

BRENDA, the enzyme information system in 2011

Maurice Scheer, Andreas Grote, Antje Chang, Ida Schomburg, Cornelia Munaretto, Michael Rother, Carola Söhngen, Michael Stelzer, Juliane Thiele and Dietmar Schomburg*

Department for Bioinformatics and Biochemistry, Technische Universität Braunschweig, Langer Kamp 19 B, 38106 Braunschweig, Germany

Received September 15, 2010; Revised and Accepted October 15, 2010

ABSTRACT

The BRENDA (BRaunschweig ENzyme Database, <http://www.brenda-enzymes.org>) enzyme information system is the main collection of enzyme functional and property data for the scientific community. The majority of the data are manually extracted from the primary literature. The content covers information on function, structure, occurrence, preparation and application of enzymes as well as properties of mutants and engineered variants. The number of manually annotated references increased by 30% to more than 100 000, the number of ligand structures by 45% to almost 100 000. New query, analysis and data management tools were implemented to improve data processing, data presentation, data input and data access. BRENDA now provides new viewing options such as the display of the statistics of functional parameters and the 3D view of protein sequence and structure features. Furthermore a ligand summary shows comprehensive information on the BRENDA ligands. The enzymes are linked to their respective pathways and can be viewed in pathway maps. The disease text mining part is strongly enhanced. It is possible to submit new, not yet classified enzymes to BRENDA, which then are reviewed and classified by the International Union of Biochemistry and Molecular Biology. A new SBML output format of BRENDA kinetic data allows the construction of organism-specific metabolic models.

INTRODUCTION

BRENDA (BRaunschweig ENzyme DATABASE) is the main repository of manually annotated enzyme data.

The development of the database began in 1987 at the former German National Research GBF (now: Helmholtz Centre for Infection Research) in Braunschweig. Initially, the enzyme data were published as a series of books (Handbook of Enzymes, Springer, 1). The data were continuously curated and improved at the University of Cologne (Institute for Biochemistry) from 1996 to 2007. In this period, the database was transformed into a publicly available database and subsequently converted from a fulltext into a relational database system. Since 2007, BRENDA is maintained and curated at the Technische Universität Braunschweig, Institute for Bioinformatics & Systems Biology.

The website <http://www.brenda-enzymes.org> is visited by more than 180 000 different users each month. All enzyme data are linked to a source organism and to a protein sequence if the sequence has been deposited. The data are manually annotated from the primary literature covering classification and nomenclature, reaction and specificity, functional parameters, occurrence, enzyme structure, application, mutant information and engineered variants, stability, disease, isolation and preparation.

An elaborate query engine provides access to all data stored in the tables of the relational database. The 'Quick search' option allows a simple search in one of the 56 data fields. More sophisticated queries can be performed using the 'Advanced Search' by combining up to 20 search categories for text or numerical data fields. The 2009 newly introduced 'Protein Search' offers a quick access to the protein-specific data in BRENDA. The 'Fulltext Search' option provides a search in all tables in BRENDA, including the 'Commentary' field. The 'Substructure Search' allows the display of enzyme–ligand interactions.

Furthermore a number of different tools afford access to the enzyme-related data, e.g. the 'Ontology Explorer' consisting of biologically and biochemically relevant ontologies, including the BTO (BRENDA Tissue

*To whom correspondence should be addressed. Tel: +49 221 470 6440; Fax: +49 221 470 5092; Email: d.schomburg@tu-bs.de

The authors wish it to be known that, in their opinion, the first four authors should be regarded as joint First Authors.

Ontology), the 'TaxTree Explorer' to search for enzymes or organism in the taxonomic tree, the 'EC Explorer' to search and browse the hierarchical tree of classified enzymes, the 'Genome Explorer' which connects the enzymes with the corresponding genome sequence, and the 'Sequence Search' which is useful for enzymes with known protein sequences.

In addition to the manually annotated data, two databases FRENDA (Full Reference ENzyme DAta) and AMENDA (Automatic Mining of ENzyme DAta) are maintained based on text-mining procedures, which are continuously improved to increase the quality and to reduce the number of false-positive entries.

Since the last publication in 2009 (2) the data content has increased substantially, and new tools and functionalities for the query and data analysis were implemented.

An essential part of BRENDA consists of information on metabolites and small molecules, which interact with enzymes as substrates and products, inhibitors, activating compounds, cofactors or bound metals. The ligands can be displayed as their 2D structures. In this context, new comprehensive ligand information, the 'Ligand View' was implemented to present a summarized view on each enzyme–ligand interaction.

BRENDA now provides a new visualization for the distribution of numeric functional parameters as histograms with the option to show special distribution diagrams for one of the six main Enzyme Commission numbers (EC-classes) or taxonomic groups such as Archaea, Bacteria or Eukaryota. In the future the statistical analysis options will be further developed.

Another new presentation is the display of protein 3D structures showing protein-specific sequence and structure features like active centres, secondary structures, binding sites, sites of post-translational modifications, etc., using the Jmol (<http://www.jmol.org/>) applet.

Since 2000, BRENDA provides enzyme disease-related information obtained from PubMed entries by text-mining procedures (3,4). The text-mining approaches were reprogrammed and upgraded with a sub-classification of these results. This led to an improved quality of the automatic search for relevant literature.

CONTENTS OF BRENDA

Summarized over all information fields BRENDA contains about 3.2 million data manually extracted from the primary literature, each connected with the respective enzyme, the source organism and a literature reference. The database covers all enzyme classes that have been classified by the IUBMB (The International Union of Biochemistry and Molecular Biology). More than 100 000 references were manually annotated resulting in an average of more than 660 single data entries for each EC class. Enzymes from 10 500 different organisms are characterized. Because enzymes frequently are not referred to by their proper recommended names issued by the IUBMB, BRENDA aims at storing complete lists of synonyms. On average, each enzyme has 15 synonymous denominations.

Table 1. BRENDA data in selected data fields

Enzyme information	Entries
Substrates and products	175 371
Inhibitors	155 586
Cofactors	17 958
Metal and ions	27 725
Activating compounds	22 568
K_M -value	103 706
K_I -value	27 263
Turnover number	40 841
Specific activity	39 974
IC_{50}	24 903
Localization and source/tissue	88 637
Enzyme names and synonyms	74 592
Citations (manually annotated)	144 863
Isolation and preparation	74 981
Enzyme structure	88 851
Mutant enzymes	42 536
Stability	40 257
Enzyme application	8 792

The number of entries refers to the combination of enzyme-organism-(protein) value.

Enzyme-catalysed reactions are one of the core areas of the database, with approximately 79 000 different reactions including approximately 13 000 *in vivo* reactions. More than 300 000 functional and kinetic parameters such as K_M -values, turnover numbers and inhibition constants are stored. Since the last publication in 2009 the number of manually annotated references increased by 30% to more than 100 000. About 169 new enzyme classes, classified by IUBMB enzyme committee in 2009 and 2010, are included. BRENDA is not restricted to a specific organism or to specific aspects in biochemistry or molecular biology, but covers a wide range of functional information: classification and nomenclature, reaction and specificity, isolation and preparation, stability and enzyme structure, functional and kinetic parameters, organism-related information, enzyme-related disease information, application and data of mutant and engineered enzymes. Table 1 gives an overview on the data content of selected data fields.

BRENDA provides information about diseases connected to anomalous enzyme function. A modified co-occurrence based text mining procedure is used for the extraction of these data from PubMed abstracts. In release 2009.2 the new text mining method yielded 745 650 distinct entries (enzyme-disease-reference) which increased by 22% to 910 897 entries in 2010.2.

The AMENDA/FRENDA text-mining approach (described in reference 2) was further developed in order to increase the specificity by using for instance improved dictionaries while maintaining a high recall. For example, the total number of enzyme-relevant references found by FRENDA increased from 1.26 million in January 2009 to 1.36 million in June 2010, which represents a growth of nearly 10%.

For viewing the enzymes in their metabolic context the BRENDA enzymes are linked to the pathway maps and modules of the KEGG (5) and MetaCyc (6) databases.

(a)

Reaction & Specificity

- Pathway
- Catalysed Reaction
- Reaction Type
- Natural Substrates and Products
- Substrates and Products
- Substrates
- Natural Substrate
- Products
- Natural Product
- Inhibitors
- Cofactors
- Metals/Ions
- Activating Compounds
- Ligands
- Ligand Views **NEW**

Enzyme Structure

(b)

Search Ligand Views

protoporphyrin begins with Search show 10 results [clear](#)

Formula: contains

:= amino acid sequences := comprehensive online version := show the catalyzed reaction

Results 1 - 9 of 9

Ligand View Link	Compound	Formula
Ligand View	protoporphyrin	C34H34N4O4
Ligand View	Protoporphyrin 4	-
Ligand View	protoporphyrin IX	C34H34N4O4
Ligand View	protoporphyrin monomethyl ester	C35H36N4O4
Ligand View	protoporphyrin-IX	C34H34N4O4
Ligand View	protoporphyrinogen	C34H52N4O4
Ligand View	Protoporphyrinogen IX	C34H40N4O4
Ligand View	Protoporphyrinogen-IX	C34H40N4O4
Ligand View	Protoporphyrinogen-XIII	C34H40N4O4

Results 1 - 9 of 9

Figure 1. (a) Access to the new ligand search form from the BRENDA homepage and (b) The new ligand search form. As an example the search results for all ligand names starting with 'protoporphyrin' are shown.

About 2365 and 1785 EC classes are linked to at least one KEGG or MetaCyc pathway, respectively. About 1538 EC classes are linked to pathways in both databases.

NEW FUNCTIONALITIES

Ligand summary page

For each enzyme, the BRENDA web interface includes an 'Enzyme View' which displays all information for a particular enzyme on a single extended page (e.g. http://www.brenda-enzymes.org/php/result_flat.php4?ecno=1.2.3.4 for EC 1.2.3.4). A similar page, the 'Ligand View', is now available for the BRENDA enzyme ligands, i.e. compounds that act as substrates, products, cofactors, activators or inhibitors in enzyme-catalysed reactions.

Access to the new ligand summary page is provided either by a normal ligand search or by the new ligand query form designated as 'Ligand View' (http://www.brenda-enzymes.org/php/search_result.php4?a=250) (Figure 1a). The full or partial name of the ligand of interest can be entered. A list of appropriate ligand entries is generated whereof each entry provides a hyperlink to its ligand view (Figure 1b).

The ligand summary page (Figure 2) consists of four parts. In the first part, basic information on the ligand is given. This includes information on the nomenclature (the preferred name (BRENDA name), and synonyms),

the molecular structure of small molecules, the molecular formula and the InchiKey (7) for further database or web searches.

The central part specifies the roles of the molecules in the interaction with one or several enzymes. Ligands acting as substrates or products are shown in the context of the enzyme-catalysed reaction. Molecules, which participate in reactions occurring *in vivo* are denoted as 'natural substrate' or 'natural product'. If the ligand acts as an activator or inhibitor this is displayed with the corresponding EC number and the literature references. The latter are listed as BRENDA reference numbers. Each reference provides a hyperlink to the full bibliographic description.

The ligand view also lists the kinetic parameters of an enzyme ligand. The turnover number (k_{cat}) and the Michaelis constant (K_M) for substrates or cofactors are specified. For molecules that act as inhibitors, the inhibition constant (K_i) and the half-maximal inhibitory concentration (IC_{50}) are listed. For each kinetic value at least one literature reference and the EC number is provided.

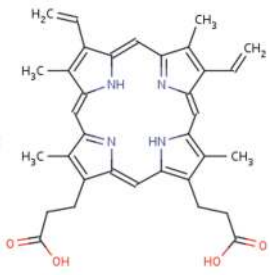
The last part contains lists of the references linked to the PubMed database and links to the PubChem (4) database. The navigation bar provides access to the content of the individual information fields of the ligand summary page (Figure 2). It displays all fields in the order of appearance in the ligand view and is grouped by the sections, 'Basic Ligand Information', 'Roles as Enzyme Ligand' and 'Enzyme Kinetic Parameters'.

BRENDA
The Comprehensive Enzyme Information System

Ligand: protoporphyrin - InChIKey: KSFOVUSSGSKXFI-UJXXFSCMSA-N

Basic Ligand Information

Molecular Structure:



Sum Formula: C₃₄H₃₄N₄O₄
BRENDA Name: protoporphyrin
Synonyms: protoporphyrin-IX, protoporphyrin IX
InChIKey: KSFOVUSSGSKXFI-UJXXFSCMSA-N

Basic Ligand Information

- Molecular Structure
- Sum Formula
- BRENDA Name
- Synonyms
- InChIKey
- Roles as Enzyme Ligand
 - Substrate
 - In Vivo Substrate
 - Product
 - In Vivo Product
 - Cofactor
 - Activator
 - Inhibitor
 - Metals and Ions
- Enzyme Kinetic Parameters
 - k_{cat} Value (Turnover Number)
 - K_M Value
 - K_i Value
 - IC₅₀ Value
- References & Links
 - References
 - Links

Roles as Enzyme Ligand

Substrate in Enzyme-catalyzed Reactions (15 results)

EC_NUMBER	REACTION_PARTNERS	LITERATURE
2.1.1.11	S-adenosyl-L-methionine + protoporphyrin = S-adenosyl-L-homocysteine + protoporphyrin monomethyl ester	485101, 687761
4.99.1.1	protoporphyrin + Fe ²⁺ = protoheme + 2 H ⁺	34938, 34959, 34964, 34962, 34967, 34963, 34966, 34960, 34961, 34968, 34969, 34971, 34949, 34965, 34970, 34972, 34933, 34935, 34939, 34942, 34943, 34945, 34946, 34973, 34936, 34947, 34948, 34932, 34934, 34937, 34944, 34941, 34974, 34930, 34931, 34940, 653206, 665534
4.99.1.1	protoporphyrin + Zn ²⁺ = Zn ²⁺ -protoporphyrin + H ⁺	34963, 34968, 34964, 660525
4.99.1.1	protoporphyrin + Cu ²⁺ = Cu ²⁺ -protoporphyrin + H ⁺	34963


Figure 2. The Ligand Summary Page displays the complete information for any ligand stored in BRENDA. As an example, the first part of the site for protoporphyrin IX is shown (http://www.brenda-enzymes.org/php/ligand_flatfile.php4?brenda_ligand_id=15923). On the upper right, the blue arrow points to an enlarged view of the navigation bar which provides hyperlinks for the direct access to the different information fields.

Visualization of 3D-structures of enzymes

Recently the visualization of 3D structures of enzymes was introduced into the BRENDA database and has now been enhanced. This can be accessed by first doing a '3D-Structure/PDB' query followed by the selection of '3D view'. All available structures of enzymes from the Protein Data Bank (8) can be viewed with the visualization tool Jmol which was integrated into the websites (Figure 3). Specific sites and sequence features from UniProt (9) are included in the representation of enzymes. The user can choose from 19 different sequence features for the visualisation (Table 2).

Functional enzyme parameter distribution

As a first step towards an extended analysis tool the visualization of the distribution of numeric functional parameters is now possible. Distributions of K_M -value, k_{cat} -value, IC_{50} -value, pH optimum, temperature optimum, etc., for selected organism groups or enzyme classes for all enzymes in the BRENDA database can be viewed. The data are shown as bar diagrams for every functional parameter. Figure 4 depicts the distribution of the temperature optimum in Archaea for all enzymes in BRENDA according to the given classification. Access to this tool is provided on the left-handed navigation bar

of the BRENDA homepage at 'Functional Enzyme Parameters' (Icon: .

SBML output form

Systems Biology Markup Language (SBML) is an XML-based format for the transfer of data or information on biological models between different modelling tools or programs (10). In BRENDA the new SBML output option allows the construction of organism-specific metabolic models from BRENDA kinetic data.

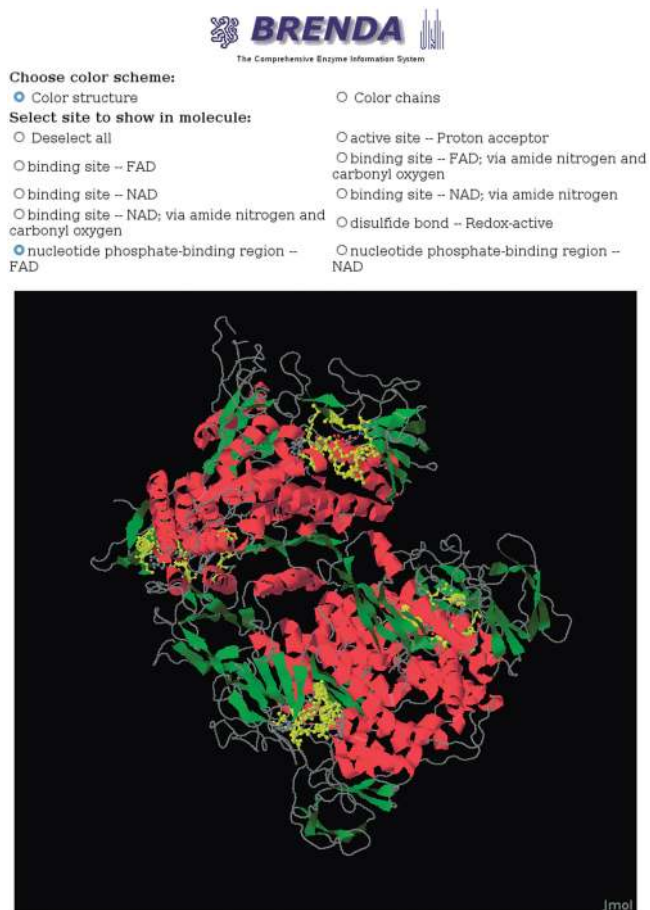
The execution of the MySQL statements is possible via a web form invoked from the browser (Figure 5). The query results are extracted from the database in the CSV format and are immediately converted into valid SBML code. The query yields best results in a species search mode whereas strain-specific searches retrieve substantially fewer results.

The query menu affords the retrieval of organism-specific information on biochemical reactions and enzyme kinetic parameters. For a selected organism the user gets a results table with up to 20 parameters:

- organism, EC Number, literature, enzyme names
- reaction (reaction-ID, substrates, products)
- cofactor

Table 2. Specific sites and sequence features, which can be displayed in the 3D structure of enzymes

Active sites	Coiled-coil regions
Calcium-binding regions	Propeptides
Binding sites	Short sequence motifs
Nucleotide phosphate-binding regions	Cross-links
DNA-binding regions	Disulfide bonds
Glycosylation sites	Domains
Lipid moiety-binding regions	Topological domains
Metal ion-binding sites	Zinc finger regions
Non-standard amino acids	Transmembrane regions
Secondary structure	

**Figure 3.** Visualization of dihydrolipoyl dehydrogenase with marked nucleotide phosphate-binding region.

- kinetic data: (K_M , k_{cat} for substrate and, where available, a commentary, temperature optimum, pH optimum, K_i for inhibitor and, where available, a commentary, temperature optimum, pH optimum).

Enzyme proposal tool

The current list of active EC classes (4299 in Sept. 2010) does not cover the full range of enzymatic activities identified so far. Each year the IUBMB biochemical

nomenclature committee adds about 100–200 new EC classes to this list which have previously undergone a careful review process. Researchers can now submit not yet classified enzymes via the new Enzyme Proposal Tool in BRENDA. It is available on the BRENDA homepage via ‘Propose new enzyme’ in the lower area of the left navigation bar or by using the internet address: http://www.brenda-enzymes.org/enzyme_proposal.

The three forms named ‘Enzyme class’, ‘Submitting person’ and ‘Reference(s)’ have to be completed. Obligatory fields in these forms are yellow coloured and marked with an asterisk. The minimum information required encompasses the name of the proposed enzyme, its catalysed reaction and at least one literature reference describing the new enzyme. For the subsequent correspondence, the name and email address of the submitting person has to be specified. For cited references that are indexed by PubMed only the PubMed ID may be entered. The auto-complete function then adds the respective citation.

After the submission, an email is automatically sent to the user acknowledging the receipt and informing about the further proceeding. The submitted enzyme data are then stored in a relational database and further examined by the BRENDA scientists. The author may be contacted for more details. If the entry finally meets all the criteria for a new EC number, the proposal is forwarded to the IUBMB, where it will be further reviewed. Currently, the proposed enzyme entries can be displayed under the section ‘Show existing submissions’ (http://www.brenda-enzymes.org/enzyme_proposal/display_submissions.php).

FUTURE ENHANCEMENTS

With the huge amount of enzyme data stored in BRENDA, it is ideally suited for model building within systems biology projects where pathways are constructed by substrate/product chains and kinetic parameters without an experimentally determined value are often estimated by elaborated procedures. In the future, BRENDA will provide tools that will strongly enhance the data access for such projects. In addition the high-quality manually extracted data will increasingly be complemented by clearly identified data obtained by text-mining approaches. Besides that, additional information fields, like information on enzyme expression, or the second order rate constant k_{cat}/K_M will be included in the data. The BRENDA team is also actively involved in the STRENDA initiative that provides standards for good practice in the publication of enzyme data.

DATABASE ACCESS AND AVAILABILITY

The BRENDA web sites are freely accessible at <http://www.brenda-enzymes.org>. In addition, all manually annotated BRENDA data can be downloaded as a single text file at the website (http://www.brenda-enzymes.org/brenda_download/index.php).

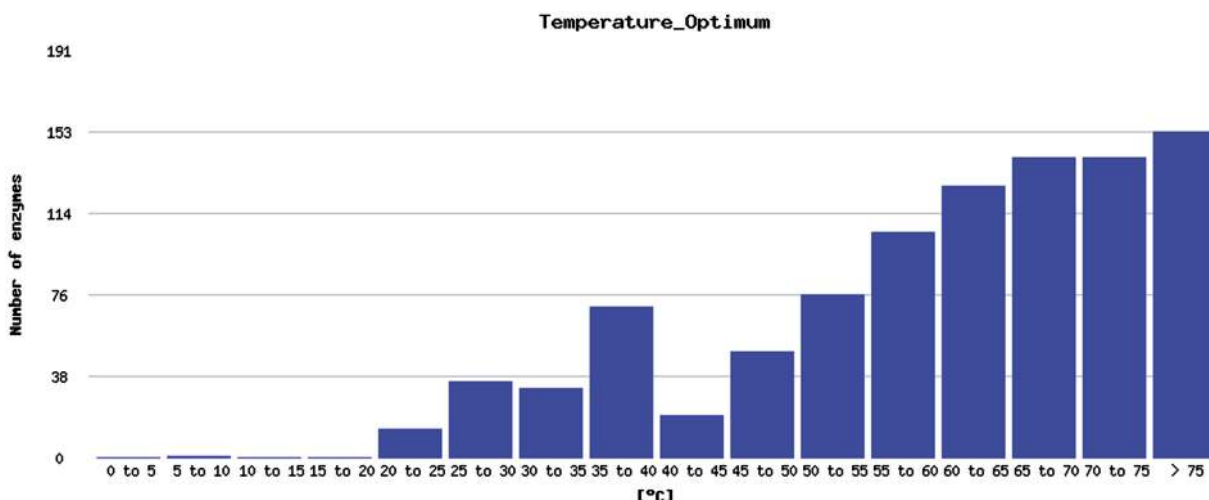


Figure 4. Distribution of the temperature optimum for all Archaea enzymes in BRENDA.

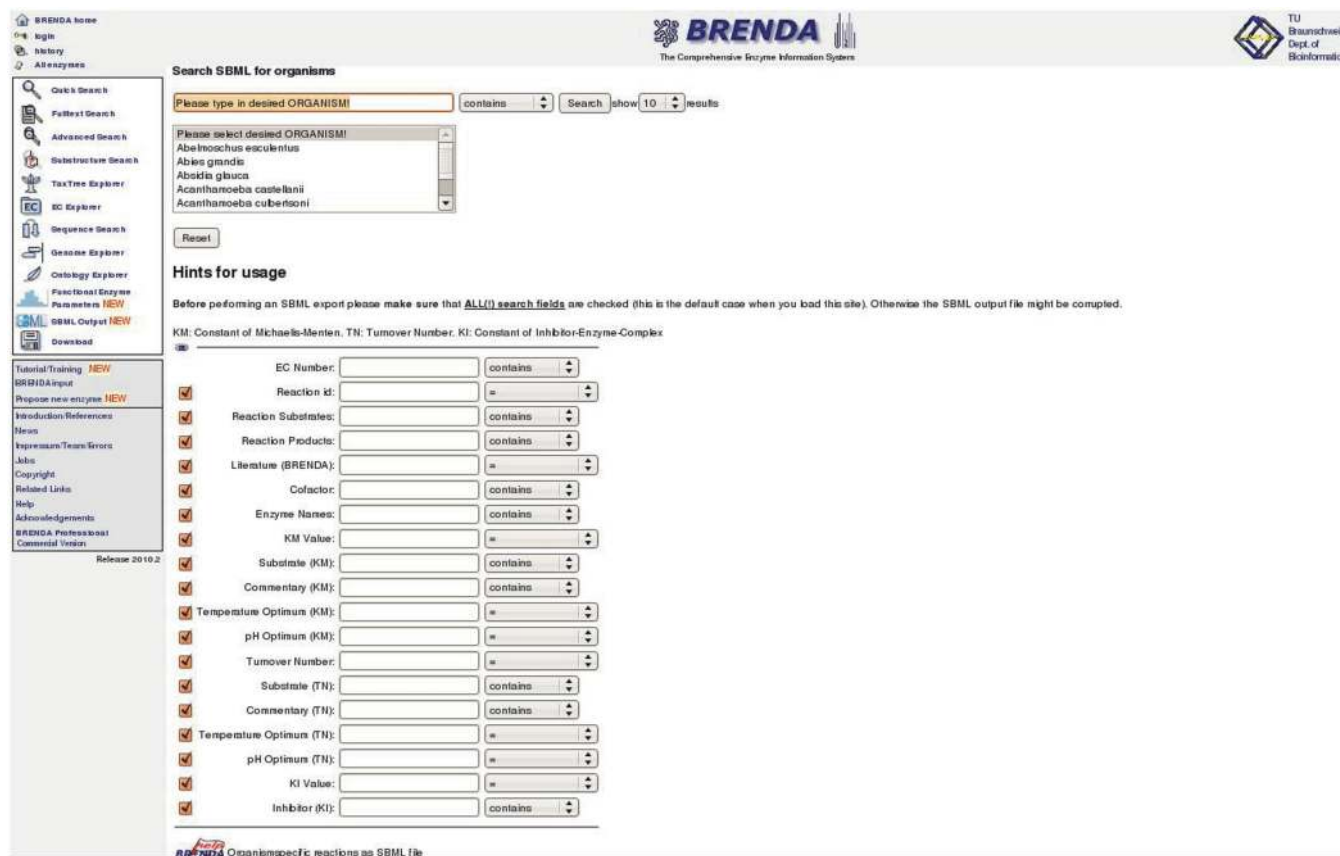


Figure 5. Link to the SBML tool within the BRENDA main menu (SBML button on left panel).

License Information can be viewed at <http://www.brenda-enzymes.org/index.php4?page=information/copy.php4>.

Computer-based access to BRENDA is possible via SOAP (<http://www.brenda-enzymes.org/soap2>) described in detail in reference 2.

Enzyme kinetic data can be obtained via the new SBML output (see above).

ACKNOWLEDGEMENTS

We wish to thank all collaborating scientists who performed the literature annotation and created the molecular structures for the enzyme ligands.

FUNDING

European Union: [FELICS: Free European Life-Science Information and Computational Services: 021902 (RII3); SLING: Serving Life-science Information for the Next Generation: 226073].

Conflict of interest statement. None declared.

REFERENCES

- Schomburg,D. and Schomburg,I. (2001–2009) *Springer Handbook of Enzymes*, 2nd edn. Springer, Heidelberg, Germany.
- Chang,A., Scheer,M., Grote,A., Schomburg,I. and Schomburg,D. (2009) BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Res.*, **37**, D588–D592.
- Schomburg,I., Hofmann,O., Baensch,C., Chang,A. and Schomburg,D. (2000) Enzyme data and metabolic information: BRENDA, a resource for research in biology, biochemistry, and medicine. *Gene Funct. Dis.*, **3–4**, 109–118.
- Sayers,E.W., Barrett,T., Benson,D.A., Bolton,E., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., Dicuccio,M., Federhen,S. *et al.* (2010) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **38**, D5–D16.
- Kanehisa,M., Goto,S., Furumichi,M., Tanabe,M. and Hirakawa,M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–D360.
- Caspi,R., Altman,T., Dale,J.M., Dreher,K., Fulcher,C.A., Gilham,F., Kaipa,P., Karthikeyan,A.S., Kothari,A., Krummenacker,M. *et al.* (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **38**, D473–D479.
- Stein,S.E., Heller,S.R. and Tchekhovskoi,D. (2003) An Open Standard for Chemical Structure Representation: The IUPAC Chemical Identifier. *Proceedings of the 2003 International Chemical Information Conference*. (Nimes), Infonortics, pp. 131–143.
- Velankar,S., Best,C., Beuth,B., Boutselakis,C.H., Cobley,N., Sousa Da Silva,A.W., Dimitropoulos,D., Golovin,A., Hirshberg,M., John,M. *et al.* (2010) PDBe: protein Data Bank in Europe. *Nucleic Acids Res.*, **38**, D308–D317.
- The UniProt Consortium. (2010) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.*, **38**, D142–D148.
- Hucka,M., Finney,A., Sauro,H.M., Bolouri,H., Doyle,J.C., Kitano,H., Arkin,A.P., Bornstein,B.J., Bray,D., Cornish-Bowden,A. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.