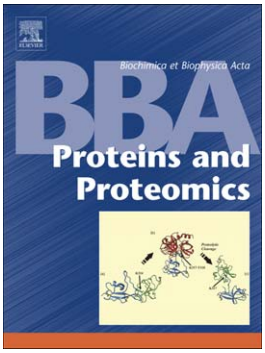


Tracing Metabolic Pathways from Enzyme Data

Andrew G. McDonald, Keith F. Tipton, Sinéad Boyce

PII: S1570-9639(09)00153-8
DOI: doi:10.1016/j.bbapap.2009.06.015
Reference: BBAPAP 38206

To appear in: *BBA - Proteins and Proteomics*

Received date: 19 April 2009
Revised date: 16 June 2009
Accepted date: 17 June 2009

Please cite this article as: Andrew G. McDonald, Keith F. Tipton, Sinéad Boyce, Tracing Metabolic Pathways from Enzyme Data, *BBA - Proteins and Proteomics* (2009), doi:10.1016/j.bbapap.2009.06.015

Tracing Metabolic Pathways from Enzyme Data

Andrew G. McDonald, Keith F. Tipton* and Sinéad Boyce

Department of Biochemistry, Trinity College, Dublin 2, Ireland

Corresponding author:
Email address: ktipton@tcd.ie
Telephone: 00353-1-8961802
Fax: 00353-1-6772400

Short title: Tracing Metabolic Pathways

Keywords: Database, Enzymes, Metabolic Pathways

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 many 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*, "Nomenclature and Classification of Enzymes by the Reactions they Catalyse," classifies enzymes in terms of their catalytic functions [see 1, 2]. It is restricted to classification and recommendations on nomenclature and is widely used by other databases as a source for avoiding ambiguity in the recognition of enzymes as catalytic entities. 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 in terms of the reaction or type of reactions catalysed. Several other databases, including *BRENDA* [3], *ExPASy* [4], *GO* [5] and *KEGG* [6], incorporate the Enzyme List data, as a primary source of information. Each enzyme is identifiable by an unique four-digit Enzyme Commission (EC) number, a systematic name and an "accepted name", which is often the name in most common use for that enzyme, provided it is not ambiguous or misleading,

The enzyme data, including associated literature references, are stored in MySQL databases that are accessed through a web application, *ExplorEnz* [2]. This allows more detailed searching facilities, including the use of Boolean algebra to include or exclude terms from the selected fields as well as searching with regular expressions. It can be downloaded in a number of different formats and is also used as a source of the flat-file version on the official IUBMB nomenclature and terminology web site [7]. Since the *Enzyme List* was designed for a specific purpose, it would be undesirable to alter it 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. A preliminary account of this work has been published previously [8].

Methods

Finding Enzyme Involvements – the "enzyme-centric" approach

The *Enzyme List* can be searched to find all of those enzymes that catalyse reactions involving a specific substrate or product. For example, a search for L-serine yields a list of 41 enzymes. It is possible to construct a simple pictorial representation from this, as shown in Fig. 1. Such "enzyme-centric" searching can be useful in predicting the possible effects of drugs or other manipulations that are targeted against a specific metabolic reaction, since they will show other pathways that might also be affected.

It is also possible to list groups of enzymes linked by common substrates and products, as shown in Fig. 2. The results of attempts to display these interactions in graphical form can, however, appear quite complex because of the multiplicity of edges that occurs when forming connections between enzymes. 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 and, for example 72,010 edges would be required to show the number of reactions in the database that involve O_2. Thus

representations, such as that shown in Fig. 2, although fully searchable, are not generally helpful as visual aids.

Clearly the situation can become more complicated than this if one considers the possibility of having several shared metabolites for each enzyme. Such a representation can readily be searched for any given enzyme. It 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 9, 10] or in the Roche Applied Science "Biochemical Pathways" wall chart, which can be searched, in segments, through *ExPASy* [11], where the separation of different metabolic systems in the display can result in the same enzyme occurring in several different places.

The substrate-centric approach

The reactions database was initially prepared by extracting all of the reactions in *ExplorEnz*, the MySQL version of the *Enzyme List* [2]. These were used to create a separate MySQL *Reactions database* and a web application, created with the PHP scripting language, was developed to provide the query interface for it [12]. The database assigns unique identifiers to reactants and reactions (C and R numbers, respectively), chemical names (primary terms) and alternative names (synonyms). The primary terms are those recommended by the IUBMB in their *Enzyme List*. Reactions are generally associated with an EC number; spontaneous reactions, or those with an unknown catalyst, are also allowed. Reactants are stored in a separate table according to their position in the formal reaction equation (left-hand or right-hand side), with an index value, a stoichiometric coefficient, and compound identifier associated with that reactant. Chemical names are not stored with the HTML markup needed for display, but are

rendered by the web application using the regular-expression based system described previously [2].

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 Table I.

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 H_2O, H^+, ATP, ADP, AMP, phosphate, diphosphate, $NADP^+$, NADPH, NAD^+, NADH, $NAD(P)^+$, NAD(P)H, A, AH_2, 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 initial searches. These might include:

ATP <=> ADP

ATP <=> AMP

AH2 <=> A

NAD+ <=> NADH

NADP+ <=> NADPH

NAD(P)+<=> NAD(P)H

CoA <=> acetyl-CoA

In the cases of H^+ and H_2O, which occurs in about 1,260 listed enzyme reactions, even if the peptidases are excluded, there is no practical alternative but to remove them from substrate/product searches.

Results & Discussion

Searching ExplorEnz for reactants is limited to the accepted name used in the *Enzyme List* plus any entries that occur in the "glossary", which relates the commonly used biochemical name to the IUPAC-approved name in some instances, and those alternative names that may occur in the references titles listed. The reaction database contains a large list of compound synonyms, and space and hyphen distinctions are ignored, for example, llysine, l-lysine and Llysine will all go to the entry for L-lysine. Similarly F6P, F-6P, F6-P, and F-6-P, whether in upper or lower case will all return the same 3 compounds, D-fructose 6-phosphate, D-hexose 6-phosphate and D-hexose phosphate, as shown in Fig. 3. Each of the primary terms used for the parent compounds is given a unique compound number (C number) internally. Links are provided for the primary term to the relevant entry in the KEGG database as well as to the PubChem entries for it and its derivatives, where such entries exist. A complete list of the primary terms of substrates and products as well as the compound numbers can be found from "Stats" on the tool bar. Selecting one of these will give the number of reactions, identified by reaction numbers, from the *Enzyme List* and linked to the relevant *ExplorEnz* entries. For example, there are 15

separate reactions involving D-fructose 6-phosphate (see Table 2) and one each for D-hexose 6-phosphate and D-hexose phosphate. Such inexact searching can be helpful in ensuring comprehensive results and can also aid those who can never remember, or do not care to remember, where the hyphens go in compounds like glucose 6-phosphate. While inexact searching is enabled by default, options for "exact" searches and those using regular expressions are also provided. If an exact search fails to find any compound, a dialogue box appears that offers the option of repeating the search using the inexact search option.

Representations

Although the *Reactions Database* can provide lists of reactions, an additional tool is needed for display purposes. This is provided by *Reaction Explorer*, which is a multi-platform application, written in REALbasic, for constructing basic metabolic network graphs. The software requires an active internet connection in order to query the *Reactions Database*. 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 and can be downloaded from the *Reaction Explorer* homepage [12]. 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 build up a pictorial representation of the process, as shown by the simple example in Fig. 4. Searching can be restarted from any previous compound. Graphs may be rearranged manually and can be saved to a file and reloaded for further exploration.

The output is designed to be simple because its purpose is to convey the basic information as efficiently clearly as possible, rather than to construct works of art. Thus, it is not designed as a competitor for the craftsman-designed Nicholson metabolic pathways charts [9,10], or the computationally-based representational systems, such as *GraphViz* [13] and the *BioCyc Pathway Tools* [14], but rather to display the essential information quickly and easily. However, different output formats to accord with those required by such systems could be provided in the future, as has already been done for *ExplorEnz* [2].

Applications

As indicated above, *Reaction Explorer* is an aid to visualize reactant interconnections in biochemical pathways. 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 Figs. 2 and 4. The ability to indicate possible metabolic pathways involving any chosen compound could be of value for assessing the possible ramifications of inhibiting, or amplifying, any single step in the pathway and for predicting the likely effects of substrate, or product, analogues on other metabolic processes. Thus the approach can readily produce models for systems biological quantitation. The system can also be of educational value in allowing students to trace known metabolic pathways and possible alternatives, from any chosen starting compound. The ability to trace the fates of each product, or substrates through different alternative pathways allows complex networks to be readily built.

Because the system is based upon the known, and verified, data in *ExplorEnz* it will not, of course, include reactions that may exist but have not yet been demonstrated.

However, its further development for use in conjunction with the substructure-search tool *BiSSCat* [15] and the Generator of Reaction Equations & Pathways, *GREP* [16], should allow its extension to predicted pathways involving orphan metabolites. Furthermore, there are limitations imposed not by the programme but by the nature of the system involved. As discussed in connection with Table 1, a tree that describes all reactions proceeding for *n* steps from any named reactant would not to be expected to be simple. Similarly, a single, unique, pathway connecting two distant metabolites does not usually occur in metabolite space. Thus the question "find the pathway from glucose to pyruvate and lactate" might be expected to yield glycolysis, but it would also give very many other pathways (>500). That is because there are very many ways in which glucose can be converted into pyruvate, including the synthesis and breakdown of compounds such as cholesterol. Similarly glucose → glycogen → glucose → pyruvate would not be an invalid pathway, despite our perceptions that glycolysis and glycogenenolysis are different processes. Neither does the *Enzyme List* contain any information on the thermodynamic feasibility of pathways, as will be discussed later. Although the possibilities might be reduced by specifying the number of steps allowed, not all metabolic pathways necessarily use the minimum number of steps.

Reaction Explorer has also proven to be useful for trouble-shooting recorded enzyme-reaction data. 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 (as shown in Fig. 4) because the reaction catalysed by aconitase (aconitate hydrolase; EC 4.2.3.1) was given as:

citrate = *cis*-aconitate + H_2O

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

cis-aconitate + H_2O = isocitrate

leading to an overall process of

citrate = isocitrate

Although this was known at the time the enzyme was first included in the *Enzyme List*, in 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 an example of the misleading inferences that can be drawn from considering isolated enzyme thermodynamics rather than system thermodynamics. The second reaction has been added to the *Reactions Database*; both it, and the overall reaction, are now included in the reaction field of the *Enzyme List* entry for aconitase.

The second example concerns the metabolism of trehalose [17], as shown in Fig. 5. The enzyme EC 2.4.1.15, α,α-trehalose-phosphate synthase (UDP-forming), was listed as catalysing the reaction:

UDP-glucose + D-glucose 6-phosphate = UDP + α,α-trehalose 6-phosphate

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

trehalose 6-phosphate + H_2O = trehalose + 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 regarded as being synonymous with *myo*-inositol and trehalose was α,α-trehalose. Since many biochemists were not familiar with these arcane conventions, some of the omitted enantiomeric designations were 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 corrected. Fig. 6 uses this pathway to illustrate nomenclature difficulties that had to be rectified.

Limitations and future developments

It is perhaps unreasonable to expect the *Enzyme List* to have functions for which it was not designed. However, there are several aspects that limit its applicability for simple adaptation to reaction pathway tracing through systems such as *Reaction Database* & *Explorer*. Some of the limitations and approaches to addressing them are described below.

(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. Work is in progress 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

(1) 3,4-dihydroxy-L-phenylalanine = dopamine + CO_2

(2) 5-hydroxy-L-tryptophan = 5-hydroxytryptamine + CO_2

but the enzyme 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

ATP + D-fructose 6-phosphate = ADP + D-fructose 1,6-bisphosphate

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, need to be incorporated into the *Reactions Database*. The comprehensive lists of substrates provided by the *BRENDA* database [3], 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

L-threonine + NAD^+ = L-2-amino-3-oxobutanoate + NADH + H^+

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

L-serine + NAD^+ = L-2-amino-3- oxopropionate + NADH + H^+

An additional complication can occur when spontaneous (uncatalysed) reactions occur. 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 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 primary-amine oxidase (EC 1.4.3.21), (*R*)-aminopropanol dehydrogenase (EC 1.1.1.75) and glycine *C*-acetyltransferase (EC 2.3.1.29).

In some cases a spontaneous reaction may compete with catalysed reactions. For example, putrescine aminotransferase (EC 2.6.1.82) catalyses the reaction

putrescine + 2-oxoglutarate = 4-aminobutanal + L-glutamate

and the 4-aminobutanal can be spontaneously converted to 1-pyrroline or oxidizes by aldehyde dehydrogenase (EC 1.2.1.3) or aminobutyraldehyde dehydrogenase (EC 1.2.1.19) to form 4-aminobutanoate.

(b) 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, some aldehydes may be substrates 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 several other enzymes.

Carnitine *O*-palmitoyltransferase (EC 2.3.1.21)

catalyses the reaction:

palmitoyl-CoA + L-carnitine = CoA + L-palmitoylcarnitine

and the comments indicate that it has a "broad specificity to acyl group, over the range C_8 to C_{18}; optimal activity with palmitoyl-CoA". The related enzyme carnitine *O*-octanoyltransferase (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 much of the necessary data are in the *Enzyme List*, supplemented by the additional information in *BRENDA* [3]. The problem is simply one of ensuring that all enzymes that may work with a given metabolite are considered.

(c) General or 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* often gives a single generic reaction. For example, the alcohol dehydrogenase reaction is given as

an alcohol + NAD^+ = an aldehyde or ketone + NADH + H^+

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

an L -amino acid + H_2O + O_2 = a 2-oxo acid + NH_3 + H_2O_2

The problem with this approach is that it does not indicate which alcohols, or L-amino acids, are not substrates. Again, the *BRENDA* database [3] can be invaluable in such cases, although it should always been 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 yet has tried it.

The example, discussed above, of searching for F6P yielding D-hexose 6-phosphate as well as D-fructose 6-phosphate is because the reaction of hexokinase (EC 2.7.1.1) is listed as:

ATP + D-hexose = ADP + D-hexose 6-phosphate

Markush terms are also used for some reactions. For example the reaction catalysed by primary-amine oxidase (EC 1.4.3.21) is given as:

$RCH_2NH_2 + H_2O + O_2 = RCHO + NH_3 + H_2O_2$

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

$R\text{-}CHOH\text{-}R' + NADP^+ = R\text{-}CO\text{-}R' + NADPH + H^+$

Although such formulations are somewhat more informative than the general reactions described 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 use of a simple reaction equation, without ambiguity. Examples of this include many reactions with oligomeric substrates, such as many of the peptides

polynucleotides and polysaccharides. For example the reaction catalysed by (α-amylase) (EC 3.2.1.1) is given as

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

and the reaction of exodeoxyribonuclease I (EC 3.1.111) is given as

> Exonucleolytic cleavage in the 3′- to 5′-direction to yield nucleoside 5′-phosphates

(d) Internal synonym inconsistencies

There have been many changes in nomenclature since the *Enzyme List* was started in 1952. Normally these have been made to correct or rationalize nomenclature, for example, the change from DPN (even before that, it was called coenzyme I) to NAD^+ caused considerable dissent at the time [see 18], whereas the change from diphosphate to bisphosphate for those compounds where the phosphate groups are attached to different groups on the same molecule went comparatively smoothly. Generally, the *Enzyme List* is punctilious about correcting entries but, as will be discussed below, a few may escape the notice of the IUBMB-IUPAC Joint Commission on Biochemical Nomenclature and of those who use the *Enzyme List*. As discussed above, pathway tracing can be of value in finding, and correcting, such inconsistencies.

(e) Reaction directionality

The reactions presented in the *Enzyme List* are, as far as possible, mass-balancing equations. They are not meant to indicate the equilibrium position of the reaction or the

direction in which the reaction is believed to operate *in vivo*. Thus the reaction is presented as A+ B = C + D, and not A+ B ⇆ C + D or A+ B → C + D. 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, also 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 for the oxidation of ethanol by alcohol dehydrogenase (EC 1.1.1.1)

$$CH_3CH_2OH + NAD^+ = CH_3CHO + NADH + H^+$$

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

$$CH_3CHO + NAD^+ + H_2O = CH_3COOH + NADH + H^+$$

Thermodynamic data for many enzymes can be found in the *GTD* Thermodynamics of Enzyme-catalysed Reactions database [19] and the kinetic data are included in the *BRENDA* database [3] may provide sufficient detail to determine reaction equilibria through Haldane relationships. However, 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 [see 20,21]. For this reason, the desktop application version of *Reaction Explorer* does not generate directed graphs (digraphs).

(f) Species differences

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 hemi-acetals; the animal, but not the yeast, enzyme acts also on cyclic secondary alcohols". The *BRENDA* database [3], 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.

(g) Overlapping specificities

It is not uncommon to find that more than one enzyme may be capable of catalysing the same reaction. Such enzymes will be classified separately 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:

palmitoyl-CoA + L-carnitine = CoA + L-palmitoylcarnitine

and the comments indicate that it has a "broad specificity to acyl group, over the range C_8 to C_{18}; optimal activity with palmitoyl-CoA". The related enzyme carnitine *O*-octanoyltransferase (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* [3]. The problem is simply one of ensuring that all enzymes that may work with a given metabolite are considered.

Conclusions

There are several other enzyme data systems that present metabolic pathways for selected systems, including *KEGG* [6], *MetaCyc* [21] and *UMBBD* [23], and the *Enzyme List*, itself, also contains some pathways and reaction systems. BioCarta [24] contains many attractively presented pathways and the PathwayExplorer website [25] collects many of these together. Whereas, these are largely static pictures, *Pathway Hunter Tool* [26] allows the user to generate *KEGG* pathways between selected metabolites. However, *Reaction Explorer* differs from these by allowing the used to follow metabolites to build up pathways *de novo* with no inbuilt constraints other than the *Enzyme List/ExplorEnz* dataset on which it is currently based.

Not all of the problems discussed above concern the *Enzyme List*. Lacunae, such as those mentioned in the previous sections, are filled as quickly as possible after they are discovered and verified. 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.

It has been frequently suggested that the *Enzyme List* should give each reaction catalysed a separate EC number. Thus, for example there would be different numbers for hexokinase catalysing the phosphorylation of D-glucose, D-mannose, D-fructose, sorbitol and D-glucosamine. Such a system would, of course, not be consistent with function of the *Enzyme List* itself, but such a parallel approach might be beneficial. Such a system could be developed from the reaction numbers used here, but the reactant-pair (RP) identifiers used by *KEGG* [6] might provide an easier approach.

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* [27] and *KEGG LIGAND* [28]. 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* [29] was searched for names to add to the *Reactions Database* and these were supplemented with information from the *Merck Index* [30]. These synonyms were then linked to the corresponding primary term for each compound. An alternative for avoiding ambiguity might be to allow searching by drawn structure or SMILES string. However, there are already several free utilities available that will convert structures or

SMILES to names [e.g., 31, 32] that can be used to generate search terms. In the future the option of searching by InChI key might be the most appropriate [33].

A major 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* [3]. Some reactions, such as those of aconitase, have already been added, and to facilitate data entry a curatorial web interface, adapted from that already employed by *ExplorEnz* [2] is being developed. It will also be necessary to address the species problem, but at least for now, but the problem of thermodynamic information may be best served by links to other sources.

Acknowledgement

We are grateful to Science Foundation Ireland for support.

References

1. S. Boyce, and K.F. Tipton. History of the enzyme nomenclature system. *Bioinformatics* 16 (2000) 34-40.
2. A.G. McDonald, S. Boyce, K.F. Tipton. ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res*. 37 (2009) D593-597. http://www.enzyme-database.org/
3. BRENDA – The Comprehensive Enzyme Information System: http://www.brenda-enzymes.info/
4. ExPASy Enzyme Nomenclature database: http://www.expasy.ch/enzyme/
5. GO - The Gene Ontology Database: http://www.geneontology.org/

6. KEGG - Kyoto Encyclopedia of Genes and Genomes: http://www.genome.jp/kegg/

7. IUBMB Recommendations on Biochemical & Organic Nomenclature, Symbols & Terminology etc: http://www.chem.qmw.ac.uk/iubmb/enzyme/

8. A.G. McDonald, K.F. Tipton, S. Boyce. From the enzyme list to pathways and back again. *Proceedings of the 3rd International Beilstein Workshop on Experimental Standard Conditions of Enzyme Characterizations* (Edited by M.G. Hicks & C. Kettner) (2008) pp. 109-122, Beilstein Institut, Frankfurt, Germany.

9. D. Nicholson. The evolution of the IUBMB-Nicholson maps, *IUBMB Life* 50 (2000) 341-344.

10. IUBMB-Nicholson Metabolic Maps, Minimaps & Animaps: http://www.iubmb-nicholson.org/

11. ExPASy Biochemical Pathways: http://www.expasy.ch/tools/pathways/

12. Reaction Explorer: http://www.reaction-explorer.org/

13. Graphviz - Graph Visualization Software: http://www.graphviz.org/

14. S.M. Paley and P.J. Karp. The Pathway Tools cellular overview diagram and Omics Viewer. *Nucleic Acids Res.* 34 (2006) 3771-3778.

15. M. Kotera, A.G. McDonald, S. Boyce and K.F. Tipton. Functional group and substructure searching as a tool in metabolomics. *PLoS ONE* 3 (2008) e1537.

16. M. Kotera, A.G. McDonald, S. Boyce and K.F. Tipton. Eliciting possible reaction equations and metabolic pathways involving orphan metabolites. *J.*

Chem. Inf. Model. 48 (2008) 2335-2349.

17. Z.Y. Wang. J.M. Jenkinson, L.J. Holcombe, D.M. Soanes, *et al.* (2005) The molecular biology of appressorium turgor generation by the rice blast fungus *Magnaporthe grisea*. *Biochem. Soc. Trans*. 33, 384-388.

18. EC Webb. Enzyme nomenclature: a personal retrospective. *FASEB J*. 7 (1993) 1192-1194.

19. R.N. Goldberg, Y.B. Tewari, T.N. Bhat. Thermodynamics of enzyme-catalyzed reactions—a database for quantitative biochemistry. *Bioinformatics* 20 (2004) 2874-2877. http://xpdb.nist.gov/enzyme_thermodynamics/

20. R.A. Alberty. Calculation of equilibrium compositions of systems of enzyme-catalyzed reactions. *J. Phys. Chem. B* 110 (2006) 24775-24779.

21. R.A. Alberty. Thermodynamics of Biochemical Reactions, (2003) John Wiley, USA.

22. R. Caspi, H. Foerster, C.A. Fulcher, P. Kaipa, et al. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res*. 36 (2008) D623-D631. http://metacyc.org/

23. L.B.M. Ellis, D. Roe and L.P. Wackett. The University of Minnesota Biocatalysis/Biodegradation Database: The First Decade. *Nucleic Acids Res.*, 34 (2006) D517-D521. http://umbbd.msi.umn.edu/

24. BioCarta. http://www.biocarta.com/

25. B. Mlecnik, M. Scheideler, H. Hackl, et al. Trajanoski Z. PathwayExplorer: web service for visualizing high-throughput expression data on biological pathways.

Nucleic Acids Res. 33 (2005) W633-W637. https://pathwayexplorer.genome.tugraz.at/

26. S.A. Rahman P. Advani, R. Schunk, R. Schrader and D. Schomburg. Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). *Bioinformatics* 21 (2005) 1189-1193.

27. K, Degtyarenko, P. de Matos, M. Ennis, J. Hastings, et al. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res*. 36 (2008) D344-350. http://www.ebi.ac.uk/chebi/

28. S. Goto, Y. Okuno, M. Hattori, T. Nishioka and M. Kanehisa. LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.* 30 (2002) 402-404. http://www.genome.jp/ligand/

29. ChemFinder. http://chemfinder.cambridgesoft.com/reference/chemfinder.asp

30. The Merck Index 13th Edition 2001 John Wiley & Sons. http://www.merckbooks.com/mindex/online.html/

31. Online SMILES Translator and Structure File Generator. http://cactus.nci.nih.gov/services/translate/

32. InChI converter. http://www.inchi.info/converter_en.html/

33. T. Kind, M. Scholz and O. Fiehn. How large is the metabolome? A critical analysis of data exchange practices in chemistry. *PLoS ONE* 4 (2009) e5440.

Legends to Figures

Fig. 1. The substrate-centric approach: a diagrammatic view of the enzymes having L-serine as a substrate or product.

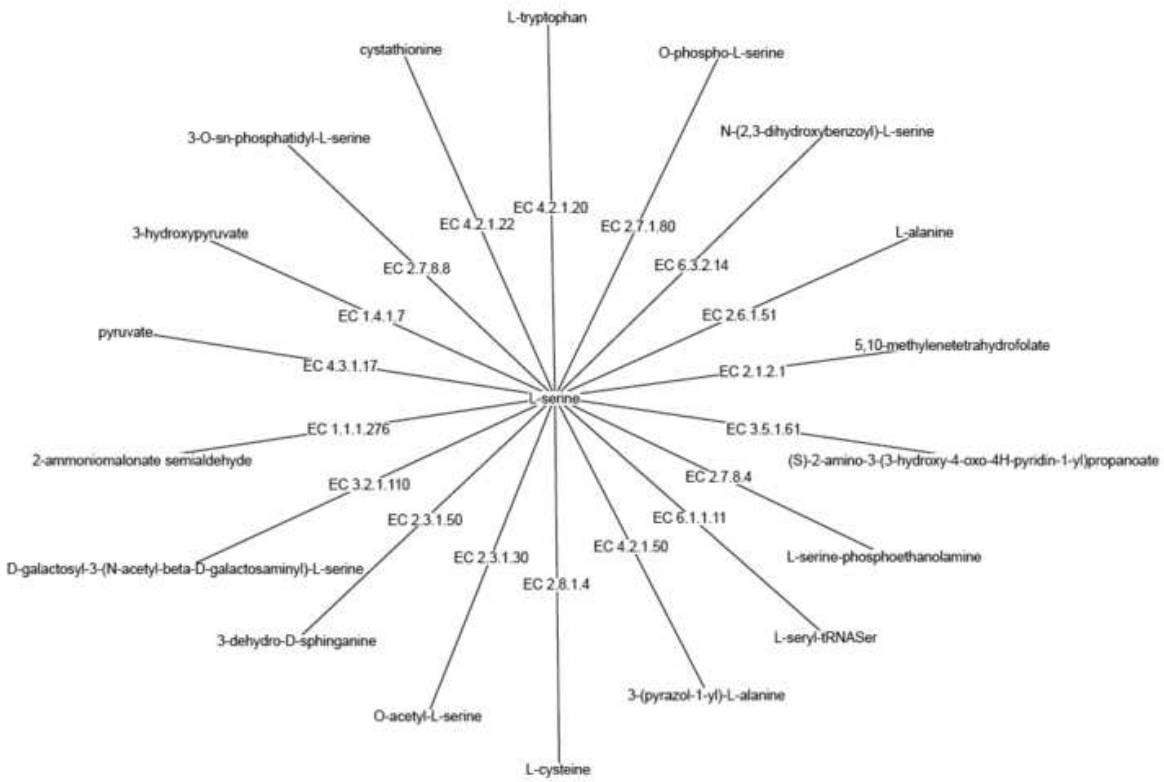
Fig 2. The enzyme-centric approach: a network of 525 enzymes with links representing shared metabolites. There are 3,985 connections made amongst the enzymes in this subset, from which H^+, H_2O, and common cofactor pairs (e.g., ATP/ADP) were excluded.

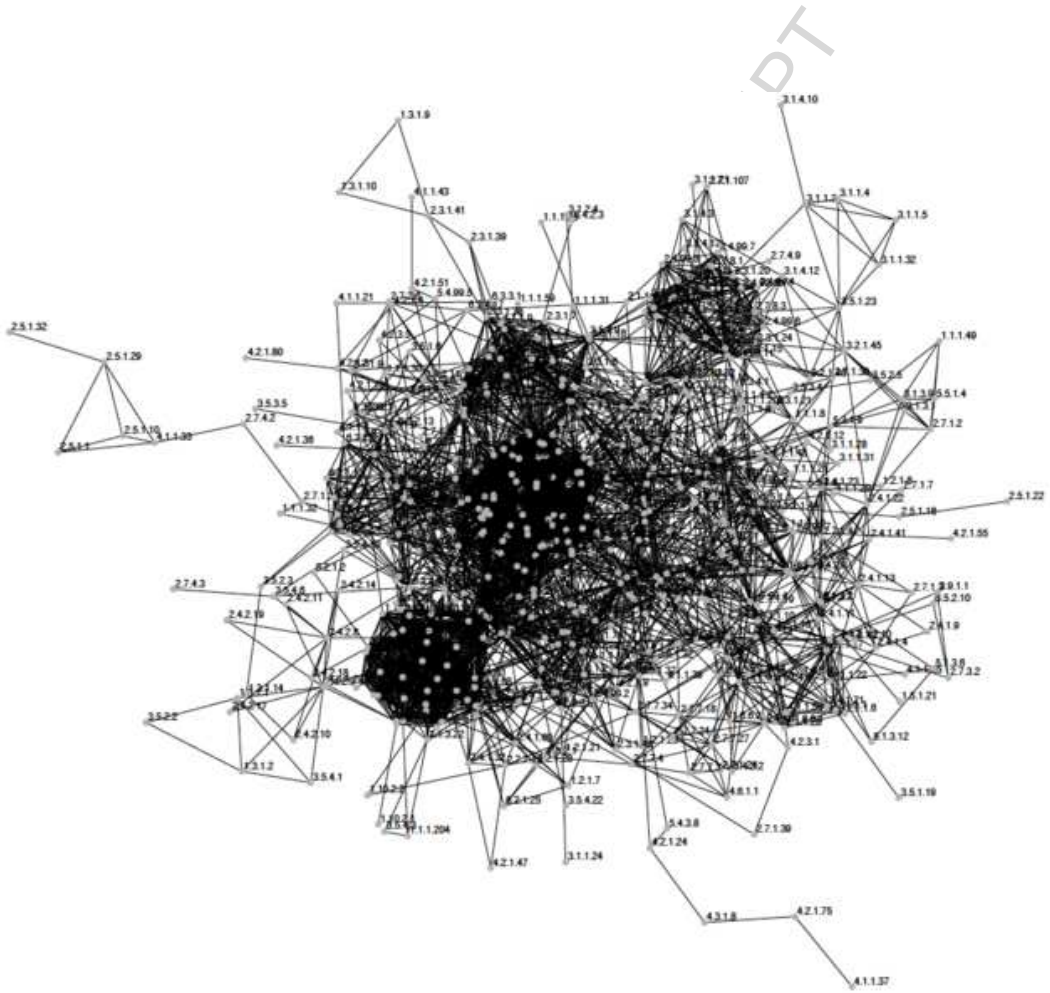
Fig. 3. Results of a search for F6P in *Reaction Explorer*.

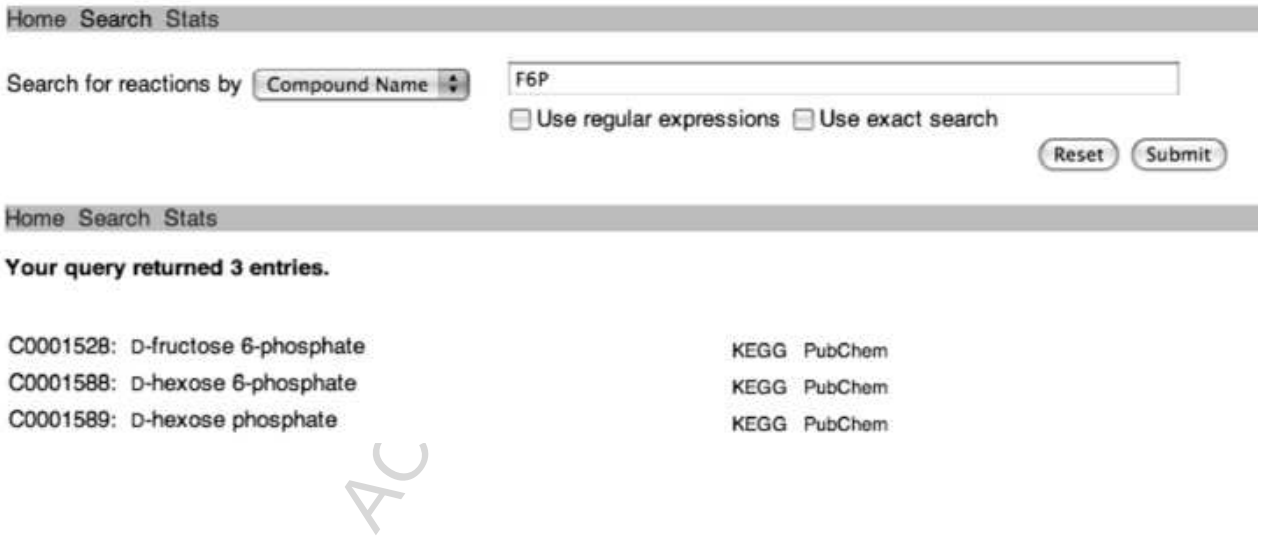
Fig. 4. Diagram of the early stages of porphyrin biosynthesis, produced by *Reaction Explorer*. The network can be extended through links from each metabolite.

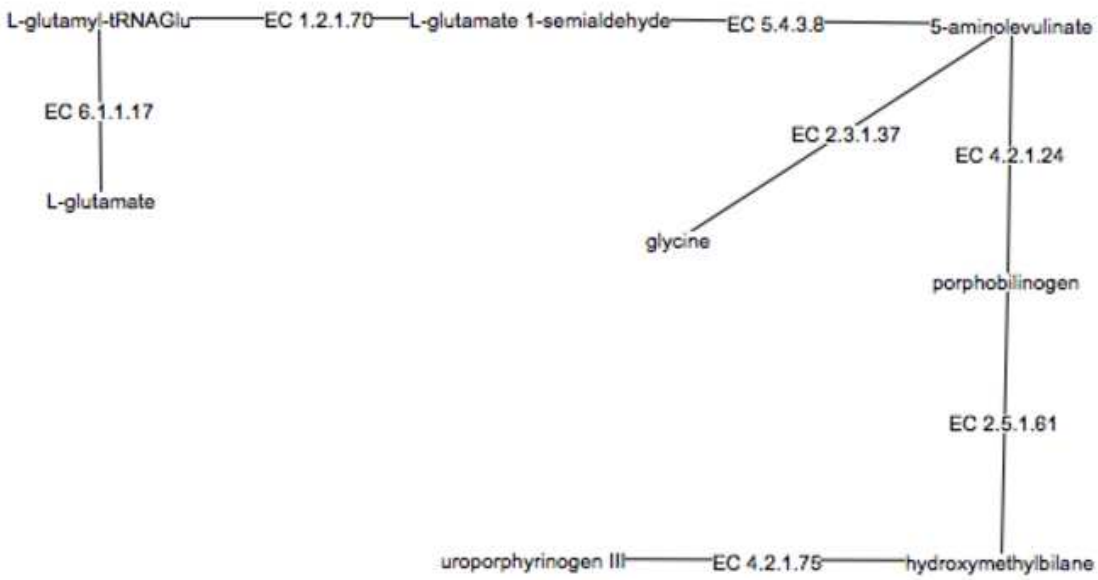
Fig 5. Trehalose metabolism in *M. grisea* (adapted from ref. 24).

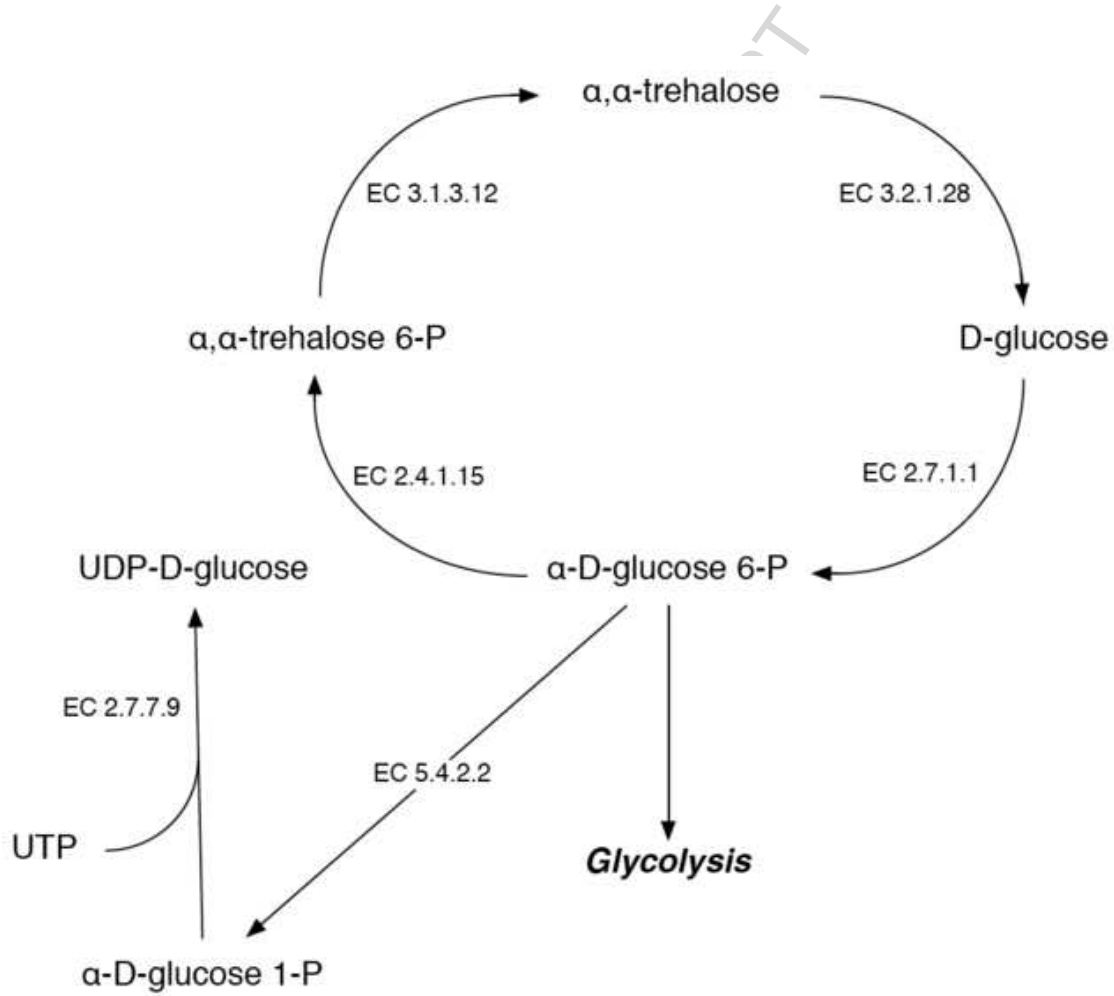
Fig. 6. A reconstruction of trehalose metabolism, drawn by *Reaction Explorer*, indicating the potential nomenclature difficulties from *ExplorEnz* data that had to be overcome.











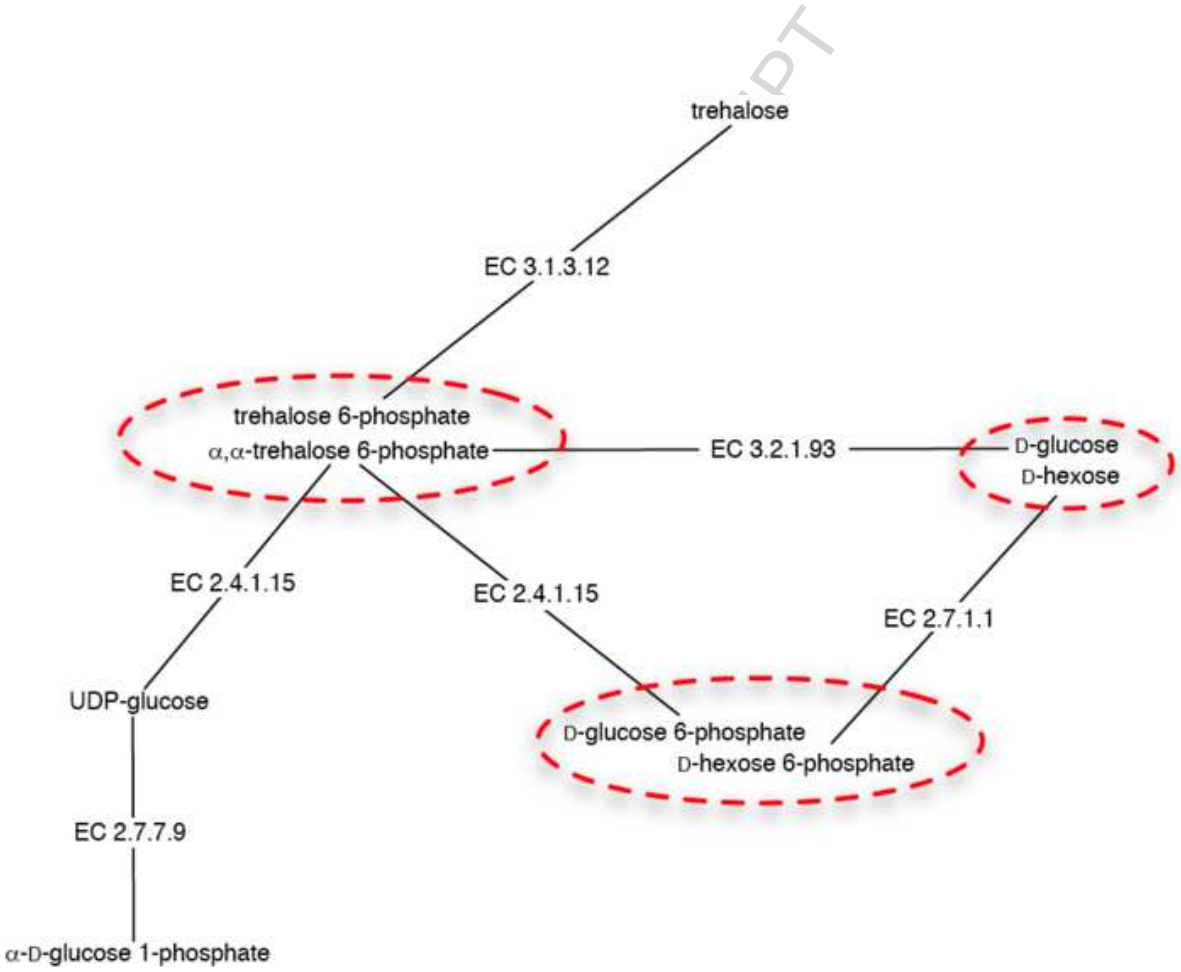


Table 1.
Some query results for successive reactions of different compounds.

Compound	Paths found		
	1 Step	2 Steps	3 Steps
L-Ascorbate	5	403	-
L-Cysteine	5	25	577
L-Tyrosine	13	676	-
Ribitol	2	50	2776

Table 2.
Results of reaction explorer search for D-fructose 6-phosphate

15 reactions found involving 'D-fructose 6-phosphate'

R0000017	D-mannitol 1-phosphate + NAD^+ = D-fructose 6-phosphate + NADH + H^+
R0000134	D-sorbitol 6-phosphate + NAD^+ = D-fructose 6-phosphate + NADH + H^+
R0001270	sedoheptulose 7-phosphate + D-glyceraldehyde 3-phosphate = D-erythrose 4-phosphate + D-fructose 6-phosphate
R0001489	UDP-glucose + D-fructose 6-phosphate = UDP + sucrose 6-phosphate
R0001827	L-glutamine + D-fructose 6-phosphate = L-glutamate + D-glucosamine 6-phosphate
R0001892	ATP + D-fructose = ADP + D-fructose 6-phosphate
R0001898	ATP + D-fructose 6-phosphate = ADP + D-fructose 1,6-bisphosphate
R0001974	diphosphate + D-fructose 6-phosphate = phosphate + D-fructose 1,6-bisphosphate
R0002024	ADP + D-fructose 6-phosphate = AMP + D-fructose 1,6-bisphosphate
R0002336	D-fructose 1,6-bisphosphate + H_2O = D-fructose 6-phosphate + phosphate
R0002370	β -D-fructose 2,6-bisphosphate + H_2O = D-fructose 6-phosphate + phosphate
R0002842	D-glucosamine 6-phosphate + H_2O = D-fructose 6-phosphate + NH_3
R0003393	D-fructose 6-phosphate + phosphate = acetyl phosphate + D-erythrose 4-phosphate + H_2O
R0003700	D-mannose 6-phosphate = D-fructose 6-phosphate
R0003701	D-glucose 6-phosphate = D-fructose 6-phosphate