

BRENDA, enzyme data and metabolic information

Ida Schomburg, Antje Chang and Dietmar Schomburg*

University of Cologne, Institute of Biochemistry, Zùlpicher Straße 47, 50674 Köln, Germany

Received August 21, 2001; Revised and Accepted October 10, 2001

ABSTRACT

BRENDA is a comprehensive relational database on functional and molecular information of enzymes, based on primary literature. The database contains information extracted and evaluated from approximately 46 000 references, holding data of at least 40 000 different enzymes from more than 6900 different organisms, classified in approximately 3900 EC numbers. BRENDA is an important tool for biochemical and medical research covering information on properties of all classified enzymes, including data on the occurrence, catalyzed reaction, kinetics, substrates/products, inhibitors, cofactors, activators, structure and stability. All data are connected to literature references which in turn are linked to PubMed. The data and information provide a fundamental tool for research of enzyme mechanisms, metabolic pathways, the evolution of metabolism and, furthermore, for medicinal diagnostics and pharmaceutical research. The database is a resource for data of enzymes, classified according to the EC system of the IUBMB Enzyme Nomenclature Committee, and the entries are cross-referenced to other databases, i.e. organism classification, protein sequence, protein structure and literature references. BRENDA provides an academic web access at <http://www.brenda.uni-koeln.de>.

INTRODUCTION

BRENDA (BRaunschweig ENzyme DATabase) was created in 1987 at the German National Research Center for Biotechnology in Braunschweig (GBF) and is now continued at the University of Cologne, Institute of Biochemistry. This enzyme information system was developed to collect and store enzyme functional data and has been an ongoing effort for >10 years. It was first published as a series of books [Enzyme Handbook, Springer (1)] with the intention from the very beginning to provide the data in a database as a retrieval system.

In the last few years all information has been transferred from a full text to a relational database system and is accessible to the academic community from <http://www.brenda.uni-koeln.de>. Commercial users have to purchase a license at <http://www.science-factory.com>.

Enzymes, the largest and most diverse group among the proteins, play an essential role in the metabolism of each organism. All chemical reactions and metabolic steps within

the cell are catalyzed and regulated by enzymes. The development and progress of projects on structural and functional genomics suggest that the systematic collection and accessibility of functional information of gene products are indispensable to understanding biological functions and the correlation between phenotype and genotype.

BRENDA represents a protein function database, containing comprehensive enzymatic and metabolic data, extracted, continuously updated and evaluated from the primary literature. The key developments in the last few years were the conversion of the database to an organism-specific information system and the improvement of the validation and the correction of data and the standardization of the entries to create prerequisites for a systematic access and analysis.

CONTENTS OF BRENDA

BRENDA contains all enzymes classified according to the system of the EC numbers, which was implemented in 1955 by the International Commission of Enzymes [now the International Union of Biochemistry and Molecular Biology, IUBMB (2)]. This nomenclature is based on the reaction the enzymes catalyze and not on the individual enzyme molecule. Presently BRENDA contains data of approximately 3900 EC numbers, which represent more than 40 000 different protein molecules, given by the combination of EC number and organism (obviously in many cases organisms have more than one enzyme with the same EC number but, as the functional data on enzymes as given in the primary literature are rarely associated to a specific sequence, a more reliable estimation is not possible in the present situation; this will change with the progress of the genome sequencing projects).

The database covers organism-specific information on functional and molecular properties, in detail on the nomenclature, reaction and specificity, enzyme structure, stability, application and engineering, organism, ligands, literature references and links to other databases (Table 1).

The data for all enzymes having the same EC number are periodically updated by manual extraction of parameters from the literature references accessible via literature databases, i.e. Chemical Abstracts and PubMed [NCBI (3)] and the full information for each EC number is continuously checked for internal inconsistencies. Depending on scientific needs and the progress in research the data fields are subject to an ongoing development.

The data and information in BRENDA are stored in 52 tables containing approximately 460 000 entries directly extracted from the primary literature in a relational database system to enable different search features. Enzymes can be searched by

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

Table 1. Data and information fields in BRENDA

Information field	Entries	Information field	Entries
Enzyme nomenclature		Functional parameters	
EC number	3869	K_m value	28 134
Recommended name	3509	Turnover number	3986
Systematic name	3182	Specific activity	11 787
Synonyms	17 707	pH optimum	14 037
CAS registry number	3552	pH range	3929
Reaction	3518	Temperature optimum	6147
Reaction type	4123	Temperature range	908
Enzyme structure		Molecular properties	
Molecular weight	12 329	pH stability	2931
Subunits	7416	Temperature stability	6825
Sequence links	33 099	General stability	5398
Post-translational modification	1112	Organic solvent stability	311
Crystallization	1003	Oxidation stability	349
3D-structure, specific PDB links	6142	Storage stability	6505
Enzyme–ligand interactions		Purification	11 176
Substrates/products	47 630	Cloned	2015
Natural substrate	7668	Engineering	797
Cofactor	6217	Renatured	199
Activating compound	6217	Application	338
Metals/ions	13 173	Organism-related information	
Inhibitors	56 336	Organism	40 027
Bibliographical data		Source tissue, organ	19 347
References	46 305	Localization	7935

their EC numbers (3870 entries), their names or synonyms (22 936 entries) or by the organisms (6921 single entries), in which the enzyme reaction is detected. All other information fields (Table 1) can be searched individually or by combination searches, which can be performed organism specifically. Therefore, it is possible to find a specific enzyme for a specific organism or even for a specific organ or tissue. Furthermore, a search for ligands, which may have a dual role (e.g. substrate/inhibitor or cofactor/inhibitor) may be performed. Kinetic data for enzyme–ligand interaction can be searched.

LIGANDS

A major part in BRENDA is the information of ligands, which function as natural or *in vitro* substrates/products, inhibitors, activating compounds, cofactors, bound metals, etc. Altogether, approximately 320 000 enzyme–ligand relationships are stored with more than 33 000 different chemical compounds functioning as 'ligand'. In BRENDA the ligands are stored as compound names, SMILES (4) strings and as Molfiles. The latter two forms are interchangeable with respect to the connectivity information. The two-dimensional chemical structures of these compounds can be displayed as images.

METABOLISM

The data in BRENDA allow the calculation or simulation of metabolic pathways by extracting the information of substrate/product chains and the corresponding kinetic data of the preceding and following enzymes in the Boehringer and KEGG metabolism (with the risk of including 'pathways' with non-natural compounds).

Based on the representation of metabolic networks as directed graphs, navigation operation will be made possible. This will give answers to questions on the structure of the metabolic paths, e.g. on shortest or alternate paths for different organisms.

ENZYME AND DISEASE INFORMATION

In order to keep up with the quickly growing scientific literature, automatic information extraction techniques were tested to include disease-related knowledge to BRENDA. References in electronic format are taken from the PubMed database, parsed for relevant key phrases and associated with correlated enzymes. Information on 789 enzymes and their associated human diseases has been included into the BRENDA database (5).

Additionally, the Online Mendelian Inheritance in Man [OMIM (2,6)] repository, a well-annotated catalog of human genes and genetic disorders, was parsed for enzyme information. In this way a total of 630 EC numbers in BRENDA could be linked to 2100 OMIM entries.

REFERENCES

1. Schomburg,D. and Schomburg,I. (2001) *Springer Handbook of Enzymes*, 2nd edn. Springer, Heidelberg, Germany.
2. Enzyme Nomenclature Committee (1992) Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes, NC-IUBMB. Academic Press, New York, NY.
3. Wheeler,D.L., Church,D.M., Lash,A.E., Leipe,D.D., Madden,T.L., Pontius,J.U., Schuler,G.D., Schriml,L.M., Tatusova,T.A., Wagner,L. and Rapp,B.A. (2001) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **29**, 11–16. Updated article in this issue: *Nucleic Acids Res.* (2002), **30**, 13–16.
4. Weininger,D. (1988) SMILES 1. Introduction and encoding rules. *J. Chem. Inf. Comput. Sci.*, **28**, 31–36.
5. 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.
6. McKusick,V.A. (1998) Mendelian Inheritance in Man. *Catalogs of Human Genes and Genetic Disorders*, 12th edn. The Johns Hopkins University Press, Baltimore, MD.