

FROM THE ENZYME LIST TO PATHWAYS AND BACK AGAIN

ANDREW G. McDONALD, KEITH TIPTON*
AND SINÉAD BOYCE

Department of Biochemistry, Trinity College, Dublin 2, Ireland

E-Mail: *ktipton@tcd.ie

Received: 25th September 2007 / Published: 20th August 2008

ABSTRACT

The IUBMB Enzyme List is widely used by other databases as a source for avoiding ambiguity in the recognition of enzymes as catalytic entities. However, it was never designed for activities such as pathway tracing, which have become increasingly important in systems biology. This is because it often relies on generic or representative reactions to show the reactions catalysed by enzymes of wide specificity. It is necessary to go to databases such as *BRENDA* to find further, more detailed, information on what is known about the range of substrates for any particular enzyme. In order to provide a framework for tracing pathways involving any specific enzyme or metabolite, we have created a *Reactions Database* from the material in the Enzyme list. This allows reactions to be searched by substrate/product and pathways to be traced from any selected starting/seed substrate. An extensive synonym glossary allows searches by any of the alternative names, including accepted abbreviations, by which a chemical compound may be known. This database was necessary for the development of the application *Reaction Explorer* (<http://www.reaction-explorer.org/>), which was written in REALbasic to search the *Reactions Database* and draw metabolic pathways from reactions selected by the user. Having input the name of the starting compound (the “seed”), the user is presented with a list of all reactions containing that compound and then selects the product of interest as the next point on the ensuing graph. The pathway diagram is then generated as the process iterates. A contextual menu is provided,

which allows the user to (i) remove a compound from the graph, along with all associated links; (ii) search the reactions database again for additional reactions involving the compound and (iii) search for the compound within the Enzyme List.

INTRODUCTION

The International Union of Biochemistry and Molecular Biology (IUBMB) Enzyme List classifies enzymes in terms of the reactions they catalyse (see [1, 2] for definitive versions). It is restricted to classification and recommendations on nomenclature. As such, the data contained within it are, as far as possible, strictly factual and should provide a system for the unambiguous identification of the enzyme(s) being studied. Because of this strictly defined function, its application to tracing metabolic systems is, perforce, limited. Although much of the data required for this application are there, the structure makes it difficult to access. Since it would be undesirable to alter the Enzyme List to meet other functions if that were to diminish its core utility, this account will discuss what can be achieved using the list itself and derivatives of it.

WHAT THE ENZYME LIST CAN DO, THE ENZYME CENTRIC APPROACH

A ExplorEnz - The Enzyme Database

Home Search Enzymes by Class New/Amended Enzymes Statistics Forms Change Log Information

Simple search
For advanced search option, click here.

Search for in
 all fields Use regular expressions [what are these?]
or
 select fields:
 EC Number Comments
 Accepted name References
 Reaction PubMed ID
 Other name(s) Glossary
 Systematic name

and display [highlight matches]
 all fields
or
 select fields:
 Accepted name Comments
 Reaction Links to other databases
 Other name(s) References
 Systematic name History
 Glossary

Sort results by displaying entries per page.

(B)

EC: 1.4.1.7: serine dehydrogenase

Reaction: L-serine + H₂O + NAD⁺ = 3-hydroxypyruvate + NH₃ + NADH + H⁺

EC: 2.3.1.30: serine O-acetyltransferase

Reaction: acetyl-CoA + L-serine = CoA + O-acetyl-L-serine

EC: 2.3.1.50: serine C-palmitoyltransferase

Reaction: palmitoyl-CoA + L-serine = CoA + 3-dehydro-D-sphinganine + CO₂

EC: 2.6.1.45: serine-glyoxylate transaminase

Reaction: L-serine + glyoxylate = 3-hydroxypyruvate + glycine

EC: 2.6.1.51: serine-pyruvate transaminase

Reaction: L-serine + pyruvate = 3-hydroxypyruvate + L-alanine

EC: 2.7.1.80: diphosphate-serine phosphotransferase

Reaction: diphosphate + L-serine = phosphate + O-phospho-L-serine

EC: 2.7.8.4: serine-phosphoethanolamine synthase

etc.

Figure 1. A specific search of *ExplorEnz* (2) for enzymes having L-serine as a substrate or product (A) with a part of the output (B).

The Enzyme List can be searched to find all of those enzymes that catalyse reactions involving a given substrate or product. This is illustrated in figure 1 using serine as an example. From the output (Fig. 1b), it is possible to construct the simple pictorial representation shown in figure 2. Such “enzyme-centric” searching can be useful in predicting the possible effects of drugs that are targeted against a specific enzyme, since they will show other enzymes that might also be affected.

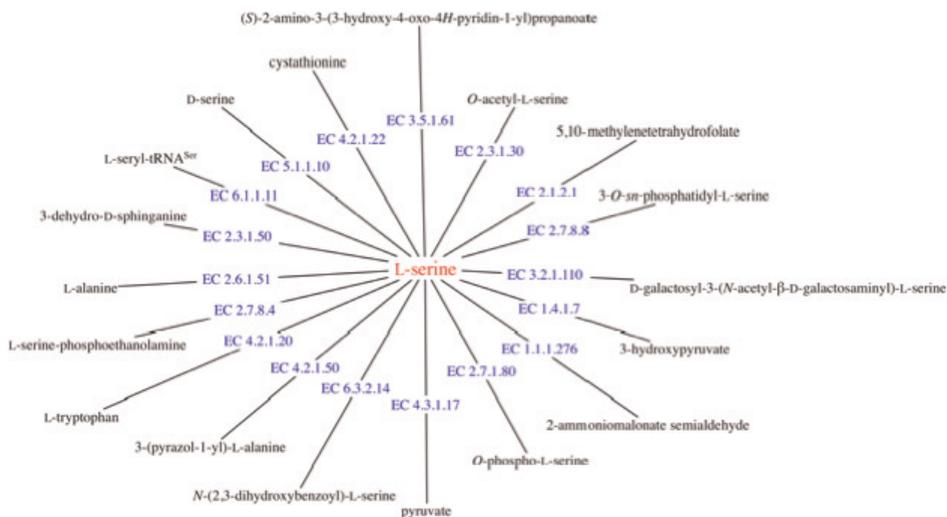


Figure 2. The substrate-centric approach: a redrawn view of the enzyme having L-serine as a substrate or product.

It is also possible to list groups of enzymes linked by common substrates and products, as shown in figure 3. The results of attempts to display these in graphical form can, however, appear quite complex because of the multiplicity of edges that occurs when forming connections between enzymes.

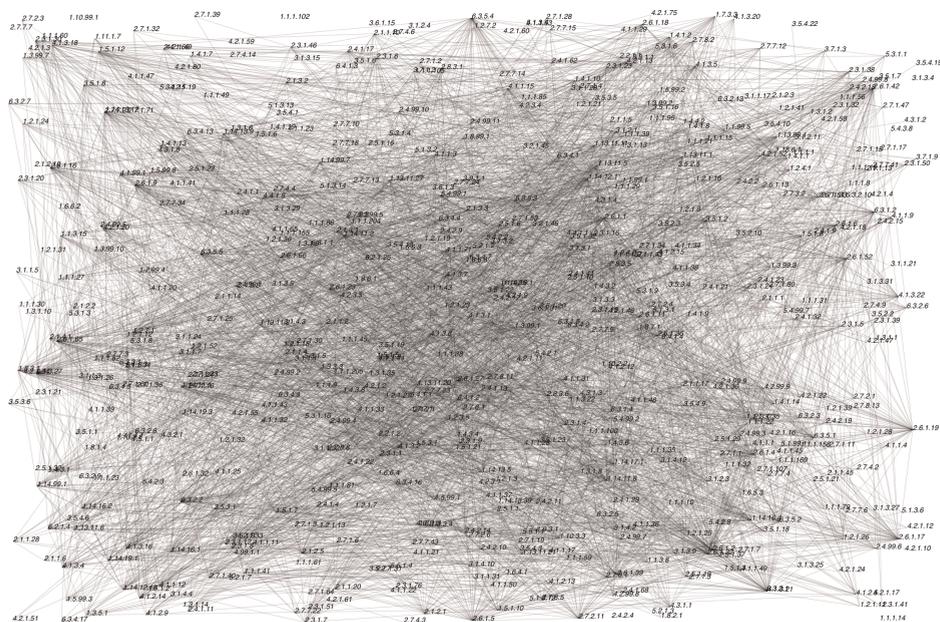


Figure 3. A network of 604 enzymes with links representing shared metabolites. There are 4,062 possible connections made amongst the enzymes in this subset, from which H^+ , H_2O , and common cofactor pairs (e.g., ATP/ADP) were excluded.

From combinatorics, the general formula for the number of ways r items can be taken from n is:

$${}^n C_r = n! / [r!(n-r)!]$$

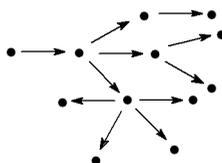
${}^n C_2$ therefore represents the total number of possible enzyme (node) pairs, where each pair shares the same metabolite ($r=2$). Thus 4 enzymes sharing a common metabolite will have to be connected by 6 edges and 380 enzyme nodes, as would be required to show the number of reactions in the database that involve O_2 , will require 72,010 edges.. Clearly, the situation can become more complicated than this if one considers the possibility of having several shared metabolites for each enzyme. Whilst such a representation can readily be searched for any given enzyme and has the advantage that each enzyme only occurs once in the diagram, as opposed to the hand-crafted, artistic, versions, such as the Nicholson metabolic pathways charts (see [3]), where the separation of different metabolic systems in the display can result in the same enzyme occurring in several different places.

REACTIONS DATABASE: THE SUBSTRATE-CENTRIC APPROACH

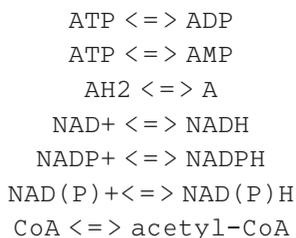
The *Reactions Database* was initially prepared by extracting all of the reactions in *Explor-Enz*, the MySQL version of the Enzyme List [2]. These were used to create a separate database (also MySQL). A web application created with PHP was developed to provide the query interface to the database. As with *ExplorEnz*, the query engine supports both case-insensitive and regular-expression substring searches. This, at least in theory, should allow all reactions involving any given substrate or product to be displayed. It should also allow one to trace the number of reactions n-steps from any given starting substrate, so that its metabolic fates can be better appreciated. However, just as in the case of edge multiplicity in the “enzyme-centric” approach, the system rapidly gains complexity. This is illustrated for some simple metabolites in the table below and by the illustrative tree structure.

Table: Some query results for successive reactions of different compounds.

Compound	Paths found		
	1 Step	2 Steps	3 Steps
L-Ascorbate	5	89	
L-Cysteine	5	25	577
L-Tyrosine	13	676	
Ribitol	2	50	2776



One major cause of this rapidly expanding complexity is the involvement of a reactant that is used or produced by several different enzymes. For example, if a reaction produces or uses ATP it will be linked to many other reactions (the kinases etc). This can be addressed by specifically excluding some compounds, such as H₂O, H⁺, ATP, ADP, AMP, phosphate, diphosphate, NADP⁺, NADPH, NAD⁺, NADH, NAD(P)⁺, NAD(P)H, A, AH₂, acceptor and reduced acceptor from the search. However, if one were to exclude, for example, NAD⁺, that would eliminate ADP-ribosylation reactions as well as oxidoreductases and excluding ATP would eliminate several adenylyltransferase reactions. This problem can be better addressed by selective elimination of reactant pairs rather than single reactants from the search. These might include:



REPRESENTATIONS: REACTION EXPLORER

Although the *Reactions database* can provide lists of reactions, an additional tool is needed for display purposes. This is provided by *Reaction Explorer* [4], which is a multi-platform application, written in REALbasic, for constructing metabolic network graphs. Versions are currently available for the following operating systems: Mac OS X 10.1 or higher, Mac OS 9.x, Linux x86 and Windows 95 or higher. Selecting any product from a reaction will automatically draw a line connecting it to its parent substrate from where one can proceed to the next step in the pathway and so on, to provide a pictorial representation of the process.

The output is designed to be basic because its purpose is to convey information and not to construct works of art. Thus, it is not designed as a competitor for representational systems, such as GraphViz [5], or the craftsman-designed Nicholson metabolic pathways charts, but rather to display the essential information quickly and easily.

WHAT REACTION DB & EXPLORER CANNOT DO

As indicated above, *Reaction Explorer* is an aid to drawing pathways so that interactions may be visualized. In fact one can generate searchable connection graphs with any dataset that is entered in the *Reaction Explorer* file format, such as those shown in figures 2 and 3. As will be discussed later, it was also found to be of value as an aid to trouble-shooting the Enzyme List.

However, there are limitations imposed not by the programme but by the nature of the system involved. As discussed in connection with the table previously shown, a simple tree that describes all reactions proceeding for n-steps from any named reactant is not to be expected. Similarly, a unique pathway connecting two distant metabolites does not usually occur in metabolite space. Thus the question “find the pathway from glucose to lactate” might be expected to yield glycolysis. Indeed it does, together with very many other pathways. That is because there are very many ways in which glucose can be converted into lactate, including the synthesis and breakdown of compounds such as cholesterol. This might be addressed by specifying the number of steps allowed, but not all major pathways necessarily use the minimum number of steps.

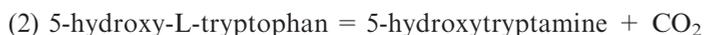
PROBLEMS WITH USING REACTIONS OF THE ENZYME LIST

It is perhaps unreasonable to expect the Enzyme List to have functions that it was not designed for. However, there are several aspects that make it unsuitable for simple adaptation to reaction pathway tracing through systems such as *Reaction Database & Explorer*.

(a) Not all reactions catalysed by a given enzyme are listed

In the past, the Enzyme List has often used a representative reaction for enzymes with broad specificities. It is intended to add a field to include additional reactions where appropriate. Reactions involving non-physiological substrates are not listed except in the case of donors and acceptors where the physiological factor has not yet been identified. However, this can result in judgements about what is, and what is not, physiologically important.

For example aromatic-L-amino-acid decarboxylase (EC 4.1.1.28) is given as catalysing two reactions



but it will also catalyse the decarboxylation of L-tyrosine, L-tryptophan and L-phenylalanine. Although these reactions may be of lesser physiological significance, they are not unimportant and can, indeed, have major significance in the responses to therapy involving some antidepressant drugs.

In some cases additional information on the specificity is also given in the “comments” associated with the Enzyme-List entry. For example, the 6-phosphofructokinase (EC 2.7.1.11) reaction is given as

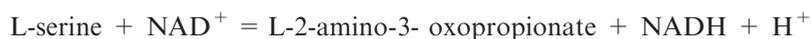


but the comments also state: “D-Tagatose 6-phosphate and sedoheptulose 7-phosphate can act as acceptors. UTP, CTP and ITP can act as donors”. Clearly such material, although readily accessible in an *ExplorEnz* search, needs to be incorporated into the *Reactions Database*. The comprehensive lists of substrates provided by *BRENDA* [6], which also contains, somewhat arbitrary, listings of “natural substrates”, can be most valuable for this purpose.

In the case of L-threonine 3-dehydrogenase (EC 1.1.1.103), only one reaction is listed



but a check of the *BRENDA* entry reveals that it also catalyses the reaction:

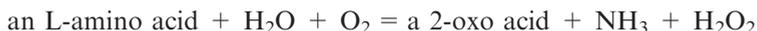


(b) General/Markush terms

In the case of enzymes with broad substrate specificities, such as alcohol dehydrogenase (EC 1.1.1.1), where the number of substrates, or potential substrates, is very large, the Enzyme List gives a single generic reaction, and *BRENDA* is an essential source of detail. For example, the alcohol dehydrogenase reaction is given as

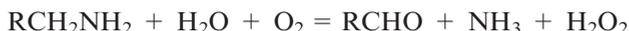


and that of L-amino-acid oxidase (EC 1.4.3.2) is given as:

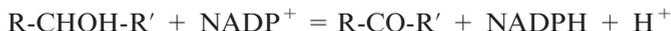


The problem with this approach is that it does not indicate which alcohols, or L-amino acids, are not substrates. Again, *BRENDA* is invaluable in such cases, although it should always be borne in mind that the absence of a compound from the substrate/product list does not necessarily mean that it is not a substrate, but may simply mean that nobody has tried it.

Markush terms are also used for some reactions. For example the reaction catalysed by amine oxidase (copper-containing) (EC 1.4.3.6) is given as:



and that of carbonyl reductase (NADPH) (EC 1.1.1.184) as:



Although such formulations are somewhat more informative than the general reactions above, and the Markush terms are searchable in *ExplorEnz* and *Reaction Explorer*, it is still necessary to revert to *BRENDA* for information on the exact substrates that are known to be used.

Another complexity occurs where it is not possible to describe the reaction catalysed by a simple reaction equation, without ambiguity. Examples of this are the reactions catalysed by (α -amylase) (EC 3.2.1.1), which is given as

Endohydrolysis of 1,4- α -D-glucosidic linkages in polysaccharides containing three or more 1,4- α -linked D-glucose units.

1,4- α -glucan branching enzyme (EC 2.4.1.18), where the reaction is described as:

Transfers a segment of a 1,4- α -D-glucan chain to a primary hydroxy group in a similar glucan chain and exodeoxyribonuclease I (EC 3.1.111)

Exonucleolytic cleavage in the 3'- to 5'-direction to yield nucleoside 5'-phosphates.

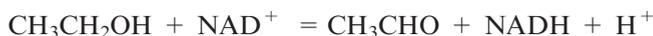
(c) Internal synonym inconsistencies

The Enzyme List has been in operation since 1952. Since then there have been many changes in nomenclature. Normally these have been made to correct or rationalize nomenclature, for example few now remember the furore caused by the change from DPN (even before that, it was called coenzyme I) to NAD, whereas the change from fructose 1,6-diphosphate to fructose 1,6-bisphosphate went relatively smoothly. Generally, the Enzyme List is punctilious about correcting entries but as will be discussed below, a few escape the notice of the IUBMB-IUPAC Joint Commission on Biochemical Nomenclature and those who use the Enzyme List. As discussed below, pathway tracing can be of value in finding such inconsistencies.

(d) No indication of *in vivo* directionality

Because the Enzyme List is restricted to providing factual data on the reaction catalysed, the reaction presented is, as far as possible, a mass-balancing equation. It is not meant to indicate the preferred equilibrium of the reaction or the direction in which the reaction is believed to operate *in vivo*. By convention, the direction chosen for the reaction in any given sub-subclass is the same for all enzymes. Systematic names are based on this written reaction and, therefore, carry no information about the favoured reaction direction. Although, this might seem to be less helpful than it might, it must be remembered that, for some enzymes, such as glutamate dehydrogenase [NAD(P)⁺] (EC 1.4.1.2) and fructose-bisphosphate aldolase (EC 4.1.2.13), the preferred reaction direction varies with cellular conditions. Furthermore, the equilibrium constant of the reaction may be misleading in terms of the direction in which it actually operates *in vivo*.

For example, the equilibrium oxidation of ethanol dehydrogenase



very much favours ethanol formation under physiological conditions, but ethanol oxidation is the dominant direction *in vivo* because acetaldehyde (ethanal) is rapidly removed in the essentially irreversible reaction catalysed by aldehyde dehydrogenase (NAD⁺) (EC 1.2.1.3)



Thermodynamic data for many enzymes can be found in the *GTD* Thermodynamics of Enzyme-catalysed Reactions database [7] and kinetic data are included in the *BRENDA* database, which may provide detail to determine reaction equilibria through Haldane relationships. It should be emphasised that only data that refer to “physiologically relevant conditions” should be used and that it is the thermodynamic properties of the overall metabolic system, not of the individual reaction, that are important in determining the flux direction [8, 9].

(e) No species information

In general the Enzyme List does not give information on the species, tissue or cell compartment in which the enzyme is found. Some information may be found in the references associated with each entry and the “comments” may refer to species in terms of behaviour that may not apply to the enzyme from all sources. For example, the entry for alcohol dehydrogenase contains the comment “Acts on primary or secondary alcohols or hemiacetals; the animal, but not the yeast, enzyme acts also on cyclic secondary alcohols”. The *BRENDA* database, however, contains extensive species data that can be used in this context, and gene and protein databases may also provide valuable information about the species in which an enzyme might be expressed.

(f) Spontaneous (uncatalysed) reactions are not listed

Although one would not expect the Enzyme List to include reactions that are not enzyme-catalysed, such reactions do occur *in vivo* and will break a metabolic chain if not added to systems, such as the *Reactions Database* and *Explorer*. For example, the Enzyme List entry for L-threonine 3-dehydrogenase (see above) includes the comment “The product spontaneously decarboxylates to aminoacetone”. This may be essential information for tracing the metabolic fates of L-threonine, since aminoacetone is known to be a substrate for the copper-containing amine oxidase (EC 1.4.3.6), (R)-aminopropanol dehydrogenase (EC 1.1.1.75) and glycine C-acetyltransferase (EC 2.3.1.29).

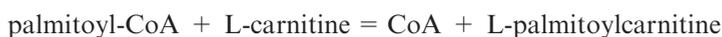
Maltose phosphorylase (EC 2.4.1.8) catalyses the reaction
maltose + phosphate = D-glucose + β -D-glucose 1-phosphate

but it would be difficult to trace the metabolic fates of the product with the information that β -D-glucose 1-phosphate spontaneously mutarotates to form α -D-glucose 1-phosphate.

(g) Overlapping specificities

It is not uncommon to find that more than one enzyme may be capable of catalysing the same reaction. These will be treated differently by the Enzyme List if they have sufficiently different substrate specificities. For example, an aldehyde may be a substrate for alcohol dehydrogenase (EC 1.1.1.1), alcohol dehydrogenase (NADP⁺) (EC 1.1.1.2), aldehyde reductase (EC 1.1.1.21) and aldehyde oxidase (EC 1.2.3.1), among many other enzymes.

Carnitine *O*-palmitoyltransferase (EC 2.3.1.21) catalyses the reaction:



and the comments indicate that it has a “Broad specificity to acyl group, over the range C₈ to C₁₈; optimal activity with palmitoyl-CoA”. The related enzyme carnitine *O*-octanoyl-transferase (EC 2.3.1.137) catalyses octanoyl-CoA + L-carnitine = CoA + L-octanoylcarnitine.

Thus both these enzymes will use octanoyl-CoA to extents that will depend on their respective activity levels, distribution and kinetic parameters.

In such cases, the necessary data are in the Enzyme List, supplemented by the additional information in *BRENDA*. The problem is simply one of ensuring that all enzymes that may work with a given metabolite are considered.

(h) Trouble-shooting through pathway reconstruction

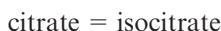
Two examples will be used to illustrate how pathway tracing may be used to reveal deficiencies in the Enzyme List data. It has been known for many years that it was not possible to use the Enzyme List data to reconstruct the citric-acid cycle because the reaction catalysed by aconitase (aconitate hydrolase; EC 4.2.3.1) was given as:



However, in the operating citric-acid cycle, the overall process catalysed includes a second reaction:



leading to an overall process of



Although this was known at the time the enzyme was first included in the Enzyme List (1961), only the first reaction was given because the equilibrium of the reaction very much favours citrate (the equilibrium mixture is 91% citrate, 6% isocitrate and 3% aconitate) and the existence of the second reaction was indicated only in the comments. Thus this is another example of the misleading inferences that can be drawn from considering isolated enzyme thermodynamics rather than system thermodynamics. The second reaction as well as the overall reaction are now included in the reaction field of the Enzyme List entry for aconitase.

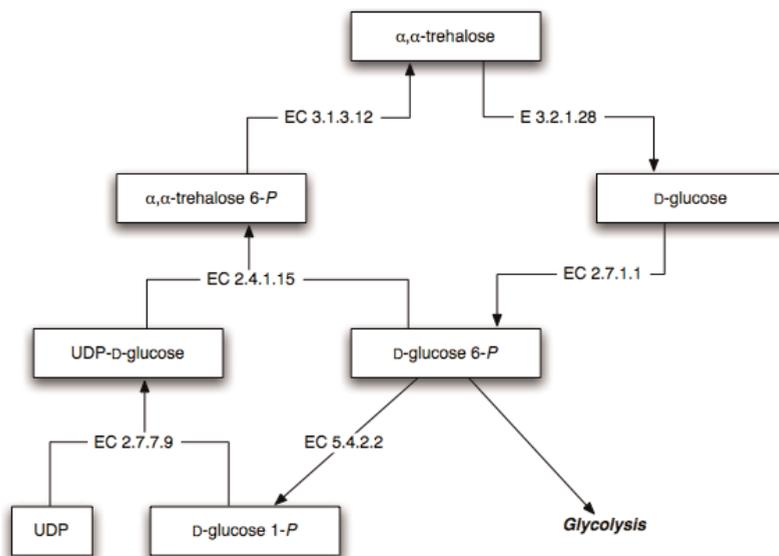
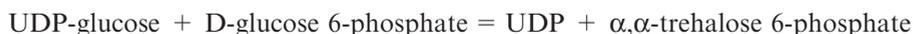


Figure 4. Trehalose metabolism in *M. grisea*.

The second example concerns the metabolism of trehalose [10], as shown in Fig. 4. The enzyme EC 2.4.1.15, α,α -trehalose-phosphate synthase (UDP-forming), was listed as catalysing the reaction:



However, the enzymes that might use this product, such as trehalose-phosphatase (EC 3.1.3.12) were shown as using trehalose 6-phosphate:



rather than α,α -trehalose 6-phosphate. Thus, any search for α,α -trehalose 6-phosphate would not reveal this, or other enzymes in the process. This was, in fact, an example of changes in nomenclature. In earlier formulations of the Enzyme List, some common enantiomeric designations were omitted. For example, it was assumed that all amino acids were L-amino acids unless otherwise specified. Similarly inositol was synonymous with *myo*-inositol and trehalose was α,α -trehalose. Since many biochemists were not familiar with these arcane conventions, the omitted enantiomeric designations have been added in more recent formulations of the Enzyme List but, somehow, this was not done for all the relevant trehalose entries. This has now been rectified.

(i) Conclusions

Not all of the problems discussed above concern the Enzyme List. Lacunae, such as those mentioned in the previous section, are filled as quickly as possible after they are discovered. While the Enzyme List primarily shows the enzyme-catalysed reaction, it is sometimes appropriate to include details of a spontaneous reaction that follows or precedes the enzyme-catalysed reaction, especially in cases where there would otherwise be a gap in a metabolic pathway.

Synonyms are important to allow the compounds to be found. The Enzyme List includes commonly used synonyms (other names) for each enzyme but it is not its function to include synonyms for all possible substrates. Synonyms are needed for searching *Reaction Database* and *Explorer* because many people use different names for the same compound and few use the often-unwieldy IUPAC-approved names. There are excellent small molecule databases, such as *ChEBI* [11] and *KEGG LIGAND* [12]. However, for convenience and since it is not uncommon to find that chemists prefer different alternative names from those favoured by biochemists and pharmacologists, *ChemFinder* [13] was searched for names to add to the *Reactions Database* and these were supplemented with information from the *Merck Index* [14]. Synonyms were then linked to the corresponding primary term for each compound, which were generally those used by the Enzyme List.

A big remaining challenge is to populate the *Reactions Database* with additional reactions that are not found in the Enzyme List, such as those provided by *BRENDA*. It will also be necessary to address the species problem, but at least for now, the problem of thermodynamic information may be best served by links to other sources.

REFERENCES

- [1] International Union of Biochemistry and Molecular Biology (IUBMB) Enzyme List, <http://www.chem.qmw.ac.uk/iubmb/enzyme/>.
 - [2] McDonald, A.G., Boyce, S., Moss, G.P., Dixon, H.B. and Tipton, K.F. (2007) ExplorEnz: a MySQL database of the IUBMB enzyme nomenclature. *BMC Biochem.* **8**:14, [<http://www.enzyme-database.org/>].
 - [3] Nicholson, D. (2000) The evolution of the IUBMB-Nicholson maps. *IUBMB Life* **50**(6):341 – 344, [<http://www.iubmb-nicholson.org/>].
 - [4] *Reaction Explorer*, <http://www.reaction-explorer.org/>
 - [5] GraphViz, <http://www.graphviz.org/>
 - [6] Schomburg, I., Chang, A., Hofmann, O., Ebeling, C., Ehrentreich, F. and Schomburg, D. (2002) BRENDA: a resource for enzyme data and metabolic information. *Trends Biochem. Sci.* **27**:54 – 56, <http://www.brenda-enzymes.info/>.
 - [7] Goldberg, R.N., Tewari, Y.B., Bhat, T.N. (2004) Thermodynamics of enzyme-catalyzed reactions – a database for quantitative biochemistry. *Bioinformatics* **20**:2874 – 2877, http://xpdb.nist.gov/enzyme_thermodynamics/.
 - [8] Alberty, R.A. (2006) Calculation of equilibrium compositions of systems of enzyme-catalyzed reactions. *J. Phys. Chem. B* **110**:24775 – 24779.
 - [9] Alberty R.A. (2003) *Thermodynamics of Biochemical Reactions*, John Wiley, U.S.A..
 - [10] Wang, Z.Y., Jenkinson, J.M., Holcombe, L.J., Soanes, D.M. *et al.* (2005) The molecular biology of appressorium turgor generation by the rice blast fungus *Magnaporthe grisea*. *Biochem. Soc. Trans.* **33**:384 – 388.
 - [11] ChEBI, <http://www.ebi.ac.uk/chebi/>.
 - [12] Goto, S., Okuno, Y., Hattori, M., Nishioka, T. and Kanehisa, M. (2002) LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.* **30**:402 – 404, <http://www.genome.jp/ligand/>.
 - [13] ChemFinder, <http://chemfinder.cambridgesoft.com/>.
 - [14] The Merck-Index, <http://www.merckbooks.com/mindex/> .
-