Perspectives in Science (2014) 1, 15–23

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/pisc

REVIEW

Standardization in enzymology—Data integration in the world's enzyme information system BRENDA ☆

Ida Schomburg, Antje Chang, Dietmar Schomburg*

Institute of Biochemistry, Biotechnology & Bioinformatics, Technische Universität Braunschweig, Braunschweig, Germany

Received 10 September 2012; accepted 29 January 2014; Available online 26 March 2014

KEYWORDS
Enzyme database;
Standardisation of experiments;
Nomenclature standards;
Ontology;
Enzyme kinetics

Abstract
In the modern life sciences literature search is mainly done electronically and huge datasets obtained by the use of diverse experimental methods have to be integrated to perform an in depth analysis of biological systems. This means that standardization is absolutely essential to allow the identification of all relevant data, their comparison and finally their integration. The main areas in enzymology where standardisation would be required but is not achieved yet are (i) use of standard nomenclature for enzymes and ligands, and (ii) the full registration and standardisation of experimental condition for function analysis. The accepted or recommended names as defined by the IUBMB biological nomenclature committee are both descriptive and unambiguous, but unfortunately not used in all papers. In addition to the enzyme names unambiguousness is needed for the ligand names, the enzyme's origin as given by the organism name, a tissue name and the description of the subcellular localisation. A comparison of enzyme functional parameters is only possible when the experimental conditions are fully characterised and ideally standardized.
The BRENDA enzyme database and its addenda (AMENDA, FRENDA, DRENDA) as the world's main information system for enzyme function and other properties makes use of standards as far as possible, but also provides non-standard names and other non-standard data, relating them to the appropriate standard. For example the enzyme nomenclature part of BRENDA includes about 82,000 synonyms for the classified enzymes, linking them to the standard accepted name. The definition of the biological enzyme sources are based on ontologies and controlled vocabularies. Kinetic data are reported together with the experimental conditions where available from the literature. For the enzyme ligands chemical structures allow an unambiguous identification.

☆This article is part of a special issue entitled "Reporting Enzymology Data—STRENDA Recommendations and Beyond". Copyright by Beilstein-Institut.

*Corresponding author.
E-mail address: d.schomburg@tu-bs.de (D. Schomburg).

http://dx.doi.org/10.1016/j.pisc.2014.02.002
2213-0209 © 2014 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

© 2014 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

Contents

Introduction . . . 16
Standards in nomenclature: good names and bad names . . . 16
The systematic name . . . 16
The recommended/accepted name . . . 17
Non-standard names in papers . . . 17
Getting the full picture of an enzyme—the elaborated synonym-identification process at BRENDA . . . 17
Names for metabolites: a chaos collection of synonyms . . . 18
Enzyme sources: organism, tissue, localisation . . . 20
Organism names . . . 20
Tissue or organ . . . 20
Cellular localisation . . . 21
Experimental standards . . . 21
Kinetic data . . . 21
Other standards for enzyme characterisation . . . 22
Conflict of interest statement . . . 22
Acknowledgements . . . 22
References . . . 23
Web references . . . 23

Introduction

Enzymes represent the largest and most diverse group of all proteins, catalysing all chemical reactions in the metabolism of all organisms. In addition to metabolism they also play a key role in the regulation of metabolic steps within the cell. With the fast progress of genomics, proteomics, structural and functional genomics and metabolomics, the information about enzymes grows quickly, but is also distributed between different disciplines, making access to the data and data integration for single researchers extremely difficult, if not impossible. Full access to relevant data would require the strict compliance with nomenclature standards in the paper; data integration and comparison of data from different labs and methods is only possible if experimental standards are used and experimental meta-data are fully documented in publications. With the current state of science the task of data integration and systematic experimental documentation can only be accomplished by databases. This article illuminates a number of principles and shortcomings in the current state of standardisation.

Standards in nomenclature: good names and bad names

Since enzymology has a long history many enzyme names are not unique. In many cases the same enzymes became known by several different names, while conversely the same name was sometimes given to different enzymes. Many names conveyed little or no information on the enzymatic function, and similar names were sometimes given to enzymes of quite different types. Recently the unfortunate habit of using gene names for enzymes has become common practice in some areas of molecular biology.

In 1956 the International Commission on Enzymes was created by the International Union of Biochemistry. Since then an elaborated enzyme classification system providing hierarchical EC numbers as well as systematic names and recommended names has been established (see also Cornish-Bowden on current IUBMB recommendations, 2014). In the EC number system an enzyme is not defined by its name but by the reaction it catalyses. In some cases where this is not sufficient, additional criteria are employed such as cofactor specificity or stereospecificity of the reaction. The EC number classifies the enzyme according to the type of reaction it catalyses. Six main classes have been established: (1) oxidoreductases; (2) transferases; (3) hydrolases; (4) lyases; (5) isomerases and (6) ligases. Each main class is attributed with sub- and sub-sub-classes further defining reaction partners, cofactors and type of substrate. Since the start of the project the list of classified enzymes has grown steadily and meanwhile comprises about 5300 (January 2014) valid EC classes plus several hundred deleted and transferred classes (McDonald et al., 2009). Detailed rules for naming an enzyme have been developed and are published on the website of the IUBMB enzyme database. Each classified enzyme receives two names:

The systematic name

This name shows the action of the enzymes as clearly as possible. Thus it often includes the name of the substrate and the type of modification which it undergoes in the course of the reaction. Very often it also includes the cofactor and the product of the reaction. Systematic names unambiguously describe an enzyme's activity. However very often they are not suitable for everyday use. If the substrate is a complex molecule the enzyme name, derived from the IUPAC name of the substrate can become very long and

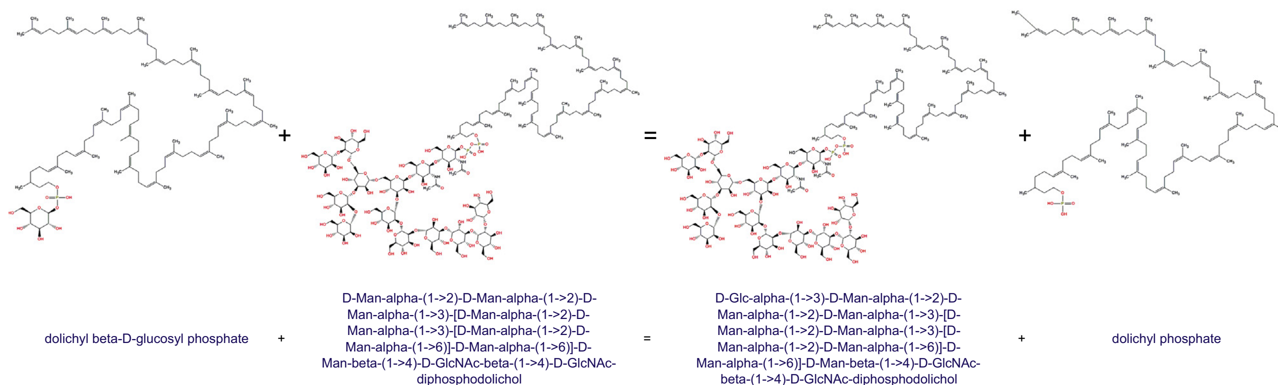


Figure 1 The reaction catalysed by EC 2.4.1.267.

almost unintelligible. An example is the recently classified enzyme EC 2.4.1.267. It specifically transfers a glucosyl residue to the growing chain of a lipid-linked oligosaccharide. In a later stage of glycoprotein biosynthesis the oligosaccharide part of the product is transferred to an asparagine side chain of the target protein (see Figure 1).

The systematic name which correctly includes both substrates is very long even though it uses the approved abbreviations for the sugar moieties:

dolichyl β-D-glucosyl phosphate:D-Man-α-(1→2)-D-Man-α-(1→2)-D-Man-α-(1→3)-[D-Man-α-(1→2)-D-Man-α-(1→3)-[D-Man-α-(1→2)-D-Man-α-(1→6)]-D-Man-α-(1→6)]-D-Man-β-(1→4)-D-GlcNAc-β-(1→4)-D-GlcNAc-diphosphodolichol α-1,3-glucosyltransferase.

Therefore this enzyme needs another name which is both descriptive and unique. The complexity of many systematic names may be the reason why they are not used consistently in the literature.

The recommended/accepted name

This name represents a unique name that either describes the enzyme function in condensed and more readable name like “alcohol dehydrogenase” for 1.1.1.1, on other, rarer cases reflects a historical name like “trypsin” for the protease 3.4.21.4. An example for a rather long recommended name is assigned to EC 2.4.1.267: dolichyl-*P*-Glc:Man9GlcNAc2-*PP*-dolichol α-1,3-glucosyltransferase. This name omits the specification of the sugar connection in the substrate and abbreviates phosphate with a simple *P*. It is applicable as long as there is no other enzyme detected which catalyses a glucosyl transfer to a lipid-linked oligosaccharide where the sugars are connected in a different way.

Many of the recommended names have been established over long years of research into a particular enzyme. As long as they are unambiguous they will be approved by the IUBMB.

Non-standard names in papers

Unfortunately many researchers do not use the defined standard names. This research represents the real problem in enzyme literature accessibility as the papers are not found if scientists search information on a certain enzyme nomenclature standardization. These non-standard names arise from multiple sources such as personal preferences, ignorance, names of individual proteins, gene names, abbreviated forms, trade names etc. The use of non-standard names is, unfortunately, widely distributed in the scientific literature because enzymes represent the only class of biological molecules where such a nomenclature system exists and most molecular biologists/biochemists/cell biologists apparently do not recognise that the use of naming standards will help scientists to find their papers. In many cases non-standard names are used more frequently than the “accepted” names. For example a Google search for EC 4.1.1.39 using the trivial name Rubisco gives more than twice a much results than the accepted name ribulose-bisphosphate carboxylase. (717,000 as compared with 342,000).

Getting the full picture of an enzyme—the elaborated synonym-identification process at BRENDA

The BRENDA enzyme information system is intended to provide comprehensive information on enzyme properties, i.e. we aim to identify all the different names in use for an enzyme and collect this information at one place: the BRENDA database (Chang et al., 2009; Scheer et al., 2011). During the manual annotation or the literature search the curators extract systematically all names and synonyms that are used for a specific enzyme except those that are totally meaningless (such as quantum for EC 3.1.3.26, or HAT for 2.3.1.32, or DDT for EC 4.1.1.84). These are in later update rounds used as search terms for the identification of relevant literature. As a result BRENDA is good source for enzyme synonyms storing about 82,000 different enzyme names for the around 5200 enzymes classified.

This number clearly shows the dramatic problems: on average each EC class is recorded with 15 different names. This means that a literature search with any particular enzyme name on average finds only 1/15, i.e., less than 8% of the relevant literature. Only 20% out of the EC classes are listed with only the accepted name plus a systematic name. 10% out of the EC classes carry only one synonym and 40% are recorded with 2–5 synonyms. Looking at these enzymes it is a general observation that enzymes with a low number of synonyms very often possess a rather narrow substrate specificity or even are specific for a single substrate. Some have been identified in the secondary metabolism of a single plant and are absent from plants in taxonomically related species.

61 EC classes are stored with more than 100 different names, where 30 have more than 150 names (see Table 1). There are different reasons for the large number of different names. If we consider the protein kinases we find very high numbers of synonyms, each for an individual protein catalysing the phosphate transfer either to tyrosine, serine, threonine or histidine. Since the reaction which is the basis for classification is identical, the enzymes are assembled under just a few EC numbers but are named for the individual role they play in different organisms. In organism 1 they could, e.g., phosphorylate a specific protein at a specific position, in organism 2 the same enzyme could phosphorylate a different protein. As long as the substrate specificity is not thoroughly analysed they are classified in the same EC-number. This could change in the future once it is proven that they have distinctly different substrate specificities. It is obvious from the table that especially for enzymes modifying proteins or other macromolecules many different names are in use.

Table 1 Enzymes with high numbers of synonyms.

ec_class	Accepteded name	Synonyms
2.7.10.1	Receptor protein-tyrosine kinase	573
2.7.11.1	Non-specific serine/threonine protein kinase	570
1.14.14.1	Unspecific monooxygenase	481
3.1.21.4	Type II site-specific deoxyribonuclease	453
2.7.10.2	Non-specific protein-tyrosine kinase	435
3.1.3.48	Protein-tyrosine-phosphatase	424
6.3.2.19	Ubiquitin-protein ligase	391
3.1.1.4	Phospholipase A2	349
3.1.3.16	Phosphoprotein phosphatase	337
2.7.13.3	Histidine kinase	297
3.5.2.6	beta-lactamase	249
4.6.1.1	Adenylate cyclase	230
3.1.1.3	Triacylglycerol lipase	221
3.2.1.4	Cellulase	220
1.6.5.3	NADH:ubiquinone reductase (H+-translocating)	219
2.7.11.24	Mitogen-activated protein kinase	208
4.2.1.1	Carbonate dehydratase	207
2.5.1.18	Glutathione transferase	206
5.2.1.8	Peptidylprolyl isomerase	198
3.2.2.22	rRNA N-glycosylase	197
2.7.7.7	DNA-directed DNA polymerase	190
3.6.5.2	small monomeric GTPase	189
3.1.4.11	Phosphoinositide phospholipase C	181
2.7.11.22	Cyclin-dependent kinase	172
2.4.1.17	Glucuronosyltransferase	168
3.1.1.1	Carboxylesterase	165
2.1.1.43	Histone-lysine N-methyltransferase	162
3.2.1.14	Chitinase	159
3.1.3.2	Acid phosphatase	154
3.2.1.8	Endo-1,4-beta-xylanase	152

A different situation is found in the cellulase case, for example. The number of different substrates accepted here is very small, being mainly amorphous or crystalline cellulose. 220 different names are presently in use in the literature. In this case the cellulose breakdown is achieved by a combination/cooperation of a number of isoenzymes. For these isoenzymes different terms are in use in the different organisms. In this case a special initiative has been started and a paper with a suggested cellulase nomenclature system has been published (Urbanowicz et al., 2007). Standards in the plant community are different from standards in the bacteria community. A separate database (http://www.cazy.org) exists for sub-classification of carbohydrate-related enzymes.

Examples for misleading or meaningless names are RACE (EC 5.1.1.3, glutamate racemase), or TIM (EC 5.3.1.1, triose-phosphate isomerase).

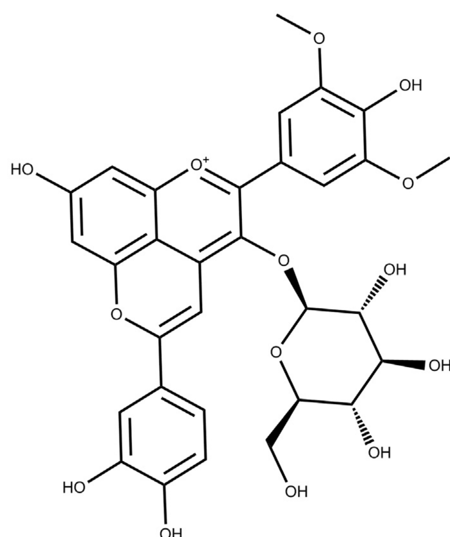
Names for metabolites: a chaos collection of synonyms

The characterisation of enzymes always includes the characterisation of the metabolites and other compounds which interact with the enzyme as cofactors, inhibitors, activators or inducers thus regulating the activity. These compounds can be large molecules such as proteins or nucleic acids or lipids. Proteins and nucleic acids can be identified by their sequence and their respective sequence identifier even though the names used in the literature are not unique.

Many compounds interacting with enzymes can be classified as "small molecules". They have a defined molecular structure and often possess stereo centres. The compounds in rare cases are named following the rules of the IUPAC (http://www.chem.qmul.ac.uk/iupac/). This organisation not only defines the rules for a fully systematic nomenclature, but also provides means for creating names based on trivial names as the systematic name is often prohibitively long. This can result in more descriptive names which give information on the compound class and the stem structure and is especially helpful for compounds composed of a common stem structure which is substituted with side chains.

An example is vitisin A which belongs to the anthocyanidins. It contains a flavylium cation as the central part and is glycosylated (Scheme 1).

A systematic name looks like: 5-(3,4-dihydroxyphenyl)-8-hydroxy-2-(4-hydroxy-3,5-dimethoxyphenyl)pyrano[4,3,2-*de*]chromen-1-ium-3-yl β-D-glucopyranoside. This name, however, does not show that the compound contains the common flavylium cation and a glucosyl residue. Thus, a



Scheme 1

name like 3-[(β-D-glucopyranosyl)oxy]-3″,4′,4″,7-tetrahydroxy-3′,5′-dimethoxypyrano[4″,3″,2″:4,5]flavylium gives much better information for the biologist whereas the trivial name vitisin A does not contain any information concerning the type of molecule or the structure.

In the biochemical literature the use of compound names for small molecules is sometimes even more inconsistent than for proteins. Most commonly the reader finds the trivial names, sometimes equipped with a systematic name in a footnote. Many compounds have however accumulated many different trivial or semi-systematic names in the course of their history or are commonly used in abbreviated forms. Acronyms are in most cases not unique and are in use for quite different compounds. One such example is THF which stands for tetrahydrofuran in the chemist's world and for tetrahydrofolate in the biologist's world. In order to compare data for metabolites it is essential to refer to unique compound names.

Apart from the situation in enzymes a cross-reference is impossible via the compound names while a comparison of the chemical structures is a method for exactly assigning synonyms. In BRENDA this is performed using the InChI codes calculated from mol-files stored in the database. Currently the BRENDA database holds 189,000 different names for compounds interacting with enzymes (referred to as “ligands” in the database). They include small molecules as well as macromolecular structures. About 145,000 of these names are currently equipped with a molecular structure. A comparison via the InChI string reveals 106,000 different structures.

Of the 106,000 different structures about 18,000 possess more than one name. 11,000 have two names. 530 compounds are cited with 10 or more names (see also Wittig et al., 2014)! Among the compounds with the highest number of synonyms are inhibitors which are frequently used such as AMP-PNP (adenosine 5′-(β,γ-imido) triphosphate) which occurs with 30 different names and is an often tested inhibitor for ligases or protein kinases (see Table 2). It becomes obvious from the table that many of the names are extremely similar; nevertheless one finds only one of them in a query.

Table 2 Synonyms for AMP-PNP.

Synonyms for the inhibitor AMP-PNP
(beta,gamma-imido)adenosine triphosphate
5-Adenoylimino phosphate
5′-Adenylyl beta,gamma-imidotriphosphate
5′-Adenylyl-beta,gamma-imidotriphosphate
Adenosine 5-(5beta,gamma-imido) triphosphate
Adenosine 5′-(beta,gamma-imido) triphosphate
Adenosine 5′-(beta,gamma-imido)-triphosphate
Adenosine 5′-(beta,gamma-imido)triphosphate
Adenosine 5′-(beta,gamma-Imino)triphosphate
Adenosine 5′-[beta,gamma-imido]-triphosphate
Adenosine 5′-[beta,gamma-imido]triphosphate
Adenosine 5′-[beta,gamma,imido]triphosphate
Adenosine 5′-beta,gamma-imido triphosphate
Adenosine 5′(beta,gamma-imino)triphosphate
Adenosine-5′-(beta,gamma-imino)-triphosphate
Adenylyl 5-imidodiphosphate
Adenylyl beta,gamma-imido diphosphonate
Adenylyl(beta-gamma-imido)triphosphonate
Adenylyl(beta,gamma-imido)triphosphonate
AMP-PNP
AMPPNP
App(NH)p
beta, gamma-imido ATP
beta,gamma-imido-adenosine-5′-triphosphate
beta,gamma-imido-ATP
beta,gamma-imidoadenosine 5-triphosphate
beta,gamma-imidoadenosine 5′-triphosphate
beta,gamma-imidoadenylyl 5′-triphosphate
beta,gamma-imidoATP
beta,gamma-imine-ATP

Table 3 Synonyms for methotrexate.

Synonyms for methotrexate
4-Amino-4-deoxy-10-methylfolate
4-Amino-N10-methylfolic acid
Amethopterin
Methopterin
Methotrexate
Methotrexate monoglutamate
N-(4-[[(2,4-diaminopteridin-6-yl)methyl](methyl)amino] benzoyl)glutamic acid
N-[(4-[[(2,4-diaminopteridin-6-yl)methyl](methyl)amino] phenyl)carbonyl]glutamic acid

For this purpose BRENDA allows a search for structural elements of compounds that are drawn by the users in a chemical editor.

Artificial substrates are frequently used in enzyme assays and appear in the literature with many different names. An example is methotrexate, which occurs in the literature with 8 synonyms (Table 3).

Table 4 Ligand structures connected to kinetic data.

Ligand structures (unique)	Items
Structures for reactants in IUBMB enzyme reactions	5338
Structures for substrates and products	45,464
Structures for cofactors	343
Structures for activating compounds	3055
structures for inhibitors	57,666
Structures linked to K_m-values	13,985
Structures linked to k_{cat}-values	8059
Structures linked to K_i-values	14,343
Structures linked to IC_{50}-values	25,144

In contrast to the BRENDA system most international databases do not allow a search for compounds by structure. When searching the literature for enzyme data, e.g., for all kinetic values for a certain substrate it is important to include all synonyms for the substrate in the search. Therefore BRENDA stores the compound name which is used in the respective citation together with a "recommended name". The BRENDA ligand recommended name is chosen manually from all available synonyms. Mostly it is the systematic name or a name that is very close to it. Sometimes, however, when a trivial name is the most abundant and when this trivial name is unique and not misleading it is designated as recommended. The chemical structure provides an unambiguous identification of the BRENDA ligands. Table 4 shows the sections where ligands are stored and the respective number of different structures.

Enzyme sources: organism, tissue, localisation

A wide range of enzyme sources are available to extract active enzymes. With the fast growing amount of enzyme data the knowledge about the enzyme source, the environmental conditions, the tissues and the intracellular localisation is important for the interpretation and evaluation of the enzyme function in the living organism. Therefore it is necessary to draw on resources with classified and unified terminology to cope with the increasing number of data.

Organism names

The NCBI Taxonomy database represents the main nomenclature and hierarchical classification resource for organism names in BRENDA (Federhen, 2012). This repository for all source organisms in the sequence databases (GenBank, ENA, DDBJ etc.) is manually curated and relies on the current taxonomic literature references and other taxonomy collections (Catalogue of Life, the Encyclopaedia of Life, WikiSpecies etc.) or more specific databases, such as IPNI for plants, Algaebase, Mycobank, Fishbase etc. to maintain a phylogenetic taxonomy corresponding to the evolutionary history of the tree of life.

The NCBI taxonomy (providing data on 846,396 species with formal names and another 491,530 with informal names) contains the scientific name and the synonyms of the organisms, including, if available, the strain information, all assigned to an taxonomy ID, e.g., the ID 4081 is assigned to tomato, the common name of *Solanum lycopersicum*, the preferred scientific name, but also to its synonyms *Lycopersicon esculentum* or *Solanum esculentum*. The enzyme data in the BRENDA database are all organism-specific. If the protein sequence is known, the respective organisms are linked to the NCBI taxonomy browser. Presently BRENDA contains enzyme data for about 10,700 different organisms. About 25% of them are not stored at the NCBI, but these are reviewed by using other databases or the original references.

Tissue or organ

The next deeper level for enzyme sources is the information on the tissue within the organisms. To evaluate the functional enzyme data, it is essential to know from which part of the organism the enzyme was extracted, e.g. lactate dehydrogenase (EC 1.1.1.27) consists of isoenzymes, which could be isolated from the heart, the liver or the lung. Each of these isoenzymes may consist of different subunits and show different functional properties. In 2003, the BRENDA Tissue Ontology, BTO, was developed to cope with the increasing number of tissue terms to provide a structured and standardized representation from all taxonomic groups covering animals, plants, fungi and prokaryotes classifying the different anatomical structures, tissues, cell types and cell lines as enzymes sources (Gremse et al., 2011). The ontology is a flexible system based on controlled and standardized vocabulary which is classified under generic categories, corresponding to the rules and formats of the Gene Ontology Consortium (GO) and organised as directed acyclic graphs (DAG) (Barrell et al., 2009). Every term in the ontology is unique. The terms are supplemented with synonyms, a definition and a literature reference. In order to correctly describe the relationships between "parent" and "child" terms four different types of relations are defined:

- is a (e.g., cardiac muscle fibre is_a muscle fibre);
- part of (e.g., muscle fibre is part_of muscle);
- develops from/derives from (e.g., myoma cell develops_from/derives_from muscle); and
- related to (e.g., electroplax is related_to muscle fibre).

Besides body or plant parts it also contains about 3200 cell lines which are used as enzyme sources. The ontology is constantly enlarged and updated. In 2014 it consists of 5478

Table 5 K_m values for the hydrolysis of delapril by carboxylesterase.

K_m [mM] delapril	Carboxylesterase EC 3.1.1.1
0.028	Enzyme from rat jejunum microsomes, pH 7.4, 37 °C
0.033	Enzyme from rat liver microsomes, pH 7.4, 37 °C
0.041	Enzyme from rat jejunum cytosol, pH 7.4, 37 °C
0.101	Enzyme from rat liver cytosol, pH 7.4, 37 °C

unique terms, 4350 synonyms and 4570 definitions. All entries in the BTO are connected to the enzyme specific information in BRENDA.

Cellular localisation

Additionally the intracellular localisation of an enzyme within the cells and the organelles has an influence on the activity. Therefore they are stored in a structured way according to the concept and rules of the Gene Ontology (GO) to represent controlled terms as sources of enzymes (Barrell et al., 2009). GO describes gene products in terms of their associated biological processes, cellular components and molecular functions in a species-independent manner.

Understanding the behaviour of enzymes depending on their localisation in tissues and organelles is essential in many applications. For example the degradation of drugs may proceed differently in different organs or organelles.

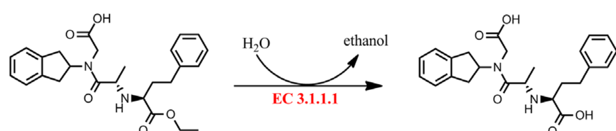


Figure 2 Hydrolysis of the antihypertensive drug delapril by carboxylesterase.

Table 5 shows the K_m values for the drug delapril, an angiotensin-converting-enzyme inhibitor (Takahashi et al., 2008). The first step in its degradation is a hydrolysis by carboxylesterase (EC 3.1.1.1) to release ethanol and *N*-[(2*S*)-1-ethoxy-1-oxo-4-phenylbutan-2-yl]-L-alanyl-*N*-(2,3-dihydro-1 *H*-inden-2-yl)glycine (Figure 2). The lowest K_m-values are observed in jejunum microsomes.

Experimental standards

Enzymatic data from different labs or even different papers from the same laboratory are only comparable when the experimental conditions are fully documented and—even better-measurements are done under standard conditions. These standard conditions should reflect the situation in the "natural environment" of the enzyme as closely as possible. As this requirement is discussed in other papers in this book (e.g., see Tipton et al., 2014) we will focus on the current state in the literature as extracted from the papers covered in BRENDA.

Kinetic data

The characteristics of an enzyme with respect to its function in the organism's metabolism are described by kinetic values such as k_{cat}, K_m, k_{cat}/K_m, V_{max}, K_i. The STRENDA Commission has issued guidelines for the reporting of these values in a

Table 6 k_{cat} values for L-lactate dehydrogenase from *S. cerevisiae* under different experimental conditions.

k_{cat} [1/s] L-lactate	L-lactate dehydrogenase (cytochrome) (EC 1.1.2.3)
7.8	Influence of anions (400 mM KBr) on mutant R289K steady-state kinetic. Reaction conditions: 100 mM phosphate buffer, 1 mM EDTA, pH 7, 5 °C, 2 mM ferricyanide and variable L-lactate concentration
8.8	Influence of anions (400 mM potassium acetate) on mutant R289K steady-state kinetic. Reaction conditions: 100 mM phosphate buffer, 1 mM EDTA, pH 7, 5 °C, 2 mM ferricyanide and variable L-lactate concentration
9.2	Influence of anions (200 mM phsophate) on mutant R289K steady-state kinetic. Reaction conditions: 100 mM phosphate buffer, 1 mM EDTA, pH 7, 5 °C, 2 mM ferricyanide and variable L-lactate concentration; influence of anions (400 mM KCl) on mutant R289K steady-state kinetic. Reaction conditions: 100 mM phosphate buffer, 1 mM EDTA, pH 7, 5 °C, 2 mM ferricyanide and variable L-lactate concentration
45.0	Influence of anions (400 mM KBr) on the FDH domain steady-state kinetic. Reaction conditions: 100 mM phosphate buffer, 1 mM EDTA, pH 7, 5 °C, constant 10 mM ferricyanide and variable L-lactate concentration
60.0	Influence of anions (400 mM KCl) on the FDH domain steady-state kinetic. Reaction conditions: 100 mM phosphate buffer, 1 mM EDTA, pH 7, 5 °C, constant 10 mM ferricyanide and variable L-lactate concentration
61.0	Influence of anions (400 mM KBr) on the wild-type steady-state kinetic. Reaction conditions: 100 mM phosphate buffer, 1 mM EDTA, pH 7, 5 °C, 1.5 mM ferricyanide and variable L-lactate concentration
71.0	Influence of anions (200 mM phsophate) on the wild-type steady-state kinetic. Reaction conditions: 100 mM phosphate buffer, 1 mM EDTA, pH 7, 5 °C, 1.5 mM ferricyanide and variable L-lactate concentration
75.0	Influence of anions (400 mM KCl) on the wild-type steady-state kinetic. Reaction conditions: 100 mM phosphate buffer, 1 mM EDTA, pH 7, 5 °C, 1.5 mM ferricyanide and variable L-lactate concentration
86.0	Influence of anions (400 mM potassium acetate) on the FDH domain steady-state kinetic. Reaction conditions: 100 mM phosphate buffer, 1 mM EDTA, pH 7, 5 °C, constant 10 mM ferricyanide and variable L-lactate concentration
101.0	Influence of anions (300 mM phsophate) on the FDH domain steady-state kinetic. Reaction conditions: 100 mM phosphate buffer, 1 mM EDTA, pH 7, 5 °C, constant 10 mM ferricyanide and variable L-lactate concentration
113.0	Influence of anions (400 mM potassium acetate) on the wild-type steady-state kinetic. Reaction conditions: 100 mM phosphate buffer, 1 mM EDTA, pH 7, 5 °C, 1.5 mM ferricyanide and variable L-lactate concentration

Table 7 BRENDA kinetic data with pH and temperature information.

	Values stored in BRENDA	With pH information (%)	With temperature information (%)
K_m value	123,269	50	46
k_{cat} value	53,545	65	64
k_{cat}/K_m value	11,425	90	80

standardized format (Apweiler et al., 2010; Gardossi et al., 2010; http://www.strenda.org). In order to allow a comparison of values these must be equipped with additional information. For obvious reasons enzyme kinetic data are measured under many different conditions:

- For the reason of convenience the activity may be measured at room temperature, not at controlled temperatures or not at the optimal temperature.
- The substrates may not be stable at the enzyme's optimal temperature. The assay temperature may not be optimal for the enzyme.
- The pH value of the assay medium may not be optimal, because the substrates need a different value for solubility or stability.
- Due to difficulties in the purification procedure the enzymes may be in different degrees of purity. Impurities may influence the efficiency of the catalytic process.
- Additives such as salts, detergents or the immobilisation on various materials may impair or accelerate the reaction.
- The enzyme or the substrates need certain additives in order to be stable. Depending on the compound used, the reaction may be modified.
- Wild-type enzymes from different strains or mutated forms may behave differently.

The kinetic data in BRENDA are extracted manually from the literature. In order to allow quick comparisons the values are recalculated to a standard unit, e.g., mM for K_m, 1/s for k_{cat}.

The experimental conditions, however, have a strong influence on the functional parameters. Therefore where possible, each value is equipped with a comment, giving the temperature, the pH and any other assay conditions if described in the original literature. As shown in Table 6 for the turnover numbers for L-lactate dehydrogenase from *Saccharomyces cerevisiae* the values can differ considerably depending on the ionic strength and the type of salt present in the assay medium.

Within the STRENDA initiative a number of obligatory conditions are defined that are necessary to characterise the experiments when enzyme kinetic data are published (Tipton et al., 2014). In this respect it is interesting to analyse the BRENDA data if at least the most important conditions, pH and temperature were given in the original paper. The analysis is shown in Table 7.

For mutant enzymes the exact sequence modifications must be given, of course. BRENDA lists more than 52,000 single kinetic data for mutant enzymes, either on natural occurring mutations or on mutations achieved by site-directed mutagenesis. Each value is connected to an organism, a protein sequence ID for the enzyme where available, and to a literature reference.

Other standards for enzyme characterisation

In addition to the mentioned cases for enzyme, ligand, organism, tissue, localisation there are a number of other information fields in BRENDA with a controlled vocabulary or a standardized form. This includes:

- Application (25 categories such as agriculture, drug development, diagnostics, environmental protection, medicine, synthesis, toxicology, veterinary medicine etc.).
- Expression (conditions for up- and down-regulation).
- Genetic engineering (positions of mutations).
- Quaternary Structure (monomer dimer etc.).

The DRENDA part of BRENDA covers information on **Enzyme/Disease** relationships, including enzymes where the function or malfunction is connected to a disease or where the enzyme is used for diagnosis or treatment (Söhngen et al., 2011). For the disease-related part of DRENDA the established **Me**dical **S**ubject **H**eadings (MeSH) standard, a comprehensive controlled vocabulary is used that was originally designed for indexing journal articles (Sewell, 1964). This includes, among other categories, 22,000 terms for diseases and metabolic disorders which are classified under the top level category “diseases”.

Conflict of interest statement

None of the authors have any conflict of interest.

Acknowledgements

We wish to thank all scientists who maintain the BRENDA website, performed the literature annotation and created the molecular structures for the enzyme ligands. In addition, this is also to thank the following funding agencies: European Union: SLING—Serving Life-Science Information for the Next Generation [226073]; Federal Ministry of Education and Research, Germany, L3S Research Centre: Advancement and Updating the Enzyme Information System BRENDA [01KX1235]; Niedersächsisches Ministerium für Wissenschaft und Kultur, Germany, MWK/TU Braunschweig [74ZN1122].

References

Apweiler, R., Armstrong, R.N., Bairoch, A., Cornish-Bowden, A., Halling, P.J., Hofmeyr, J.-H.S., Kettner, C., Leyh, T.S., Rohwer, J.M., Schomburg, D., Steinbeck, C., Tipton, K.F., 2010. A large-scale protein-function database. Nat. Chem. Biol. 6, 785.

Barrell, D., Dimmer, E., Huntley, R.P., Binns, D., O'Donovan, C., Apweiler, R., 2009. The GOA database in 2009—an integrated Gene Ontology Annotation resource. Nucleic Acids Res. 37, D396–D403.

Chang, A., Scheer, M., Grote, A., Schomburg, I., Schomburg, D., 2009. BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. Nucleic Acids Res. 37, D588–D592.

Cornish-Bowden, A., 2014. Current IUBMB recommendations on enzyme nomenclature and kinetics. Perspect. Sci. 1, 74-87.

Federhen, S., 2012. The NCBI Taxonomy database. Nucleic Acids Res. 40, D136–D143.

Gardossi, L., Poulsen, P.B., Wohlgemuth, R., Hult, K., Svedas, V.K., Vasić-Racki, D., Carrea, G., Magnusson, A., Schmid, A., Halling, P.J., European Federation of Biotechnology Section on Applied Biocatalysis, 2010. Guidelines for reporting of biocatalytic reactions. Trends Biotechnol. 28, 171–180.

Gremse, M., Chang, A., Schomburg, I., Grote, A., Scheer, M., Ebeling, C., 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.

McDonald, A.G., Boyce, S., Tipton, K.F., 2009. ExplorEnz: the primary source of the IUBMB enzyme list. Nucleic Acids Res. 37, D593–D597.

Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Söhngen, C., Stelzer, M., Thiele, J., Schomburg, D., 2011. BRENDA, the enzyme information system in 2011. Nucleic Acids Res. 39, D670–D676.

Sewell, W., 1964. Medical subject headings in medlars. Bull. Med. Libr. Assoc. 52, 164–170.

Söhngen, C., Chang, A., Schomburg, D., 2011. Development of a classification scheme for disease-related enzyme information. BMC Bioinforma. 12, 329.

Takahashi, S., Katoh, M., Saitoh, T., Nakajima, M., Yokoi, T., 2008. Allosteric kinetics of human carboxylesterase 1: species differences and interindividual variability. J. Pharm. Sci. 97, 5434-5445.

Tipton, Keith, Armstrong, Richard N., Bakker, Barbara, Bairoch, Amos, Cornish-Bowden, Athel, Halling, Peter, Hofmeyr, Jan-Hendrik, Leyh, Thomas S., Kettner, Carsten, Raushel, Frank M., Rohwer, Johann, Schomburg, Dietmar, Steinbeck, Christoph, 2014. Standards for reporting enzyme data: the STRENDA Consortium: what it aims to do and why it should be helpful. Perspect. Sci. 1, 131–137.

Urbanowicz, B.R., Bennett, A.B., Del Campillo, E., Catalá, C., Hayashi, T., Henrissat, B., Höfte, H., McQueen-Mason, S.J., Patterson, S.E., Shoseyov, O., Teeri, T.T., Rose, J.K., 2007. Structural organization and a standardized nomenclature for plant endo-1,4-β-glucanases (cellulases) of glycosyl hydrolase family. Plant Physiol. 9, 144.

Wittig, Ulrike, Kania, Renate, Bittkowski, Meik, Wetsch, Elina, Shi, Lei, Jong, Lenneke, Golebiewski, Martin, Rey, Maja, Weidemann, Andreas, Rojas, Isabel, Müller, Wolfgang, 2014. Data extraction for the reaction kinetics database SABIO-RK. Perspect. Sci. 1, 33–40.

Web references

Algaebase: ⟨www.algaebase.org⟩.
BRENDA: ⟨www.brenda-enzymes.org⟩.
Catalogue of Life: ⟨www.catalogueoflife.org⟩.
DDBJ—DNA Data Bank of Japan: ⟨www.ddbj.nig.ac.jp⟩.
ENA—European Nucleotide Archive: ⟨http://www.ebi.ac.uk/ena/⟩.
Encyclopaedia of Life: ⟨eol.org⟩.
Fishbase: ⟨www.fishbase.org⟩.
GenBank: ⟨www.ncbi.nlm.nih.gov/genbank/⟩.
IPNI—The International Plant Names Index: ⟨www.ipni.org⟩.
Mycobank: ⟨http://www.mycobank.org⟩.
NCBI Taxonomy: ⟨www.ncbi.nlm.nih.gov/taxonomy⟩.
WikiSpecies: ⟨species.wikimedia.org⟩.