Beilstein-Institut ESCEC, Oct. 5th - 8th 2003, Rüdesheim, Germany

CONTROLLED VOCABULARIES AND ONTOLOGIES IN ENZYMOLOGY

KIRILL DEGTYARENKO

European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, U.K.

E-Mail: kirill@ebi.ac.uk

Received: 8th January 2004 / Published: 1st October 2004

ABSTRACT

The diversity of objects and concepts in enzymology can be reflected in the number of possible classifications ('ontologies') needed to describe an 'elementary' biochemical event such as an enzymatic reaction: for example, the overall enzymatic reaction (including direction) taking place under physiological conditions; any other enzymatic reaction catalysed by the same enzyme observed *in vivo* or *in vitro*; the biochemical pathway of which the reaction is part; the mechanism of the enzymatic reaction; an enzyme itself; any of the subunits of a multimeric enzyme. These are all different classes of entities and as such have to be given their own terms and/or identifiers. In reality, the terminology used in publications or biological databases often is a mixture of terms borrowed from orthogonal (or contradicting) classifications.

In this respect, the Enzyme Nomenclature should provide the ultimate reference, whereas in fact it suffers the same problem. EC numbers form a strict hierarchy of *IsA* relationships and the enzymes often require re-classification. It is unlikely that in its current form, the Enzyme Nomenclature can cope with the growing demands of the biological and bioinformatics community in the 21st century. A more flexible, but at the same a time a more strictly defined, approach has been pioneered by the Gene Ontology Consortium, which provides controlled vocabulary for *molecular functions* used to annotate *gene products*. I am going to discuss the building of an Enzyme Ontology. Here, novel relationships unique to chemical ontologies have to be introduced.

Published in „Experimental Standard Conditions of Enzyme Characterizations“, M.G. Hicks & C. Kettner (Eds.), Proceedings of the 1st Int'l Beilstein Symposium on ESCEC, Oct. 5th - 8th 2003, Rüdesheim, Germanytp

Degtyarenko, K.

A FEW NOTES IN LIEU OF AN INTRODUCTION

> *The Hitch Hiker's Guide to the Galaxy is an indispensable companion to all those who are keen to make sense of life in an infinitely complex and confusing Universe, for though it cannot hope to be useful or informative on all matters, it does at least make the reassuring claim, that where it is inaccurate it is at least* definitely *inaccurate. In cases of major discrepancy it's always reality that's got it wrong.*
>
> (Douglas Adams)

'Ontology' is a formal definition of concepts (such as entities and relationships) of a given area of knowledge, described in a standardized form [1]. It can be organized as a structured vocabulary in the form of a directed acyclic graph or a network in which each term may be a 'child' of one or more 'parent' [2].

Naturally, sequences form the *core data* of biological sequence databases such as EMBL [3] and Swiss-Prot (SW) [4], while 3-D coordinates form the core data of structural databases such as the Protein Data Bank (PDB) [5]. The other data typically found in a database entry, such as the name of a gene or protein, the name of the organism, or literature references, are called *annotation*. In contrast, the Enzyme Nomenclature [6] and Gene Ontology (GO) [2] contain a fundamentally different kind of core data: terms and definitions. Thus GO contains free text definitions as in a dictionary, while in the Enzyme Nomenclature it is the chemical reactions themselves which characterize the enzymatic function.

Throughout this paper, I provide examples from the biological databases in the form (DatabaseName:AccessionNumber), e.g. SW:P15335, GO:0016610, PDB:1H4J. The database entries themselves contain comprehensive literature and database references.

VOCABULARIES AND ONTOLOGIES FOR BIOCHEMICAL REACTIONS

Enzyme nomenclature

The Enzyme Nomenclature provides the oldest controlled vocabulary for biochemical function. It was originally developed by the Enzyme Commission (EC) of the International Union of Biochemistry and now is published by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) [6]. The EC number serves as both unique identifier (ID) and descriptor of the enzyme place in hierarchy. Not surprisingly, EC numbers are often used (and misused) for annotation of gene products.

It is important to remember that the basis of the Enzyme Nomenclature is the *overall reaction catalysed* [7], but not the reaction mechanism or any other specific property of an enzyme. Nevertheless, other biological catalysts such as ribozymes [8], deoxyribozymes [9] or catalytic antibodies (abzymes) [10] do not form a part of the Enzyme Nomenclature. EC numbers form a *taxonomy*, a strict hierarchy of parent-child relationships known as *IsA* ('is kind of').

IntEnz

At the EBI, enzyme classification is collected in the Integrated relational Enzyme database (IntEnz) [11], a joint project with Trinity College Dublin, the Swiss Institute of Bioinformatics and the University of Cologne, supported by the NC-IUBMB. Currently, IntEnz contains enzyme data curated and approved by the members of NC-IUBMB. The goal is to create a single relational database containing all the relevant enzyme data, including those from the ENZYME [12] and BRENDA [13] databases.

GO

The Gene Ontology Consortium [2] has been developing controlled vocabularies which are widely used by the bioinformatics community. GO comprises three domains: 'molecular function', 'biological process' and 'cellular component'. In each domain, the GO terms are organized as a directed acyclic graph (DAG), which differs from a taxonomy in that a child term can have one or more parent terms. GO uses two parent-child relationships, *IsA* and *IsPartOf*. Throughout the text, I will use symbols for these relationships as specified in the GO File Format Guide [14]:

%	=	*IsA*
<	=	*IsPartOf*

GO is a part of the wider initiative known as OBO (Open Biology Ontologies). A list of freely available ontologies that are relevant to genomics and proteomics and which are structured in same way as GO can be found at the OBO website [15].

Specialized enzyme function resources

These include MEROPS, the Peptidase Database [16], REBASE, The Restriction Enzyme Database [17], CAZy, the database of carbohydrate-active enzymes [18], and UM-BBD,

The University of Minnesota Biocatalysis/Biodegradation Database [19]. The first three databases aim to provide comprehensive in-depth information about corresponding groups of enzymes. For example, REBASE (as on the 16 December 2003) contains information on 3631 restriction enzymes, which correspond to only three EC numbers: EC 3.1.21.3, EC 3.1.21.4 and EC 3.1.21.5. UM-BBD is not focused on any particular groups of enzymes. Currently, it lists 573 enzymes, many of which do not have EC numbers assigned, and 900 (enzymatic and non-enzymatic) reactions of xenobiotic biodegradation in micro-organisms.

WHAT IS THE MEANING OF AN ENZYME NAME?

Both in the literature and in biological databases, enzyme names and EC numbers are used to describe various concepts such as:

- an enzyme
- any subunit of a multimeric enzyme
- an enzyme system
- the overall enzymatic reaction (including the direction) taking place under physiological conditions
- any of the enzymatic reactions catalysed by the same enzyme observed *in vivo* or *in vitro*
- the reaction in the metabolic pathway context

Note that the first three concepts are biochemical objects while the last three are biochemical events. To illustrate the difference between these concepts, consider several database entries containing the term "nitrogenase" (EC 1.18.6.1), starting with the IntEnz entry (Example 1).

Example 1. EC 1.18.6.1 in IntEnz

IUBMB Enzyme Nomenclature
EC 1.18.6.1

Common name:	Nitrogenase
Reaction:	**3** reduced ferredoxin + **6** H^+ + N_2 + **n** ATP = **3** oxidized ferredoxin + **2** NH_3 + **n** ADP + **n** phosphate
Systematic name:	reduced ferredoxin:dinitrogen oxidoreductase (ATP-hydrolysing)
Comments:	An iron-molybdenum protein. Acetylene can also act as acceptor; in the absence of other acceptros, H^+ is reduced to H_2; **n** is about 12 - 18. Formerly EC 1.18.2.1
Links to other databases:	BRENDA, EXPASY, GO, KEGG, WIT CAS registry number: 9013-04-1
References:	1. Zumft, W.G., Paneque, A., Aparicio, P.J. and Losada, M. Mechanism of nitrate reduction in *Chlorella*. *Biochem. Biophys. Res. Commun.* 36(1969) 980-986.[Medline UI: 70008605]

[EC 1.18.6.1 created 1978 as EC 1.18.2.1, transferred 1984 to EC 1.18.6.1]

The comment to this entry mentions that nitrogenase is "an iron-molybdenum protein". Now let us see how this enzyme is presented in a protein sequence database. Swiss-Prot entries are species-specific. For instance, there are 11 entries for EC 1.18.6.1 from *Azotobacter vinelandii* (Example 2). These constitute three enzymes: iron-molybdenum nitrogenase or dinitrogenase (SW:P07328, SW:P07329, SW:P00459), iron-vanadium nitrogenase or dinitrogenase 2 (SW:P16855-P16857, SW:P15335) and iron-iron nitrogenase or dinitrogenase 3 (SW:P16266-P16268). Note that only the first name in the description line provides an unequivocal name for a polypeptide in this species; the other names may be shared between several entries (e.g. eight instances of "Nitrogenase component I") while all of them are referred to as EC 1.18.6.1.

Example 2. EC 1.18.6.1 from *Azotobacter vinelandii* in Swiss-Prot.

ID	AC	Description	Cofactor
NIFD_AZOVI	P07328	Nitrogenase molybdenum-iron protein alpha chain (EC 1.18.6.1) (Nitrogenase component I) (Dinitrogenase).	FeMoco
NIFK_AZOVI	P07329	Nitrogenase molybdenum-iron protein beta chain (EC 1.18.6.1) (Nitrogenase component I) (Dinitrogenase).	Fe_8S_7
NIH1_AZOVI	P00459	Nitrogenase iron protein 1 (EC 1.18.6.1) (Nitrogenase component II) (Nitrogenase Fe protein 1) (Nitrogenase reductase).	Fe_4S_4
VNFD_AZOVI	P16855	Nitrogenase vanadium-iron protein alpha chain (EC 1.18.6.1) (Nitrogenase component I) (Dinitrogenase 2 alpha subunit).	FeVco
VNFK_AZOVI	P16856	Nitrogenase vanadium-iron protein beta chain (EC 1.18.6.1) (Nitrogenase component I) (Dinitrogenase 2 beta subunit).	Fe_8S_7?
VNFG_AZOVI	P16857	Nitrogenase vanadium-iron protein delta chain (EC 1.18.6.1) (Nitrogenase component I) (Dinitrogenase 2 delta subunit).	
NIH2_AZOVI	P15335	Nitrogenase iron protein 2 (EC 1.18.6.1) (Nitrogenase component II) (Nitrogenase Fe protein 2) (Nitrogenase reductase).	Fe_4S_4
ANFD_AZOVI	P16266	Nitrogenase iron-iron protein alpha chain (EC 1.18.6.1) (Nitrogenase component I) (Dinitrogenase 3 alpha subunit).	Feco?
ANFK_AZOVI	P16267	Nitrogenase iron-iron protein beta chain (EC 1.18.6.1) (Nitrogenase component I) (Dinitrogenase 3 beta subunit).	Fe_8S_7?
ANFG_AZOVI	P16268	Nitrogenase iron-iron protein delta chain (EC 1.18.6.1) (Nitrogenase component I) (Dinitrogenase 3 delta subunit).	
NIH3_AZOVI	P16269	Nitrogenase iron protein 3 (EC 1.18.6.1) (Nitrogenase component II) (Nitrogenase Fe protein 3) (Nitrogenase reductase).	Fe_4S_4

Example 3. Nitrogenase in GO.

```
   Molecular function
   %oxidoreductase activity\, acting on reduced ferredoxin as donor\,
dinitrogen as acceptor ; GO:0016732 ; EC:1.18.6.-
     %nitrogenase activity ; GO:0016163 ; EC:1.18.6.1 ;
MetaCyc:NITROGENASE-RXN ; UM-BBD_enzymeID:e0395
   Cellular component
   <intracellular ; GO:0005622 ; synonym:protoplasm
     <nitrogenase complex ; GO:0016610
       %iron-iron nitrogenase complex ; GO:0016611
       %molybdenum-iron nitrogenase complex ; GO:0016612
       %vanadium-iron nitrogenase complex ; GO:0016613
```

In GO, the EC number is mapped to its molecular function (Example 3).

This is reflected in the term 'activity' now customarily added to the enzyme name. In the GO molecular function ontology, *nitrogenase activity* (GO:0016163) is defined as:

"Catalysis of the reaction: 8 reduced ferredoxin + 8 H^+ + nitrogen + 16 ATP = 8 oxidized ferredoxin + 2 ammonia + 16 ADP + 16 phosphate". In the GO cellular component ontology, nitrogenase complex (GO:0016610) is defined as "An enzyme complex composed of two proteins, dinitrogenase and nitrogenase reductase; dinitrogenase is tetrameric with an $alpha_2$-$beta_2$ structure and nitrogenase reductase is a homodimer, and both are associated with metal ions, which differ between species. Both proteins are required for the enzyme activity of the complex, the formation of oxidized ferredoxin and ammonia from reduced ferredoxin and nitrogen". The only missing logical link is the one between the object (nitrogenase complex) and its function (nitrogenase activity).

Apart from recommended names, the enzyme names used in the literature and databases annotation often include various 'functional qualifiers' (see Example 4, underscored terms). Many organisms have more than one enzyme with the same EC number, and therefore such qualifiers are needed to provide a less ambiguous description of the enzyme function.

Example 4. Some functional qualifiers for enzyme names in Swiss-Prot.

ID	AC	Description	EC number
NASC_BACSU	P42434	Assimilatory nitrate reductase catalytic subunit	EC 1.7.99.4
SIR_DEVSVH	Q05805	Sulfite reductase, assimilatory-type	EC 1.8.-.-
DSVA_DESVH	P45574	Sulfite reductase, dissimilatory-type alpha subunit	EC 1.8.99.3
3DHQ_ACICA	Q59087	Catabolic 3-dehydroquinate dehydratase	EC 4.2.1.10
MBHL_ALCEU	P31891	Uptake hydrogenase large subunit	EC 1.12.99.6
OTCC-PSESH	P23752	Ornithine carbamoyltransferase, phaseolotoxin-insensitive, catabolic	EC 2.1.3.3
CN8B_HUMAN	O95263	High-affinity cAMP-specific and IBMX-insensitive 3',5'-cyclic phosphodiesterase 8B	EC 3.1.4.17
PBPB_VIBCH	Q9KUCo	Penicillin-binding protein 1B [Includes: Penicillin-insensitive transglycosylase Penicillin-sensitive transpeptidase]	 EC 2.4.1.129 EC 3.4.-.-
ABP_HUMAN	P19801	Amiloride-sensitive amine oxidase [copper-containing] precursor	EC 1.4.3.6
NFSA-ECOLI	P17117	Oxygen-insensitive NADPH nitroreductase	EC 1.-.-.-
AOX2_ARATH	O22049	Alternative oxidase 2, mitochondrial precursor	EC 1.-.-.-
COOH-RHORU	P31895	Carbon monoxide-induced hydrogenase	
CDCLZ_ECOLI	P52095	Lysine decarboxylase, constitutive	EC 4.1.1.18

Some myths about EC numbers

- every EC number appearing in the literature is approved by NC-IUBMB
- every EC number is a unique and stable identifier for an enzymatic reaction
- EC numbers always correspond to reactions taking place under physiological conditions
- EC numbers can be predicted by similarity with known enzymes
- EC stands for European Community

I am sure that many participants of this workshop can deliver more fascinating examples of the 'EC number mythology', together with an exhaustive critique on each myth. However, my point is that we cannot be content with the current situation and should not mostly blame a user for misunderstanding what an EC number is. If the Enzyme Nomenclature assigns EC numbers to *proteins*, we should not really be surprised by bioinformaticists' claims such as "for functional annotation, 40% sequence identity can still be used as a confident threshold to transfer the first three digits of an EC number" [20], or that "the enzyme code similarity" can be used to increase the accuracy of protein fold recognition [21].

To clarify the meanings of the enzyme names and their usage in biological databases, I am going to revisit the three general principles of the Enzyme Nomenclature which can be summarized as follows [22]:

1. The names of enzymes, especially those ending in *-ase*, should be used only for single enzymes (single catalytic entities).
2. The enzymes are principally classified and named according to the reaction they catalyse.
3. The enzymes are divided into groups on the basis of the type of reaction catalysed, and this, together with the name(s) of the substrate(s) provides a basis for naming individual enzymes. It is also the basis for classification and code numbers.

Enzyme-polypeptide relationships

"The first *general principle* … is that names purporting to be names of enzymes, especially those ending in *-ase*, should be used only for single enzymes, i.e. single catalytic entities. They should not be applied to systems containing more than one enzyme" [22]. I dwell on three problems with realization of this principle. The first is that there is no clear definition of what constitutes a "single enzyme". Let us come back to our example of nitrogenase. Component II of nitrogenase (an iron-sulfur protein, also known as "nitrogenase reductase") dissociates from component I (whether it is a molybdenum-iron, vanadium-iron or iron-iron protein) several times during its reaction cycle, but it reacts only with component I. The nitrogenase experts agree that it would be meaningless to classify the two proteins as separate enzymes: although the proteins do dissociate, they are part of a single enzyme (Richard Cammack, personal communication). The second problem is that there are no authoritative recommendations on naming systems apart from the notion that "the word *system* should be included in the name".

Finally, what is classified today as a single enzyme may be found to be a system tomorrow, and vice versa. For example, several P450 mono-oxygenases were classified as members of EC 1.14.13 sub-subclass ("With NADH or NADPH as one donor, and incorporation of one atom of oxygen"). Their corresponding reactions uniformly contain NADPH, while it is known that apart from P450nor (nitric oxide reductase), P450s do not interact directly with NADPH. Eukaryotic P450s are reduced by either a flavoprotein (NADPH-cytochrome P450 reductase; EC 1.6.2.4), or an iron-sulphur protein (adrenodoxin), which in turn is reduced by a flavoprotein (NADPH-adrenodoxin reductase) [23]. Therefore, if one has to follow Principle 1, these enzymes should be re-classified as soon as there is enough evidence about their cognate reductases. The question is, why not assign an identifier to the enzyme system if there is a good reason to do so, as is done already for a number of 'systems'? For instance, fatty-acid synthase (EC 2.3.1.85) has a comment: "The animal enzyme is a multi-functional protein catalysing the reactions of EC 2.3.1.38 [acyl-carrier-protein] *S*-acetyltransferase, EC 2.3.1.39 [acyl-carrier-protein] *S*-malonyltransferase, EC 2.3.1.41 3-oxoacyl-[acyl-carrier-protein] synthase, EC 1.1.1.100 3-oxoacyl-[acyl-carrier-protein] reductase, EC 4.2.1.61 3-hydroxypalmitoyl-[acyl-carrier-protein] dehydratase, EC 1.3.1.10 enoyl-[acyl-carrier-protein] reductase (NADPH, B-specific) and EC 3.1.2.14 oleoyl-[acyl-carrier-protein] hydrolase".

The reaction catalysed by the system may be found also to be catalysed by a single fusion protein. The well studied example is P450 BM-3 from *Bacillus megaterium*, which is a fusion of P450 mono-oxygenase haem domain with NADPH-cytochrome P450 reductase (CPR). In plants and animals, P450 enzymes and CPR exist as separate proteins. Owing to its domain arrangement, P450 BM-3 has the highest catalytic activity of any known P450 mono-oxygenase system [24].

The *Nomenclature for multi-enzymes* [25] gives recommendations for proteins "possessing more than one catalytic function contributed by distinct parts of a polypeptide chain ('domains'), or by distinct subunits, or both." The concept of 'catalytic entity' of Principle 1 is thus extended towards domains and subunits. These recommendations are adopted in Swiss-Prot for the description of multi-enzymes. The description line of the previously mentioned P450 BM-3 (SW:P14779) contains: "[Includes: Cytochrome P450 102 (EC 1.14.14.1); NADPH-cytochrome P450 reductase (EC 1.6.2.4)]".

It is less clear how to deal with a novel 'monodomain' enzyme, which corresponds to a domain in a previously known enzyme. For instance, animal nitric oxide synthase (NOS; EC 1.14.13.39) includes an N-terminal haem-containing oxygenase domain and a C-terminal reductase domain, homologous to CPR. Bacterial nitric oxide synthase oxygenase protein (SW:O34453) is homologous to the NOS oxygenase domain and is able to synthesize nitric oxide upon reduction; however, its physiological reductase remains to be discovered. Table 1 summarizes the possible scenarios of enzyme-polypeptide relationships.

Table 1. Enzyme-polypeptide relationships.

Enyzme	Polypeptide	Protein	Example
1	1	Monomeric monofunctional enzyme	SW:P50578 (EC 1.1.1.2)
1	>1	Multisubunit enzyme	PDB:1H4J (EC 1.1.99.8)
>1	1	Multienzyme polypeptide	SW:Q9DCL9 (EC 6.3.2.6; EC 4.1.1.21)
>1	1	Polyprotein	SW:P08291(EC 3.4.22.29; EC 3.4.22.28; EC 2.7.7.48)
>1	>1	Multienzyme complex	SW:Q04728 (EC 2.3.1.35 ; EC 2.3.1.1) SW:P40939 (EC 4.2.1.17; EC 1.1.1.35) and SW:P55084 (EC 2.3.1.16)

I) 1 enzyme: 1 polypeptide

The famous "one gene-one enzyme" hypothesis of Beadle and Tatum [26] is still very much with us, being found in about every biochemistry textbook around, even though neither "one gene-one polypeptide" nor "one polypeptide-one enzyme" models are always true. How much easier would life be for database authors if one enzyme always did correspond to one polypeptide! Many monomeric enzymes apparently have evolved from oligomeric enzymes via gene duplication/fusion events [27-29].

II) 1 enzyme: >1 polypeptide

This is the case of a multisubunit (oligomeric) enzyme. The difference between a monomeric enzyme and a subunit of an oligomeric enzyme is that the latter is not active on its own. Therefore, one can argue that an EC number should not be assigned to any of the subunits (as is customarily done in sequence databases) but only to a functional complex. The counter-argument is that to do this would be not helpful since in the protein sequence databases, as a rule, one entry corresponds to one polypeptide.

III) >1 enzyme: 1 polypeptide

a) Multi-enzyme polypeptide

A multi-enzyme polypeptide contains several domains associated with separate enzymatic activities (see EC 2.3.1.85 example above).

b) Polyprotein

In this case, one polypeptide is a precursor of more than one mature protein. Many viral polyproteins give rise to several separate enzymes. For example, POL polyprotein from human immunodeficiency virus type 1 is a precursor of HIV-1 retropepsin (EC 3.4.23.16), RNA-directed DNA polymerase (EC 2.7.7.49) and ribonuclease H (EC 3.1.26.4); all three enzymes were structurally characterized (SW:P03366).

IV) >1 enzyme: >1 polypeptide

Here, each subunit has one or more separate enzymatic activities. For instance, yeast fatty acid synthase (EC 2.3.1.86) is a multi-enzyme complex consisting of two types of multifunctional subunits organized as an $\alpha_6\beta_6$ hetero-oligomer (SW:P19097; SW:P07149).

Some other examples do not fit easily into the above four scenarios. The arginine biosynthesis bifunctional protein ARG7 (SW:Q04728) is a heterodimer of a large and a small subunit that are proteolytically processed from a single polypeptide precursor within the mitochondrion. In the case of mitochondrial fatty acid β-oxidation trifunctional protein, the α-subunit is bifunctional (EC 4.2.1.17 and EC 1.1.1.35) while the β-subunit is monofunctional (EC 2.3.1.16) [30].

Enzyme-reaction relationships

Let us now consider enzyme-reaction relationships (Table 2). Three possible scenarios are:

Table 2. Enzyme-reaction relationships.

Enzyme	Reaction	Protein	Example
1	1	Monofunctional enzyme	
1	>1	Promiscuous enzyme	EC 1.4.3.19
1	>1	Enzyme catalyses sequential reactions	EC 3.5.3.21
>1	1	Enzymes are assigned different EC numbers because of different mechanism or by "historical reasons"	EC 1.4.3.4 and EC 1.4.3.6EC 3.4.16.4 and EC 3.4.17.14

I) 1 enzyme: 1 reaction

This is the 'typical' case, which is in accord with Principle 2 of Enzyme Nomenclature which states that "enzymes are principally classified and named according to the reaction they catalyse". In a way, this principle is responsible for confusion arising time and again when the enzyme (i.e. protein) is equated with an EC number.

II) 1 enzyme: >1 reaction

a) Promiscuous enzyme

One enzyme may catalyse more than one reaction. Example 5 shows the entry for glycine oxidase (EC 1.4.3.19) containing four alternative reactions, only one of which involves glycine. Cf. Rule 15 of [22]: "When an enzyme catalyses more than one type of reaction, the name should normally refer to one reaction only."

Example 5. Alternative reactions.

	EC 1.4.3.19
Common name:	glycine oxidase
Reaction:	(1) glycine + H_2O + O_2 = glyoxylate + NH_3 + H_2O_2 (2) D-alanine + H_2O + O_2 = pyruvate + NH_3 + H_2O_2 (3) sarcosine + H_2O + O_2 = glyoxylate + methylamine + H_2O_2 (4) *N*-ethylglycine + H_2O + O_2 = glyoxylate + ethylamine + H_2O_2
Systematic name:	glycine:oxygen oxidoreductase (deaminating)
Comments:	A flavoenzyme containing non-covalently bound FAD. The enzyme from *Bacillus subtilis* is active with glycine, sarcosine, *N*-ethylglycine, D-alanine, D-α-aminobutyrate, D-proline, D-pipecolate and *N*-methyl-D-alanine. It differs from EC 1.4.3.3, D-amino acid oxidase, due to its activity on sarcosine and D-pipecolate.

b) Sequential reactions

Example 6 shows the entry for methylenediurea deaminase (EC 3.5.3.21) containing three consecutive reactions; only reaction 1 is enzymatic while reactions 2 and 3 are described as 'spontaneous' (while the meaning is 'non-catalytic').

Example 6. Sequential reactions.

EC 3.5.3.21

Common name:	mehtylenediurea deaminase
Reaction:	(1) NH_2-CO-NH-CH_2-NH-CO-NH_2 + H_2O = *N*-(carboxyaminomethyl)urea + NH_3 (2) *N*-(carboxyaminomethyl)urea = *N*-(aminomethyl)urea + CO_2 (spontaneous) (3) *N*-(aminomethyl)urea + H_2O = *N*-(hydroxymethyl)urea + NH_3 (spontaneous)
Systematic name:	methylenediurea aminohydrolase
Comments:	The methylenediurea is hydrolysed and decarboxylated to give an aminated methylurea, which then spontaneously hydrolyses to hydroxymethylurea. The enzyme from *Ochrobactrum anthropi* also hydrolyses dimethylenetriurea and trimethylenetetraurea as well as ureidoglycolate, which is hydrolysed to urea and glyoxylate, and allantoate, which is hydrolysed to ureidoglycolate, ammonia and carbon dioxide.

Cf. "Special problems attend the classification and naming of enzymes catalysing complicated transformations that can be resolved into several sequential or coupled intermediary reactions of different types, all catalysed by a single enzyme (not an enzyme system). Some of the steps may be spontaneous non-catalytic reactions, while one or more intermediate steps depend on catalysis by the enzyme" [22].

III) >1 enzyme: 1 reaction

Rule 16 of [22] states: "A group of enzymes with closely similar specificities should normally be described by a single entry. <…> Separate entries are also appropriate for enzymes having similar catalytic functions, but known to differ basically with regard to reaction mechanism or to the nature of the catalytic groups, *e.g. amine oxidase (flavin-containing)* (EC 1.4.3.4) and *amine oxidase (copper-containing)* (EC 1.4.3.6)." It is clear that Rule 16 is but a built-in violation of Principle 2, for the reaction mechanism does not change the overall reaction.

Enzyme classification

"A *third general principle* adopted is that the enzymes are divided into groups on the basis of the type of reaction catalysed, and this, together with the name(s) of the substrate(s) provides a basis for naming individual enzymes. It is also the basis for classification and code numbers."

Somewhat confusingly, Principle 3 in part repeats Principle 2, viz. "enzymes are principally classified and named according to the reaction they catalyse." However, if Principle 2 is mostly relevant to the naming of the enzyme, it is Principle 3 which is responsible for the current classification of enzymes.

Principle 3 allows one and only one way of classification. (Any one EC number belongs to one and only one sub-subclass, which belongs to one and only one subclass, which belongs to one and only one class.) Since an enzyme cannot be simultaneously a member of two different classes (subclasses, sub-subclasses), many enzymes keep being renamed over and over again [7] (Example 7).

Example 7. The history line of EC 3.4.21.103.

```
[EC 3.4.21.103 created 1992 as EC 3.4.23.27 (EC 3.4.23.6 created 1992
(EC 3.4.23.6 created 1961 as EC 3.4.4.17, transferred 1972 to EC
3.4.23.6, modified 1981 [EC 3.4.23.7, EC 3.4.23.8, EC 3.4.23.9, EC
3.4.23.10, EC 3.4.99.1, EC 3.4.99.15 and EC 3.4.99.25 all created 1972
and incorporated 1978], part incorporated 1992), transferred 2003 to EC
3.4.21.103]
```

Historically, the EC number served as both unique ID and descriptor of the enzyme place within the hierarchy. For example, in oxidoreductases (EC 1), the subclass indicates the type of donor (e.g. EC 1.1 "acting on the CH-OH group of donors") and the sub-subclass indicates the type of acceptor (e.g. EC 1.1.1 "with NAD^+ or $NADP^+$ as acceptor"). Of course, the placement of the donor in the second position and the acceptor in the third position is completely arbitrary. (For a reverse reaction they should swap places.) This dual function of EC numbers is fairly limiting because it requires the enzyme to have a *unique* place within the hierarchy. However, an enzyme can be *correctly* classified in more than one way. For example, intramolecular oxidoreductases (EC 5.3) are as much oxidoreductases (EC 1) as isomerases (EC 5). In every subclass of oxidoreductases, the acceptors form repeating series, and therefore the alternative grouping is feasible, e.g. EC 1.1.1, EC 1.2.1, …, EC 1.18.1 can be classified in an 'EC 1.x.1' subclass of 'oxidoreductases with NAD^+ or $NADP^+$ as acceptor'.

The systematic names, in general, are created according to the same conventions as EC numbers. However, not every enzyme entry has a systematic name.

Even within one sub-subclass, some additional hierarchical *IsA* relationships can be suggested on the basis of a natural hierarchy of chemical compound classes.

For instance, 3-hydroxybenzyl-alcohol dehydrogenase reaction (EC 1.1.1.97) can be considered to be a kind of aryl-alcohol dehydrogenase ($NADP^+$) reaction (EC 1.1.1.91) that, in turn, can be considered as a kind of alcohol dehydrogenase ($NADP^+$) reaction (EC 1.1.1.2).

```
%EC 1.1.1.2 alcohol dehydrogenase (NADP+)
  %EC 1.1.1.91 aryl-alcohol dehydrogenase (NADP+)
    %EC 1.1.1.97 3-hydroxybenzyl-alcohol dehydrogenase
```

Another problem of enzyme classification concerns the classes of overall transformations themselves. As Table 3 shows, the six classes of enzyme-catalysed reactions are mostly based on fundamental overall transformations of organic chemistry [31]. On the one hand, all EC 3 (Hydrolases), many EC 1 (Oxidoreductases) and some EC 4 (Lyases) can be considered as EC 2 (Transferases). EC 6 (Ligases) also can be considered as a kind of EC 3 (Hydrolases) because all ligase reactions involve hydrolysis of a diphosphate bond in ATP or other triphosphate. On the other hand, there are no classes for some fundamental reaction types, e.g. addition not to double bonds.

Table 3. Overall transformation classes in enzymology and organic chemistry.

Enzyme classes	**Transformations in organic chemistry** from [31]
EC 1, Oxidoreductases $AH_2 + B \leftrightarrow A + BH_2$ $AH_2 + B^+ \leftrightarrow A + BH + H^+$ EC 2, Transferases A-X + B-H ⇔ A-H + B-X EC 3, Hydrolases A-B + HOH → A-H + B-OH EC 6, Ligases A + B + XTP → A-B + XDP + P_i A + B + XTP → A-B + XMP + PP_i	Substitution A-X + B-Y → A-Y + B-X σ-bound atom or group is replaced by another σ-bound atom or group
EC 4 (synthases, or reverse lyases) X=Y + Z → X-Y-Z	Addition A + B → A-B usually, one π bond in A replaced by two new σ bonds
EC 4, Lyases X-Y-Z →X=Y + Z	Elimination A-B → A + B usually, two σ bonds in A-B replaced by a new π bond
EC 5, Isomerases A ⇔ B	Rearrangement A → B

Example 8. Misclassified enzymes?

Enzyme	**Reaction**
EC 4.99 Other Lyases	
EC 4.99.1.1 Ferrochelatase	protoporphyrin + Fe^{2+} = protoheme + **2** H^+
EC 4.99.1.2 alkylmercury lyase	$RHg^+ + H^+ = RH + Hg^{2+}$
EC 6.6.1 Forming coordination complexes Magnesium chelatase	ATP + protoporhyrin IX + Mg^{2+} + H_2O = ADP + phosphate + Mg-protoporphyrin IX + **2** H^+

Some enzymes are classified together simply because there is no suitable (sub)subclass. Example 8 shows that EC 4.99 (Other Lyases) hosts EC 4.99.1.1 (ferrochelatase) and EC 4.99.1.2 (alkylmercury lyase), although neither enzyme fits the definition of lyases (Keith Tipton, personal communication). Ferrochelatase is involved in the biosynthesis of a coordination compound (*creates* metal-N bond) while alkylmercury lyase catalyses breakdown of an organometallic compound (*breaks* metal-C bond). On the other hand, EC 6.6.1.1 (magnesium chelatase) couples ATP hydrolysis (1) with creation of a metal-N bond (2):

$$ATP + H_2O \rightarrow ADP + phosphate \quad (1)$$

$$protoporphyrin\ IX + Mg^{2+} \rightarrow \quad Mg\text{-}protoporphyrin\ IX + \mathbf{2}\ H^{+} \quad (2)$$

$$ATP + protoporphyrin\ IX + Mg^{2+} + H_2O \rightarrow ADP + phosphate + Mg\text{-}protoporphyrin\ IX + \mathbf{2}\ H^{+}$$

The use of partial reactions provides a logical solution for building a classification of enzymatic reactions. In this particular example, both EC 4.99.1.1 and EC 6.6.1.1 should be classified as metallochelatases (create metal-N bond), while magnesium chelatase also belongs to the ATP hydrolases.

Ligases exemplify one remarkable feature of enzymes, viz. their ability to couple different reactions. Apart from EC 6, many EC 1 (e.g. NAD^+/$NADP^+$-dependent) and EC 3 (e.g. EC 3.6.4 "Acting on acid anhydrides; involved in cellular and subcellular movement") enzymes catalyse coupled reactions. In my opinion, the Enzyme Nomenclature will benefit from explicit descriptions of coupled reactions. However, one cannot define coupled reactions while ignoring the reaction directionality.

In most of the Enzyme Nomenclature entries, "the direction in which the reaction occurs is not specified (i.e. an equals sign is used rather than an arrow) so, even if a reaction has only been observed in the reverse direction, it is usually written in the direction that is common to the subclass to which it belongs" [7]. (In the other entries, the verbal description of the reaction often does convey the direction, e.g. EC 3.1.6.7 "Hydrolysis of the 2- and 3-sulfate groups of the polysulfates of cellulose and charonin".) Of course, this practice contrasts with what a chemist would intuitively expect, i.e. that a reaction is written in the direction it occurs. It also affects the *systematic* names, for these are "derived from a written reaction, even though only the reverse of this has been actually demonstrated experimentally" [22].

As *The Hitch Hiker's Guide to the Galaxy* says, "it's always reality that's got it wrong" [32]. But it is also in contrast with the higher order Enzyme Nomenclature itself. Three class names (EC 3 Hydrolases, EC 4 Lyases and EC 6 Ligases) and numerous subclass names (e.g. EC 6.4 "Forming Carbon-Carbon Bonds") *do* imply the direction of the reaction.

This poses little problem for irreversible reactions or when the reaction can be catalysed by the same enzyme in both directions. However, it matters in cases when the opposite reactions are catalysed by different enzymes which are nevertheless given the same EC number (Example 9).

Example 9. Same EC numbers for the reverse reactions.

Enzyme	Reaction
EC 1.3.5.1	
Succinate dehydrogenase	succinate + Q $\rightarrow$ fumarate + QH_2
Fumarate reductase	fumarate + QH_2 $\rightarrow$ succinate + Q
EC 1.18.1.2	
Ferredoxin–$NADP^+$ reductase	reduced ferredoxin + $NADP^+$ $\rightarrow$ oxidized ferredoxin + NADPH + H^+
NADPH–adrenodoxin reductase	oxidized ferredoxin + NADPH + H^+ $\rightarrow$ reduced ferredoxin + $NADP^+$

Therefore, the three fundamental principles of the Enzyme Nomenclature deal satisfactorily only with the simplest cases (one enzyme:one polypeptide; one enzyme:one reaction; one enzyme:one way of classification).

Other relationships between enzyme entities

EC numbers form a strict hierarchy of *IsA* relationships. However, other relationships can exist between EC numbers. The Enzyme Nomenclature includes a number of these relationships.

I have divided them into 'structural and functional', 'historical' and 'dodgy' relationships (Tables 4-6).

Table 4. Structural and functional relationships between EC entries.

I to J	Example	Comment
(reaction) **I** is kind of (reaction) **J**	EC 1.1.1.1 is kind of EC 1.1.1	*IsA* (is kind of) relationship
I is a component of **J**	EC 1.3.99.1 is a component of EC 1.3.5.1	*IsPartOf* relationship
I and **J** are involved in metabolic process **K**	EC 1.1.1.252 is involved with EC 4.2.1.94 in the biosynthesis of melanin in pathogenic fungi	(unspecified) metabolic relationship
product of **I** is a substrate of **J**	The substrate of EC 3.5.4.26 is the product of EC 3.5.4.25	direct metabolic interaction
I activates **J**	EC 3.4.24.29 activates EC 3.4.21.19	positive regulatory relationship
I inactivates **J**	EC 3.1.2.16 inactivates EC 4.1.3.6	negative regulatory relationship
I phosphorylates **J**	EC 2.7.1.110 phosphorylates and activates EC 2.7.1.109	enzyme-substrate relationship
EC **I** dephosphorylates EC **J**	EC 3.1.3.44 dephosphorylates and activates EC 6.4.1.2	enzyme-substrate relationship

Table 5. Historical relationships between EC entries.

I to J	Example	Comment
I is transferred to **J**	EC 1.6.4.2 transferred to EC 1.8.1.7	Entry transfer
(deleted) **I** reinstated as **J**	EC 1.1.1.249 reinstated as EC 2.5.1.46	Entry reinstatement
I is incorporated in **J**	EC 4.2.1.21 incorporated in EC 4.2.1.22	Entries are merged
I is part incorporated in **J1** **I** is part incorporated in **J2** ...	EC 1.1.1.182 created 1983, part incorporated in EC 1.1.1.198, EC 1.1.1.227 and EC 1.1.1.228	Entry is split

Table 6. "Dodgy" relationships between EC entries.

I to J	Example	Comment
I is not identical with **J**	EC 2.8.2.5 is not identical with EC 2.8.2.17	In theory, all non-deleted EC entries should be not identical
I may be identical with **J**	EC 1.2.3.1 may be identical with EC 1.2.3.11	If proven, one of the entries should be deleted as "identical with"
I is possibly identical with **J**	EC 3.9.1.1 is possibly identical with EC 3.1.3.9 or EC 3.1.3.16	If proven, one of the entries should be deleted as "identical with"
I is related to **J**	EC 3.1.8.2. related to EC 3.1.8.1	unspecified relationship

'Structural and functional relationships' are mostly found in the comments. There is no official or otherwise formalized way of relating the EC entries functionally. 'Historical relationships' are those derived from the history line found at the very bottom of the enzyme entry. Therefore they are rather official. 'Dodgy relationships' are found in comments in the enzyme entries. In many cases, no meaningful information can be derived.

Incomplete EC numbers

One unfortunate consequence of the dual nature of EC numbers is the continuing practice of providing *incomplete EC numbers* in literature and databases. For example, Swiss-Prot contains many entries described as enzymes that have not been assigned EC numbers by NC-IUBMB, but which are assigned a provisional class (subclass, sub-subclass), the final digit(s) being '-'. Such assignments are thought to be useful for providing some functional information about the enzymes. The problem is that incomplete EC numbers are non-identifiers. As Example 4 shows, one EC 1.-.-.- is not equal to another EC 1.-.-.- at all!

BIOREACTION ONTOLOGY

In the words of Robert Rankin, it is a tradition, or old charter, or something, that biologists divide all reactions occurring *in vivo* into two groups: 'enzymatic' and 'non-enzymatic'. Of course, this is an extreme simplification which nevertheless summarizes the view of the world according to the modern molecular biologist: whatever is not in the genome is irrelevant.

Although enzymology, not surprisingly, deals with enzymatic reactions, it would be useful to put such reactions in the context of other biological reactions.

Thus the reactions could be classified according to type of catalyst. Chemists deal with homogeneous and heterogeneous catalysis; examples of both are also abundant in living cell. Catalytic macromolecules include catalytic proteins (enzymes and abzymes) and catalytic nucleic acids (deoxyribozymes and ribozymes). Macromolecular catalysis has features of both homogeneous and heterogeneous catalysis, for, even though reactants and catalytic macromolecules usually exist in the same phase, the reactant adsorption is crucial for the reaction mechanism.

The orthogonal classification of bioreactions is by overall reaction class, as presented in Fig. 1. These reaction classes correspond to the type of bond involved (Table 7). Most of the reactions with which the Enzyme Nomenclature deals are biotransformations.

```
%biochemical reaction
  %by overall reaction
    %binding reaction
    %biotransformation rection
    %conformation change reaction
    %molecular transport reaction
    %electron transfer reaction
    %excitation-energy transfer reaction
  %by nature of catalyst
    %catalytic macromolecule reaction
      %enzymatic reaction
        %abzymatic reaction
      %deoxyribozymatic reaction
      %ribozymatic reaction
    %heterogenous catalytic reaction
    %homogenous catalytic reaction
  %non-catalytic reaction
    %photoinduced reaction
    %thermal reaction ('spontaneous' reaction)
  %intramolecular catalysis reaction
```

Figure 1. Ontology of biochemical reactions in GO file format.

Table 7. Fundamental bioreaction classes.

Bioreaction class	Reaction	Bond involved
Biotransformation	A + B → C + D	Covalent
Binding	A + M → AM M = macromolecule	Ionic, hydrogen, van der Waals
Conformation change	A → B A, B = conformers	Hydrogen, van der Waals
Molecular transport	$A_{compartment} \rightarrow A_{compartment\ Y}$	_
Electron transfer	$A^{n} + B^{n-m} \rightarrow A^{n-m} + B^{n}$ *m* = number of transferred electrons	_
Exciton transfer reactions	A* + B → A + B*	_

Note that these reactions, like those of the Enzyme Nomenclature, are overall transformations. All enzymatic reactions include binding and conformational change as intermediate steps. Many enzymatic reactions are coupled, e.g. active transport is coupled with ATP hydrolysis. Protein folding is mostly conformational change but also may involve covalent and ionic bonds.

Electron transfer and excitation-energy transfer reactions

The term "pure electron transferase" was originally introduced as a name for a class of flavoproteins (exemplified by flavodoxins) where the flavin is reduced and reoxidized in one-electron steps [33, 34]. The meaning can be naturally extended to cover all the proteins that catalyse electron transfer reactions only, such as cytochromes, ferredoxins and cupredoxins. Although proteins involved in electron transfer are usually classified as oxidoreductases, none of the 'pure electron transferases' is assigned an EC number. Similarly, excitation-energy transfer processes, as in the antenna systems of photosynthetic organisms, are not covered by the Enzyme Nomenclature.

Analogous to 'traditionally understood' metabolic pathways that consist of separate enzymatic reactions, electron/exciton transfer reactions form electron/exciton transfer pathways, that form an integral part of the metabolic pathways.

Transmembrane Transport

A great number of fundamental biochemical reactions can be represented as:

$$X_{\text{compartment A}} \rightarrow X_{\text{compartment B}}$$

Since solutes cannot cross biological membranes, the transmembrane transport has to be facilitated by specific carriers or transporters [35, 36]. Importantly, a distinct class (Energases) has been proposed to cover the enzymes that catalyse conversion of chemical energy into mechanical energy [37]. Energases include primary active transporters (directly utilizing covalent bond energy to transport solutes against a concentration gradient) and rotational molecular motors such as ATP synthase. However, there is no reason to deny the other transporters their place in the enzyme classification. Some electron transferases are also transmembrane transporters (class TC 5 in [35]).

Non-catalytic reactions

Examples of biologically relevant non-catalytic reactions include chain reactions of lipid peroxidation [38] and photoinduced transformation of ergosterol to previtamin D3 and its subsequent thermal isomerisation to vitamin D3 [39]. The non-catalytic thermal reactions often are referred to as "spontaneous" in the biological literature, including the Enzyme Nomenclature (in fact, every process with a negative change of Gibbs free energy at constant pressure is spontaneous, whether it is catalysed or not). Chain reactions of lipid oxidation can be initiated photochemically (UV radiation) or catalytically, e.g. via a Fenton mechanism, and can be terminated non-catalytically or catalytically [38].

Yet other reactions

Some biochemical reactions do not easily fit into 'catalytic' or 'non-catalytic' categories. For example, the comment for methylated-DNA-[protein]-cysteine *S*-methyltransferase (EC 2.1.1.63) reads: "Since the acceptor protein is the 'enzyme' itself and the *S*-methyl-L-cysteine derivative formed is relatively stable, the reaction is not catalytic." The reaction proceeds through suicidal alkyl transfer from the guanine O6 to the cysteine residue of the enzyme; therefore the enzyme should be present in stoichiometric, not catalytic, amounts. The reaction catalysed by EC 2.1.1.63 fits the IUPAC "Gold Book" definition of intramolecular catalysis [40].

On the one hand, the intramolecular catalyst is a kind of catalyst, since "the catalyst is both a reactant and product of the reaction" [41]. But if the direct result of the reaction is an inactivated catalyst, it makes the whole process non-catalytic (according to the Gold Book).

The term "autocatalytic reaction" is often used in a meaning not consistent with the Gold Book definition [42], e.g., *"autocatalytic quinone-methide mechanism of protein flavinylation"* [43] or *"autocatalytic formation of a thioether cross-link between the active-site residues"* in galactose oxidase [44]. These are in fact intramolecular catalysis events.

Mechanism of biochemical reaction

At least two orthogonal types of mechanism can be considered: one is inherent to the reaction or some reaction steps, another is specified by the catalyst. The 'inherent' mechanisms, such as polar reactions, free-radical reactions and pericyclic reactions [31], are shared with organic chemistry. The catalyst-specific mechanisms often feature particular metals or prosthetic groups (e.g. "copper-dependent" or "pterin-dependent" enzymes) or kinetic properties (e.g. "Ping Pong mechanism") [45]. Note that coenzymes, in contrast to prosthetic groups, do enter the overall transformations and therefore are covered by reaction classifications.

ONTOLOGY FOR ENZYME-CATALYSED REACTION: SOME WORKING PRINCIPLES

Organization: Similar to GO, terms should be organized as a directed acyclic graph (DAG). A child term can have many parent terms.

A *unique identifier* in form of accession number (AC) should be assigned to every node of DAG so we can be sure that we find data with AC even if the term is changed. When entries are merged, ACs do not disappear. Unlike EC numbers, ACs are devoid of any other meaning.

Traceability: Evidence should be assigned to every node (and possibly to every edge) of a DAG. Such evidence could be a literature reference, the database entry, or a curator judgement. The history line of an IntEnz entry in Example 9 provides some but not all of the information about an enzyme's history (it is important to know not only when, but also why and how the entries were merged, split or modified in any other way). Modern tools allow all events of this sort to be tracked (provided, of course, that editing is done within a database and not at an external source).

Scalability: "An architecture is considered to be scalable if, unchanged, it can handle increasingly complex problems that demand a greater amount of knowledge" [46]. The enzyme ontology should be designed in such a way that it can cope with the growing amount of data. Since our knowledge is always incomplete, the system should be adaptable for higher/lower levels of granularity.

It seems obvious that if the reactions are classified by overall transformation, considerations such as the nature of the catalyst should not be used; yet we saw exactly this in the Enzyme Nomenclature. So, the next principle is that the *orthogonal classifications* should not mix. The top level of the ontology of enzymatic reactions in a 'GO file format' is shown in Fig. 2.

```
%enzymatic reaction
   %by overall reaction (aka Enzyme Nomenclature)
   %by partial reaction
   %by nature of enzyme
   %by reaction mechanism
   %by enzyme regulation
      %activation
      %inhibition
```

Figure 2. Ontology of enzymatic reactions in GO file format

An important step towards development of a bioreaction ontology is made by introduction of the REACTION database, which is developed as an integral part of the LIGAND database [47]. The reactions are completely independent from the ENZYME entries. Therefore, it is possible to link one ENZYME entry to more than one REACTION entry as well as to include non-enzymatic reactions and, perhaps even more importantly, the reactions catalysed by yet unclassified enzymes. Currently, only 'parseable' reactions (i.e. those expressed by an equation) are included. REACTION does not take account of reaction directionality, the reactions being written as if they were reversible.

If the directed partial reactions are assigned their own ACs, the coupled reactions could be constructed as linear combinations of partial reactions; these new reactions are assigned their own ACs, etc.

Finally, in creating the ontology of enzymatic reactions, we *should avoid preconceived opinions regarding experimental conditions, efficiency or physiological role*. In many cases, the physiological role cannot be elucidated before we know enough about the physiology of the organism in question; however this should not prevent us from describing the enzyme function.

NEW RELATIONSHIPS

Let us have a look at Fig. 3. The universes of biochemical objects and of biochemical processes do not overlap. Most biochemical objects relate to each other via familiar *IsA* and *IsPartOf* relationships. The same is true for most biochemical processes. However, for some objects and processes novel relationships (e.g. *IsTransformedTo*, *Accelerates*, *SlowsDown*) have to be introduced. Since the reaction (process) uniquely specifies the relationship between 'reactant' and 'product', the logical way to link the two universes is via the overall reaction. Indeed, 'reactant', 'product' and 'solvent' are functional roles of chemical compounds in a particular reaction (in our example, A → B + C takes place in solvent D) and may play different roles elsewhere. All the relationships are directed, *including reactions*. As our knowledge of enzyme structure and function grows, further links relating active site (object) with catalytic mechanism (process), compound (object) with inhibition or activation (process), etc., may be required.

CONCLUSION

Although the Enzyme Nomenclature is the only accepted systematic nomenclature for biochemical reactions, it often violates its own fundamental principles. Moreover, it suffers from inherent confusion between enzyme (object) and enzymatic reaction (process) concepts that propagate into the scientific literature and biological databases. For instance, EC numbers are often used as synonyms of protein names. This strict hierarchy and depth limit may have been justified for a book layout but is otherwise too rigid. The limit of four levels does not allow additional hierarchical *IsA* relationships which otherwise can be introduced on the basis of a natural hierarchy of chemical compound classes. The extension of Enzyme Nomenclature beyond the traditional six classes of overall transformations will include, e.g., reactions affecting non-covalent bonds and transport phenomena.

Further modification of the Enzyme Nomenclature is required to enable multiple ancestry for enzymatic reactions. Ultimately, the Enzyme Nomenclature, cleansed of inconsistencies, should become a part of the new Enzyme Ontology.

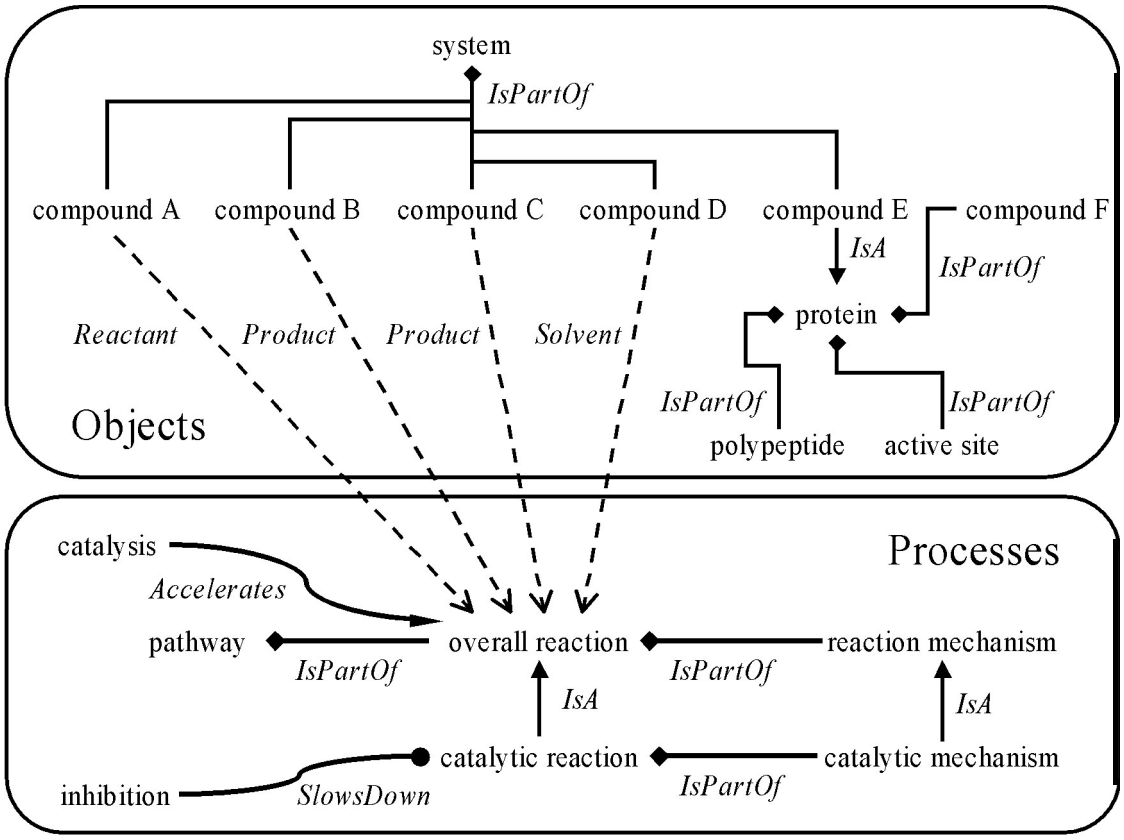

Figure 3. Some relationships between biochemical objects and biochemical processes.

ACKNOWLEDGMENTS

I wish to thank Rolf Apweiler, who suggested that I should submit an abstract for this workshop, and Carsten Kettner, who read through it and asked me to give this presentation. I thank my colleagues at the EBI, Evelyn Camon, Marcus Ennis and Jane Lomax, for their helpful comments and suggestions on the manuscript. This work is supported by the European Commission grant QLRT-2000-00981.

List of Abbreviations

AC	accession number
CPR	NADPH-cytochrome P450 reductase
DAG	directed acyclic graph
EC	Enzyme Commission
GO	Gene Ontology
ID	identifier
NC-IUB	Nomenclature Committee of the International Union of Biochemistry
NC-IUBMB	Nomenclature Committee of the International Union of Biochemistry and Molecular Biology
NOS	nitric oxide synthase
PDB	Protein Data Bank

References

[1] Carugo, O., Pongor, S. (2002) The evolution of structural databases. *Trends Biotechnol.* **20**: 498-501.

[2] The Gene Ontology Consortium, http://www.geneontology.org/

[3] The EMBL Nucleotide Sequence Database, http://www.ebi.ac.uk/embl/

[4] The Swiss-Prot Protein Knowledgebase, http://www.ebi.ac.uk/swissprot/

[5] The Protein Data Bank, http://www.pdb.org/

[6] IUBMB (1992) *Enzyme Nomenclature: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes.* Academic Press, San Diego.

[7] Tipton, K., Boyce, S. (2000) History of the enzyme nomenclature system. *Bioinformatics* **16**: 34-40.

[8] Lilley, D.M.J. (2003) The origins of RNA catalysis in ribozymes. *Trends Biochem. Sci.* **28**: 495-501.

[9] Emilsson, G.M., Breaker, R.R. (2002) Deoxyribozymes: new activities and new applications. *Cell. Mol. Life Sci.* **59**: 596-607.

[10] Nevinsky, G.A., Buneva, V.N. (2003) Catalytic antibodies in healthy humans and patients with autoimmune and viral diseases. *J. Cell. Mol. Med.* **7**: 265-276.

[11] IntEnz: Integrated relational Enzyme database, http://www.ebi.ac.uk/IntEnz/

[12] The ENZYME database, http://www.expasy.org/enzyme/

[13] BRENDA: The Comprehensive Enzyme Information System, http://www.brenda.uni-koeln.de/

[14] GO File Format Guide, http://www.geneontology.org/doc/GO.format.html

[15] Open Biology Ontologies, http://obo.sourceforge.net/

[16] MEROPS, the Peptidase Database http://www.merops.ac.uk/

[17] REBASE, The Restriction Enzyme Database, http://rebase.neb.com/

[18] CAZy, Carbohydrate-Active enZYmes, http://afmb.cnrs-mrs.fr/CAZY/

[19] UM-BBD, The University of Minnesota Biocatalysis/Biodegradation Database, http://umbbd.ahc.umn.edu/

[20] Tian, W., Skolnick, J. (2003) How well is enzyme function conserved as a function of pairwise sequence identity? *J. Mol. Biol.* **333**: 863-882.

[21] Bindewald, E., Cestaro, A., Hesser, J., Heiler, M., Tosatto, S.C.E. (2003) MANIFOLD: protein fold recognition based on secondary structure, sequence similarity and enzyme classification. *Protein Engng* **16**: 785-789.

[22] NC-IUBMB (1992) Classification and Nomenclature of Enzyme-Catalysed Reactions. http://www.chem.qmul.ac.uk/iubmb/enzyme/rules.html

[23] Degtyarenko, K.N., Kulikova, T.A. (2001) Evolution of bioinorganic motifs in P450-containing systems. *Biochem. Soc. Trans.* **29**: 139-147.

[24] Munro, A.W., Leys, D.G., McLean, K.J., Marshall, K.R., Ost, T.W.B., Daff, S., Miles, C.S., Chapman, S.K., Lysek, D.A., Moser, C.C., Page, C.C., Dutton, P.L. (2002) P450 BM3: the very model of a modern flavocytochrome. *Trends Biochem. Sci.* **27**: 250-257.

[25] NC-IUB (1989) Nomenclature for multienzymes. Recommendations 1989. http://www.chem.qmul.ac.uk/iubmb/misc/menz.html

[26] Beadle, G.W., Tatum, E.L. (1941) Genetic control of biochemical reactions in Neurospora. *Proc. Natl. Acad. Sci. USA* **27**: 499-506.

[27] Fani, R., Liò, P., Lazcano, A. (1995) Molecular evolution of the histidine biosynthetic pathway. *J. Mol. Evol.* **41**: 760-774.

[28] Hawkins, A.R., Lamb, H.K. (1995) The molecular biology of multidomain proteins. Selected examples. *Eur. J. Biochem.* **232**: 7-18.

[29] Yasutake, Y., Watanabe, S., Yao, M., Takada, Y., Fukunaga, N., Tanaka, I. (2002) Structure of the monomeric isocitrate dehydrogenase: evidence of a protein monomerization by a domain duplication. *Structure* **10**: 1637-1648.

[30] Kamijo, T., Aoyama, T., Komiyama, A., Hashimoto, T. (1994) Structural analysis of cDNAs for subunits of human mitochondrial fatty acid b-oxidation trifunctional protein. *Biochem. Biophys. Res. Commun.* **199**: 818-825.

[31] Grossman, R.B. (1999) *The Art of Writing Reasonable Organic Reaction Mechanisms.* Springer-Verlag, New York.

[32] Adams, D. (1980) *The Restaurant at the End of the Universe*. p.35. Pan Books Ltd, London.

[33] Hemmerich, P., Massey, V., Fenner, H. (1972) Flavin and 5-deazaflavin: A chemical evaluation of 'modified' flavoproteins with respect to the mechanisms of redox biocatalysis. *FEBS Lett.* **84**: 5-21.

[34] NC-IUB (1991) Nomenclature of electron-transfer proteins. Recommendations 1989. http://www.chem.qmul.ac.uk/iubmb/etp/

[35] Transport Protein Database, http://tcdb.ucsd.edu/tcdb/

[36] NC-IUBMB (2002) Membrane transport proteins. Recommendations 2002. http://www.chem.qmul.ac.uk/iubmb/mtp/

[37] Purich, D.L. (2001) Enzyme catalysis: a new definition accounting for noncovalent substrate- and product-like states. *Trends Biochem. Sci.* **26**: 417-421.

[38] Girotti, A.W. (1998) Lipid hydroperoxide generation, turnover, and effector action in biological systems. *J. Lipid Res.* **39**: 1529-1542.

[39] Holick, M.F. (1995) Defects in the synthesis and metabolism of vitamin D. Exp. Clin. Endocrinol. *Diabetes* **103**: 219-227.

[40] McNaught, A.D., Wilkinson, A., Eds. (1997) *Compendium of Chemical Terminology: IUPAC Recommendations* ("The Gold Book"), 2nd Edition, p.206. Blackwell Scientific Publications, Oxford.

[41] Ibid., p. 58.

[42] Ibid., p. 34.

[43] Edmondson, D.E., Newton-Vinson, P. (2001) The covalent FAD of monoamine oxidase: structural and functional role and mechanism of the flavinylation reaction. *Antioxid. Redox Signal.* **3**: 789-806.

[44] Firbank, S.J., Rogers, M., Hurtado-Guerrero, R., Dooley, D.M., Halcrow, M.A., Phillips, S.E.V., Knowles, P.F., McPherson, M.J. (2003) Cofactor processing in galactose oxidase. *Biochem. Soc. Trans.* **31**: 506-509.

[45] Silverman, R.B. (2002) *The Organic Chemistry of Enzyme-Catalyzed Reactions.* Academic Press, San Diego.

[46] Lemon, B., Pynadath, D., Taylor, G., Wray, R.E. Cognitive architectures: a hypertext analysis of architectures for intelligence. http://ai.eecs.umich.edu/cogarch4/

[47] LIGAND database of chemical compounds and reactions in biological pathways, http://www.genome.ad.jp/ligand/