Review TRENDS in Parasitology Vol.23 No.11

Metabolic reconstruction and analysis for parasite genomes

John W. Pinney[1], Balázs Papp[1,2], Christopher Hyland[3], Lillian Wambua[3], David R. Westhead[3] and Glenn A. McConkey[3]

[1] Faculty of Life Sciences, The University of Manchester, Oxford Road, Manchester M13 9PT, UK
[2] Institute of Biochemistry, Biological Research Center, H-6726 Szeged, Hungary
[3] Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK

With the completion of sequencing projects for several parasite genomes, efforts are ongoing to make sense of this mass of information in terms of the gene products encoded and their interactions in the growth, development and survival of parasites. The emerging science of systems biology aims to explain the complex relationship between genotype and phenotype by using network models. One area in which this approach has been particularly successful is in the modeling of metabolism. With an accurate picture of the set of metabolic reactions encoded in a genome, it is now possible to identify enzymes or transporters that might be viable targets for new drugs. Because these predictions greatly depend on the quality and completeness of the genome annotation, there are substantial efforts in the scientific community to increase the numbers of metabolic enzymes identified. In this review, we discuss the opportunities for using metabolic reconstruction and analysis tools in parasitology research, and their applications to protozoan parasites.

Metabolic reconstruction in post-genomic parasitology

Genome sequencing projects have been undertaken for numerous protozoan parasites, and much sequence information is now publicly available. Complete genome sequences have been published for several Apicomplexa (including two *Plasmodium* species [1,2], *Cryptosporidium parvum* [3] and *Cryptosporidium hominis* [4]) and the trypanosomatids *Trypanosoma cruzi* [5], *Trypanosoma brucei* [6] and *Leishmania major* [7]. Sequencing of several other protozoan, helminth and parasitic nematode genomes is ongoing, and data are publicly available. Published genome annotations provide initial predictions of the metabolic enzymes, and preliminary genome-scale metabolic reconstructions (see Glossary) for *Plasmodium falciparum* and the trypanosomatids have been produced [6,8].

Genome annotation and metabolic reconstruction can be considered complementary efforts: isolated enzymatic functions are likely to indicate false-positive annotations, whereas missing enzymes in otherwise complete pathways might represent as-yet unassigned gene functions. The annotation of parasitic genomes is difficult owing to their great divergence and compositional bias, but funds are lacking for complete manual curation of several neglected organisms. Owing to these difficulties, the scientific community has been enlisted to assist curation of the parasite genomes available at present [9].

For some industrially significant prokaryotes, the analysis of reconstructed metabolic networks has already proved useful; for example, in increasing the production of a desired metabolic product [10]. In the context of a pathogenic organism, however, metabolic reconstruction has the potential to play an important part in the drug discovery

Glossary

Analogous enzymes: Two enzymes in different species that share the same catalytic function but have unrelated amino acid sequences.
Bidirectional best hits: Two genes from different organisms that are each other's closest matches in reciprocal sequence similarity searches and thus might be supposed to be functional orthologs.
Constraint-based analysis: A methodology for analysis of metabolic networks that uses various constraints (physico-chemical, topological, environmental or regulatory) to limit the range of achievable functional states that are accessible to the organism.
EC number: Enzyme Commission number for the classification of enzymatic function.
Flux balance analysis (FBA): A type of constraint-based analysis that imposes mass balance and flux capacity constraints on the network and uses an optimization procedure to find a steady-state flux distribution that maximizes the biomass yield of the cell without the need for measurements of kinetic parameters. FBA can be used to identify essential reactions and to simulate growth in different environmental conditions.
Functional orthologs: Two genes from different organisms that share the same function, having both arisen from the same ancestral gene.
Horizontal gene transfer: A mechanism by which genes might be transferred between different species, for example, from a prokaryote to a protozoan through endosymbiosis.
Machine learning: A field of computer science in which existing knowledge is used to train a program to recognize or to classify new instances within the same domain.
Metabolic control analysis: A method for calculating the degree of sensitivity of the flux through a pathway to variations in the concentrations of each of its enzymes. Metabolic control analysis requires detailed knowledge of the kinetic parameters for each reaction.
Metabolic reconstruction: A method for predicting an organism's complement of metabolic enzymes and the network of reactions that it supports; the prediction is derived from a combination of genome sequence analysis, experimental data and reasoning based on network structure.
Missing enzyme or pathway hole: An enzyme that is presumed to be present in a genome on the basis of the results of metabolic reconstruction, but for which a gene has yet to be identified.
Phylogenetic profiling: A comparative genomics method for suggesting biological processes associated with a gene of unknown function; phylogenetic profiling is based on the observation that sets of genes that form functional modules tend to be gained or lost from the genome together.
Profile model: A bioinformatic model for identifying a protein functional domain; the model represents the frequency of amino acids observed at each position in a multiple alignment of protein sequences taken from a range of organisms.

Corresponding author: McConkey, G.A. (g.a.mcconkey@leeds.ac.uk).

 1471-4922/$ – see front matter doi:10.1016/j.pt.2007.08.013

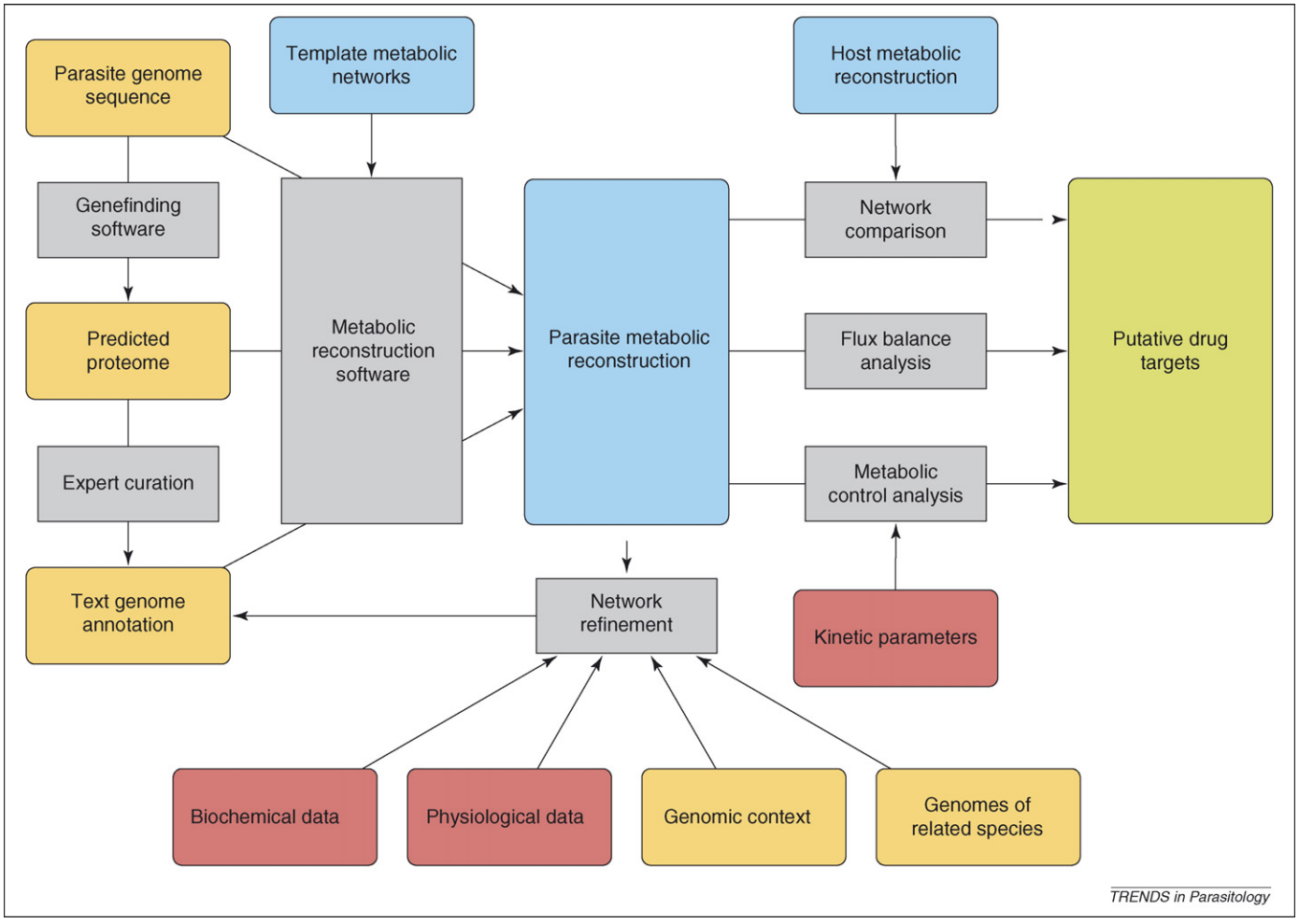

Figure 1. Summary of metabolic reconstruction for a parasite genome. Flowchart showing the various data structures (rounded boxes) and operations (grey rectangles) involved in the preparation of a metabolic reconstruction and subsequent prediction of drug targets for a parasitic organism. The colours of the boxes indicate the nature of the data source: red for experimental data, orange for information derived from genome sequences, blue for network structures and green for drug targets.

process [11]. With advances in metabolic network reconstruction and analysis, and the increasing data sets of parasite and host genome sequences, metabolic reconstruction in parasitology, although challenging, is now a realistic goal. In this review, we discuss tools for this field and some of their applications (Figure 1).

Reconstruction of metabolic networks

General methods for metabolic reconstruction applicable to prokaryotic genomes have been recently reviewed elsewhere [12]. Here, we focus on the issues surrounding the application of these methods to eukaryotic parasites.

Most techniques for predicting the enzyme complement of an organism rely on the availability of a fully annotated genome. An example of this approach is the Pathway Tools software [13], which uses text-mining methods to produce a preliminary reconstruction that must be refined by manual curation. The reliability of these metabolic reconstructions therefore depends crucially on the accuracy and completeness of the genome annotation. Pathway Tools has been used by several consortia to produce metabolic pathway databases for fully sequenced organisms, including *Homo sapiens* [14], *Arabidopsis thaliana* [15] and *Plasmodium falciparum* [8] (see the BioCyc Database, http://biocyc.org). Researchers are encouraged to support the curation of these databases; continual revision is necessary as more organism-specific pathways are added and further information about gene function is discovered.

To assist in this type of metabolic reconstruction, several integrated genome annotation systems have been developed that can incorporate information from both computational tools and manually curated sources. A key feature of the most recent software is the application of comparative genomics techniques across multiple sequenced genomes to improve functional annotations [16,17].

Extending metabolic reconstruction beyond model species

With genomes for which an expert annotation is unavailable or substantially incomplete, but for which a reliable set of gene structure predictions has been obtained, it is possible to produce a 'first-pass' metabolic reconstruction by using automated methods (e.g. see Refs [18,19]). In the simplest case, this reconstruction consists of performing BLAST [20] searches against annotated genomes and looking for bidirectional best hits that are assumed to be functional orthologs, allowing the assignment of enzymatic annotations [21]. This method works well as long as an annotation of a closely related organism is available, but is

Table 1. Global analysis of reconstructed metabolic networks[a]

Genome	Enzymes	Metabolites	Produced and consumed[b]	Produced only (not consumed)[c]	Consumed only (not produced)[d]	%Miss[e]
P. falciparum	346	685	409	148	128	40.29
C. parvum	281	520	298	110	112	42.69
T brucei	476	869	569	168	132	34.52
T. cruzi	583	1040	692	196	152	33.46
L. major	560	986	673	179	134	31.74
S. cerevisiae	626	1078	754	175	149	30.06
E. coli	767	1183	876	167	140	25.95

[a]This table was constructed with SHARKhunt software [23], which is based on the reference metabolic network of KEGG [73]. The numbers of enzymes and metabolites in the reconstructed network are given.
[b]Metabolites in the reconstructed network can be predicted to be both produced and consumed.
[c]Enzymes can be predicted to produce the metabolite, but no enzyme can be predicted to consume it.
[d]There is a prediction of an enzyme that consumes the metabolite but not an enzyme that produces it. Note that the fifth and sixth columns refer only to metabolites in these categories for the species concerned and that are both produced and consumed in the KEGG reference network. Metabolites in these two categories are likely to be adjacent to network holes (i.e. missing enzymes that would produce or consume them).
[e]Percentage of metabolites next to potential holes.

not sensitive enough to detect some of the highly divergent enzymes within a parasite genome. A more sophisticated search technique is to use a library of profile models derived from multiple alignments of enzyme functional domains to predict the enzymatic functions of each protein in the target proteome, as implemented in the PRIAM software [22].

Reliance on completed genomes, however, can unnecessarily delay the investigation of metabolic function for unfinished and ongoing genome projects. For prokaryotic organisms, a set of open reading frames is easily obtained and can be used as input into a protein-sequence-based tool. For eukaryotes, however, the presence of introns means that analysis of a set of open reading frames is usually inappropriate, and *ab initio* gene-finding tools will often produce unsatisfactory results unless a suitable training set of gene structures is available. The SHARKhunt software [23] addresses this issue by using the PRIAM [22] library of profile models to search through DNA sequences for regions with significant similarity to known enzymes. Even at less than a single coverage of the genome, sufficient enzyme-encoding genes can be detected by this method to give a reasonable idea of the metabolic capabilities of an organism.

Despite these advances in search techniques, the phylogenetic distances from model organisms are often so great that successful metabolic reconstruction for parasites remains extremely challenging [24]. Enzymes might have functions that are not represented in sequence databases, they might be functional analogs unrelated to other proteins catalyzing the same reaction, or they might simply have diverged too far to be recognizable. For example, the only remaining unidentified step in folate synthesis in *Plasmodium* was found, through biochemical approaches, to be mediated by an unusual ortholog of an enzyme normally associated with tetrahydrobiopterin synthesis (S. Dittrich and J. Hyde, personal communication). Additional complications such as the high A + T content of the *Plasmodium* genomes and the general proliferation of low-complexity regions in apicomplexans will further reduce the sensitivity of sequence-based methods for enzyme identification.

Pathway holes in metabolic networks

With the difficulty of assigning specific enzymatic function in parasite genomes, initial automated metabolic reconstructions usually produce networks with many 'holes' – reactions essential for a complete biochemical pathway, but for which no enzyme has been annotated in the genome. Table 1 shows an analysis of holes in metabolic networks predicted on the basis of genome data with the SHARKhunt system. In the apicomplexans, more than 40% of predicted metabolites identified on the basis of enzyme functions are either 'substrates' or 'products' without a predicted enzyme to catalyze the reaction. This percentage is reduced to approximately 33% in the

Table 2. Reconstruction of pantothenate and coenzyme A biosynthesis for various parasites[a]

	P. falciparum					*C. parvum*		*T. gondii*	*E. tenella*	*T. annulata*		*L. major*		*T. brucei*		*T. cruzi*		*E. histolytica*	
EC	MP	PtdCho	KG	P2	MS[b]	KG	MS	MS[b]	MS[b]	KG	MS[b]	KG	MS[b]	KG	MS[b]	KG	MS[b]	KG	MS[b]
2.1.2.11								X	X										
1.1.1.169								X	X										
6.3.2.1								X	X										
2.7.1.33	X				X			X	X	X		X	X	X	X	X	X		X
6.3.2.5	X				X		X	X	X		X		X		X		X		
4.1.1.36	X				X		X	X	X		X	X	X		X	X	X	X	X
2.7.7.3	X				10^{-7}			10^{-5}	10^{-7}		10^{-7}		10^{-8}		10^{-5}		10^{-4}	X	10^{-3}
2.7.1.24	X	X		X	X	X		X	X		X	X	X	X	X	X	X	X	X

[a]This table shows the results of metabolic annotation of the biosynthesis pathways of pantothenate (first three EC number steps) and coenzyme A (last five EC number steps) using several software systems and websites (KG, KEGG [73]; PtdCho, PlasmoCyc [8]; P2, PUMA2 [16]; MS, metaSHARK [23]; and MP, Malaria Metabolic pathways [25]). Enzymes are labeled by their EC numbers: 2.1.2.11, 3-methyl-2- oxobutanoate hydroxymethyltransferase; 1.1.1.169, 2-dehydropantoate 2- reductase; 6.3.2.1, pantoatebeta-alanine ligase; 2.7.1.33, pantothenate kinase; 6.3.2.5, phosphopantothenate-cysteine ligase; 4.1.1.36, phosphopantothenoylcysteine decarboxylase; 2.7.7.3, pantetheinephosphate adenylyltransferase; and 2.7.1.24, dephospho-CoA kinase. An 'X' indicates that the enzyme has been located in the genome. Note that annotations are not available for all organisms with all software packages.
[b]In these examples, the statistical support (profile match *E* value) is given because it is above(weaker than)the default cut-off (10^{-10}); in all other examples, software was used with default parameters.

trypanosomes, and better characterized organisms have baseline values of 25%–30%. Many of these holes can be attributed to failures of the automated method to find all enzymes, but some might represent metabolites obtained from the host or other more complex deviations from standard metabolic pathways (as described for the folate pathway earlier).

Table 2 examines these effects in more detail for the linked pathways synthesizing pantothenate and coenzyme A. Several software systems and websites are compared, although PlasmoCyc [8] and Malaria Metabolic Pathways (MMP) [25] involve substantial human input. The systems do not agree perfectly, and there are many holes. Software systems might disagree simply through the use of different statistical cut-offs to define significant hits or for more complex reasons. For EC numbers 2.7.1.33 and 6.3.2.5 in *P. falciparum*, for example, the genes predicted by meta-SHARK and MMP contain large (~200 amino acid) low-complexity regions that can confuse identification.

The results for *P. falciparum* in Table 2 illustrate the value of manual input and distant homology detection in metabolic reconstruction. Only MMP has a hole-free pathway from pantothenate to coenzyme A. The automated SHARKhunt system finds all enzymes except 2.7.7.3, but closer inspection of the results reveals a hit for this enzyme with lower statistical support. In this case, the profiles used by SHARKhunt detected a potential distant evolutionary relationship that would elude simple BLAST-based methods. Alone this evidence would be weak, but in the presence of strong support for the other enzymes a reasoned manual analysis would add this gene to complete the pathway. Thus, despite the holes, the results in Table 2 provide evidence that most of these organisms have the ability to synthesize coenzyme A from pantothenate, but probably depend (with the exception of *Eimeria tenella* and *Toxoplasma gondii*) on their hosts for a supply of this substrate.

Of the manually informed metabolic reconstructions for *P. falciparum*, MMP is the most current and informed. The PlasmoCyc [8] reconstruction contains 798 reactions and 679 enzymes, but has 241 holes, many of which are in pathways where only 1–2 enzymes have been annotated in the genome. The presence of such a small percentage of the enzymes in a pathway does not usually provide support for the presence of that pathway. Nevertheless, current biochemical knowledge of *Plasmodium* is heavily biased towards the erythrocytic stages, and it is clear from the number of orphan enzyme activities (in pathways where few other enzymes have been found) that our knowledge of the malaria metabolic network will be improving for the foreseeable future.

Searching for the 'missing' enzymes

Although the problem of filling pathway holes remains a difficult one, some techniques are available that can help to identify candidate genes wherever a specific enzyme is thought to be present but is not yet annotated. Comparative genomics methods include the analysis of cross-species patterns of gene retention ('phylogenetic profiling'), gene fusion events, expression profiles, shared regulatory sites, and proximity to functionally related genes on the genome [26]. Enzymes that are evolutionarily unrelated to known proteins with the same function ('functional analogs') might also be discovered by analyzing anticorrelations in the presence of genes across genomes of different species [27]. Machine learning techniques (e.g. see Ref. [28]) can be used to integrate these diverse sets of observations to give an overall value of the likelihood that a candidate gene has the required function.

An alternative explanation for an apparent pathway hole in a parasitic organism is that either the enzymatic function or the metabolite product is supplied by the host or by the vector. These cases might be difficult to demonstrate by genome analysis alone, particularly because specific functional annotation of metabolite transporters is difficult [29]. Detailed comparison of host and pathogen enzyme complements, however, has suggested plausible schemes for metabolite exchange – for example, in coenzyme A biosynthesis in intracellular bacterial pathogens and symbionts [17,30,31] and in the parasites *Encephalitozoon cuniculi*, *Giardia lamblia* and *P. falciparum* [32]. Interference in metabolite dependence on the host is considered a very effective strategy for the future development of novel antimicrobial drugs [29,31].

Quality control for reconstructed networks

Biochemical accuracy is crucial if the network is to be used for computational analyses and, despite advances in automatic reconstruction methods, these methods still have limitations that can be overcome only by manual curation based on the integration of genomic, biochemical, physiological and comparative data. Potential problems include incorrect substrate specificity, cofactor usage, reaction reversibility, treatment of enzyme subunits as separate enzymes, and missing reactions for which no genes have been assigned so far [33].

Where parasite genomes have lost metabolic functions relatively recently, there is also the possibility of false-positive identification of an enzyme on the basis of its remnant sequence. It is promising, however, that high-quality reconstructions amenable to computational analysis have been obtained for an increasing number of bacterial pathogens, including *Haemophilus influenzae* [34], *Helicobacter pylori* [35] and *Staphylococcus aureus* [36,37].

Computational identification of essential enzymes

One approach to discover new drug targets is to find genes that are essential during the targeted life-cycle stage of the parasite. Although large-scale gene disruption screens have been conducted in several model organisms [38,39] and some pathogens [40–42], systematic experimental identification of essential genes is not yet available for most parasites. As a result, computational tools to predict the set of indispensable genes on the basis of genome data are in great demand. The availability of high-quality metabolic network reconstructions paves the way for such *in silico* approaches.

In principle, the behavior of biochemical pathways can be simulated computationally by solving a set of equations describing all enzyme-catalyzed reaction steps with formulae taken from enzyme kinetics (e.g. the Michae-

lis–Menten rate equation) [43]. The application of this type of dynamic analysis to genome-scale networks is, however, currently hindered by a lack of detailed information on kinetic parameters. Because of these limitations, a new modeling framework known as the 'constraint-based approach' has been introduced to deduce the metabolic phenotype from the genotype [44]. This approach attempts to narrow – on the basis of the successive imposition of governing physicochemical and biological constraints – the range of possible phenotypes that a metabolic system can display [44,45].

The widely used flux balance analysis (FBA) method assumes that the metabolic network will reach a steady state that satisfies certain constraints (e.g. mass balance and flux limitations) and maximizes biomass production [46]. Despite its simplicity, FBA has been shown to predict essential genes with good accuracy in yeast [47] and *Escherichia coli* [48]; thus, it might prove to be a useful tool for rational identification of drug targets in microbes [49]. Recently, FBA has been applied to the mycolic acid pathway of *Mycobacterium tuberculosis* to find antitubercular drug targets [50].

A conceptual shortcoming of FBA is the assumption that the flux distribution is optimal even in mutant cells. As a result, attempts have been made to introduce more realistic assumptions to increase the prediction accuracy of gene deletion phenotypes [51,52]. Other approaches based on the investigation of network topology have also been put forward to predict vulnerable points of metabolic networks (e.g. see Refs [8,53]).

Predicting multiple targets for combination therapy

Owing to both partial inhibition and the grave potential for drug resistance in the field, it is important to inhibit multiple targets [54]. A relevant issue in chemotherapy is whether incomplete inactivation of enzyme activity will be sufficient to compromise the fitness of the parasite. It is well established that a partial decrease in enzyme activity at most steps in a pathway will have only negligible effect on the flux through the pathway [55].

The framework of 'metabolic control analysis' has been developed to calculate the sensitivity of flux to inhibition of any individual enzyme in the pathway. For example, the function of the plasma membrane glucose transporter is predicted to be the most sensitive step in the glycolytic pathway of *Trypanosoma brucei*, and this prediction is supported by experimental measurements [56]. Unfortunately, such calculations can be performed only when a detailed kinetic model is available; therefore, the application of control analysis is restricted at present to a few well-characterized pathways of a limited number of organisms.

Lastly, it has been proposed that partial inhibition of multiple targets might be more efficient than the complete inhibition of a single target [57] – an idea that is supported by the fact that several highly efficient drugs affect many targets simultaneously. Indeed, the potential to identify combinations of target enzymes whose inhibition acts in synergy is a compelling motivation for the development of accurate metabolic models for parasitic organisms. In the simplest case, synergy would be expected for drugs targeting two independent pathways for the biosynthesis of an essential metabolite product, but there could be many effective drug combinations that might be uncovered only by computational analysis of the complete network or by other systems biology approaches [58,59]. Empirically determined combinations of existing monotherapies are widely used against parasitic infections to mitigate the rate of evolved resistance [60]. As our understanding of the molecular mechanisms of drug resistance improves, computational analysis coupled with laboratory evolution experiments [61] will have immense value in the rational design of drug combinations that will remain effective for as long as possible.

Comparative genomics of pathogen and host metabolism

Comparisons of parasite and host metabolic networks have both evolutionary and chemotherapeutic interest, and an important goal is to assemble integrated networks of pathogen and host metabolic reactions that provide information for interactions between the two organisms [62,63]. Differences in parasite and host metabolic gene contents are due to gene losses and gene gain events such as horizontal gene transfer [64]. Many parasites have lost pathways that are essential in free-living species. For example, *Mycoplasma genitalium* contains fewer than 500 genes [65] and the obligate anaerobes *Entamoeba histolytica* and *C. parvum* lack Kreb's cycle enzymes and other metabolic pathways [66,67]. Parasite genomes, however, can also encode metabolic functions that are lacking in humans [68]. For example, hemoglobin degradation in *Plasmodium* species utilizes specialized proteases (e.g. falcipain), and trypanosomes contain a novel mechanism for inactivating free radicals [69].

The resemblance of some parasite pathways to parts of prokaryotic metabolism suggests that these pathways have been acquired by horizontal gene transfer (e.g. *Cryptosporidium* thymidine kinase [64] and ferrodoxins in *E. histolytica* and *G. lamblia* [70]). Indeed, the acquisition of the apicomplexan plastid by secondary endosymbiosis of an algal cell [71] represents an extreme horizontal gene transfer event that contributed an essential component of metabolism [72]. These surviving pathways are unique in eukaryotes and thus represent attractive drug targets.

In combination with a metabolic database (e.g. KEGG [73]), sequence similarity searches can rapidly predict pathways that differ between host and parasite [74]. An approach motivated by set theory has also been proposed to compare metabolic network topology between multiple organisms [75]. An important next step will be the application of metabolic modeling and the addition of functional (e.g. metabolomic) data to decipher the physiological importance of such network differences.

Concluding remarks

The reconstruction and analysis of metabolic networks is at the forefront of systems biology, and developments are ongoing at several different levels from control engineering to computer science to comparative genomics. As these techniques are evolving and our understanding of metabolism is increasing, *in silico* models are starting to generate

real, testable predictions of metabolic function. Although the number of completely novel broad-spectrum drug targets seems lower than previously thought [76], computational methods have great potential to help in the discovery of new parasite-specific targets and possibly to forecast therapies to combat resistance. With continuing support from the research and drug-development communities, post-genomic parasitology will be well placed to make the best use of these opportunities in the continuing fight against parasitic diseases.

Acknowledgements

We thank those who have participated in these studies and offer apologies to any whose work has been omitted due to the space restrictions. We greatly appreciate J. Hyde's generosity in sharing unpublished results from his laboratory. We acknowledge funding from the MRC and BBSRC for this work. J.P. and B.P. are supported by the BBSRC and the International Human Frontier Science Program Organization, respectively.

References

1 Gardner, M.J. *et al.* (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* 419, 498–511
2 Carlton, J.M. *et al.* (2002) Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* 419, 512–519
3 Abrahamsen, M.S. *et al.* (2004) Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* 304, 441–445
4 Xu, P. *et al.* (2004) The genome of *Cryptosporidium hominis*. *Nature* 431, 1107–1112
5 El-Sayed, N.M. *et al.* (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309, 409–415
6 Berriman, M. *et al.* (2005) The genome of the African trypanosome *Trypanosoma brucei*. *Science* 309, 416–422
7 Ivens, A.C. *et al.* (2005) The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 309, 436–442
8 Yeh, I. *et al.* (2004) Computational analysis of *Plasmodium falciparum* metabolism: organizing genomic information to facilitate drug discovery. *Genome Res.* 14, 917–924
9 Berry, A.E. *et al.* (2004) Curation of the *Plasmodium falciparum* genome. *Trends Parasitol.* 20, 548–552
10 Rohlin, L. *et al.* (2001) Microbial pathway engineering for industrial processes: evolution, combinatorial biosynthesis and rational design. *Curr. Opin. Microbiol.* 4, 330–335
11 Cowman, A.F. and Crabb, B.S. (2003) Functional genomics: identifying drug targets for parasitic diseases. *Trends Parasitol.* 19, 538–543
12 Francke, C. *et al.* (2005) Reconstructing the metabolic network of a bacterium from its genome. *Trends Microbiol.* 13, 550–558
13 Karp, P.D. *et al.* (2002) The Pathway Tools software. *Bioinformatics* 18 (Suppl. 1), S225–S232
14 Romero, P. *et al.* (2005) Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.* 6, R2
15 Mueller, L.A. *et al.* (2003) AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiol.* 132, 453–460
16 Maltsev, N. *et al.* (2006) PUMA2 – grid-based high-throughput analysis of genomes and metabolic pathways. *Nucleic Acids Res.* 34, D369–D372
17 Overbeek, R. *et al.* (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33, 5691–5702
18 Arakawa, K. *et al.* (2006) GEM system: automatic prototyping of cell-wide metabolic pathway models from genomes. *BMC Bioinformatics* 7, 168
19 Notebaart, R.A. *et al.* (2006) Accelerating the reconstruction of genome-scale metabolic networks. *BMC Bioinformatics* 7, 296
20 Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402
21 Bono, H. *et al.* (1998) Reconstruction of amino acid biosynthesis pathways from the complete genome sequence. *Genome Res.* 8, 203–210
22 Claudel-Renard, C. *et al.* (2003) Enzyme-specific profiles for genome annotation: PRIAM. *Nucleic Acids Res.* 31, 6633–6639
23 Pinney, J.W. *et al.* (2005) metaSHARK: software for automated metabolic network prediction from DNA sequence and its application to the genomes of *Plasmodium falciparum* and *Eimeria tenella*. *Nucleic Acids Res.* 33, 1399–1409
24 McConkey, G.A. *et al.* (2004) Annotating the *Plasmodium* genome and the enigma of the shikimate pathway. *Trends Parasitol.* 20, 60–65
25 Ginsburg, H. (2006) Progress in *in silico* functional genomics: the Malaria Metabolic Pathways database. *Trends Parasitol.* 22, 238–240
26 Osterman, A. and Overbeek, R. (2003) Missing genes in metabolic pathways: a comparative genomics approach. *Curr. Opin. Chem. Biol.* 7, 238–251
27 Morett, E. *et al.* (2003) Systematic discovery of analogous enzymes in thiamin biosynthesis. *Nat. Biotechnol.* 21, 790–795
28 Green, M.L. and Karp, P.D. (2004) A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. *BMC Bioinformatics* 5, 76
29 Martin, R.E. *et al.* (2005) The 'permeome' of the malaria parasite: an overview of the membrane transport proteins of *Plasmodium falciparum*. *Genome Biol.* 6, R26
30 Ye, Y. *et al.* (2005) Automatic detection of subsystem/pathway variants in genome analysis. *Bioinformatics* 21 (Suppl. 1), i478–i486
31 Gerdes, S.Y. *et al.* (2002) From genetic footprinting to antimicrobial drug targets: examples in cofactor biosynthetic pathways. *J. Bacteriol.* 184, 4555–4572
32 Genschel, U. (2004) Coenzyme A biosynthesis: reconstruction of the pathway in archaea and an evolutionary scenario based on comparative genomics. *Mol. Biol. Evol.* 21, 1242–1251
33 Reed, J.L. *et al.* (2006) Towards multidimensional genome annotation. *Nat. Rev. Genet.* 7, 130–141
34 Schilling, C.H. and Palsson, B.O. (2000) Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J. Theor. Biol.* 203, 249–283
35 Schilling, C.H. *et al.* (2002) Genome-scale metabolic model of *Helicobacter pylori* 26695. *J. Bacteriol.* 184, 4582–4593
36 Becker, S.A. and Palsson, B.O. (2005) Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiol.* 5, 8
37 Heinemann, M. *et al.* (2005) In silico genome-scale reconstruction and validation of the *Staphylococcus aureus* metabolic network. *Biotechnol. Bioeng.* 92, 850–864
38 Gerdes, S.Y. *et al.* (2003) Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriol.* 185, 5673–5684
39 Giaever, G. *et al.* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418, 387–391
40 Salama, N.R. *et al.* (2004) Global transposon mutagenesis and essential gene analysis of *Helicobacter pylori*. *J. Bacteriol.* 186, 7926–7935
41 Glass, J.I. *et al.* (2006) Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. U. S. A.* 103, 425–430
42 Knuth, K. *et al.* (2004) Large-scale identification of essential *Salmonella* genes by trapping lethal insertions. *Mol. Microbiol.* 51, 1729–1744
43 Wiechert, W. (2002) Modeling and simulation: tools for metabolic engineering. *J. Biotechnol.* 94, 37–63
44 Covert, M.W. *et al.* (2003) Identifying constraints that govern cell behavior: a key to converting conceptual to computational models in biology? *Biotechnol. Bioeng.* 84, 763–772
45 Edwards, J.S. *et al.* (2002) Metabolic modelling of microbes: the flux-balance approach. *Environ. Microbiol.* 4, 133–140
46 Kauffman, K.J. *et al.* (2003) Advances in flux balance analysis. *Curr. Opin. Biotechnol.* 14, 491–496
47 Kuepfer, L. *et al.* (2005) Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res.* 15, 1421–1430
48 Edwards, J.S. and Palsson, B.O. (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5528–5533
49 Trawick, J.D. and Schilling, C.H. (2006) Use of constraint-based modeling for the prediction and validation of antimicrobial targets. *Biochem. Pharmacol.* 71, 1026–1035
50 Raman, K. *et al.* (2005) Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs. *PLoS Comput. Biol.* 1, e46
51 Segrè, D. *et al.* (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15112–15117

52 Shlomi, T. *et al.* (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc. Natl. Acad. Sci. U. S. A.* 102, 7695–7700
53 Rahman, S.A. and Schomburg, D. (2006) Observing local and global properties of metabolic pathways: 'load points' and 'choke points' in the metabolic networks. *Bioinformatics* 22, 1767–1774
54 Barrett, M.P. *et al.* (1999) Recent advances in identifying and validating drug targets in trypanosomes and leishmanias. *Trends Microbiol.* 7, 82–88
55 Kacser, H. and Burns, J.A. (1973) The control of flux. *Symp. Soc. Exp. Biol.* 27, 65–104
56 Bakker, B.M. *et al.* (2000) Metabolic control analysis of glycolysis in trypanosomes as an approach to improve selectivity and effectiveness of drugs. *Mol. Biochem. Parasitol.* 106, 1–10
57 Csermely, P. *et al.* (2005) The efficiency of multi-target drugs: the network approach might help drug design. *Trends Pharmacol. Sci.* 26, 178–182
58 Keith, C.T. *et al.* (2005) Multicomponent therapeutics for networked systems. *Nat. Rev. Drug Discov.* 4, 71–78
59 Yeh, P. *et al.* (2006) Functional classification of drugs by properties of their pairwise interactions. *Nat. Genet.* 38, 489–494
60 Shanks, G.D. (2006) Treatment of falciparum malaria in the age of drug resistance. *J. Postgrad. Med.* 52, 277–280
61 Elena, S.F. and Lenski, R.E. (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* 4, 457–469
62 Forst, C.V. (2002) Network genomics – a novel approach for the analysis of biological systems in the post-genomic era. *Mol. Biol. Rep.* 29, 265–280
63 Forst, C.V. (2006) Host–pathogen systems biology. *Drug Discov. Today* 11, 220–227
64 Striepen, B. *et al.* (2004) Gene transfer in the evolution of parasite nucleotide biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3154–3159
65 Fraser, C.M. *et al.* (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* 270, 397–403
66 Henriquez, F.L. *et al.* (2005) The unusual mitochondrial compartment of *Cryptosporidium parvum*. *Trends Parasitol.* 21, 68–74
67 Yarlett, N. (2004) Anaerobic protists and hidden mitochondria. *Microbiology* 150, 1127–1129
68 Chaudhary, K. and Roos, D.S. (2005) Protozoan genomics for drug discovery. *Nat. Biotechnol.* 23, 1089–1091
69 Fairlamb, A.H. and Cerami, A. (1992) Metabolism and functions of trypanothione in the Kinetoplastida. *Annu. Rev. Microbiol.* 46, 695–729
70 Nixon, J.E. *et al.* (2002) Evidence for lateral transfer of genes encoding ferredoxins, nitroreductases, NADH oxidase, and alcohol dehydrogenase 3 from anaerobic prokaryotes to *Giardia lamblia* and *Entamoeba histolytica*. *Eukaryot. Cell* 1, 181–190
71 Roos, D.S. *et al.* (1999) Origin, targeting, and function of the apicomplexan plastid. *Curr. Opin. Microbiol.* 2, 426–432
72 Ralph, S.A. *et al.* (2004) Tropical infectious diseases: metabolic maps and functions of the *Plasmodium falciparum* apicoplast. *Nat. Rev. Microbiol.* 2, 203–216
73 Kanehisa, M. *et al.* (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280
74 Anishetty, S. *et al.* (2005) Potential drug targets in *Mycobacterium tuberculosis* through metabolic pathway analysis. *Comput. Biol. Chem.* 29, 368–378
75 Forst, C.V. *et al.* (2006) Algebraic comparison of metabolic networks, phylogenetic inference, and metabolic innovation. *BMC Bioinformatics* 7, 67
76 Becker, D. *et al.* (2006) Robust *Salmonella* metabolism limits possibilities for new antimicrobials. *Nature* 440, 303–307