Cell. Mol. Life Sci. 63 (2006) 517–523
1420-682X/06/050517-7
DOI 10.1007/s00018-005-5520-6

Cellular and Molecular Life Sciences

Visions & Reflections

Puzzling over orphan enzymes

O. Lespinet and B. Labedan*

Institut de Génétique et Microbiologie, CNRS UMR 8621, Université Paris Sud, Bâtiment 400, 91405 Orsay Cedex (France), Fax: +33 1 69 15 72 96, e-mail: bernard.labedan@igmors.u-psud.fr

Received 31 October 2005; received after revision 8 December 2005; accepted 20 December 2005
Online First 7 February 2006

Abstract. Despite the current availability of several hundreds of thousands of amino acid sequences, more than 39% of the well-defined enzyme activities (EC numbers) are not associated with any sequence in major public databases. This wide gap separating knowledge of biochemical function and sequence information is found in nearly all classes of enzymes. Thus, there is an urgent need to explore the 1525 orphan enzymes (EC numbers without associated sequences), in order to progressively bridge this unwanted gap. Improving genome annotation could unveil a significant proportion of sequenceless enzymes. Peptide mass mapping and further genome mining would be useful to identify proper sequence for enzymes found in species for which genetic tools are missing. Finally, the whole community must help major public databases to begin addressing the problem of missing or incomplete information.

Keywords. Orphan enzyme, EC number, protein sequence, protein function, gene annotation, database exactitude, hidden knowledge.

Our vision of the protein universe is presently fuzzy, although its basic organization and some of its properties are beginning to be illuminated [1 and references therein]. The distribution of proteins inside this universe does not appear to be stochastic. Rather, proteins with similar functions can be found close together [2], and there is a colocalization of structure (the protein fold universe [3]) and function (all possible types of functions that are listed by the Gene Ontology (GO) Consortium [4]). However, two terrae incognitae are currently obscuring this protein universe, each one being made of orphan entities.

The profusion of genes without known homologues was one of the greatest surprises unveiled by genome sequencing. We now know that a huge number of open reading frames, the so-called orphan genes, have been missed by classical approaches of molecular genetics and biochemistry. For instance, the last release of the OrphanMine database (http://www.genomics.ceh.ac.uk/orphan_mine/orphanmine.php) has compiled 44,752 orphans from the 150 genomes analyzed (430,826 predicted proteins), and this mean percentage (around 10% of the total genes per genome) is maintained at a steady pace despite the availability of the complete sequences of more than 275 genomes in public databases [5].

Enzymes form a major group within the protein function universe. For instance, enzymes correspond to nearly half of all defined functions in *Escherichia coli*, one of the most studied model organisms [6]. However, despite the deluge of various genome sequences, two groups have recently observed that no amino acid sequence is available in the public databases for a surprisingly high proportion of well-defined enzyme activities [7, 8]. Such an unexpected finding defines another kind of orphan. Here, we try to understand this failure and the problem it creates of an unwanted gap separating knowledge of biochemical function and sequence information.

* Corresponding author.

A deep gap between biochemistry and genetics/ genomics: introducing the concept of orphan enzymes

Biochemists have studied numerous enzymatic processes in various organisms for many decades. In an effort to organize all the pieces of information gathered and to check their validity and consistency, the International Enzyme Commission (EC) has been classifiying enzyme activities since 1956 (http://www.chem.qmul.ac.uk/iubmb/enzyme/index.html). This standardization effort is based on the definition of the so-called EC numbers, which comprise four digits. The first one (from 1 to 6) delineates the broad type of activity: Oxidoreductase, Transferase, Hydrolase, Lyase, Isomerase and Ligase, respectively. The second and third digits indicate the reaction that an enzyme catalyzes. For example, among the class Transferase (EC 2), glycosyltransferases form the subclass EC 2.4 and hexosyltransferases the sub-subclass EC 2.4.1, respectively. The last digit stands for substrate specificity. For instance, EC 2.4.1.34 corresponds to the 1,3-beta-glucan synthase that is involved in cell wall synthesis in fungi, while EC 2.4.1.80 defines the ceramide glucosyltransferase that catalyzes the first step of glycosphingolipid metabolism. It is very important to grasp that an EC number qualifies an enzyme activity but not the macromolecule itself that catalyzes this activity.

The EC categorization is constantly evolving as its curators decide, based on their specific expertise, what the best definition of each enzymatic activity is. Presently (September 2005), 4499 EC numbers have been indexed, but only 3877 correspond to a defined unambiguous activity. The 622 others are in the process of either being deleted or being assigned to another EC number.

Unexpectedly, Peter Karp [7] and we [8] independently observed that a significant portion of these well-defined EC numbers have no protein sequence in the public databases. A recent update of our initial results confirms this very large gap. There are presently only 2352 EC numbers that have at least one associated sequence in release 5.8 (30 August 2005) of the UniProt Knowledgebase [9]. We have dubbed orphan enzymes [8] the 1525 EC numbers found to be sequenceless. Remarkably, these orphan enzymes currently represent 39.3% of the total of the retained EC numbers, a figure close to that we found last year (42.5% [8]) despite the definition of 169 new EC numbers and the reannotation of many newly sequenced genes.

Orphan enzymes are widely distributed in the main functional categories. This is the case, for instance, of a significant number of enzymes that are involved in various metabolic pathways, despite a multitude of labs worldwide having studied them intensively and extensively for many years. Table 1 summarizes the data we compiled when looking at the main classes of pathways (as defined in KEGG [10] and MetaCyc [11]). For example, among the 11 pathways belonging to the class 'Cofactors and Vitamins' only one (that of folate) is complete, i.e., the sequence of the enzymes mediating each step is known,

Table 1. Relative percentage in the main classes of metabolic pathways.

Pathway class	Total EC numbers	Total orphans	Percentage orphans	Pathway with the highest percentage of orphans in the class	Number of complete pathways	Example of complete pathways
Cofactors and vitamins	231	54	23.4%	retinol metabolism (80.0%)	1/11	folate
Carbohydrate	669	156	23.3%	C5-branched dibasic acid metabolism (61.9%)	1/17	inositol
Amino acid	755	169	22.4%	lysine degradation (44.4%)	2/25	urea cycle
Lipid	241	51	21.2%	bile acid biosynthesis (47.8%)	2/10	fatty acids
Nucleotide	156	28	17.9%	purine metabolism (20%)	0/2	–
Glycan	153	26	17.0%	chondroitin/heparan sulfate biosynthesis (33.3%)	6/14	LPS, peptidoglycan
Energy	156	21	13.4%	sulfur metabolism (31.0%)	4/8	oxidative phosphorylation
Secondary metabolites	195	73	37.4%	alkaloid biosynthesis i (51.4%)	3/14	penicillin and cephalosporin biosynthesis
Biodegradation of xenobiotics	200	55	27.4%	2,4-dichlorobenzoate degradation (59.1%)	2/18	1,2-dichloroethane degradation
Polyketides and nonribosomal peptides	13	2	15.4%	biosynthesis of 12-, 14- and 16-membered macrolides (50.0%)	3/5	biosynthesis of siderophore group nonribosomal peptides

Table 2. The 20 pathways displaying the highest content of orphan EC numbers.

Pathway name	EC numbers	Number of orphans	Percentage of orphans
Retinol metabolism	10	8	80
C5-Branched dibasic acid metabolism	21	13	61.9
Vitamin B6 metabolism	20	12	60
2,4-Dichlorobenzoate degradation	22	13	59.1
Indole and ipecac alkaloid biosynthesis	22	13	59.1
Alkaloid biosynthesis I	37	19	51.4
D-Arginine and D-ornithine metabolism	8	4	50
Monoterpenoid biosynthesis	16	8	50
Bile acid biosynthesis	23	11	47.8
Nitrobenzene degradation	11	5	45.5
Flavonoid biosynthesis	20	9	45
Caprolactam degradation	9	4	44.4
Lysine degradation	45	20	44.4
Ascorbate and aldarate metabolism	25	11	44
Nucleotide sugars metabolism	28	12	42.9
Stilbene, coumarine and lignin biosynthesis	19	8	42.1
D-Glutamine and D-glutamate metabolism	12	5	41.7
Glutathione metabolism	27	11	40.7
beta-Alanine metabolism	32	13	40.6
Toluene and xylene degradation	17	6	35.3
Total	424	205	48.35

whereas 80% of the enzymes involved in the retinol pathway are sequenceless. Extending this analysis to all of metabolism, Table 1 further shows that, remarkably, only 24 out of 124 pathways (corresponding to 185 EC numbers) are devoid of any orphan EC numbers. The remaining 100 pathways contain 635 orphan enzymes amounting to 24.6% of their 2600 EC numbers. This mean percentage rises to 48.35% in the case of the 20 pathways that have the highest content in orphan EC numbers and as high as 55.8% in the case of the top ten (Table 2).

It was already surprising to find so many orphans among the enzymes involved in metabolism. However, it appears even more astonishing that 52.1% of the enzymes that are not involved in general metabolism are sequenceless (Table 3). For instance, we find a significant proportion of orphan enzymes in the case of Hydrolase class despite intensive studies both in fundamental and applied research and their industrial use [12]. This is the case of 51 among the 304 identified peptidases (EC 3.4) and of near half (46 out of 93) of the recognized glycosylases (EC 3.2).

Table 3. General distribution of orphan EC numbers.

Retained EC numbers		Orphans	
Type	Number	Proportion	%
Involved in metabolism	1994	544/1994	27.3
Not involved in metabolism	1883	981/1883	52.1
Total	3877	1525/3877	39.3

This observed imbalance in the distribution of orphan EC numbers dramatically increased over the last 2 years: we found that the respective proportion of non-metabolic orphans rises to 65% of the EC numbers created in 2004 and lacking any sequence. This figure goes up to 100% in 2005!

Defining protein function is a complex task: the intricacies of amino acid sequence/EC number relations

The eruption of both orphan genes and orphan EC numbers on differing sides of the protein universe is underlining our deep ignorance in fundamental domains of biology despite advances in genomics.

Presently, one of the most challenging problems we have is defining protein function. These definitions must be as accurate and meaningful as possible when annotating genes and genomes. In particular, a major difficulty is to delineate the relationships between molecular function, cellular function(s), primary (main) and secondary (accessory) functions and so on. In an ideal world, one expects a cascade of simple relations such as one enzyme

corresponding to one EC number that corresponds to one molecular function that corresponds to one cellular function. As usual in biology, things are far more complex, and different kinds of intricate relationships have been described.

First, there are cases where a wide array of sequences and structures exhibit the same catalytic activity. This is the case of the 1232 proteins that are annotated in the current release of Uniprot as displaying kinase activity (EC 2.7.1.37). Besides this universal catalytic mechanism, a very few features are common to kinases, for instance, *Bacillus subtilis* SpoIIAB, which acts as an anti-sigma F factor [13], and the human STE20/SPS1-related proline-alanine-rich kinase, which performs as a mediator of stress-activated signals [14].

The detailed analysis of such enzymes sharing the same catalytic activity may actually lead to creating orphan enzymes *de novo*. For a while, EC 2.1.1.23 defined protein-arginine *N*-methyltransferase (PRMT) activity [15]. More recently, the curators at the International Enzyme Commission attempted to better define the precise function of these diverse PRMTs that play a crucial role in post-translational modification of proteins. The entry EC 2.1.1.23 has been deleted and replaced with three new entries, EC 2.1.1.124 defining [cytochrome c]-arginine *N*-methyltransferase, EC 2.1.1.125 defining histone-arginine *N*-methyltransferase and EC 2.1.1.126 defining [myelin basic protein]-arginine *N*-methyltransferase, respectively. The newly created EC 2.1.1.125 was attributed to the sequences previously assigned to EC 2.1.1.23. Therefore, this improvement in function definition led to the paradoxical production of two orphans, EC 2.1.1.124 and EC 2.1.1.126. Note, moreover, that Uniprot currently annotates PRMT either as EC 2.1.1.- or without an associated EC number. This underscores how partially qualified EC numbers are ambiguous and thus misleading, as recently pointed out [16].

Another intricate case is that of a unique molecule (corresponding to a unique domain) endowed with several EC numbers. Initially, three different activities were independently described, that of tartrate dehydrogenase (EC 1.1.1.93), tartrate decarboxylase (EC 4.1.1.73) and D-malate dehydrogenase [decarboxylating] (EC 1.1.1.83), respectively. For two of them, EC 1.1.1.83 and EC 4.1.1.73, for a while there was no available sequence in Uniprot. More recently, based on the findings described in [17], Uniprot curators have reassigned the two orphan enzyme activities to the molecule that was already annotated as endowed with the EC 1.1.1.93 activity. This is a striking case illustrating how difficult it can be to unmask the protein responsible for the experimentally observed catalytic activity.

On the contrary, the same catalytic mechanism may correspond to diverse EC numbers, depending on the nature of their coenzyme or other interacting compounds. For instance, as shown in Table 4, five different EC numbers have been qualified up to now to define glyceraldehyde 3-phosphate dehydrogenase (GAPDH), including one orphan. Furthermore, a protein may have one molecular function but several, unrelated cellular functions. Such proteins have been dubbed moonlighting proteins [18–20]. Besides its traditional (housekeeping) key role in energy production as a glycolytic enzyme, GAPDH plays a unexpected variety of cellular roles such as regulation of the cytoskeleton, membrane fusion and transport, glutamate accumulation into presynaptic vesicles, binding to low-molecular-weight G proteins [21 and references therein]. GAPDH seems also to participate in activation of transcription in neurons, exportation of nuclear RNA and DNA repair. Particularly intriguing are the observations that GAPDH would be an intracellular sensor of oxidative stress during early apoptosis and may participate in neuronal death in some neurodegenerative diseases [21 and references therein].

Improving annotation is essential

We already examined two interesting cases where improving annotation and definition of enzyme activity may lead either to specifying the molecule responsible for the orphan EC number (the case of the trifunctional tartrate dehydrogenase [17]) or to the artificial creation of new orphans (case of PRMT). We can further list supplementary cases of misannotation to illustrate other unwanted cases of creation of orphan EC numbers.

For example, there may be cases of 'trivial' annotation error. We noticed that there are an undetermined number of misapprehensions in public databases, including highly supervised databases such as Uniprot. We have already described a particular – and exemplary – case where several putrescine carbamoyltransferases were erroneously annotated as ornithine carbamoyltransferases [22].

Another instance of missed annotation is that of the recently discovered glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR, EC 1.2.7.6) that is reported in various enzyme databases (BioCyc, Brenda, Enzyme, KEGG) to have no available sequence (Table 4). This enzyme is thought to have a glycolytic role and to function in place of the regular GAPDH (EC 1.2.1.59) and possibly phosphoglycerate kinase in the novel Embden-Meyerhof-type glycolytic pathway found in *Pyrococcus furiosus* [23, 24]. However, three species of this euryarchaeal genus have been entirely sequenced, the sequence of GAPOR was identified as early as 1998 [24] and several homologues were later detected in closely related euryarchaeaota [see for instance, 25]. Yet, for an unknown reason, the GAPOR sequences were introduced in the Uniprot database without their cognate EC number (EC 1.2.7.6). It is highly probable that a significant number of

Table 4. The different glyceraldehyde 3-phosphate dehydrogenases.

EC number	Coenzyme	Type of activity	Short name	Number of sequences in Uniprot
EC 1.2.1.9	NADP+	non-phosphorylating activity	GAPN	36
EC 1.2.1.12	NAD+	phosphorylating activity	GAPDH	660
EC 1.2.1.13	NADP+	phosphorylating activity	GAPDH	39
EC 1.2.1.59	NAD(P)+	phosphorylating activity	GAPDH	46
EC 1.2.7.6	NADP+	ferredoxin	GAPOR	0

similar omissions currently pollute databases. As a test, we used a text-mining approach to check sets of newly created EC numbers for the last 3 years. We discovered that, in those years, Uniprot curators have introduced a significant amount of the sequences of the so-called orphan enzymes without their cognate EC numbers. The respective percentage values found are 36, 31 and 35 for the years 2003, 2004 and 2005, respectively [O. Lespinet and B. Labedan, unpublished]. Because of this unexpected Uniprot failure, the dedicated databases (e.g. BioCyc, Brenda, Enzyme and KEGG) perpetuate the dramatic misannotations that artificially inflate the proportion of orphan EC numbers. At any rate, the whole community must make a major effort to help the main public databases double-check annotations concerning enzyme features. If achieved, this could decrease by a fourth the relative proportion of sequenceless EC numbers.

Unmasking orphan EC is no easy task

There is an urgent need to explore the 1525 EC numbers without associated sequences, to progressively bridge the gap between function and sequence. Several approaches are presently feasible, as summarized below.

Text mining

First, we assume that a significant amount of information relating to these elusive molecules could be found by simply mining the literature and especially unpublished works (doctoral theses, colloquial papers, even old lab books) as already proposed [8, 22]. As previously underscored [7, 8], it seems to us crucial that the whole community help to identify genes that synthesize orphan enzymes. Indeed, many of the orphan enzymes appear to be neglected. Table 5 shows the distribution of the number of publications for all orphans. It may be revealing that among the 1525 EC numbers found as being sequenceless, 1119 have no recent citations when querying PubMed with the respective EC numbers. However, not all orphan enzymes are cast aside. Table 6 lists the top 10 most cited orphan enzymes in PubMed (October 2005). For instance, as many as 10 papers were published on the mechanism of action of phosphogluconate 2-dehydrogenase (EC 1.1.1.43) for the first half of the year 2005, despite the absence of any associated amino acid sequence. Another striking case of orphaning fashioned by a confusing annotation is that of the AcpD family. EC 3.1.4.14 defines an acyl carrier protein phosphodiesterase that is described as being encoded by *acpD* gene in *Escherichia coli* and many other bacteria. However, it was demonstrated 5 years ago that the *acpD* gene product is an azoreductase [26]. Strikingly, no sequence is associated with this last activity (EC 1.7.1.6). Very recently, the proper gene encoding the acyl carrier protein phosphodiesterase was experimentally characterized in *E. coli* as *acpH* [27].

Table 5. Distribution of citations of orphan enzymes in PubMed (October 2005).

Number of citations	EC Numbers
0	1119
1– 3	190
4–50	205
>50	36

Another occurrence of orphaning would be the case of enzyme activities that occur in organisms for which no genetic approaches are available. We observed such an occurrence for near half of the orphan EC numbers. This trend expanded considerably recently with an increase in the proportion of orphan EC numbers found in a unique species to a mean of 78% (79, 78 and 77 for 2003, 2004 and 2005, respectively). This may reflect the specific metabolic requirements of a peculiar lifestyle as in the case of the chlorophenol *O*-methyltransferase (EC 2.1.1.136), which is involved in the formation of chloroanisoles and in cork taint in wines. This enzyme has been characterized exclusively in the fungus *Trichoderma longibrachiatum* [28]. Such peculiarities combined with the absence of reliable genetic tools could explain the dif-

Table 6. The 10 most cited orphan enzymes (October 2005).

EC numbers	Enzyme name	Functional categories	PubMed entries
EC 1.1.1.43	Phosphogluconate 2-dehydrogenase	Pentose phosphate pathway, glutathione metabolism	2429
EC 2.3.2.12	peptidyltransferase	ribosome machinery	2054
EC 3.2.1.32	xylan endo-1,3-beta-xylosidase		681
EC 2.7.4.4	nucleoside-phosphate kinase	pyrimidine metabolism	609
EC 2.7.1.109	[hydroxymethylglutaryl-CoA reductase (NADPH)] kinase	n.a.	523
EC 1.3.1.22	cholestenone 5 alpha-reductase	n.a.	398
EC 3.1.6.9	chondro-4-sulfatase	n.a.	288
EC 3.1.3.53	myosin-light-chain-phosphatase	n.a.	288
EC 2.3.1.23	1-acylglycerophosphocholine *O*-acyltransferase	glycerophospholipid metabolism	259
EC 2.4.1.22	lactose synthase	galactose metabolism	231

ficulties of getting cognate sequences. Although several genomic projects are in progress for various *Trichoderma* species, one cannot be sure that these sequenced species, endowed with different metabolic properties, would encode such an unusual enzyme.

Peptide mass mapping and genome mining

However, similar cases appear less extreme. In the fungus *Schizophyllum commune*, tryptic peptide mass mapping disclosed a significant portion of the sequence of the trehalose phosphorylase (EC 2.4.1.231), opening the way to find close structural similarities with homologues in other basidiomycete fungi [29].

The availability of genomic sequences is crucial to going a step further. For instance, determining the N-terminal amino acid sequence of the *p*-hydroxyphenylacetate decarboxylase helped to identify the full sequence of the orphan enzyme by mining the unfinished genome of *Clostridium difficile* [30].

There is some hope that genome mining will be easier in the near future. Indeed, identifying the complete genomic sequence of any microorganism has long seemed to be mission impossible. The recent advent of new and fast sequencing methods [31] that are supposed to read entire genomes of fungi within 4 hours [32] now makes this task reachable.

Conclusions

During the last 60 years, biochemists and geneticists have been working on parallel tracks, using their specific experimental approaches in an effort to disclose the functional role(s) afforded by a specific gene or a peculiar protein. Such reductionist approaches have been very successful with the advent of molecular biology. Mixing genetic, biophysical and biochemical concepts and methodologies helped to disclose complex molecular mechanisms in very fine detail [33, 34]. In the last decade, however, more holistic approaches such as genomics, integrative biology and systems biology [35] have unveiled unexpected weaknesses of the previous work of biochemistry, genetics and molecular biology. We have underlined one of them in this paper. Clearly, the traditional approaches of biochemistry and genetics ignored many of the mechanisms of the living cell. It is now essential to bridge this gap separating function and sequence. A concerted effort would help to increase database precision while exploiting the huge amount of hidden knowledge they contain.

1 Vendruscolo M. and Dobson C. M. (2005) A glimpse at the organization of the protein universe. Proc. Natl. Acad. Sci. USA **102:** 5641–5642
2 Hou J. T., Jun S. R., Zhang C and Kim S. H. (2005) Global mapping of the protein structure space and application in structure-based inference of protein function. Proc. Natl. Acad. Sci. USA **102:** 3651–3656
3 Coulson A. F. W. and Moult J. (2002) A unifold, mesofold and superfold model of protein fold use. Proteins **46:** 61–71
4 Ashburner M., Ball C. A., Blake J. A., Botstein D., Butler H., Cherry J. M. et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. **25:** 25–29
5 Wilson G. A., Bertrand N., Patel Y., Hughes J. B., Feil E. J. and Field D. (2005) Orphans as taxonomically restricted and ecologically important genes. Microbiology **151:** 2499–2501
6 Serres M. H., Goswami S. and Riley M. (2004) GenProtEC: an updated and improved analysis of functions of Escherichia coli K-12 proteins. Nucleic Acids Res. **32:** D300–D302

7 Karp P. D. (2004) Call for an enzyme genomics initiative. Genome Biol. **5:** 401
8 Lespinet O and Labedan B. (2005) Orphan enzymes? Science **307:** 342
9 Bairoch A., Apweiler R., Wu C. H., Barker W. C., Boeckmann B., Ferro S. et al. (2005) The Universal Protein Resource (UniProt) Nucleic Acids Res. **33:** D154–D159
10 Kanehisa M. and Goto S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. **28:** 27–30
11 Krieger C. J., Zhang P., Mueller L. A., Wang A., Paley S., Arnaud M. et al. (2004) MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res. **32:** D438–D442
12 Kirk O., Borchert T. V. and Fuglsang C. C. (2002) Industrial enzyme applications. Curr. Opin. Biotechnol. **13:** 345–351
13 Min K. T., Hilditch C. M., Diederich B., Errington J. and Yudkin M. D. (1993) Sigma F., the first compartment-specific transcription factor of *B. subtilis*, is regulated by an anti-sigma factor that is also a protein kinase. Cell **74:** 735–742
14 Johnston A. M. Naselli G., Goneg L. J., Martin R. M., Harrison L. C. and DeAizpurua H. J. (2000) SPAK, a STE20/SPS1-related kinase that activates the p38 pathway. Oncogene **19:** 4290–4297
15 Bedford M. T. and Richard S. (2005) Arginine methylation, an emerging regulator of protein function. Mol. Cell **18:** 263–272
16 Green ML and Karp P. D. (2005) Genome annotation errors in pathway databases due to semantic ambiguity in partial EC numbers. Nucleic Acids Res. **33:** 4035–4039
17 Karsten W. E., Tipton P. A. and Cook P. F. (2002) Tartrate dehydrogenase catalyzes the stepwise oxidative decarboxylation of D-malate with both NAD and thio-NAD. Biochemistry. **41:** 12193–12199
18 Jeffery C. J. (1999) Moonlighting proteins. Trends Biochem. Sci. **24:** 8–11
19 Jeffery C. J. (2002) Multifunctional proteins: examples of gene sharing. Ann. Med. **35:** 28–35
20 Jeffery C. J. (2003) Moonlighting proteins: old proteins learning new tricks. Trends Genet. **19:** 415–417
21 Chuang D. M., Hough C. and Senatorov V. V. (2005) Glyceraldehyde-3-phosphate dehydrogenase, apoptosis and neurodegenerative diseases. Annu. Rev. Pharmacol. Toxicol. **45:** 269–290
22 Naumoff D. G., Xu Y., Glansdorff N. and Labedan B. (2004) Retrieving sequence of enzymes experimentally characterized but erroneously annotated: the case of putrescine carbamoyltransferase. BMC Genomics **5:** 52
23 Mukund S. and Adams M. W. (1995) Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon Pyrococcus furiosus. J. Biol. Chem. **270:** 8389–8392
24 van der Oost J., Schut G. J., Kengen S. W. M., Hagen W. R., Thomm M. and de Vos. W. M. (1998) The ferredoxin-dependent conversion of glyceraldehyde-3-phosphate in the hyperthermophilic archaeon *Pyrococcus furiosus* represents a novel site of glycolytic regulation. J. Biol. Chem. **273:** 28149–28154
25 Fukui T., Atomi H., Kanai T., Matsumi R., Fujiwara S. and Imanaka T. (2005) Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. Genome Res. **15:** 352–363
26 Nakanishi M., Yatome C., Ishida N. and Kitade Y. (2001) Putative ACP phosphodiesterase gene (*acpD*) encodes an azoreductase. J. Biol. Chem. **276:** 46394–46399
27 Thomas J. and Cronan J. E. (2005) The enigmatic acyl carrier protein phosphodiesterase of *Escherichia coli*: genetic and enzymological characterization. J. Biol. Chem. **280:** 34675–34683
28 Coque J. J., Alvarez-Rodriguez M. L. and Larriba G. (2003) Characterization of an inducible chlorophenol O-methyltransferase from *Trichoderma longibrachiatum* involved in the formation of chloroanisoles and determination of its role in cork taint of wines. Appl. Environ. Microbiol. **69:** 5089–5095
29 Eis C., Watkins M., Prohaska T. and Nidetzky B. (2001) Fungal trehalose phosphorylase: kinetic mechanism, pH-dependence of the reaction and some structural properties of the enzyme from *Schizophyllum commune.* Biochem. J. **356:** 757–767
30 Selmer T. and Andrei P. I. (2001) p-Hydroxyphenylacetate decarboxylase from Clostridium difficile: a novel glycyl radical enzyme catalysing the formation of p-cresol. Eur. J. Biochem. **268:** 1363–1372
31 Margulies M., Egholm M., Altman W. E., Attiya S., Bader J. S., Bemben L. A., et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:** 376–380
32 Rogers Y. H. and Venter J. C. (2005) Genomics: massively parallel sequencing. Nature **437:** 326–327
33 Morange M. (1998) A History of Molecular Biology, Harvard University Press, Cambridge, MA
34 Darden L. and Tabery J. (2005) Molecular biology. In: The Stanford Encyclopedia of Philosophy, Zalta E. N. (ed.), http://plato.stanford.edu/archives/spr2005/entries/molecular-biology/
35 Hood L. E. Galas D., Dewey G., Veres R. Wilson J. (2006) Biological Information and the Emergence of Systems Biology, Roberts, Denver, CO