Potter,S.C., Punta,M., Qureshi,M., Sangrador-Vegas,A. *et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.*, **44**, D279–D285.
25. Finn,R.D., Clements,J., Arndt,W., Miller,B.L., Wheeler,T.J., Schreiber,F., Bateman,A. and Eddy,S.R. (2015) HMMER web server: 2015 update. *Nucleic Acids Res.*, **43**, W30–W38.