D588–D592 *Nucleic Acids Research, 2009, Vol. 37, Database issue* *Published online 4 November 2008*
doi:10.1093/nar/gkn820

BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009

Antje Chang, Maurice Scheer, Andreas Grote, Ida Schomburg and Dietmar Schomburg*

Technical University Braunschweig, Institute for Bioinformatics and Biochemistry, Langer Kamp 19 B, 38106 Braunschweig, Germany

Received September 12, 2008; Revised and Accepted October 13, 2008

ABSTRACT

The BRENDA (BRaunschweig ENzyme DAtabase) (http://www.brenda-enzymes.org) represents the largest freely available information system containing a huge amount of biochemical and molecular information on all classified enzymes as well as software tools for querying the database and calculating molecular properties. The database covers information on classification and nomenclature, reaction and specificity, functional parameters, occurrence, enzyme structure and stability, mutants and enzyme engineering, preparation and isolation, the application of enzymes, and ligand-related data. The data in BRENDA are manually curated from more than 79 000 primary literature references. Each entry is clearly linked to a literature reference, the origin organism and, where available, to the protein sequence of the enzyme protein. A new search option provides the access to protein-specific data. FRENDA (Full Reference ENzyme DAta) and AMENDA (Automatic Mining of ENzyme DAta) are additional databases created by continuously improved text-mining procedures. These databases ought to provide a complete survey on enzyme data of the literature collection of PubMed. The web service via a SOAP (Simple Object Access Protocol) interface for access to the BRENDA data has been further enhanced.

INTRODUCTION

The development of the BRENDA (BRaunschweig ENzyme DAtabase) enzyme information system was started in 1987 at the former German National Research Centre for Biotechnology (now: Helmholtz Centre for Infection Research) in Braunschweig. Originally published as a series of books (1) it was curated and continuously improved at the University of Cologne from 1996 to 2007. In this period it was transformed into a publicly available database (2). Since 2007, BRENDA is maintained and curated at the Technische Universität Braunschweig, Institute of Bioinformatics & Systems Biology. Since the last publication (3), major new developments and improvements have been integrated. A new search option offers direct querying of protein-specific data. AMENDA (Automatic Mining of ENzyme DAta) and FRENDA (Full Reference ENzyme DAta) have been developed as additional databases based on text-mining procedures and are completely reprogrammed to improve the quality and to reduce false positive entries.

BRENDA provides an intuitive search engine with a number of widely different possibilities to access the data, stored in a relational database system. The Quick Search performs a direct search in one of the 53 data fields; the Advanced Search allows a combinatorial search of up to 20 search categories for text or numerical data fields. The Fullltext Search performs a search in all sections of the database including comments. Furthermore the Substructure Search allows to explore enzyme–ligand interactions. Enzyme catalyzed reactions can be viewed as graphical representations.

The Ontology Explorer allows to simultanously search in all biochemically relevant ontologies, among them BTO (BrendaTissueOntology), which is constantly curated by the BRENDA team and contains now approximately 3400 terms on tissues, organs and cell types. The Genome Explorer connects enzymes to genome sequences. The locations of classified enzymes are displayed in their genomic context. The Taxonomy Tree Explorer provides a search for enzymes or organisms in the taxonomic tree. The EC Explorer can be used to browse or search the hierarchical tree of enzymes. The Sequence Search is

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

useful for enzymes with a known protein sequence. It is also possible to search specifically to membrane proteins using the program TMHMM (4).

CONTENTS OF BRENDA

BRENDA contains functional data for all enzyme classes (~4800 entries in six main classes in 2008) that have been classified according to the EC scheme of the IUBMB [International Union of Biochemistry and Molecular Biology, (5)] irrespectively of the enzyme's source.

The range of data in BRENDA is not restricted to specific aspects but includes a wide area of biochemical and molecular properties of enzymes such as

- Classification and nomenclature
- Reaction and specificity
- Functional parameters
- Organism-related information
- Enzyme structure
- Isolation and preparation
- Literature references
- Application and engineering
- Enzyme–disease relationships

All data and information are manually extracted from the primary literature and are connected to the biological source of the enzyme, i.e. the organism, the tissue, the subcellular localization and/or the protein sequence (if available). Since 2007, the amount of manually annotated data has increased by ~18%. An overview on the data in the various sections is displayed in Table 1.

NEW DEVELOPMENTS AND NEW DATA FIELDS

Since the last publication (3) major developments and improvements have been included. The annotation speed has been increased and thus each EC class is updated with manually curated data every 2–2.5 years (compared with 4–4.5 years since 2006). A newly integrated search option offers direct access to protein-specific data. A new numerical data field (IC_{50}, see below) has been added.

Table 1. Data entries in BRENDA

Enzyme information	Single data[a]
Names and synonyms	102 540
Isolation and preparation	68 126
Stability	38 196
Reaction and specificity	475 510
Enzyme structure (including sequences)	716 397
Functional and kinetic parameters	253 795
Organism-related information	106 218
References	113 626
Enzyme application	7718
Mutant enzymes	31 421
IC50 (NEW)	8473

[a]The numbers refer to the combination of enzyme-organism-(protein-) value.

Protein-specific search

In order to search for enzyme properties specific to a certain gene product and for the discrimination of different isoforms of enzymes a new option, the protein-specific search, was introduced to the BRENDA database. The new feature is based on the UniProtKB/Swiss-Prot accession codes (6). Different isoforms can be selected directly by applying the UniProtKB/Swiss-Prot accession code or by using the search mask of the BRENDA database with recommended name, EC number or organism of the protein as keywords. Thus, specific information can be obtained for each isoform of the enzyme in question. Currently, BRENDA contains different isoforms for more than 3300 enzymes in nearly 500 different organisms. As an example Figure 1 shows the numbers of different isoforms of human protein kinases as a diagram.

IC_{50} value

The IC_{50} value is a measure of the inhibitory strength of a compound with respect to an enzyme and represents the concentration of a drug that is required for a 50% inhibition *in vitro*. These values are commonly used for the comparison of antagonist drug potency in pharmacological research. BRENDA now stores IC_{50} values for inhibitors in the relationship inhibitor-enzyme-organism (protein sequence if available). The chemical structures of the inhibitors can be displayed or downloaded in the molfile format.

FRENDA/AMENDA SUPPLEMENTS

While the manually curated BRENDA database aims at keeping up with the newly published literature references, it is, of course, impossible to extract all relevant information for well-investigated enzyme classes. In these cases BRENDA tries to be comprehensive but cannot be complete.

Therefore, since release 2006 FRENDA and AMENDA are provided as additional resources. They represent a

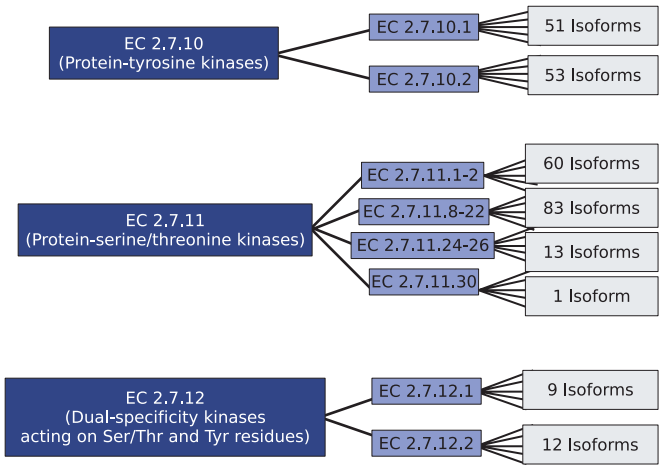

Figure 1. Depicted are the numbers of different isoforms for human protein kinases in BRENDA.

complete survey on enzyme data derived from the titles and abstracts of all articles from the literature database PubMed (7) and are the results of an automated text-mining procedure (3) that was revised and improved during the last months.

FRENDA

FRENDA provides links to all literature references indexed by PubMed that cover enzyme-specific information in combination with the name of the organism or one of its synonyms. Likewise, for each enzyme the EC number and its alternative names are considered which can constitute up to several hundreds for some enzymes.

In order to create term dictionaries, enzyme names and their synonyms from BRENDA and organism names from the NCBI taxonomy (7) are used.

The dictionaries are then employed for screening articles of the PubMed database in a co-occurrence approach [see Figure 2 and ref. (3)].

Due to adaptations in the text-mining procedure FRENDA now also covers abbreviated scientific organism names such as *H. sapiens*, *E. coli* or *A. thaliana* since these are frequently used throughout the literature. Furthermore, due to a significant optimization of the text-mining computing time performance, the MeSH (Medical Subject Heading)-term-based prefiltering step was removed (3) so that nearly 18 million PubMed abstracts were interpreted, thus minimizing the loss of enzyme-specific articles. Consequently, a manual evaluation with over 500 randomly chosen PubMed records yields a comparatively high recall of 84% whereas the corresponding precision of nearly 40% is still acceptable.

Another enhancement of FRENDA is the consequent exclusion of ambiguous enzyme and organism names by manually compiled exclusion lists.

In total more than 1.5 million distinct references (Table 2) are present in the current FRENDA release (June 2008) which means an increase by a factor of three compared with the last described release of June 2006 (3). FRENDA now comprises more than 400 000 organism-specific enzyme hits (Table 2).

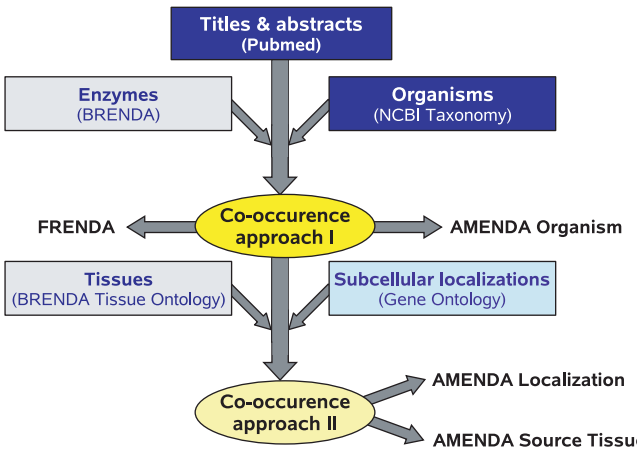

Figure 2. Input data and their source databases used for the co-occurence based text-mining approach that generates the BRENDA supplements FRENDA and AMENDA.

AMENDA

AMENDA is a subset of FRENDA comprising information on enzyme occurrence in organism, localization and source tissue. It includes the most reliable organism-specific enzyme information from FRENDA (AMENDA Enzyme–Organism). In addition, it comprises data on the subcellular localization of enzymes (AMENDA Localization) and the source tissues in which the enzymes are active (AMENDA Source Tissue).

Whereas the information in FRENDA is completely based on co-occurrence of enzyme names and organism names in title and abstract of a paper, for AMENDA a refined and more rigorous text-mining procedure is used (3). The corresponding enzyme–organism combinations are stored in the new supplement AMENDA Enzyme–Organism (see Figure. 2) and rated according to four reliability ranks which reflect the degree of co-occurrence of the enzyme and organism name. The best rank is assigned when both, enzyme and organism name, occur in the title plus in the same sentence of the abstract and, additionally, the EC number is found in the abstract or among the MeSH terms.

As of June 2008 AMENDA Enzyme–Organism comprises more than 225 000 organism-specific enzyme hits.

The tissues for AMENDA Source Tissue and the localizations for AMENDA are obtained from the BRENDA Tissue Ontology (http://www.obofoundry.org/cgi-bin/detail.cgi?id=brenda) and the Gene Ontology (8), respectively [see Figure. 2 and (3)]. From these two resources dictionaries are constructed which are used in addition to the enzyme and organism name dictionary that are employed for building up FRENDA and AMENDA Enzyme–Organism (see above).

Thus, the underlying PubMed reference for each enzyme- and organism-specific hit in FRENDA is further analyzed for co-occurring localization and source tissue names or synonyms. Reliability ranks are assigned to AMENDA Source Tissue and to AMENDA Localization data according to the same principles as for AMENDA Enzyme–Organism. AMENDA Source Tissue contains more than 30 000 and AMENDA Localization more than 60 000 organism-specific hits, respectively (Table 2). The same manual evaluation as mentioned

Table 2. Number of AMENDA and FRENDA entries obtained by the text-mining approach

BRENDA supplement	Number of entries
AMENDA Enzyme–Organism	225 336
AMENDA Localization	29 773
AMENDA Source Tissue	60 890
FRENDA	410 749

AMENDA: 1 036 066 distinct references.
FRENDA: 1 577 622 distinct references.

Table 3. Data fields supported by the BRENDA SOAP interface

Activating_Compound	Molecular_Weight	Reaction_Type
Application	Natural_Product	Recommended_Name
CAS_Registry_Number	Natural_Substrate	Renatured
Cloned	Natural_Substrates_Products	Sequence
Cofactor	Organic_Solvent_Stability	Source_Tissue
Crystallization	Organism	Specific_Activity
Disease	Organism_synonyms	Storage_Stability
EC_Number	Oxidation_Stability	Substrate
Engineering	PDB	Substrates_Products
Enzyme_Names	pH_Optimum	Subunits
General_Stability	pH_Range	Synonyms
Inhibitors	pH_Stability	Systematic_Name
KEGG_Pathway	pI_Value	Temperature_Optimum
KI_Value	Posttranslational_Modification	Temperature_Range
KM_Value	Product	Temperature_Stability
Ligands	Purification	Turnover_Number
Localization	Reaction	
Metals_Ions	Reference	

above yields a precision of 75% and a recall of 25% for enzyme-organism combinations for the two highest reliability ranks (AMENDA reliability $+++$ and $++++$). When the lower AMENDA reliability rank $++$ is also included, the precision is reduced to 56%, whereas the recall increases to 55%.

As the BRENDA enzyme name dictionary contains on average approximately 10 different names per enzyme class (with several hundred names in use for some enzyme classes) the information contents in FRENDA is much more complete than a simple PubMed search using one or two names for an enzyme known to the scientist.

INTEROPERABILITY BY SOAP-BASED WEB SERVICES

A web service is a software system which allows interoperable machine-to-machine communication via a computer network such as the Internet. Thus, it facilitates access to databases hosted on remote servers that offer the web service and even allows executing remote query operations on these machines. The advantage for the web service using computer is that no download and parsing of the complete database is required and changes of the database structure are negligible if the API (Application Programming Interface) remains the same.

Since the release of June 2006 (3) BRENDA offers a SOAP-based web service API (http://www.brenda-enzymes.org/soap). It covers more than 50 different data fields (Table 3) and provides access to them via almost 150 different remote methods.

For all data fields in the record sets the corresponding literature references can be obtained. The methods for querying the individual data fields accept either the organism, the EC number or—where meaningful—the ligand identifier or combinations of all three as input parameters.

A critical point of a web service is the complexity of its API and the interoperability with many different programming languages. Following the user feedback, the SOAP interface of BRENDA has been simplified with respect to the used data types. The data types of the return values of the new SOAP API (http://www.brenda-enzymes.org/soap2) which will be made available from January 2009 use exclusively strings and which are supported by almost every programming language.

For compatibility reasons the old SOAP interface (http://www.brenda-enzymes.org/soap) will remain online for some time but is no longer actively supported. Examples of client code snippets and a full documentation of the web service's capabilities can be found on the above specified URLs.

FUNDING

European Union (FELICS, Free European Life-Science Information and Computational Services: 021902 (RII3)). Funding for open access charge: FELICS, Free European Life-Science Information and Computational Services.

Conflict of interest statement: None declared.

REFERENCES

1. Schomburg,D. and Schomburg,I. (2001–2007) *Springer Handbook of Enzymes*, *2 edn*., Springer, Heidelberg, Germany.
2. Schomburg,I., Chang,A., Ebeling,C., Gremse,M., Heldt,C., Huhn,G. and Schomburg,D. (2004) BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res*., **32**, D431–D433.
3. 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.
4. Sonnhammer,E.l.L., von Heijne,G. and Krogh,A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. In Glasgow,J., Littlejohn,T., Major,F., Lathrop,R., Sankoff,D. and Sensen,C. (eds), *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology*, AAAI Press, Menlo Park, CA, pp. 175–182.
5. Webb,E.C. and NC-IUBMB. (1992) *Enzyme Nomenclature: Recommendations of the Nomenclature Committee of the International Union*

of Biochemistry and Molecular Biology on the Nomenclature and Classifcation of Enzymes. Academic Press, New York, NY.
6. Boutet,E., Lieberherr,D., Tognolli,M., Schneider,M. and Bairoch,A. (2007) UniProtKB/Swiss-Prot: the manually annotated section of the UniProt knowledgeBase. *Methods Mol. Biol.*, **406**, 89–112.
7. Wheeler,D.L., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., Dicuccio,M., Edgar,R., Federhen,S. *et al*. (2008) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **36**, D13–D21.
8. Gene Ontology Consortium. (2008) The Gene Ontology project in 2008. *Nucleic Acids Res.*, **36**, D440–D444.