# New Developments at the Brenda Enzyme Information System

**Jens Barthelmes, Christian Ebeling, Antje Chang, Ida S Chomburg and Dietmar Schomburg**

Technical University Braunschweig, Bioinformatics and Systems Biology,
Langer Kamp 19b, 38106 Braunschweig, Germany

**E-Mail:** d.schomburg@tu-bs.de

*Received: 31st October 2006 / Published: 31st August 2007*

## Abstract

The BRENDA enzyme information system (http://www.brenda.uni-koeln.de) is the largest publicly available enzyme information system worldwide. The major part of its content is manually extracted from primary literature. It is not restricted to specific groups of enzymes, but includes information on all identified enzymes irrespective of the source of the enzyme. The range of data encompasses functional, structural, sequence, localization, disease-related, isolation, stability information on enzyme and ligand-related data. Each single entry is linked to the enzyme source and to a literature reference. Recently the data repository was complemented by text mining data which is stored in AMENDA and FRENDA. A genome browser, membrane protein prediction and full text search capacities were added. The newly implemented web service provides instant access to the data for programmers via a SOAP interface. The BRENDA data can be downloaded in the form of a text file from the beginning of 2007.

## Introduction

The BRENDA (BRaunschweig ENzyme DAtabase) enzyme information system [1, 2] is a manually annotated repository for enzyme data. Originally published as a series of books [3] in 1987, it was integrated into a publicly available database in 1998 and has been curated and continuously improved and updated at the University of Cologne since then. Its contents are not restricted to specific groups of enzymes, but include information on all enzymes that have been classified in the EC scheme of the IUBMB (International Union of Biochemistry and Molecular Biology) irrespective of the enzyme's source. The range of data includes the catalysed reaction, detailed description of the substrate, cofactor and inhibitor specificity, kinetic data, structure properties, information on purification and crystallization, properties of mutant enzymes, participation in diseases, and amino acid sequences. Each single entry is linked to the enzyme source (organism and, if applicable, the tissue, and/or the protein sequence) and to the literature reference. Data queries can be performed in a number of different ways, including an EC-tree browser, a taxonomy-tree browser, an ontology browser, and a combination query of up to 20 parameters. However the huge amount of literature on enzymes does not allow the manual annotation of the complete literature for all enzymes. The capacity for manual annotation has been restricted to ~8,000 references per year. To be able to include more literature, text-mining programs have been developed. Recently, two additional databases (AMENDA and FRENDA) which contain the results of these procedures, have been added to the BRENDA host. They complement the existing database with respect to organisms, tissues and references.

## Contents of BRENDA

At present, BRENDA contains ~1.9 million manually annotated data for more than 4,000 EC-numbers, on average 500 single entries per EC-number. These data are stored in ~120 tables in a relational database system enabling extensive search modes, i.e. quick search, full text search, advanced search, substructure search, sequence search, TaxTree search, ECTree browser, searches in the Genome browser, and searches in more than 20 different ontologies.

### *Functional parameters*

In total BRENDA holds data for > 140,000 kinetic parameters (Table 1). In addition to the numeric values, the experimental conditions are given in a commentary as a text in order to account for the different procedures for enzyme characterization in the laboratory. A web portal for the deposition of enzyme kinetic parameters has been developed in cooperation with the STRENDA commission (http://www.strenda.org/) [4]. This will increase the availability of well-defined kinetic parameters that are essential for systems biology approaches. Each entry in BRENDA is linked to a literature reference. This makes it possible to retrieve detailed information from the original literature (provided the literature is accessible as online version).

The pI-value has recently been included into the section of functional parameters. The isoelectric point provides information about the pH at which the protein carries no net electrical charge. This value is of significance for the purification procedure allowing conclusions about the solubility of the enzyme and its motility in electrophoretic procedures. Presently BRENDA contains more than 1,700 pI-values.

The reactivity of mutant enzymes can reveal detailed insights into the catalytic process and may give valuable clues about the active sites, the mechanism of the reaction, or the regulation. Meanwhile ~19,000 engineered enzymes are described in the database. For each single modification of the protein sequence, the properties of the resulting enzyme are described. Kinetic data for these enzymes are included in the respective database sections.

**Table 1.** Data statistics for the various sections of the database.

| **Enzyme Information** | **Single Data*** |
|---|---:|
| Nomenclature | 70972 |
| Isolation & preparation | 53364 |
| Stability | 34532 |
| Reaction & specificity | 396760 |
| Enzyme structure | 232824 |
| Functional & kinetic parameters | 191134 |
| Km Value | 76894 |
| Ki Value | 14014 |
| pI Value | 1745 |
| Turnover Number | 20493 |
| Specific Activity | 30070 |
| pH Optimum | 26220 |
| pH Range | 6344 |
| Temperature Optimum | 13354 |
| Temperature Range | 2000 |
| Organism-related information | 80964 |
| Source Tissue | 56557 |
| Localization | 24407 |
| References | 91403 |
| Enzyme application | 3854 |
| Enzyme-related diseases | 52558 |
| Mutant enzymes | 18194 |

* These numbers refer to the combination of enzyme–organism–(protein-)value.

### *Organism-related information*

Because enzymes and their properties vary greatly depending on the organism (e. g. eukaryotic or prokaryotic) it is highly important to link enzyme data to their source organism. Presently BRENDA covers information on enzymes of more than 7,500 different organisms (Fig. 1). With ~170,000 single data human enzymes are the most thoroughly described in the literature, followed by enzymes of the rat (~132,000 entries) and *Escherichia coli* (~93,000 entries).

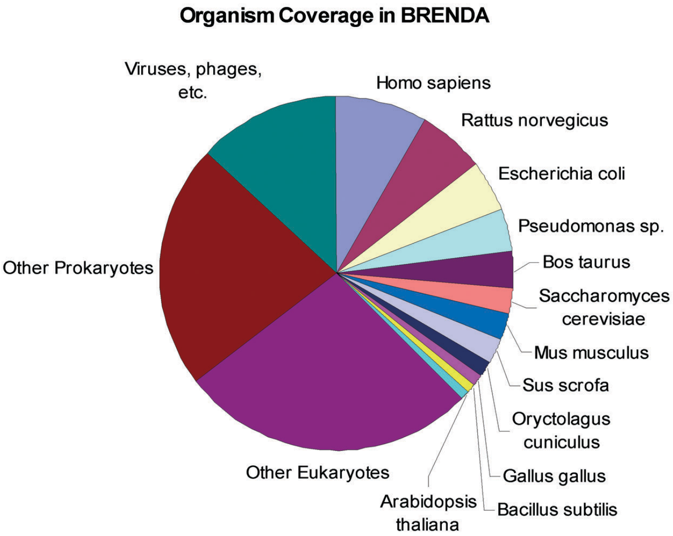
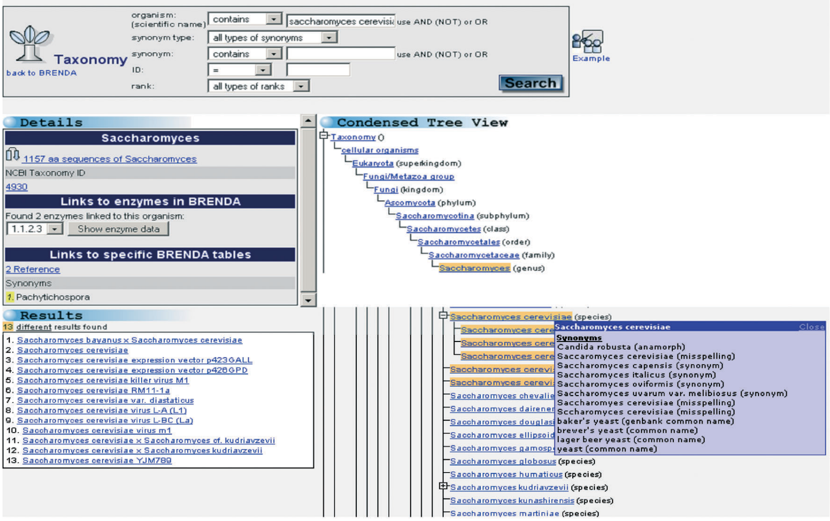


**Figure 1.** Organism coverage in BRENDA data.

All organisms are integrated into the BRENDA TaxTree (Fig. 2). The researcher may search along the TaxTree or switch to higher or lower branches to get an overview in e.g. a class or family or may focus the search on a specific species. Most of the TaxTree entries are linked to the NCBI taxonomy database. A small number of organisms cannot be linked to this tree because they do not appear in the NCBI taxonomy tree.



organism: (scientific name) contains saccharomyces cerevisia use AND (NOT) or OR
synonym type: all types of synonyms
synonym: contains use AND (NOT) or OR
ID: =
rank: all types of ranks

Details

Saccharomyces

1157 aa sequences of Saccharomyces

NCBI Taxonomy ID

4930

Links to enzymes in BRENDA

Found 2 enzymes linked to this organism:

1.1.2.3

Links to specific BRENDA tables

2 Reference

Synonyms

1. Pachytichospora

Results

13 different results found

1. Saccharomyces bayanus x Saccharomyces cerevisiae
2. Saccharomyces cerevisiae
3. Saccharomyces cerevisiae expression vector p423GALL
4. Saccharomyces cerevisiae expression vector p426GPD
5. Saccharomyces cerevisiae killer virus M1
6. Saccharomyces cerevisiae RM11-1a
7. Saccharomyces cerevisiae var. diastaticus
8. Saccharomyces cerevisiae virus L-A (L1)
9. Saccharomyces cerevisiae virus L-BC (La)
10. Saccharomyces cerevisiae virus m1
11. Saccharomyces cerevisiae x Saccharomyces cf. kudriavzevii
12. Saccharomyces cerevisiae x Saccharomyces kudriavzevii
13. Saccharomyces cerevisiae YJM789

Condensed Tree View

```
Taxonomy ()
 cellular organisms
  Eukaryota (superkingdom)
   Fungi/Metazoa group
    Fungi (kingdom)
     Ascomycota (phylum)
      Saccharomycotina (subphylum)
       Saccharomycetes (class)
        Saccharomycetales (order)
         Saccharomycetaceae (family)
          Saccharomyces (genus)

           Saccharomyces cerevisiae (species)
            Saccharomyces cere
            Saccharomyces cere
            Saccharomyces cere
           Saccharomyces cerevi
           Saccharomyces cerevi
           Saccharomyces chevalie
           Saccharomyces dairener
           Saccharomyces douglasi
           Saccharomyces ellipsoid
           Saccharomyces gamosp
           Saccharomyces globosus (species)
           Saccharomyces humaticus (species)
           Saccharomyces kudriavzevii (species)
           Saccharomyces kunashirensis (species)
           Saccharomyces martiniae (species)
```

Saccharomyces cerevisiae

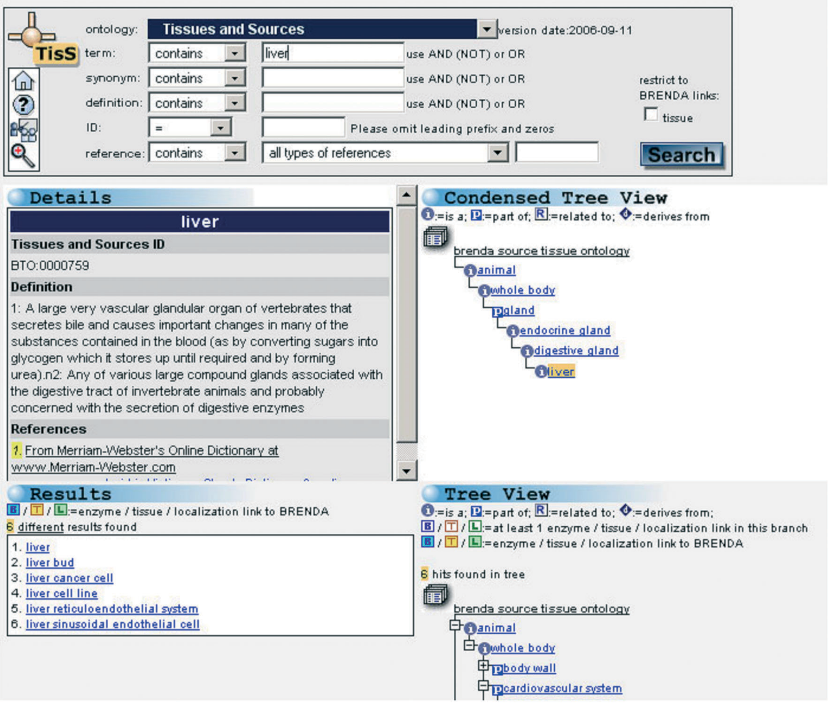
Synonyms
Candida robusta (anamorph)
Saccaromyces cerevisiae (misspelling)
Saccharomyces capensis (synonym)
Saccharomyces italicus (synonym)
Saccharomyces oviformis (synonym)
Saccharomyces uvarum var. melibiosus (synonym)
Saccharomyes cerevisiae (misspelling)
Sccharomyces cerevisiae (misspelling)
baker's yeast (genbank common name)
brewer's yeast (common name)
lager beer yeast (common name)
yeast (common name)

**Figure 2.** Sample of the search in the TaxTree.

### *BRENDA tissue ontology*

For multicellular organisms it is not sufficient to relate enzyme data to the organism alone. The biochemical and molecular properties of one enzyme in different tissues or cell types can vary enormously. The information about the source of an enzyme, i.e. the tissue or cell-lineage therefore is vitally important. The occurrence of enzymes can be restricted to a specific cell type, cell line, or tissue from uni- and multicellular organisms, or can occur ubiquitously. BRENDA has developed its own ontology [1] in which the tissues are sorted hierarchically, corresponding to the format and rules of the Gene Ontology Consortium [5]. The tissue tree in BRENDA is divided into four areas, i.e. animal, plant, fungi and other sources, separated into subtrees. Most of the terms have definitions and synonyms which all can be displayed in the hierarchical tree.

In addition to the occurrence of the tissue, the localization of the enzyme within the cell is given. BRENDA provides a controlled vocabulary in cooperation with the GO consortium. A common shared vocabulary of the cellular components terms has been developed.
Both, the tissue and localization terms are classified in a concise ontology and the localization vocabulary is consistent with the GO terms.



ontology: Tissues and Sources version date:2006-09-11
term: contains liver use AND (NOT) or OR
synonym: contains use AND (NOT) or OR
definition: contains use AND (NOT) or OR
ID: = Please omit leading prefix and zeros
reference: contains all types of references
restrict to BRENDA links: □ tissue
Search

**Details**

**liver**

**Tissues and Sources ID**

BTO:0000759

**Definition**

1: A large very vascular glandular organ of vertebrates that secretes bile and causes important changes in many of the substances contained in the blood (as by converting sugars into glycogen which it stores up until required and by forming urea).n2: Any of various large compound glands associated with the digestive tract of invertebrate animals and probably concerned with the secretion of digestive enzymes

**References**

1. From Merriam-Webster's Online Dictionary at www.Merriam-Webster.com

**Condensed Tree View**

i:=is a; P:=part of; R:=related to; ◊:=derives from

brenda source tissue ontology
- animal
  - whole body
    - gland
      - endocrine gland
        - digestive gland
          - liver

**Results**

B / T / L:=enzyme / tissue / localization link to BRENDA

6 different results found

1. liver
2. liver bud
3. liver cancer cell
4. liver cell line
5. liver reticuloendothelial system
6. liver sinusoidal endothelial cell

**Tree View**

i:=is a; P:=part of; R:=related to; ◊:=derives from;
B / T / L:=at least 1 enzyme / tissue / localization link in this branch
B / T / L:=enzyme / tissue / localization link to BRENDA

6 hits found in tree

brenda source tissue ontology
- animal
  - whole body
    - body wall
    - cardiovascular system

**Figure 3.** Sample of the search in the BRENDA Ontology (BTO).

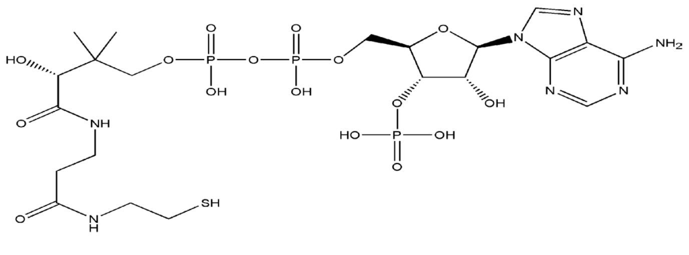
### *Ligands and metabolites*

Enzymes interact with ligands in manifold ways. These can be substrates, products, prosthetic groups, cofactors, but also activating, stabilizing or inhibiting compounds. The present version contains ~88,500 different ligand names. Of these 52,250 molecules are stored as 2D structures in MOL-format. Generic compound names (e. g. “dextrans” or “carboxylic acid”) amount to ~10,000 entries. Applying the organism-specific search option ligands occur in:

- 737,240 enzyme/ligand relationships
- 424,186 enzyme/substrate relationships
- 396,270 enzyme/product relationships
- 16,010 enzyme/cofactor relationships
- 107,331 enzyme/inhibitor relationships
- 17,563 enzyme/activating compound relationships
- 26,303 enzyme/metal or ion relationships

When searching for enzyme ligands or response modifiers two different query procedures are possible:

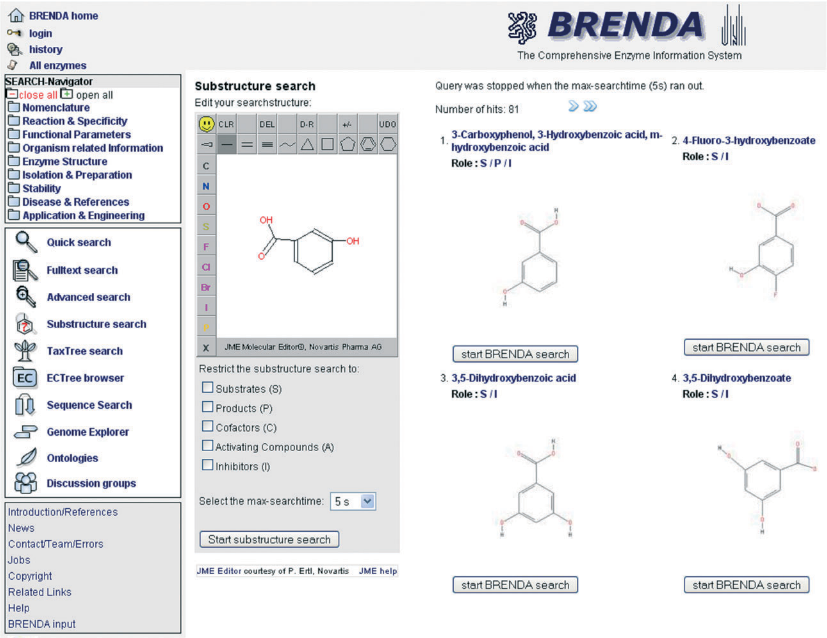
- Using the name of the compound: This option returns not only the data stored for the ligand under the given name but applies the integral molecular thesaurus. The newly generated thesaurus is based on the InChI (IUPAC International Chemical Identifier) [4] codes of the molecular structures stored as molfiles. An InChI is a non-proprietary identifier for chemical substances that can be used in printed and electronic data sources thus enabling easier linking of diverse data compilations. In earlier versions of the database unique isomeric SMILES [7, 8] were used for the calculation of the thesaurus. This procedure has been abandoned since it sometimes caused problems with complex structures.



InChI = 1/C 21H36N7O16P3S/c1 – 21(2,16(31)19(32)24 – 4-3 – 12(29)23 – 5-6 –
48)8 – 41 – 47(38,39)44 – 46(36,37)40 – 7-11 – 15(43 – 45(33,34)35)14(30)20(42 –
11)28 – 10 – 27 – 13 – 17(22)25 – 9-26 – 18(13)28 /h9 – 11,14 – 16,20,30 – 31,48 H,3 –
8H2,1 – 2H3,(H,23,29)(H,24,32)(H,36,37)(H,38,39)(H2,22,25,26)(H2,33,34,35)/t11-
,14-,15-,16+,20-/m1 /s1 /f/h23 – 24,33 – 34,36,38 H,22H2

**Figure 4.** Structure and InChI code for coenzyme A.

- Performing a substructure search (Fig. 5) with the integrated JME Editor [9]. This is an easy to use Java application for drawing molecules. The search can be restricted to a specific function (e. g. substrates). The results page displays the images, names, and synonyms of the found compounds, their function when interacting with the enzyme and also provides a button for an immediate BRENDA search.



Substructure search

Edit your searchstructure:

Restrict the substructure search to:

□ Substrates (S)

□ Products (P)

□ Cofactors (C)

□ Activating Compounds (A)

□ Inhibitors (I)

Select the max-searchtime: 5 s

Query was stopped when the max-searchtime (5s) ran out.

Number of hits: 81

1. 3-Carboxyphenol, 3-Hydroxybenzoic acid, m-hydroxybenzoic acid
Role : S / P / I

2. 4-Fluoro-3-hydroxybenzoate
Role : S / I

3. 3,5-Dihydroxybenzoic acid
Role : S / I

4. 3,5-Dihydroxybenzoate
Role : S / I

**Figure 5.** Substructure search.

## New Databases at the BRENDA Host

For the BRENDA enzyme database the references for manual annotation are chosen from the results of database searches in literature databases such as PubMed [10] and Chemical Abstracts (SciFinder) [11]. For some enzyme classes it is possible to include the complete literature that has been published for a specific enzyme. For the vast majority of enzymes, however, this is impossible for several reasons.

- the number of annually published references is too large to keep up with in the manual annotation capacities
- The literature on enzymes also covers aspects which are not in the focus of the BRENDA database. These may be reports on the genome-annotation, global expression of proteins, and literature in which the enzyme is used in a standard

assay as a tool without any information on the enzyme's properties. References of this kind are not taken into consideration for BRENDA since they would only increase the statistic number of references per enzyme without providing more information and may even reduce the conciseness

For specific projects the user however might wish to retrieve a complete list of references for an enzyme. This would require a PubMed [10] search not only with the recommended name of the enzyme, but also with all the synonyms which are used. Conducting a single search for each synonym might be very time-consuming because most enzymes are used with different names, some even with hundreds of names as can be seen from Table 2

**Table 2.** Multiple synonyms for enzymes.

| EC-number | Recommended Name | No. of Synonyms |
|---|---|---|
| 2.7.10.1 | receptor protein-tyrosine kinase | 416 |
| 3.1.21.4 | type II site-specific deoxyribonuclease | 368 |
| 1.6.5.3 | NADH dehydrogenase (ubiquinone) | 169 |
| 3.1.3.48 | protein-tyrosine-phosphatase | 176 |
| 5.2.1.8 | Peptidylprolyl isomerase | 161 |

Similarly, searching for the complete literature on an enzyme in a specific organism or tissue would require searching with all known synonyms, common and scientific names. Since especially organism names have changed frequently because of taxonomic requirements a search in PubMed [10] with organism synonyms would require much time and a good knowledge in taxonomy. Similarly a search for an enzyme in a specific tissue would require a detailed knowledge of animal or plant anatomy.

In order to provide complete sets of references for all enzymes two databases were added at the BRENDA host.

### *FRENDA*

**FRENDA** (**F**ull **RE**ference **EN**zyme **DA**ta) is an additional database to BRENDA available to the academic community with BRENDA release 6.2 (June 2006). FRENDA aims at providing an exhaustive collection of indexed literature references containing organism-specific enzyme information. Compared to a standard PubMed [10] query, FRENDA also returns all references on the enzyme published under one of its synonyms.

FRENDA currently covers 1.4 million enzyme/organism combinations from 550,000 distinct references, automatically extracted from more than 16 million PubMed abstracts (June 2006) [10]. The scientific articles are pre-filtered using MeSH terms [12] – only references declared as ”enzyme” hits are used (1.6 million remaining abstracts). FRENDA uses a dictionary-based approach for recognizing named entities (enzymes, organisms) in titles and abstracts. The dictionaries are compiled from BRENDA and NCBI Taxonomy [10]. The text-mining proceeds in a two-step approach:

1. Identification of the enzyme names (recommended names and synonyms) in title, abstract or MeSH terms,
2. Searching for co-occurring organism names (scientific names and synonyms) in title, abstract or MeSH terms.

The results of this indexing process were classified into 4 reliability categories depending on the occurrence of search terms in title and/or abstract and/or MeSH terms.

- Enzyme name and organism occur in the title or abstract but not in the same sentence. These hits are discarded.
- + Enzyme name and organism occur in the same sentence in the abstract or they both occur in the title
- ++ EC-number occurs in the MeSH-Terms or in the abstract, the organism occurs in the title or in the Abstract
- +++ Enzyme name and organism occur in the same sentence in the abstract and they both occur in the title
- ++++ Enzyme name and organism occur in the same sentence in the abstract, they both occur in the title and the EC-number is found in the abstract or in the MeSH terms

This classification is provided with the commentaries in the FRENDA database.
The manual evaluation of the quality of the FRENDA approach using 250 randomly chosen results indicates a precision of 64.8% with a recall of 72% from a set of 250 manually annotated enzyme-related literature references.

### *AMENDA*

As a subset of FRENDA, AMENDA (Automatic Mining of ENzyme DAta) currently covers organism-specific information on enzyme localization (more than 30,000 records, compared to 17,000 records in BRENDA) and source tissues (roughly 150,000 records, compared to 38,000 records in BRENDA) from a text-mining procedure (to be published).

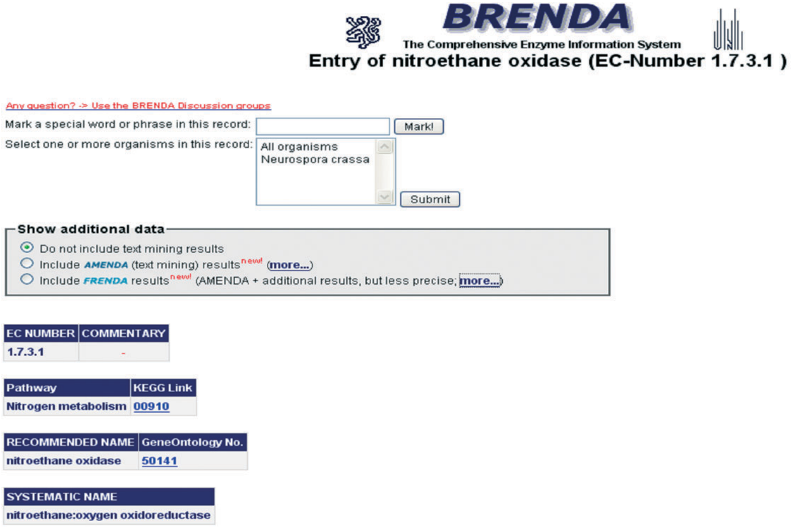
Search terms for enzyme names, organism names, localization, and sources and tissues are compiled from BRENDA enzyme synonyms, the BRENDA tissue-tree (http://obo.sourceforge.net/cgi-bin/detail.cgi?_brenda) and the NCBI Taxonomy [10]. AMENDA is based on the FRENDA co-occurrence approach. Protozoa, viruses, and bacteria are excluded for tissue search. References with enzyme/organism hits are searched for occurrences of tissue terms (singular and plural) and localization terms in title, abstract, and MeSH terms and further evaluated based on text-mining criteria.

- + Enzyme name, localization (or tissue), and organism (or the corresponding synonyms) occur in the title or in the same sentence in the Abstract

- ++ Enzyme name, localization (or tissue) and organism (or the corresponding synonyms) occur in the title. EC-number is contained in the MESH terms assigned to this article or EC-number occurs in the Abstract
- +++ Enzyme name, localization (or tissue), and organism (or the corresponding synonyms) occur in the title and in the same sentence in the Abstract
- ++++ Enzyme name, localization (or tissue) and organism (or the corresponding synonyms) occur in the title and in the same sentence in the abstract. EC-number is contained in the MESH terms assigned to this article or EC-number occurs in the Abstract

The text mining approach described above was tested on 200 randomly selected results. A precision of approximately 76.0% for the combined search terms enzyme–organism–tissue/localization was achieved. In a way similar to FRENDA, the commentaries indicate the individual reliability level for each data set.

When searching for enzyme data the user can choose which data should be displayed. In the default selection only the manually annotated BRENDA data are displayed. With each data set an additional box is displayed which gives the choice to display FRENDA resp. AMENDA results. Entries from these databases are specifically flagged in order to distinguish them from the BRENDA data.



BRENDA
The Comprehensive Enzyme Information System
Entry of nitroethane oxidase (EC-Number 1.7.3.1 )

Any question? -> Use the BRENDA Discussion groups

Mark a special word or phrase in this record: Mark!

Select one or more organisms in this record: All organisms<br>Neurospora crassa

Submit

Show additional data

● Do not include text mining results
○ Include AMENDA (text mining) results new! (more...)
○ Include FRENDA results new! (AMENDA + additional results, but less precise; more...)

| EC NUMBER | COMMENTARY |
|---|---|
| 1.7.3.1 | - |

| Pathway | KEGG Link |
|---|---|
| Nitrogen metabolism | 00910 |

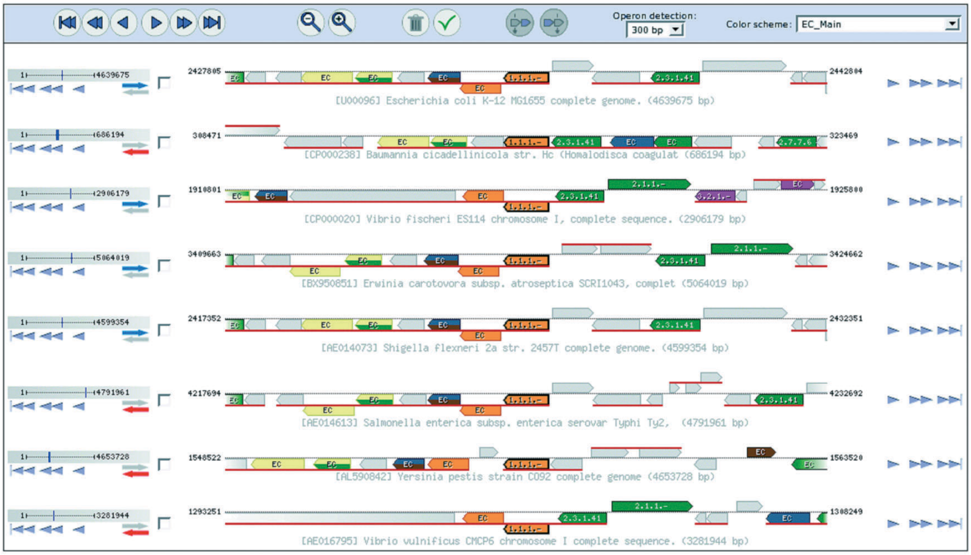
| RECOMMENDED NAME | GeneOntology No. |
|---|---|
| nitroethane oxidase | 50141 |

| SYSTEMATIC NAME |
|---|
| nitroethane:oxygen oxidoreductase |

**Figure 6.** The databases AMENDA and FRENDA can be displayed simultaneously with the BRENDA data.

## BRENDA Genome Explorer

The BRENDA Genome Explorer is an enzyme-centred genome visualization tool for browsing and comparing enzyme annotations in full genomes. It closes the gap between genomic and enzymatic data and allows the alignment of genomes at a given enzyme-coding gene and its orthologs, thus allowing visual comparison of the genomic environment of the gene in different organisms (Fig. 2). The underlying genome database is compiled from EBI Genomes [13] and ENSEMBL [14] and supplemented by UniProt [15] annotations. It can be searched for specific proteins via names, EC-numbers, or UniProt accessions, allowing for a highly target-oriented search.



Operon detection: 300 bp

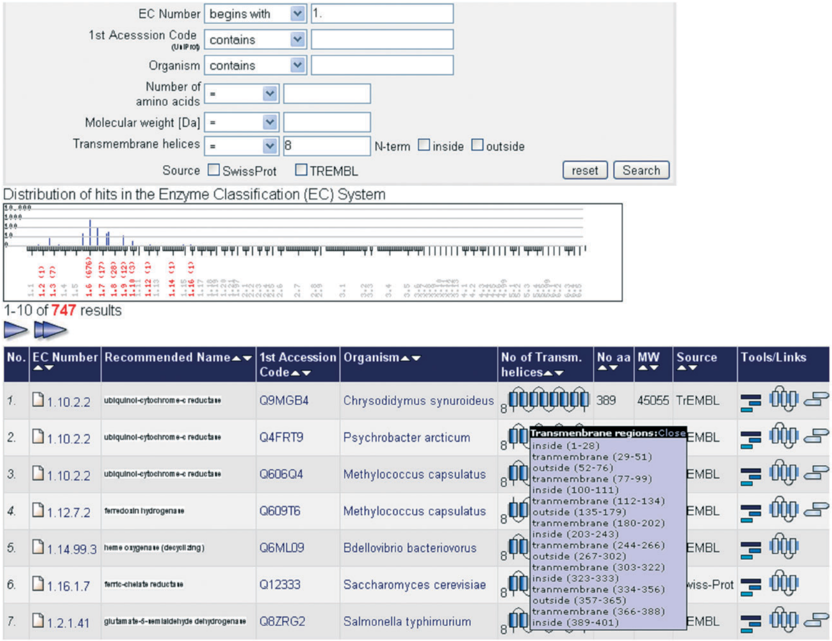
Color scheme: EC_Main

2427805
2442804
[U00096] Escherichia coli K-12 MG1655 complete genome. (4639675 bp)
308471
323469
[CP000238] Baumannia cicadellinicola str. Hc (Homalodisca coagulat (686194 bp)
1910801
1925800
[CP000020] Vibrio fischeri ES114 chromosome I, complete sequence. (2906179 bp)
3409663
3424662
[BX950851] Erwinia carotovora subsp. atroseptica SCRI1043, complet (5064019 bp)
2417352
2432351
[AE014073] Shigella flexneri 2a str. 2457T complete genome. (4599354 bp)
4217694
4232692
[AE014613] Salmonella enterica subsp. enterica serovar Typhi Ty2, (4791961 bp)
1548522
1563520
[AL590842] Yersinia pestis strain CO92 complete genome (4653728 bp)
1293251
1308249
[AE016795] Vibrio vulnificus CMCP6 chromosome I complete sequence. (3281944 bp)
EC
1.1.1.-
2.3.1.41
2.1.1.-
2.7.7.6
3.2.1.-

**Figure 7.** BRENDA Genome Explorer showing a part of a genome alignment for *Escherichia coli* erythronate-4-phosphate dehydrogenase, EC 1.1.1.290.

## Transmembrane Protein Prediction

Transmembrane helices for enzymes are predicted with TMHMM (TransMembrane Hidden Markov Model) developed by Sonnhammer *et al.* [16]. With the aid of this tool it is possible to predict the number, the size and the location of trans-membrane helices, thereby discriminating soluble and membrane-bound enzymes.



EC Number: begins with ▾ | 1.

1st Acesssion Code (UniProt): contains ▾ |

Organism: contains ▾ |

Number of amino acids: = ▾ |

Molecular weight [Da]: = ▾ |

Transmembrane helices: = ▾ | 8 N-term □ inside □ outside

Source □ SwissProt □ TREMBL

1-10 of 747 results

| No. | EC Number ▲▼ | Recommended Name ▲▼ | 1st Accession Code ▲▼ | Organism ▲▼ | No of Transm. helices ▲▼ | No aa ▲▼ | MW ▲▼ | Source ▲▼ | Tools/Links |
|---|---|---|---|---|---|---|---|---|---|
| 1. | 1.10.2.2 | ubiquinol-cytochrome-c reductase | Q9MGB4 | Chrysodidymus synuroideus | 8 | 389 | 45055 | TrEMBL | |
| 2. | 1.10.2.2 | ubiquinol-cytochrome-c reductase | Q4FRT9 | Psychrobacter arcticum | 8 | | | EMBL | |
| 3. | 1.10.2.2 | ubiquinol-cytochrome-c reductase | Q606Q4 | Methylococcus capsulatus | 8 | | | EMBL | |
| 4. | 1.12.7.2 | ferredoxin hydrogenase | Q609T6 | Methylococcus capsulatus | 8 | | | EMBL | |
| 5. | 1.14.99.3 | heme oxygenase (decyclizing) | Q6ML09 | Bdellovibrio bacteriovorus | 8 | | | EMBL | |
| 6. | 1.16.1.7 | ferric-chelate reductase | Q12333 | Saccharomyces cerevisiae | 8 | | | wiss-Prot | |
| 7. | 1.2.1.41 | glutamate-5-semialdehyde dehydrogenase | Q8ZRG2 | Salmonella typhimurium | 8 | | | EMBL | |

```
Transmenbrane regions:Close
inside (1-28)
tranmembrane (29-51)
outside (52-76)
tranmembrane (77-99)
inside (100-111)
tranmembrane (112-134)
outside (135-179)
tranmembrane (180-202)
inside (203-243)
tranmembrane (244-266)
outside (267-302)
tranmembrane (303-322)
inside (323-333)
tranmembrane (334-356)
outside (357-365)
tranmembrane (366-388)
inside (389-401)
```

**Figure 8.** Characteristic output of the trans-membrane prediction tool.

## Accessibility

BRENDA is accessible via the various search options (quick search, advanced search, ontologies, sequence search, Genome Explorer etc.). The database will be downloadable as a text file from January 2007 on. Access to AMENDA and FRENDA requires a registration.

## SOAP-Based Web Service

Web services provide a simple way to access the data collection without the need for downloading, parsing, and preparing an entire database for local queries. Web services are independent of the internal organization of the database and avoid parsing problems caused by changes in the text file structure.

BRENDA now provides a SOAP (Simple Object Access Protocol, http://www.w3.org/TR/soap) based web service comprised of 148 methods covering 52 data fields. Flexible queries can be performed directly from programs written in different programming languages (Perl, Java, C++, Python, PHP) on data fields such as substrate, $K_m$-value and pH-optimum. For any given record returned, a set of complete literature references can be retrieved using unique reference identifiers. Every data field may be queried by providing at least one of the three parameters EC-number, organism, or – if applicable – ligand structure identifier. The ligand structure identifier, which can be queried with the name of a chemical compound, is used to ensure that all synonyms for a given molecular structure are also retrieved.

The BRENDA web service also gives access to the data using identifiers from other databases like UniProt [14] or NCBI Taxonomy [10], as well as ontologies like Gene Ontology [5] or BRENDA Tissue Ontology [1]. The ontology-based search allows for queries based on entire branches of the hierarchy, avoiding a complex search for all leaves in the given branch. For example, an ontology-based search for the term 'brain' or the respective Gene Ontology identifier will return all tissues and cell types under the umbrella term 'brain'. The same method can also be applied to search for whole groups of organisms. The documentation of the BRENDA web service including examples in different programming languages is available at http://www.brenda.uni-koeln.de/soap.

## Conclusions

The BRENDA enzyme information system has made a big step forward not only by a formidable increase in the annotation speed but also by inclusion of data based on text-mining approaches and by the development of different new methods for data access. The new funding by an EU grant allows the annotation speed to be increased even further to bring the backlog down to less than one year and will also allow a substantial increase in the percentage of ligands with full structural information.

## Acknowledgements

This work was supported by the European Union (FELICS, Free European Life-Science Information and Computational Services: 0121902 (RII3)) and the German BMBF, Federal Ministry of Education, Science, Research and Technology (NGFN, Nationales Genomforschungsnetz: 0313398B)

## References

[1] Schomburg,I., Chang,A., Ebeling,C., Gremse,M., Heldt,C., Huhn,G., Schomburg,D. (2004) BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res.* **32:**D 431 –D 433.

[2] Schomburg,I., Chang,A., Schomburg,D. (2002) BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.* **30:**47–49.

[3] Schomburg,D., Schomburg,I. (2001–2006) *Springer Handbook of Enzymes.* 2nd Edn. Springer, Heidelberg, Germany.

[4] Kettner,C., Hicks,M.G. (2005) The dilemma of modern functional enzymology. *Curr. Enzyme Inhib.* **1:**171–181.

[5] Gene Ontology Consortium (2006) The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.* **34:**D 322 –D 326.

[6] Stein, S.E., Heller,S.R., Tchekhovski, D. (2003) An Open Standard for chemical structure representation – The IUPAC Chemical Identifier. *Nimes International Chemical Information Conference Proceedings*, pp. 131–143.

[7] Weininger,D. (1988) SMILES, a chemical language and information system.1. Introduction to methodology and encoding rules. *J. Chem. Inform. Comput. Sci.* **28:**31–36.

[8] Weininger, D., Weininger, A., Weininger, J. (1989) SMILES. 2. algorithm for generation of unique SMILES notation. *J. Chem. Inform. Comput. Sci.* **29:**97–101.

[9] Csizmadia, P. (2000) MarvinSketch and MarvinView: molecule applets for the World Wide Web. *Proceedings of ECSOC-3 and Proceedings of ECSO-4,* 367–369.

[10] 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.* (2006) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **34:**D 173–180.

[11] Ridley,D.D. (2002) *SciFinder and SciFinder Scholar.* J. Wiley & Sons, New York.

[12] National Library of Medicine. (1960) Medical subject headings: main headings, subheadings, and cross references used in the Index Medicus and the National Library of Medicine Catalog. 1st Edn. Washington, DC: U.S. Department of Health, Education, and Welfare.

[13] Cochrane, G., Aldebert, P., Althorpe, N., Andersson, M., Baker, W., Baldwin, A., Bates, K., Bhattacharyya, S., Browne, P., v. d. Broek, A. *et al.* (2006) EMBL Nucleotide Sequence Database: developments in 2005. *Nucleic Acids Res.* **34:**D 10 –D 15.

[14] Birney, E., Andrews, D., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cox, T., Cunningham, F., Curwen, V., Cutts, T. *et al*. (2006) Ensembl 2006. *Nucleic Acids Res*. **34:**D556–561.

[15] Wu, C.H., Apweiler, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R. *et al*. (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res*. **34:**D187–191.

[16] Sonnhammer, E. l. L., von Heijne, G., Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. In: *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology.* (Glasgow, J., Littlejohn, T., Major, F., Lathrop, R., Sankoff, D., Sensen, C. Eds), pp.175–182. AAAI Press, Menlo Park, CA.