THE ROYAL SOCIETY

Published online 4 January 2002

Metabolic pathway analysis in trypanosomes and malaria parasites

Alan H. Fairlamb

Division of Biological Chemistry and Molecular Microbiology, The Wellcome Trust Biocentre, University of Dundee, Dundee DD1 5EH, UK (*a.h.fairlamb@dundee.ac.uk*)

Identification of novel drug targets is required for the development of new classes of drugs to overcome drug resistance and replace less efficacious treatments. In theory, knowledge of the entire genome of a pathogen identifies every potential drug target in any given microbe. In practice, the sheer complexity and the inadequate or inaccurate annotation of genomic information makes target identification and selection somewhat more difficult. Analysis of metabolic pathways provides a useful conceptual framework for the identification of potential drug targets and also for improving our understanding of microbial responses to nutritional, chemical and other environmental stresses. A number of metabolic databases are available as tools for such analyses. The strengths and weaknesses of this approach are discussed.

Keywords: metabolic pathways; drug discovery; bioinformatics; metabolomics; *Mycobacterium tuberculosis*

1. INTRODUCTION

Viral, bacterial and parasitic infections continue to be one of the major obstacles to improving human health in developing countries. Millions die each year from HIV–AIDS, tuberculosis and protozoal diseases such as malaria, sleeping sickness and leishmaniasis (World Health Organization 2000). Better tools, including new drugs, vaccines and diagnostics, are urgently required to control these 'Great Neglected Diseases'. The therapeutic value of many of the existing drugs has been seriously eroded due to the widespread emergence of drug resistance. Due to chronic under-investment in drug discovery over the last two decades, there are perilously few replacements in the development pipeline. An analysis of pharmaceutical activity over the last 25 years revealed that only a handful ($< 1\%$) of new drugs were marketed for tropical diseases. Of the dozen or so that did reach the clinic, more than half had been initially developed with specific veterinary or cancer therapies in mind, rather than the treatment of tropical diseases (Trouiller & Olliaro 1999). Although the feasibility of experimental immunization of humans against malaria has been demonstrated (Clyde *et al.* 1973), despite nearly 30 years of intense research, there is still no malaria vaccine. The prospects of developing effective vaccines against trypanosomes responsible for African sleeping sickness and Chagas disease are even more remote.

Genome sequencing has led to fresh optimism about the future prospects for the control of infectious diseases. In a commentary in the journal *Nature* in 1995, Barry Bloom predicted: 'Sequencing bacterial and parasite pathogens…could buy the sequence of every virulence determinant, every protein antigen and every drug target…for all time' (Bloom 1995). The complete genome sequence of *Mycobacterium tuberculosis* was published in 1998 (Cole *et al.* 1998), and genome sequences of *Plasmodium falciparum* (malaria), *Trypanosoma brucei* (sleeping sickness), *Trypanosoma cruzi* (Chagas disease) and *Leishmania major* (cutaneous leishmaniasis) are imminent. Thus, the challenge for the post-genomic era is to exploit this wealth of information to gain deeper insights into the biology of these organisms with the ultimate goal of developing new therapeutic tools for the control of these devastating diseases.

2. THE METABOLOME

Just as the full complement of genes within an organism has been defined as the genome, the full cellular complement of RNA molecules, proteins and metabolites (including complex macromolecules) have been termed transcriptome, proteome and metabolome, respectively. Unlike the DNA sequence of a genome, expression of the latter molecules varies in cells, tissues and organs dependent on the temporal, developmental and environmental context of the organism (Oliver 2000; see also Oliver 2002).

Metabolism is one of the better-documented core biological processes in cells. Small organic compounds (metabolites) are sequentially interconverted by protein catalysts (enzymes) where each enzyme has a defined molecular function according to the particular chemical reaction they catalyse. Most enzymes, but not all, have been assigned a specific EC number. Likewise thousands of metabolic intermediates (the substrates and products of each enzyme catalysed reaction) have a unique compound registration number. A list of EC numbers and compound numbers, *per se,* is not particularly informative to the biologist. However, assembly of sequential reactions into metabolic pathways and integration of metabolic pathways into metabolic networks represents a convenient and useful method of defining specific processes at the biological level. To take *Escherichia coli* as an example, there are about 4390 genes in the genome, of which 97% are pre-

DOI 10.1098/rstb.2001.1040
© 2002 The Royal Society

Table 1. Metabolic pathway databases available on-line.

acronym	database name	URL
KEGG	Kyoto Encyclopaedia of Genes and Genomes	http://www.genome.ad.jp/kegg
EcoCyc/MetaCyc	Encyclopaedia of *Escherichia coli* Genes and Metabolism	http://ecocyc.pangeasystems.com
EMP	Enzymes and Metabolic Pathways Project	http://emp.mcs.anl.gov
WIT	What is there?	http://wit.mcs.anl.gov/WIT2
ExPASy (SWISS-PROT)	Boehringer Mannheim Metabolic Pathways Chart	http://www.expasy.ch/cgi-bin/search-biochem-index
aMAZE	Protein Function and Biochemical Pathways	http://www.ebi.ac.uk/research/pfbp
	Malaria Parasite Metabolic Pathways	http://sites.huji.ac.il/malaria/

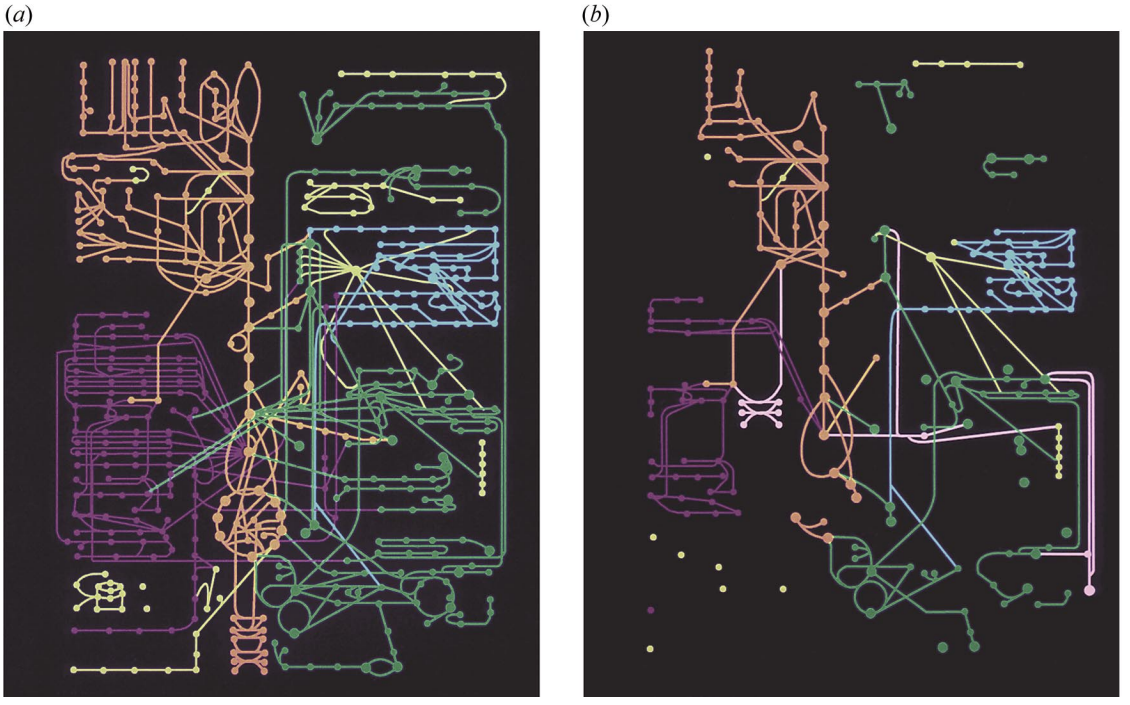

Figure 1. Global comparison of metabolomes: (*a*) a typical mammalian cell; (*b*) bloodstream stage of the African trypanosome, *T. brucei*. Pathways are coloured as follows: orange for carbohydrate metabolism and electron transport; purple for fatty acid and sterol metabolism; green for amino acids and related metabolites; light blue for purine, pyrimidine and nucleic acid metabolism; yellow for co-enzyme biosynthesis. In (*b*), pathways not found in mammals are shown in pink (cytochrome-independent electron transport via the glycerophosphate oxidase system; synthesis of acetyl-CoA from threonine; trypanothione biosynthesis). Based on Fairlamb (1989).

dicted to encode proteins. Of these, 21% (905) have been assigned an enzymatic function and 4% (162) have been assigned a transport function (data from EcoCyc, v. 5.6, (Karp *et al.* 2000)). These enzymes and transporters have been further assembled into 164 metabolic pathways in EcoCyc. Thus about one-quarter of this bacterial genome has already been mapped to the metabolome, allowing a global analysis of the metabolic capabilities of *E. coli* (Ouzounis & Karp 2000). Interestingly, even in such an extensively studied organism as *E. coli*, about the same proportion of unidentified ORFs have yet to be assigned a molecular function. A similar situation exists in *Saccharomyces cerevisiae* where 994 (16%) out of 6368 proteins have been assigned an EC number in the MIPS database (Mewes *et al.* 1999).

3. USES OF METABOLOMICS

Assembly of genes into metabolic pathways (metabolic construction) or mapping genes onto metabolic pathways (metabolic analysis) has a number of potential applications as follows.

(i) It can be used in the functional annotation of genomes to predict the full potential biosynthetic capability of an organism. This sort of analysis can be used to predict incorrect functional assignments of enzymes (false positives) and, conversely, to identify missing steps in a pathway (false negatives).
(ii) It can be used to map cellular gene-expression profiles onto metabolic pathways. Such metabolic displays could be used to study the global cellular

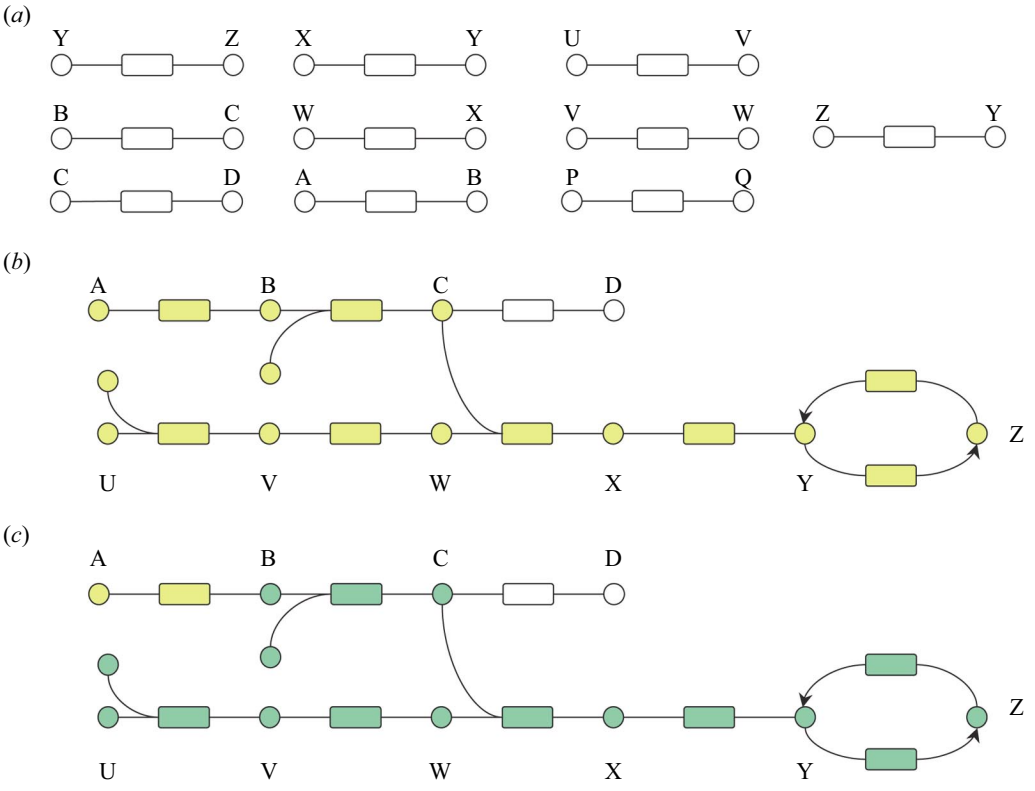

Figure 2. Metabolic pathway analysis for: (*a*) whole genome; (*b*) generic metabolome; (*c*) context-specific metabolome. Individual metabolites are indicated by circles and enzymes by lozenges. Genes that can be mapped onto a pathway of all known possible metabolic interconversions are coloured yellow (*b*). Genes that are expressed in a particular cell are coloured green (*c*).

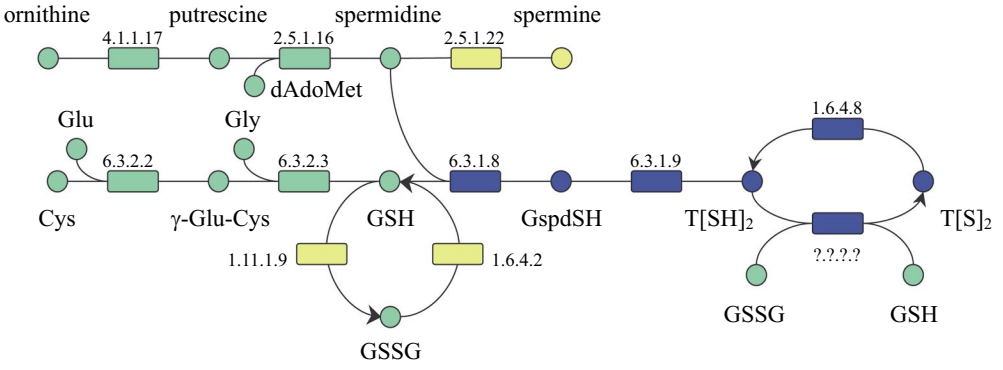

Figure 3. Differential metabolic analysis of polyamine and glutathione metabolism in host and parasite. Enzymes unique to humans are coloured yellow and those unique to the African trypanosome are coloured blue. Enzymes or metabolites common to both organisms are coloured green. EC numbers for the individual reactions are indicated. For clarity, not all metabolites and pathways are shown (e.g. the polyamine retro-conversion pathway is present in humans, but absent in trypanosomes).

response of a host cell to invasion by an intracellular parasite, to study parasite responses to a variety of chemical, nutritional or other environmental stresses or to elucidate the mode of action of a drug.

(iii) It can be used to compare metabolic profiles between cells, tissues or developmental stages.

(iv) When comparison is made between species (i.e. host and parasite) it can be used in the discovery of potential drug targets (Karp *et al.* 1999).

(v) In the biotechnology arena, it could be used for engineering of novel metabolic pathways in plants, bacteria and fungi to generate novel natural products for drug discovery (Dixon 2001).

4. BIOINFORMATIC TOOLS FOR STUDYING METABOLISM

There are two principal ways in which metabolic pathways can be conveniently presented in a graphical format. The first is based on the well-known metabolic map (designed by Dr E. Nicholson) that adorns the walls of

most biochemical laboratories. This is also available online in electronic format from ExPASy (see table 1 for URL). Using this representation it is possible to present a global view of most of the metabolome, where metabolites are represented by dots, enzymes as lines joining dots and pathways identified by particular colours, rather like a map of the London underground (figure 1). Provided there is sufficient biochemical data available, it is possible, though somewhat laborious, to create a reasonably accurate metabolic profile by hand. Figure 1*b* shows a possible metabolome of the bloodstream form of the African trypanosome. When compared with the metabolome of a typical mammalian cell (figure 1*a*), the most notable feature is the apparent absence of many core metabolic pathways in the parasite. For example, comparison of the pathways in orange representing carbohydrate metabolism and electron transport shows that most enzymes of the citric acid cycle (the circle at the bottom of figure 1*a*) are missing in the bloodstream of the African trypanosome (figure 1*b*). However, the procyclic (insect) life cycle stage has a functional citric acid cycle and a full complement of cytochromes (Fairlamb & Opperdoes 1986), indicating that these genes should be present in the trypanosome genome. Indeed, BLAST searches at National Center for Biotechnology Information (NCBI) using yeast citric acid cycle enzymes against unfinished sequence data from the *T. brucei* genome sequencing project revealed significant hits for several of these genes. Thus, the metabolome of the procyclic stage will more probably resemble the metabolome depicted in figure 1*a*, emphasizing the context-dependent nature of metabolomes. Since the asexual erythrocytic stages of the malaria parasite also lack a citric acid cycle (Sherman 1998*a*), purine biosynthesis (Sherman 1998*b*) and steroid biosynthesis (Vial & Ancelin 1998), and have a limited capacity to metabolize amino acids (Rosenthal & Meshnick 1998), it can be predicted that the metabolome of this stage of the parasite's life cycle will also resemble figure 1*b*. Again, BLAST searches identified several candidate genes for enzymes of the citric acid cycle, suggesting that malaria parasites can express a more extensive metabolome in other stages in their complex life cycle. Not surprisingly, the other feature to emerge from a global comparison of metabolomes is that truly unique pathways in parasites are rare (pink in figure 1*b*). A more detailed description of the individual features of these maps has been published elsewhere (Fairlamb 1989).

A major disadvantage of these types of representations of metabolism is that they have to be constructed by hand and are dependent on the availability of reliable and comprehensive biochemical data. They are therefore not useful for organisms or life cycle stages that have not been extensively studied biochemically. Neither are they useful for mapping genomic, transcriptomic or proteomic datasets onto a generic set of metabolic pathways to produce organism-, stage- or tissue-specific metabolomes. A number of databases and computer programs are under development to try to address these and other needs, particularly in relating molecular function to higher order function (Ashburner *et al.* 2000; Kanehisa 2000; Karp *et al.* 2000; Kuffner *et al.* 2000; Van Helden *et al.* 2000). Some of those databases with graphical pathway representations are listed in table 1. The most comprehensive of these is KEGG which has 53 gene catalogues (but no parasite genomes) (Kanehisa & Goto 2000), whereas EcoCyc is specifically focused on the biochemistry of *E. coli* (Karp *et al.* 2000). MetaCyc, which is also at the EcoCyc URL, provides a useful reference source for individual metabolic reactions from a wide variety of species, including some from trypanosomatids, but does not provide genomic data such as genome maps or sequences (Karp *et al.* 2000). The reader is recommended to visit these Web sites in order to assess their capabilities, strengths and weaknesses for their particular research needs.

The basic approach adopted in most of these pathway databases is illustrated schematically in figure 2. In figure 2*a*, the two circles connected via a lozenge represent each individual enzyme-catalysed reaction identified as present in a fully sequenced genome. The circles represent the substrates and products of the reaction (metabolites) linked to a compound database. The lozenge represents the enzyme catalysing a particular reaction and is defined by its EC number and linked to a gene catalogue for the genome of interest. In figure 2*b*, all known interacting metabolites are assembled into a reference pathway, representing known biochemical pathways in all organisms. The genes contained within the genome of interest can then be overlaid onto the reference pathway to display what might be called a generic metabolome (metabolites and enzymes highlighted in yellow). In figure 2*c*, genes that are expressed in a cell under specific conditions (e.g. data from EST databases, microarrays or proteomic analysis) are coloured green and mapped onto the reference pathway to produce a context-dependent metabolome.

In the hypothetical generic metabolome given in figure 1, it can be seen that precursors A and U can be converted into end-products Y and Z and that A can also be converted to C, but not D. If there were biochemical evidence for the biosynthesis of D in the cell, then this would suggest that the gene had either not been identified or incorrectly annotated in the genome. Assuming metabolite D was essential for growth or survival of a pathogen, then identification of this gene could be of interest from a drug discovery perspective since it may be sufficiently different from homologues in the human host to make it amenable to selective inhibition. Comparison of the generic metabolome with the context-dependent metabolome in figure 2 reveals a second difference—the enzyme for converting A to B is apparently absent. This could mean:

(i) the enzyme is not expressed in this stage of the life cycle of the parasite;
(ii) it is not expressed under the prevailing environmental or temporal conditions; or
(iii) the level of expression is below the limits of detection of the assay method.

A variety of experiments could be carried out to resolve such possibilities. For example, a more systematic search of the unknown open reading frames (ORFs) in a genome might reveal a candidate for the missing metabolic step. Expression of the candidate gene and demonstration of enzymatic activity in the recombinant protein would provide strong proof of its existence. At a broader level, metabolic-labelling studies could establish whether a complete pathway is operative in an organism under a particular set

of conditions. Demonstrating that key metabolic intermediates are labelled and that enzyme activity for each of the individual metabolic steps is present would provide further evidence. Reverse genetics (e.g. targeted gene deletion) combined with metabolic labelling could be informative and in some cases demonstrate that alternative enzymes or alternative routes to a particular metabolite are present. Conversely reverse genetics (e.g. episomal expression) could be used to provide proof of absence of an enzyme activity under a particular set of conditions.

Comparisons between genomes could also be displayed on metabolomes. As an example, in figure 3 I have overlaid the known polyamine and glutathione pathways of the African trypanosome and the human host to indicate enzymes that are either unique to the host (e.g. glutathione reductase, EC 1.6.4.2, is shown in yellow), common to host and parasite (shown in green) or unique to the parasite (e.g. trypanothione reductase, EC 1.6.4.8, shown in blue). As discussed in § 5 this potentially useful target discovery tool is not yet available from any metabolic database.

5. CURRENT LIMITATIONS OF METABOLIC ANALYSIS

None of the existing metabolic databases are perfectly accurate and one should be aware of the limitations of these tools, particularly for drug discovery.

One major problem arises from the limitations of EC numbering system. For example, a single reaction may be catalysed by an enzyme complex, in which case the same EC number is assigned to different subunits without distinction of their different functional roles. Conversely, a single gene may encode an enzyme with multiple independent catalytic domains (e.g. dihydrofolate reductase–thymidylate synthase), in which case different EC numbers should be assigned to a single gene. Moreover, the specificity of reactions catalysed by a particular EC number is not unequivocal. For example, spermidine synthase (EC 2.5.1.16) from mammalian sources is highly specific for putrescine as the acceptor of the aminopropyl group donated from decarboxylated AdoMet and has a distinct enzyme for the synthesis of spermine (figure 3). In contrast, the orthologue from *E. coli* can also synthesize spermine from spermidine, yet is not assigned an additional EC number for the reaction catalysed by spermine synthase (EC 2.5.1.22).

Incorrect assignment of an EC number by homology in a genome annotation can result in incorrect pathway assignment in a database. Take the case of mycothione reductase (Patel & Blanchard, 1999), an NADPH flavoprotein disulphide oxidoreductase that shows 29 and 27% identity to glutathione reductase from humans and *E. coli*, respectively. In the GenBank–EMBL–DDJB database, this enzyme is described as a *homologue* of glutathione reductase (i.e. EC 1.6.4.–), yet has been incorrectly assigned the EC number 1.6.4.2 (glutathione reductase) and therefore appears in the KEGG metabolic pathway for glutathione metabolism in *M. tuberculosis*. Examination of glutathione metabolism in KEGG (map 00480) for glutathione metabolism in *M. tuberculosis* should alert the researcher to a potential problem. The complete absence of any enzymes for the biosynthesis of glutathione (γ-glutamylcysteine synthetase EC 6.3.2.2 and glutathione synthetase EC 6.3.2.3) or other glutathione-dependent pathways should raise the possibility that the database may be in error. EcoCyc Pathway Tools reports a similar error and also indicates that the gene product may function as a mercuric reductase, another member of the flavoprotein disulphide oxidoreductase family. At present, mycothione reductase does not have an EC number (EC 1.6.4.–) and therefore cannot be retrieved from any enzyme databases. Thus, another difficulty concerns enzymes that have not yet been classified by the EC.

Another problem arises from the metabolites themselves. Unusual metabolites are not commonly represented in databases and thus mycothiol (or its disulphide, mycothione) from *Mycobacterium* sp. (Spies & Steenkamp 1994), and its biosynthetic enzymes cannot be identified by pathway analysis in KEGG. Neither can trypanothione biosynthesis or trypanothione reductase, a key drug target in trypanosomatids (Fairlamb & Cerami 1992). Biochemical pathways in which not all the intermediate steps are known are also difficult to represent. For example, the biosynthetic pathway of the redox active ovothiols (1-methyl-4-mercaptohistidines), which are found in abundant amounts in the eggs of sea urchins and other marine organisms (Shapiro & Hopkins 1991) and in trypanosomatids (Spies & Steenkamp 1994; Steenkamp & Spies 1994; Ariyanayagam & Fairlamb 2001), is not known with certainty. Finally, their utility for discovery of new drug targets for the protozoan parasites is limited since only complete genomes or whole chromosomes tend to be annotated.

6. USE OF METABOLIC ANALYSIS IN DRUG DISCOVERY

In order to fully exploit the vast potential of genome sequence information for drug discovery, it is important to bear in mind the general characteristics of an antimicrobial drug target. The most important of these characteristics are:

(i) a target must be essential for growth or survival of the parasite; and
(ii) a target must either be absent from the host, or, if present, either sufficiently different to allow selective inhibition or non-essential for host survival.

Both of these conditions must be met if the principle of 'selective toxicity' is to be achieved. Thus, the classical tools of biochemistry, enzymology, molecular genetics and structural biology will continue to underpin target discovery and target validation through database mining.

Despite the current limitations and shortcomings of tools for metabolic analysis, there have been some significant successes in the discovery of potential drug targets from pathogen genome sequencing projects. For example, the 'thiol-specific antioxidant proteins' identified in EST databases of trypanosomes and leishmania are now known to belong to a class of peroxidases, the peroxiredoxins, with specificity for tryparedoxin (a homologue of thioredoxin). These proteins form part of the trypanothione peroxidase system and could represent novel targets for drug discovery (Flohé *et al.* 1999). The structural details

of all components in this pathway have been determined rendering structure-based drug design an attractive option (Alphey *et al*. 1999, 2000; Bond *et al*. 1999; Hofmann *et al*. 2001). Likewise, in malaria, database mining has revealed the existence of hitherto unsuspected biochemical pathways with similarity to bacterial metabolism, such as the DOXP pathway of isoprenoid biosynthesis (Jomaa *et al*. 1999) and type II or 'dissociative' fatty acid biosynthesis in the apicoplast (Waller *et al*. 1998; Surolia & Surolia 2001). Both of these latter discoveries have resulted in promising drug leads such as fosmidomycin, an antibacterial antibiotic (Jomaa *et al*. 1999), and Triclosan, a component of toothpastes and mouthwashes (McLeod *et al*. 2001; Surolia & Surolia 2001). Database mining has also provided important insights into the metabolic functions of the apicoplast, that may reveal other interesting potential drug targets (Roos *et al*. 1999, 2002).

7. CONCLUSION AND FUTURE PROSPECTS

In theory, knowledge of the entire genome identifies every potential drug and vaccine target in any given microbial pathogen. In practice, this is clearly not the case. Better bioinformatics tools are urgently needed to retrieve and query information. More accurate and rigorous annotation of databases is necessary if scientists are to be able to retrieve biologically robust information. Even then, clues to the biological functions of many parasite gene products will not be elucidated by bioinformatics alone, since a significant proportion of these unknown genes will have no orthologues in other model organisms such as yeast and *E. coli*. Many of these pathogen-specific genes will probably play key roles in parasitism, allowing particular species to successfully colonize specific intracellular and extracellular biological niches in a restricted range of invertebrate and vertebrate hosts. These will include many interesting genes involved in virulence, immune evasion and survival that could be exploited ultimately for the development of new drugs, vaccines and diagnostics. Thus, fundamental research will continue to underpin all of our efforts to understand the functions of any genome. Only then can the ultimate goal of developing drugs, vaccines and diagnostics to control these devastating diseases be achieved.

The author thanks the Wellcome Trust for financial support as a Wellcome Principal Research Fellow.

REFERENCES

Alphey, M. S., Leonard, G. A., Gourley, D. G., Tetaud, E., Fairlamb, A. H. & Hunter, W. N. 1999 The high resolution crystal structure of recombinant *Crithidia fasciculata* tryparedoxin-I. *J. Biol. Chem.* **274**, 25 613–25 622.

Alphey, M. S., Bond, C. S., Tetaud, E., Fairlamb, A. H. & Hunter, W. N. 2000 The structure of reduced tryparedoxin peroxidase reveals a decamer and insight into reactivity of 2Cys-peroxiredoxins. *J. Mol. Biol.* **300**, 903–916.

Ariyanayagam, M. R. & Fairlamb, A. H. 2001 Ovothiol and trypanothione as antioxidants in trypanosomatids. *Mol. Biochem. Parasitol.* **115**, 189–198.

Ashburner, M. (and 19 others) 2000 Gene ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29.

Bloom, B. R. 1995 Genome sequences—a microbial minimalist. *Nature* **378**, 236.

Bond, C. S., Zhang, Y. H., Berriman, M., Cunningham, M. L., Fairlamb, A. H. & Hunter, W. N. 1999 Crystal structure of *Trypanosoma cruzi* trypanothione reductase in complex with trypanothione, and the structure-based discovery of new natural product inhibitors. *Structure* **7**, 81–89.

Clyde, D. F., Most, H., McCarthy, V. C. & Vanderberg, J. P. 1973 Immunization of man against sporozite-induced falciparum malaria. *Am. J. Med. Sci.* **266**, 169–177.

Cole, S. T. (and 41 others) 1998 Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**, 537–544.

Dixon, R. A. 2001 Natural products and plant disease resistance. *Nature* **411**, 843–847.

Fairlamb, A. H. 1989 Novel biochemical pathways in parasitic protozoa. *Parasitology* **99S**, 93–112.

Fairlamb, A. H. & Cerami, A. 1992 Metabolism and functions of trypanothione in the Kinetoplastida. *A. Rev. Microbiol.* **46**, 695–729.

Fairlamb, A. H. & Opperdoes, F. R. 1986 Carbohydrate metabolism in African trypanosomes, with special reference to the glycosome. In *Carbohydrate metabolism in cultured cells* (ed. M. J. Morgan), pp. 183–224. New York: Plenum.

Flohé, L., Hecht, H. J. & Steinert, P. 1999 Glutathione and trypanothione in parasitic hydroperoxide metabolism. *Free Radic. Biol. Med.* **27**, 966–984.

Hofmann, B. (and 11 others) 2001 Structures of tryparedoxins revealing interaction with trypanothione. *Biol. Chem.* **382**, 459–471.

Jomaa, H. (and 11 others) 1999 Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* **285**, 1573–1576.

Kanehisa, M. 2000 Pathway databases and higher order function. *Adv. Protein Chem.* **54**, 381–408.

Kanehisa, M. & Goto, S. 2000 KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30.

Karp, P. D., Krummenacker, M., Paley, S. & Wagg, J. 1999 Integrated pathway-genome databases and their role in drug discovery. *Trends Biotechnol.* **17**, 275–281.

Karp, P. D., Riley, M., Saier, M., Paulsen, I. T., Paley, S. M. & Pellegrini-Toole, A. 2000 The EcoCyc and MetaCyc databases. *Nucleic Acids Res.* **28**, 56–59.

Kuffner, R., Zimmer, R. & Lengauer, T. 2000 Pathway analysis in metabolic databases via differential metabolic display (DMD). *Bioinformatics* **16**, 825–836.

McLeod, R. (and 12 others) 2001 Triclosan inhibits the growth of *Plasmodium falciparum* and *Toxoplasma gondii* by inhibition of Apicomplexan Fab I. *Int. J. Parasitol.* **31**, 109–113.

Mewes, H. W., Heumann, K., Kaps, A., Mayer, K., Pfeiffer, F., Stocker, S. & Frishman, D. 1999 MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **27**, 44–48.

Oliver, S. 2000 Guilt-by-association goes global. *Nature* **403**, 601–603.

Oliver, S. 2002 Functional genomics: lessons from yeast. *Phil. Trans. R. Soc. Lond.* B **357**, this issue.

Ouzounis, C. A. & Karp, P. D. 2000 Global properties of the metabolic map of *Escherichia coli*. *Genome Res.* **10**, 568–576.

Patel, M. P. & Blanchard, J. S. 1999 Expression purification and characterization of *Mycobacterium tuberculosis* mycothione reductase. *Biochemistry* **38**, 11 827–11 833.

Roos, D. S., Crawford, M. J., Donald, R. G. K., Kissinger, J. C., Klimczak, L. J. & Striepen, B. 1999 Origin, targeting, and function of the apicomplexan plastid. *Curr. Opin. Microbiol.* **2**, 426–432.

Roos, D. S. 2002 Mining the *Plasmodium* genome database to define organellar function: what does the apicoplast do? *Phil. Trans. R. Soc. Lond.* B **357**, this issue.

Rosenthal, P. J. & Meshnick, S. R. 1998 Hemoglobin pro-

cessing and the metabolism of amino acids, heme, and iron. In *Malaria: parasite biology, pathogenesis, and protection* (ed. I. W. Sherman), pp. 145–158. Washington, DC: ASM Press.

Shapiro, B. M. & Hopkins, P. B. 1991 Ovothiols—biological and chemical perspectives. *Adv. Enzymol. Related Areas Mol. Biol.* **64**, 291–316.

Sherman, I. W. 1998*a* Carbohydrate metabolism of asexual stages. In *Malaria: parasite biology, pathogenesis and protection* (ed. I. W. Sherman), pp. 135–143. Washington, DC: ASM Press.

Sherman, I. W. 1998*b* Purine and pyrimidine metabolism of asexual stages. In *Malaria: parasite biology, pathogenesis, and protection* (ed. I. W. Sherman), pp. 177–184. Washington, DC: ASM Press.

Spies, H. S. C. & Steenkamp, D. J. 1994 Thiols of intracellular pathogens: identification of ovothiol A in *Leishmania donovani* and structural analysis of a novel thiol from *Mycobacterium bovis*. *Eur. J. Biochem.* **224**, 203–213.

Steenkamp, D. J. & Spies, H. S. C. 1994 Identification of a major low-molecular-mass thiol of the trypanosomatid *Crithidia fasciculata* as ovothiol A: facile isolation and structural analysis of the bimane derivative. *Eur. J. Biochem.* **223**, 43–50.

Surolia, N. & Surolia, A. 2001 Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of *Plasmodium falciparum*. *Nature Med.* **7**, 167–173.

Trouiller, P. & Olliaro, P. L. 1999 Drug development output: what proportion for tropical diseases? *Lancet* **354**, 164.

Van Helden, J., Naim, A., Mancuso, R., Eldridge, M., Wernisch, L., Gilbert, D. & Wodak, S. J. 2000 Representing and analysing molecular and cellular function using the computer. *Biol. Chem.* **381**, 921–935.

Vial, H. J. & Ancelin, M. L. 1998 Malarial lipids. In *Malaria: parasite biology, pathogenesis, and protection* (ed. I. W. Sherman), pp. 159–175. Washington, DC: ASM Press.

Waller, R. F., Keeling, P. J., Donald, R. G. K., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A. F., Besra, G. S., Roos, D. S. & McFadden, G. I. 1998 Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **95**, 12 352–12 357.

World Health Organization, 2000 *The World Health report 2000: health systems: improving performance*. Geneva: World Health Organization. 164–169. See http://www.who.int/whr/2000/index.htm.