ELSEVIER

Orphan enzymes could be an unexplored reservoir of new drug targets

Olivier Lespinet and Bernard Labedan

Institut de Génétique et Microbiologie, CNRS UMR 8621, Université Paris Sud, Bâtiment 400, 91405 Orsay Cedex, France

Despite the immense progress of genomics, and the current availability of several hundreds of thousands of amino acid sequences, >39% of well-defined enzyme activities (as represented by enzyme commission, EC, numbers) are not associated with any sequence. There is an urgent need to explore the 1525 orphan enzymes (enzymes having EC numbers without an associated sequence) to bridge the wide gap that separates knowledge of biochemical function and sequence information. Strikingly, orphan enzymes can even be found among enzymatic activities successfully used as drug targets. Here, knowledge of sequence would help to develop molecular-targeted therapies, suppressing many drug-related side-effects.

Biology is exploring numerous and diverse fields, each of which is very complex and difficult to study in its entirety. For many years, immense advances in disclosing molecular functions have been made using reductionist approaches, such as molecular biology. Combining genetic, biophysical and biochemical concepts and methodologies has helped to disclose details of complex molecular mechanisms such as DNA replication. However, it became clear that understanding how a living cell functions requires more detail than simply adding up the operations of each of the cellular constituents. Reductionist approaches are, therefore, inefficient methods for fully comprehending the relationships between all of the components of a cell and, to go a step further, we need a more integrative view of the basic mechanisms that are working together in a smooth, tuned process. Systems approaches, such as genomics, help to bring into place an integrative view of interactions between all genes and their products.

This trend of using holistic approaches (rather than reductionist ones) has unearthed unexpected weaknesses in previous biochemistry, genetics and molecular biology research. This has become obvious, for example as soon as the first chromosome to be sequenced, the yeast chromosome III, was analyzed and annotated [1]. It came as a big surprise that the complete sequence contained so many open reading frames (ORFs) that had not been discovered before, despite intensive research by thousands of people studying the genetics and biochemistry of *Saccharomyces cerevisiae* over the past 50 years. This observation has been confirmed repeatedly, with the cohort of genomes that have been sequenced, at a steady pace, over the past ten years. We now know that many genes have been missed by reductionist approaches and we do not have any information about the actual role of some of these genes in cell metabolism, leading to the concept of orphan (sometimes dubbed ORFan) genes. The OrphanMine database has been built to allow the study of taxonomically restricted microbial genes that have no homologues in any organism (www.genomics.ceh.ac.uk/orphan_mine/orphanmine.php). The most recent release of OrphanMine has compiled 44,752 orphan genes from 150 analyzed genomes (430,826 predicted proteins). This means that, on average, ~10% of the total number of genes per genome are orphan genes, and this percentage still holds true now that the complete sequences of ~300 genomes are available in public databanks [2].

It appears clear that the traditional approaches of biochemistry and genetics failed to gain access to large parts of living-cell mechanisms. In this short review, we describe another important flaw that appears because of the absence of interplay between biochemistry and genetic advances. We discuss another kind of orphan, orphan enzymes, and underline their relevance for applied research, for example in pharmacology, medicine and agronomy.

Corresponding author: Labedan, B. (bernard.labedan@igmors.u-psud.fr)

1359-6446/06/$ - see front matter doi:10.1016/j.drudis.2006.02.002

Defining enzyme activities and their putative sensitivity to drugs

Enzymatic processes have been analyzed in various organisms, by biochemists. These studies span a wide catalogue of data that have been gathered over several decades – from the mere description of a newly detected activity to studies that disclose many features of an enzyme in thorough detail. As early as the 1960s, an international effort was launched to organize all the pieces of information and to check its validity and consistency. Accordingly, the International Enzyme Commission (EC) classifies enzyme activities using EC numbers (www.chem.qmul.ac.uk/iubmb/enzyme/index.html). EC numbers comprise four digits. The first digit (from numbers 1–6) delineates the broad type of activity: (1) oxidoreductase, (2) transferase, (3) hydrolase, (4) lyase, (5) isomerase, and (6) ligase. The second and third digits detail the reaction that an enzyme catalyzes. For example, among the transferases (EC 2), glycosyltransferases form the subclass EC 2.4 and hexosyltransferases form the sub-subclass EC 2.4.1. The final digit represents the substrate specificity. For instance, EC 2.4.1.80 is the ceramide glucosyltransferase that catalyzes the first step of glycosphingolipid metabolism, and is the target of miglustat (1,5-(butylimino)-1,5-dideoxy-d-glucitol), a drug used for the treatment of the mild–moderate type 1 Gaucher disease in humans [3].

Introducing the concept of orphan enzymes

The EC classification of enzyme activities is constantly evolving because the curators of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology are deciding on the best definition for each enzymatic activity (in relation to their specific expertise). Note that, an EC number denotes an enzyme activity but not the enzyme itself that catalyzes this activity. Thus, different sequences and structures can share the same EC number (see later).

Currently (September 2005), 4499 EC numbers are indexed but only 3877 correspond to a defined unambiguous activity. The 622 others are either in the process of deletion or have been found to be identical to (or included with) another EC number.

It came as a big surprise when Karp *et al.* [4] and Lespinet and Labedan [5] independently observed that a significant proportion of EC numbers was not associated with any protein sequence. An update of our results confirms this finding. At present, only 2352 EC numbers have at least one associated sequence in the UniProt Knowledgebase (Release 5.8 30th August 2005). We have assigned orphan enzymes the 1525 EC numbers found to be sequence-less in Uniprot and in public databases specializing in EC numbers (see Box 1). Remarkably, these orphan enzymes currently represent 39.3% of the total of the retained EC numbers – a figure close to one that we reported in 2005 (42.5%) – despite the description, during this past year, of 169 new enzyme activities and the re-annotation of many newly sequenced genes.

BOX 1

Searching information about EC numbers in specialized databases

Uniprot (www.expasy.uniprot.org/index.shtml)
ENZYME (www.expasy.org/enzyme)
NC-IUBMB (www.chem.qmul.ac.uk/iubmb/enzyme)
BIOCYC (www.biocyc.org)
KEGG (www.genome.ad.jp/kegg)
BRENDA (www.brenda.uni-koeln.de)
BBD (http://umbbd.ahc.umn.edu/index.html)
EC→PDB (www.ebi.ac.uk/thornton-srv/databases/enzymes)

TABLE 1

Listing orphan EC numbers in the androgen and estrogen metabolic pathway

EC number	Activity
EC_1.1.1.63	Testosterone 17β-dehydrogenase
EC_1.1.1.64	Testosterone 17β-dehydrogenase (NADP+)
EC_1.1.1.148	Estradiol 17α-dehydrogenase
EC_1.1.1.152	3 α-Hydroxy-5 β-androstane-17-one 3 α-dehydrogenase
EC_1.1.1.239	3 α (17b)-Hydroxysteroid dehydrogenase (NAD+)
EC_1.14.99.11	Estradiol 6 β-monooxygenase
EC_1.14.99.12	4-Androstene-3,17-dione monooxygenase
EC_4.1.2.30	17α-Hydroxyprogesterone aldolase

Interestingly, many orphan enzymes could be drug targets and identifying their sequences would be a first step for disclosing new molecular therapies. For instance, no sequences are associated with the two testosterone 17-β-dehydrogenases (EC 1.1.1.63 and EC 1.1.1.64), despite the fact that these enzymatic activities have been studied intensively over the years. A deficiency in these dehydrogenases leads to male pseudohermaphroditism with post-pubertal virilization and, sometimes, gynecomastia [6]. It has been proposed that these enzymes should be used as targets for the treatment of prostate cancer, because they are competitively inhibited by anticancer drugs [7,8].

These two orphan enzymes belong to the androgen and estrogen metabolic pathway comprising 23 EC numbers that correspond to well-characterized enzyme activities. Interestingly, besides EC 1.1.1.63 and EC 1.1.1.64, this pathway contains six more orphan enzymes (Table 1). Considering the importance of hormonal regulations in many aspects of human health, some of these orphan enzymes could be potential drug targets. Indeed, knowing their sequences could help in designing drugs that are more efficient.

Are orphan enzymes currently used as drug targets?

In a recent review, Robertson [9] listed 71 EC numbers for enzymes that are targets for marketed drugs. These enzyme activities are present in humans (48), bacteria (13), viruses (5), fungi (4) and protists (1).

Unexpectedly, we found that there were orphans even among enzymes used as drug targets. Three of the 71 reported activities inhibited by drugs have no associated sequences. Here is the detailed list of these orphan EC numbers used as targets for pharmaceutical drugs:

- The vitamin-K-epoxide reductase EC 1.1.4.1 is involved in the biosynthesis of steroids. In humans, the protein responsible for this activity is known to be sensitive to coumarin anticoagulants (such as dicumarol and warfarin), which help to combat thrombosis. For a recent review, see [10].
- The xanthine oxidase EC 1.17.3.2 is involved in the final steps of the breakdown pathway of purines, this leads to the

production of uric acid. In humans, high levels of uric acid (hyperuricaemia) trigger chronic gout. Allopurinol, a structural analogue of hypoxanthine and, thus, a specific inhibitor of xanthine oxidase (mainly through direct substrate competition), is a drug routinely used to prevent the overproduction of uric acid [11].

- The dolichyl-phosphatase EC 3.1.3.51 is involved in N-glycan biosynthesis and seems to be the target of the antibiotic bacitracin, produced by *Bacillus subtilis*. Topical therapy with bacitracin is currently widely used in dermatology and the treatment of bacterial conjunctivitis. However, the molecular mode of action of dolichyl-phosphatase is not well-characterized. Actually, bacitracin is better-known as being a nonspecific protease inhibitor [12].

It is highly conceivable that knowing the sequences of these enzyme activities could help to define the molecular interaction between drugs and these target enzymes more accurately, and to improve their efficiency.

Looking for orphan enzymes in metabolic pathways that could be used as new drug targets

A recent review [13] proposed the comparison of the metabolic pathways in protozoan parasites with those of their human hosts, to facilitate the identification of new drug targets against major diseases such as malaria, leishmaniasis, Chagas disease and trypanosomiasis. Indeed, human cells synthesize only ten amino acids (Y, A, S, G, E, Q, P, D, N and C) and many of the pathways involved are absent in protozoan parasites. To cope with these auxotrophies, the parasites developed alternative pathways. Such parasite-specific pathways provide attractive targets for drug development where no existing therapy is available. For instance, the genomes of two species of *Cryptosporidium*, responsible for acute gastroenteritis and diarrhoea, have been sequenced recently [14,15]. *Cryptosporidium* is auxotrophic for Y, A, S, E, D, N and C. We find that several of those alternative pathogen's pathways are rich in orphan EC numbers, the respective figures being 1, 7, 3, 1, 7, 16 and 5 orphans for each amino acid metabolic pathway. For instance, among the five orphans of the cysteine–cystine pathway, three enzyme activities are involved in glutathione metabolism: EC 1.8.1.10 (CoA-glutathione reductase); EC 1.8.4.3 (glutathione-CoA-glutathione transhydrogenase); and EC 1.8.4.4 (glutathione-cystine transhydrogenase). The two other orphans in this cysteine–cystine pathway are EC 2.6.1.3 (cysteine transaminase) and EC 4.4.1.10 (cysteine lyase).

The same kind of methodological approach using, as targets, enzymes that are involved in metabolic pathways in parasites (but not in the host) has been suggested for sulfonucleotide reductases. This diverse family of enzymes catalyzes the first committed step of reductive sulfur assimilation [16]. Bacteria and plants use the sulfite generated in this reaction for producing cysteine and coenzyme A (CoA). Humans do not possess a homologous metabolic pathway and, thus, these sulfonucleotide reductases represent attractive targets for therapeutic intervention. As indicated in Tables 2 and 3, sulfur metabolism is rich in orphan enzymes that are present in bacteria but not in humans. In particular, there are three sulfotransferases, EC 2.8.2.3, EC 2.8.2.5 and EC 2.8.2.6 (Table 2).

TABLE 2

Listing orphan EC numbers in sulfur metabolism

EC number	Activity
EC_1.13.11.18	Sulfur dioxygenase
EC_2.7.7.5	Sulfate adenylyltransferase (ADP)
EC_2.8.1.5	Thiosulfate-dithiol sulfurtransferase
EC_2.8.2.3	Amine sulfotransferase
EC_2.8.2.5	Chondroitin 4-sulfotransferase
EC_2.8.2.6	Choline sulfotransferase
EC_3.6.2.1	Adenylylsulfatase
EC_3.6.2.2	Phosphoadenylylsulfatase
EC_3.12.1.1	Trithionate hydrolase

More generally, many metabolic pathways appear to contain a significant number of orphan EC numbers, as summarized in Table 3 for the ten major pathway classes. Remarkably, only 24 out of 124 pathways are devoid of orphan EC numbers. An example of such a completely defined pathway is the 14 enzymes involved in pathway 1 of fatty acid biosynthesis [17]. Note that two of the enzymes in this pathway are used as drug targets for tuberculosis treatment: the enoyl-[acyl-carrier-protein]-reductase (NADH, EC 1.3.1.9), which is sensitive to isoniazid, and the fatty-acid synthase EC 2.3.1.85, which is sensitive to pyrazinamide [18].

The 100 pathways that contain orphan enzymes are a huge reservoir of potential drug targets. For example, Biosynthesis of Secondary Metabolites contains 73 orphans out of 195 EC numbers and Biodegradation of Xenobiotics contains 55 orphans from 201 EC numbers.

Looking for orphan enzymes, from nonmetabolic pathways, that could be used as new drug targets

Table 3 describes a large panel of orphan EC numbers representing enzymes, involved in cell metabolism, with the potential to be putative drug targets. However, the relative proportion of EC numbers without sequence goes from 27.3% (among the enzymes involved in metabolism) to 52.1% (in the case of nonmetabolic enzymes). It might be interesting to look at these nonmetabolic enzymes despite the fact that they represent less than one-third of the total population of defined enzymes. A few examples, that highlight orphan nonmetabolic enzymes that could be attractive for drug development, are described in this review.

It has been proposed that researchers attempt to combat malarial parasites using specific proteases as potential targets [19]. The so-called histoaspartic protease (HAP)1 is one of several proteases involved in the hemoglobin degradation pathway, which is essential for the survival and reproduction of *Plasmodium falciparum*. Because the mechanism of action of this protease is unknown, computational tools have been used for modelling substrate-binding events, opening the way to drug-targeting against malaria [20]. To date, 37 aspartic endopeptidases (EC 3.4.23) can be found in the databases and six of them are identified as orphans (Table 4), including four proteinases (rhodotorulapepsin, acrocylindropepsin, pycnoporopepsin and scytalidopepsin A) made by pathogenic fungi. Note that two other proteinases are

TABLE 3

Relative percentage in main classes of metabolic pathways

Pathway class	Total EC numbers	Total orphans	Percentage of 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 from 11	Folate
Carbohydrate	669	156	23.3	C5-Branched dibasic acid metabolism (61.9)	1 from 17	Inositol
Amino acid	755	169	22.4	Lysine degradation (44.4)	2 from 25	Urea cycle
Lipid	241	51	21.2	Bile acid biosynthesis (47.8)	2 from 10	Fatty acids
Nucleotide	156	28	17.9	None	0 from 2	Not given
Glycan	153	26	17.0	Chondroitin–heparan sulfate biosynthesis (33.3)	6 from 14	LPS, peptidoglycan
Energy	156	21	13.4	Sulfur metabolism (31.0)	4 from 8	Oxidative phosphorylation
Secondary metabolites	195	73	37.4	Alkaloid biosynthesis I (51.4)	3 from 14	Penicillins and cephalosporins biosynthesis
Biodegradation of xenobiotics	200	55	27.4	2,4-Dichlorobenzoate degradation (59.1)	2 from 18	1,2-Dichloroethane degradation
Polyketides and nonribosomal peptides	13	2	15.4	Biosynthesis of 12-, 14- and 16-membered macrolides (50.0)	3 from 5	Biosynthesis of siderophore group nonribosomal peptides
Total	**2769**	**635**	**22.9**		**24 from 124**	

already used to combat HIV infection. Indeed, HIV-1 retropepsin (EC 3.4.23.16) and HIV-2 retropepsin (EC 3.4.23.47) are targets of nelfinavir.

A recent paper demonstrated that the amino-6-boronohexanoic acid (ABH), an analogue of its boronic acid substrate, is a highly potent inhibitor of human arginase I [21]. This enzyme plays a significant role in modulating the immune response and preliminary experiments demonstrate that ABH attenuates proliferation in an estrogen-receptor-negative human breast cancer cell line [22]. Thus, arginase I is explored as a potential target for new therapies aiming at inflammatory and immunological disorders. Note, moreover, that boronic-acid-based inhibitors establish a niche in the broader family of enzyme-targeted drugs [9] because they were also found to potentiate the activity of β-lactam antibiotics [23]. We observe that the sequence of l-arginase (EC 3.5.3.1) has been determined in 63 organisms, but no sequence is available for d-arginase (EC 3.5.3.10), including in humans. It could be interesting to search for a role and for a sequence for human d-arginase. Note, furthermore, that the d-arginine–d-ornithine metabolic pathway is abundant in orphan EC numbers (four out of eight enzymes).

The mechanism of action of agrocin 84, a natural drug encoded by the *agnB2* plasmid gene in *Agrobacterium radiobacter* that is important for controlling plant tumours, has been recently illuminated [24]. The structure of the toxic moiety of agrocin 84 is similar to that of leucyl-adenylate (Leu-AMP), a crucial enzyme-bound reaction intermediate. Plausibly, the stable toxic moiety of agrocin 84 could impart its antibiotic effect by binding to the catalytic domain of the *Agrobacterium tumefaciens* genomic-encoded LeuRS as a Leu-AMP mimic. The discovery of this remarkable process encourages the identification of analogous cases. Accordingly, it seems worthwhile to notice that the glutamine-hydrolyzing asparaginyl-tRNA synthase (EC 6.3.5.6), discovered previously as an unusual tRNA synthase in *Deinococcus radiodurans* and other organisms lacking a specific enzyme for asparagine synthesis [25], is an orphan EC number. Such an enzyme could be a good target for new antibiotics.

The difficulties in targeting orphan enzymes

One might wonder why there are still so many orphan enzymes, especially because the genomes of many species have been completely sequenced. As well as the difficulty that is inherent to annotation problems, we would like to emphasize other points that complicate the process of targeting enzyme macromolecules to specific therapeutic drugs. Some of these hindrances are conceptual; others are relevant to applied research.

TABLE 4

Listing orphan EC numbers among aspartic endopeptidases

EC number	Activity	Specificity
EC_3.4.23.12	Nepenthesin	Carnivorous plants
EC_3.4.23.17	Pro-opiomelanocortin converting enzyme	Bovine pituitary secretory vesicle
EC_3.4.23.26	Rhodotorulapepsin	Imperfect fungi
EC_3.4.23.28	Acrocylindropepsin	Imperfect fungi
EC_3.4.23.30	Pycnoporopepsin	Fungi
EC_3.4.23.31	Scytalidopepsin A	Imperfect fungi

First, it is very important to understand that an EC number represents an enzyme activity, and not the enzyme itself that catalyzes this activity. Indeed, proteins that differ in sequence and/or structure [26,27] can catalyze the same enzymatic reaction and, therefore, have the same EC number. For instance, EC 2.7.1.37 defines the protein kinase activity. Protein kinases are important drug targets, as, for example, recently outlined in the case of parasitic protozoa [28]. There are currently 1232 sequences that are classified as protein kinases, corresponding to a wide array of proteins present in every organism from bacteria to humans. However, these protein kinases could be as different as the *B. subtilis* SpoIIAB protein (www.expasy.org/uniprot/P10728), which acts as an anti-sigma F factor [29], and the human STE20/SPS1-related proline-alanine-rich protein (www.expasy.org/uniprot/Q9UEW8), which acts as a mediator of stress-activated signals [30].

Second, the same catalytic activity could correspond to different EC numbers depending on the coenzyme or other interacting compounds. For instance, five different glyceraldehyde 3-phosphate dehydrogenases (GAPDHs) have been described (Table 5). Furthermore – and this is of a great interest – a protein can have one molecular function but several, unrelated cellular functions. Such proteins have been described as moonlighting proteins [31–33]. As well as its traditional (housekeeping) key role in energy production as a glycolytic enzyme, GAPDH plays a variety of additional roles, such as regulation of the cytoskeleton, membrane fusion and transport, glutamate accumulation into presynaptic vesicles and binding to low-molecular-weight G proteins [34]. GAPDH also seems to participate in the activation of transcription in neurons, exportation of nuclear RNA and DNA repair. Particularly intriguing are the observations that GAPDH might be an intracellular sensor of oxidative stress during early apoptosis and might participate in neuronal death in some neurodegenerative diseases [34]. Such a variety of roles for one unique EC number could explain the difficulty in identifying the correct amino acid sequence when studies have been limited to only a few, badly defined, physiological events.

Third, there is too much uncertainty in public databases, even highly supervised ones. For instance, in databases, the most recently discovered glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR, EC 1.2.7.6) is reported to have no sequence. This enzyme (Table 5) is thought to have a glycolytic role and to function in place of GAPDH (EC 1.2.1.59), and possibly phosphoglycerate kinase, in the novel Embden–Meyerhof-type glycolytic pathway seen in *Pyrococcus furiosus* [35,36]. However, three species of this archaeal genus have been entirely sequenced, the sequence of GAPOR was identified as early as 1998 [35] and several homologues of GAPOR were later detected in various archaea [37]. Nevertheless, for unknown reasons, the GAPOR sequences that have been entered in databases have never been equated with their cognate EC number. It is highly probable that a significant number of similar omissions are currently polluting databases. For instance, we found, by text mining (Lespinet and Labedan, unpublished results), that one-third of the so-called orphan enzyme sequences (added in 2005) were introduced into Uniprot without their cognate EC numbers. Because of this unexpected failure, the dedicated databases (Box 1) carry on the dramatic misannotations that artificially inflate the proportion of orphan EC numbers. At any rate, the whole community has to make a major effort to help public databases to improve the annotations regarding enzyme features.

TABLE 5

The different glyceraldehyde 3-phosphate dehydrogenases

EC number	Coenzyme	Type of activity	Short name
EC_1.2.1.9	NADP+	Non-phosphorylating activity	GAPN
EC_1.2.1.12	NAD+	Phosphorylating activity	GAPDH
EC_1.2.1.13	NADP+	Phosphorylating activity	GAPDH
EC_1.2.1.59	NAD(P)+	Phosphorylating activity	GAPDH
EC_1.2.7.6	NADP+	Ferredoxin	GAPOR

For the rest of the orphans, we expect that a significant amount of information relating to these elusive molecules could be found by simply mining the literature and, additionally, unpublished works (e.g. doctoral theses, congress papers, laboratory books), as already demonstrated in a specific and exemplary case [38]. This would really help people interested in drug targeting to perform automated searches of databases using data mining tools.

Conclusions

Most diseases are of a complex nature and, over the years, it became more and more evident that reductionist approaches were insufficient tools to deliver a complete understanding. Previous reductionist approaches have suffered from drawbacks, such as focusing on the putative drug target without completely understanding its role in the pathophysiology of the disease. As we previously stated, this was evident in the case of moonlighting proteins [31–33]. Recent studies have shown that holistic, systems-biology approaches are more appropriate for understanding the complexity of most animal models [39,40].

It was surprising to discover that the amino acid sequences of three of the enzymes currently used as pharmaceutical drugs are still unknown. This suggests that it is possible to combat a disease without any knowledge of the molecular properties of the drug target. However, even if such a situation was justified because of the urgent need of medical treatment, it is not satisfying. For instance, drugs developed using past paradigms attack cancerous and healthy cells, often causing devastating short- and long-term side effects. Moreover, individual patient responses to drugs vary, even in cases where the type of cancer the patients suffer from appears to be the same. Molecular-targeted therapies hold the promise of being more highly selective but it is necessary to bridge the current gap between the gene sequence and the classical biochemistry (specifically enzymology) of a product.

Genome sequencing projects are producing vast amounts of information in the form of annotated genes. The functional characterization is progressing at an increasing rate for many organisms, which could help to identify sequences of orphan enzymes. For instance, the sequence of the *p*-hydroxyphenylacetate decarboxylase (EC 4.1.1.83) from *Clostridium difficile* has been identified by a two-step procedure; *C. difficile* causes gastrointestinal infections in humans and is very common in hospitalized patients. The determination of the N-terminal amino acid sequence of this protein (by tryptic peptide mass mapping) helped in identifying the full ORF (by mining the unfinished genome) [41].

There is some hope that genome mining will be easier in the near future – with the advent of new and very fast sequencing methods [42,43]. It is also possible that resolving more sequences of orphan enzymes could, at the same time, shine a light on the nature of a few other ORFans [1,2] if, by chance, the sequences are matching up.

References

1 Oliver, S.G. *et al.* (1992) The complete DNA sequence of yeast chromosome III. *Nature* 357, 38–46
2 Wilson, G. *et al.* (2005) Orphans as taxonomically restricted and ecologically important genes. *Microbiology* 151, 2499–2501
3 Platt, F.M. *et al.* (1994) N-Butyldeoxynojirimycin is a novel inhibitor of glycolipid biosynthesis. *J. Biol. Chem.* 269, 8362–8365
4 Karp, P.D. (2004) Call for an enzyme genomics initiative. *Genome Biol.* 5, 401
5 Lespinet, O. and Labedan, B. (2005) Orphan enzymes? *Science* 307, 42
6 Wilson, J.D. (2001) Androgens, androgen receptors, and male gender role behavior. *Horm. Behav.* 40, 358–366
7 Lombardo, M.E. *et al.* (1993) *In vitro* studies on the inhibition of testosterone synthesis in the human testis by atamestane. *J. Steroid Biochem. Mol. Biol.* 44, 287–290
8 Le Lain, R. *et al.* (2001) Inhibitors of human and rat testes microsomal 17beta-hydroxysteroid dehydrogenase (17beta-HSD) as potential agents for prostatic cancer. *J. Enzyme Inhib.* 16, 35–45
9 Robertson, J.G. (2005) Mechanistic basis of enzyme-targeted drugs. *Biochemistry* 44, 5561–5571
10 Kresge, N. *et al.* (2005) Hemorrhagic sweet clover disease, dicumarol, and warfarin: the work of Karl Paul Link. *J. Biol. Chem.* 280, e5
11 Pittman, J.R. and Bross, M.H. (1999) Diagnosis and management of gout. *Am Fam Physician* 59, 1799–1806
12 Vitković, L. and Sadoff, H.L. (1977) Purification of the extracellular protease of *Bacillus licheniformis* and its inhibition by bacitracin. *J. Bacteriol.* 131, 891–896
13 Chaudhary, K. and Roos, D.S. (2005) Protozoan genomics for drug discovery. *Nat. Biotechnol.* 23, 1089–1091
14 Abrahamsen, M.S. *et al.* (2004) Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 304, 441–445
15 Xu, P. *et al.* (2004) The genome of *Cryptosporidium hominis*. *Nature* 431, 1107–1112
16 Carroll, K.S. *et al.* (2005) A conserved mechanism for sulfonucleotide reduction. *PLoS Biol.* 3, e250
17 Wakil, S.J. *et al.* (1983) Fatty acid synthesis and its regulation. *Annu. Rev. Biochem.* 52, 537–579
18 McNeill, L. *et al.* (2003) Pyrazinamide and rifampin vs. isoniazid for the treatment of latent tuberculosis. *Chest* 123, 102–106
19 Rosenthal, P.J. (1998) Proteases of malaria parasites: new targets for chemotherapy. *Emerg. Infect. Dis.* 4, 49–57
20 Bjelic, S. and Aqvist, J. (2004) Computational prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site. *Biochemistry* 43, 14521–14528
21 Di Costanzo, L. *et al.* (2005) Crystal structure of human arginase I at 1.29-A resolution and exploration of inhibition in the immune response. *Proc. Natl. Acad. Sci. U. S. A.* 102, 13058–13063
22 Metcalf, B.W. (1998) The impact of genomics on medicinal chemistry. *Pure Appl. Chem.* 70, 359–363
23 Weston, G.S. *et al.* (1998) Structure-based enhancement of boronic acid-based inhibitors of AmpC beta-lactamase. *J. Med. Chem.* 41, 4577–4586
24 Reader, J.S. *et al.* (2005) Major biocontrol of plant tumors targets tRNA synthetase. *Science* 309, 1533
25 Min, B. *et al.* (2002) Transfer RNA-dependent amino acid biosynthesis: an essential route to asparagine formation. *Proc. Natl. Acad. Sci. U. S. A.* 99, 2678–2683
26 Hegyi, H. and Gerstein, M. (1999) The relationship between protein structure and function: A comprehensive survey with application to the yeast genome. *J. Mol. Biol.* 288, 147–164
27 Todd, A.E. *et al.* (2001) Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.* 307, 1113–1143
28 Doerig, C. *et al.* (2002) Protein kinases as drug targets in parasitic protozoa. *Trends Parasitol.* 18, 366–371
29 Min, K.T. *et al.* (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
30 Johnston, A.M. *et al.* (2000) SPAK, a STE20/SPS1-related kinase that activates the p38 pathway. *Oncogene* 19, 4290–4297
31 Jeffery, C.J. (1999) Moonlighting proteins. *Trends Biochem. Sci.* 24, 8–11
32 Jeffery, C.J. (2003) Multifunctional proteins: examples of gene sharing. *Ann. Med.* 35, 28–35
33 Jeffery, C.J. (2003) Moonlighting proteins: old proteins learning new tricks. *Trends Genet.* 19, 415–417
34 Chuang, D.M. *et al.* (2005) Glyceraldehyde-3-phosphate dehydrogenase, apoptosis, and neurodegenerative diseases. *Annu. Rev. Pharmacol. Toxicol.* 45, 269–290
35 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
36 van der Oost, J. *et al.* (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
37 Fukui, T. *et al.* (2005) Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res.* 15, 352–363
38 Naumoff, D.G. *et al.* (2004) Retrieving sequence of enzymes experimentally characterized but erroneously annotated: the case of putrescine carbamoyltransferase. *BMC Genomics* 5, 52
39 Hood, L. *et al.* (2004) Systems biology and new technologies enable predictive and preventative medicine. *Science* 306, 640–643
40 Hood, L.E. *et al.* (2006) *Biological information and the emergence of systems biology*. Roberts and Company Publishers, Denver, Colorado, U. S. A
41 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
42 Margulies, M. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380
43 Rogers, Y.H. and Venter, J.C. (2005) Genomics: massively parallel sequencing. *Nature* 437, 326–327