THE ROYAL SOCIETY

Published online 15 January 2002

Mining the *Plasmodium* genome database to define organellar function: what does the apicoplast do?

David S. Roos*, Michael J. Crawford, Robert G. K. Donald, Martin Fraunholz, Omar S. Harb, Cynthia Y. He, Jessica C. Kissinger, Michael K. Shaw and Boris Striepen

Department of Biology, University of Pennsylvania, Philadelphia, PA 19104-6018, USA

Apicomplexan species constitute a diverse group of parasitic protozoa, which are responsible for a wide range of diseases in many organisms. Despite differences in the diseases they cause, these parasites share an underlying biology, from the genetic controls used to differentiate through the complex parasite life cycle, to the basic biochemical pathways employed for intracellular survival, to the distinctive cell biology necessary for host cell attachment and invasion. Different parasites lend themselves to the study of different aspects of parasite biology: *Eimeria* for biochemical studies, *Toxoplasma* for molecular genetic and cell biological investigation, etc. The *Plasmodium falciparum* Genome Project contributes the first large-scale genomic sequence for an apicomplexan parasite. The *Plasmodium* Genome Database (http://PlasmoDB.org) has been designed to permit individual investigators to ask their own questions, even prior to formal release of the reference *P. falciparum* genome sequence. As a case in point, PlasmoDB has been exploited to identify metabolic pathways associated with the apicomplexan plastid, or 'apicoplast'—an essential organelle derived by secondary endosymbiosis of an alga, and retention of the algal plastid.

Keywords: apicomplexan parasites; *Toxoplasma gondi*; functional genomics; organellar genomes; plastid evolution

1. APICOMPLEXAN PARASITES: DIVERSITY AND SIMILARITY

The phylum Apicomplexa includes many thousands of protozoan species, all of which are parasites, and many of which cause important clinical and/or veterinary diseases (Levine 1988). *Plasmodium* species are responsible for malaria, and more than one million deaths per year are attributable to infection by *P. falciparum* alone (World Health Organization 1997). The afflictions of parasitic disease are not limited to the tropical world, however: congenital infection with *Toxoplasma gondii* is among the most common sources of neurological birth defects worldwide, and this parasite is also associated with a variety of immunosuppressive diseases and treatments (Frenkel 1973; Remington & Desmonts 1989; Luft & Remington 1992; McAuley *et al.* 1994). Parasites of the genus *Eimeria* hamper the growth of poultry and livestock, severely affecting commercial viability, while other apicomplexans infect organisms as diverse as lizards, whales, and shellfish (Levine 1988). Beyond the inherent biological fascination engendered by parasites in general, the rise and spread of drug-resistant *Plasmodium* and *Eimeria*, and the emergence of *Toxoplasma*, *Cryptosporidium*, and other parasites as opportunistic infections associated with AIDS, has served to focus attention on the identification of novel drug targets and treatment strategies (Laughon *et al.* 1991).

Apicomplexan parasites differ in many important respects (Frenkel 1973; Levine 1988). Consider, for example, the diseases they cause: malaria is characterized by periodic fevers, toxoplasmosis by focal lesions in the central nervous system, and coccidiosis by diarrhoea and intestinal malabsorption. Some (but not all) of these differences can be rationalized as a consequence of differences in transmission strategy. The ease of faecal/oral transmission in commercial poultry houses means that *Eimeria* propagates readily as an intestinal infection without an intermediate host. The fastidiousness of cats (the definitive host for *T. gondii*) may make the faecal/oral route a less efficient means of transmission, but systemic infection in mice and other animals leads to the development of long-lived 'bradyzoite' tissue cysts which are capable of propagating the infection via carnivores. Human–human transmission of *T. gondii* is therefore rare, limiting the spread of drug-resistant parasites. Mosquito vectors make human–human transmission of *Plasmodium* possible, setting the stage for the spread of drug-resistant parasites.

These pathogens can best be described as having adopted distinctive life history strategies. *Plasmodium* is the consummate specialist: the range of mosquito species capable of carrying any individual parasite species is highly restricted; each parasite species infects only a single (or limited number of) avian, reptilian, or mammalian host species; and different stages of the parasite life cycle exhibit exquisite tissue specificity (Sinden 1983). *Plasmodium falciparum* sporozoites, for example, only infect hepatocytes, while merozoites infect only mature erythrocytes. In contrast, *Toxoplasma* is the consummate generalist: a single species is capable of infecting virtually any vertebrate, and virtually

*Author for correspondence (droos@sas.upenn.edu).

any nucleated cell in any tissue within these organisms (Frenkel 1973).

While the behaviour of these parasites is strikingly different, their basic biochemistry, genetics, and subcellular architecture are strikingly similar. All apicomplexan parasites synthesize pyrimidines *de novo,* making antifolate treatment a mainstay for anti-parasite chemotherapy; all are purine auxotrophs, raising interest in purine salvage enzymes as possible targets for new drug treatments (Ullman & Carter 1995). All replicate in haploid form, but undergo sexual recombination in the definitive host (mosquitoes for *Plasmodium,* cats for *Toxoplasma,* etc.) (Frenkel 1973; Levine 1988). All of these parasites exhibit a distinctive 'apical complex' of organelles, for which the phylum Apicomplexa is named: including specialized secretory organelles (rhoptries and micronemes), and cytoskeletal elements (the conoid and associated structures of unknown composition or function) (Chobotar & Scholtyseck 1982). The precise function of these organelles is unknown, but they are thought to be involved in host cell attachment and invasion. Microneme contents are secreted upon host cell contact, and rhoptry contents are released coincident with the establishment of the intracellular parasitophorous vacuole within which replicating parasites reside (Carruthers & Sibley 1997).

Apicomplexan parasites also harbour a variety of other unusual organelles. The inner membrane complex—a flattened patchwork of vesicles—is central to the peculiar mode of division observed in these parasites (endodyogeny, endopolyogeny, schizogeny) (Sheffield & Melton 1968; Aikawa 1971; Chobotar & Scholtyseck 1982; Shaw & Tilney 1992; Roos et al. 1999*a*). Daughter parasites are assembled and packaged within the mother, ultimately capturing the mother's plasma membrane and sloughing off a residual body consisting of indigestible components of the mother, and other products. Perhaps the strangest of all intracellular organelles is the plastid—a chloroplast-like organelle acquired by secondary endosymbiosis of an algal ancestor, and retention of the algal plastid as a distinct (albeit non-photosynthetic) organelle (Köhler *et al.* 1997; McFadden *et al.* 1997; Wilson & Williamson 1997; Blanchard & Hicks 1999; McFadden 1999; Roos *et al.* 1999*b*). The origin and function of the apicoplast has been the subject of much interest, and provides the focus of this paper, as characterizing this organelle serves as a useful case study for genome database mining efforts and effective integration of results from diverse parasites and experimental systems.

2. *TOXOPLASMA GONDII:* MOLECULAR GENETIC AND CELL BIOLOGICAL EXPLORATION OF APICOMPLEXAN PARASITES

Despite their importance as pathogens, and the fascinating biology they present, apicomplexan parasites have not always been easy to study in the laboratory. *In vitro* growth conditions for many of these parasites have only recently been established, and many important apicomplexans remain refractory to cultivation (e.g. *Plasmodium vivax*, *Cryptosporidium parvum*, *Eimeria* sp.). Axenic culture conditions have not been established for any apicomplexan parasite save the highly divergent molluscan parasite *Perkinsus marinus* (Gauthier *et al.* 1995). Although most, if not all, apicomplexans exhibit a sexual cycle, no known culture conditions permit sexual recombination for any apicomplexan parasite at any experimentally useful efficiency. Sexual crosses therefore require mosquitoes, cats, chickens, or other inconvenient laboratory animals (Levine 1988).

Biochemical studies on apicomplexan parasites are often limited by the difficulty of obtaining sufficient material for detailed analysis. For example, it is impractical to obtain more than 1 g (wet weight) of *T. gondii* tachyzoites (the rapidly replicating form of this parasite) (Boothroyd *et al.* 1994; Roos *et al.* 1994). Large numbers of *P. falciparum*-infected red cells can be produced, but extracellular merozoites die within minutes (Schlichterle *et al.* 2000). In cell biological terms, many apicomplexan parasites, including *Plasmodium*, *Babesia*, and *Theileria* exhibit a stripped down morphology that makes it impossible to recognize even the most familiar eukaryotic landmarks (mitochondria, Golgi, etc.) (Aikawa 1971; Chobotar & Scholtyseck 1982; Shaw & Tilney 1992; Hager *et al.* 1999).

The genome of the most important apicomplexan parasite—*P. falciparum*—is notorious for exhibiting the highest A/T content known for any eukaryote, making the manipulation of recombinant molecules a challenge (Fletcher 1998). *Plasmodium falciparum* DNA fragments of even modest size are subject to extensive deletions and rearrangements in bacterial plasmids. Sequencing of the homopolymer runs that are common in the *P. falciparum* genome requires specialized protocols. The A/T-rich *P. falciparum* genome results in extreme codon biases, which complicate protein synthesis in standard expression systems.

The above litany of experimental difficulties poses serious challenges, but a remarkable amount has nevertheless been accomplished, taking advantage of individual systems for the particular opportunities that they offer. For example, experimental crosses have been conducted for both *T. gondii* and *P. falciparum* (Pfefferkorn & Pfefferkorn 1980; Wellems 1991). The ability to produce copious amounts of *Eimeria tenella* oocysts has enabled careful biochemical study of this parasite (Schmatz 1997).

Toxoplasma gondii provides a particularly attractive experimental system. This parasite is readily cultivated in virtually any cell type, permitting analysis of what the parasite can do for itself, versus what is required from the host (through cultivation in suitable host cell mutants) (Pfefferkorn 1988). Ultrastructural studies of *T. gondii* allow easy recognition of the nucleus, mitochondrion, ER, Golgi apparatus, and specialized apical complex organelles. Indeed, this parasite has been advanced as a model eukaryote for cell biological studies (Hager *et al.* 1999). Virtually all of the parasite's subcellular organelles can be labelled with variously coloured fluorescent reporters for quantitative analysis in living cells, permitting molecular dissection of organellar targeting signals (Striepen *et al.* 1998, 2000). Targeting into the secretory pathway is accomplished using a classical N-terminal signal sequence, while targeting to the rhoptries and micronemes may use either internal or C-terminal signals (Di Cristina *et al.* 2000; Striepen *et al.* 2001).

The molecular genetics of *T. gondii* are similarly well developed, offering a wide range of positive and negative

selectable markers, integrating and episomal vectors, and strategies for both homologous and non-homologous recombination (Boothroyd *et al.* 1994; Roos *et al.* 1994). Perhaps the most valuable aspect of *T. gondii* molecular genetics has been the extraordinarily high frequency of stable transformation observed, reported as more than 5% using certain vectors and conditions (Donald & Roos 1994). As a consequence, non-homologous recombination can be used to saturate the parasite genome with insertional transgenes (which can readily be rescued to clone the tagged loci) (Roos *et al.* 1997); homologous recombination can be used for targeted gene knock-outs, knock-ins, and allelic replacements (Donald & Roos 1998); and mutant genes can be cloned by complementation (Black & Boothroyd 1998).

It is not clear why *Toxoplasma* has proved so amenable to experimental analysis, while *Plasmodium* (for example) has proved so difficult. Although any answer to such a question must be purely speculative, it is tempting to suggest that the generalist life history strategy adopted by *Toxoplasma* means that generic approaches developed for use in other systems will also be effective in this organism. By contrast, the specialist's life history strategy adopted by *Plasmodium* means that this parasite is highly divergent (a 'long branch' on virtually any phylogenetic tree). Regardless of the reason, *T. gondii* has provided an invaluable system for studying apicomplexan biology.

3. DISCOVERY OF THE APICOMPLEXAN PLASTID: HORIZONTAL GENETIC TRANSFER FROM A PLANT (ALGA)!

(a) *The target of macrolide antibiotics (and other compounds traditionally known as antibacterial agents)*

One striking example of how studies on *T. gondii* have elucidated the basic aspects of apicomplexan biology is provided by the discovery of the apicomplexan plastid, or apicoplast (Köhler *et al.* 1997; McFadden *et al.* 1997; Wilson & Williamson 1997). These studies are derived from work originally intended to identify the target and mechanism of action of macrolide and lincosamide antibiotics—prokaryotic protein synthesis inhibitors that act by targeting transpeptidation on bacterial ribosomes (Cundliffe 1990). These antibiotics would not normally be expected to be effective against a eukaryotic organism. However, azithromycin, clindamycin, and spiramycin all inhibit the replication of *T. gondii*, both *in vitro* and *in vivo*, and chloramphenicol (another transpeptidation inhibitor specific for prokaryotic ribosomes) has been shown to be effective against *Plasmodium* (Geary *et al.* 1989; Beckers *et al.* 1995; Fichera *et al.* 1995; Fichera & Roos 1997). All of these structurally distinct antibiotics kill parasites with a distinctive 'delayed death' phenotype, arguing for a common mechanism of action (Fichera *et al.* 1995; Fichera & Roos 1997).

However, sequence analysis revealed that neither the cytoplasmic nor the mitochondrial ribosomes are likely to be the target for these antibiotics (Beckers *et al.* 1995). Direct assays have also failed to reveal any effect of these drugs on cytoplasmic protein synthesis or mitochondrial function (it has thus far proved difficult to examine mitochondrial protein synthesis directly). Another possible ribosomal target was suggested by a mysterious 35 kb circular DNA element found in both *Plasmodium* and *Toxoplasma,* and originally thought to comprise the mitochondrial genome (Feagin 1994; Wilson & Williamson 1997). Identification of the true mitochondrial DNA in *Plasmodium* left this 35 kb element without a home or function. The 35 kb element was known to encode ribosomal genes, however, and sequences from the predicted transpeptidation domain indicated that ribosomes encoded by this DNA provide a plausible target for macrolides and lincosamides (Beckers *et al.* 1995).

Proof of the role of the 35 kb element in drug sensitivity first came through the use of fluoroquinolone antibiotics. Under conditions that specifically eliminate the 35 kb circular DNA without affecting nuclear or mitochondrial DNA replication, the DNA gyrase inhibitor ciprofloxacin killed parasites with delayed death kinetics, directly linking this unusual phenotype with the 35 kb circle (Fichera & Roos 1997; Weissig *et al.* 1997). Cell biological experiments have provided a non-pharmacological method for eliminating the 35 kb DNA, with the same effect (He *et al.* 2001*a*,*b*). More recently, a clindamycin-resistant mutant has been shown to harbour point mutations in the transpeptidation domain of ribosomal genes encoded on the 35 kb element (J. C. Boothroyd, personal communication).

(b) *Phylogenetic origins of the 35 kb DNA*

Phylogenetic analysis of the *tufA* gene from the 35 kb circle (encoding elongation factor Tu, a well conserved and broadly sampled gene that is useful for phylogeny reconstruction) revealed that among the prokaryotic world, this gene is most closely related to cyanobacteria—the blue-green algal ancestor of plant chloroplasts (Köhler *et al.* 1997). Given the strong support for apicomplexan affinity with dinoflagellates and ciliates from many phylogenetic studies (Gajadhar *et al.* 1991; Cavalier-Smith 1993), and the equally strong support for affinity of the 35 kb element with cyanobacteria, this element is presumed to have arisen through horizontal genetic transfer. The 35 kb element clusters with plastid genomes rather than with free-living cyanobacteria (Köhler *et al.* 1997; McFadden 1999), arguing that this DNA was acquired by secondary endosymbiosis, as shown in figure 1 (Palmer & Delwiche 1996). In other words, a common ancestor of all apicomplexan parasites 'ate' a eukaryotic alga, whose ancestor had previously engulfed a cyanobacterium, and progeny parasites retained the genome of the algal plastid!

Although the *tufA* phylogeny suggests a green algal ancestry for this DNA, other work indicates that a red algal origin is more probable (McFadden *et al.* 1997; Blanchard & Hicks 1999). More recent studies hint that the secondary endosymbiotic transfer of plastid DNA may have pre-dated the divergence of apicomplexans, dinoflagellates and ciliates (the Alveolata) (Fast *et al.* 2001). This result argues that the apicoplast may be related to some perdinin-containing dinoflagellate plastids (plastids have probably been acquired many times in dinoflagellate evolution), and that the absence of plastids in ciliate lineages is due to secondary loss(es). It also appears that certain apicomplexans (e.g. *C. parvum*) may have secondarily lost their plastid (Zhu *et al.* 2000). Regardless of the precise relationship between the apicomplexan 35 kb DNA

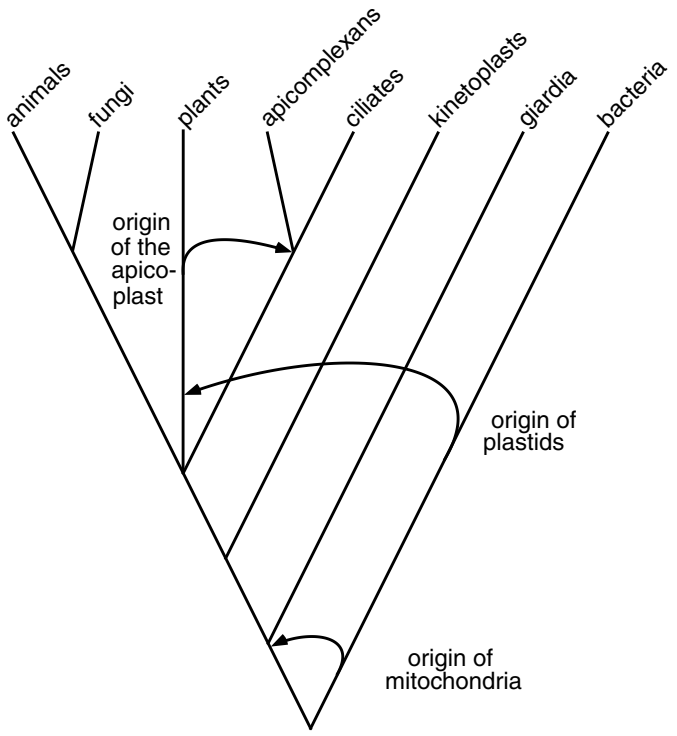

Figure 1. The evolutionary origins of the apicoplast. All plastids are thought to have arisen from a single primary endosymbiotic event, when the ancestor of plants and algae ate (or was invaded by) a free-living cyanobacterium (Delwiche *et al.* 1995; Bhattacharya 1997). The apicoplast arose by secondary endosymbiosis: a common ancestor of apicomplexan parasites engulfed a eukaryotic alga (probably a red alga), and retained the algal plastid (Palmer & Delwiche 1996). This event may have occurred prior to the divergence of apicomplexans, dinoflagellates and ciliates (Fast *et al.* 2001).

and the plastid genomes of algae, plants, and other alveolates, it is clear that apicomplexan parasites harbour a true plastid, rather than a 'plastid-like organelle' (Köhler *et al.* 1997; Blanchard & Hicks 1999).

(c) *Subcellular localization of the apicoplast*

In situ hybridization studies have been difficult in *Plasmodium*, for several reasons. The A/T-rich nature of the apicoplast genome results in high background hybridization with nuclear DNA (particularly for *P. falciparum*). The presence of haemoglobin in infected erythrocytes complicates hybridization in frozen sections. The relatively poor ultrastructural resolution in *Plasmodium*, even under the best of circumstances, makes it difficult to resolve subcellular architecture under the conditions required for nucleic acid hybridization. Fortunately, *in situ* hybridization has been relatively straightforward in *Toxoplasma*, permitting identification of the subcellular home of the 35 kb element—an organelle lying just apical to the nucleus, and adjacent to (but distinct from) the mitochondrion and Golgi apparatus (McFadden *et al.* 1996; Köhler *et al.* 1997). This structure has previously been given many names: the spherical body, the Golgi adjunct, the Hohlzylinder, the organelle plurimembranaire (Siddall 1992), but it is now clear that this organelle is the apicoplast. This organelle is surrounded by four membranes, which is consistent with a secondary endosymbiotic origin (McFadden & Roos 1999; Roos *et al.* 1999*b*).

(d) *Apicoplast proteins*

The efficacy of macrolide antibiotics validates the apicoplast as an essential organelle, and therefore a promising target for parasiticidal chemotherapy ... but what does the apicoplast do? The 35 kb apicoplast genome itself provides no clue (Wilson & Williamson 1997; http://www.sas.upenn.edu/~jkissing/toxomap.html). Apicoplast DNA encodes two large-subunit and two small-subunit ribosomal RNAs, a complete set of tRNAs, ribosomal proteins (17 in *T. gondii*; 18 in *P. falciparum*), three subunits of RNA polymerase, a protein implicated in DNA replication, elongation factor Tu, and a *clp* protease. Five or six small ORFs remain unidentified (in *T. gondii* versus *P. falciparum*, respectively), but their codon bias differs from other genes in the apicoplast genome, suggesting that these ORFs may not be translated. Proteins encoded in the apicoplast genome appear to be synthesized in infinitesimally small amounts, as they have been virtually impossible to detect directly (M. A. Fichera and D. S. Roos, unpublished data). In summary, the apicoplast genome claims the dubious distinction of having the lowest information content of any organellar genome—its 35 kb genome is probably required only for the *clp* gene, to facilitate the import of metabolic enzymes encoded in the nuclear genome.

The vast majority of proteins destined for endosymbiotic organelles (e.g. mitochondria, chloroplasts) are synthesized in the cytoplasm and imported post-translationally (Cline & Henry 1996; Haucke & Schatz 1997; Neupert 1997; Soll & Tien 1998). Examination of random cDNA and genomic sequences emerging from the *T. gondii* EST and *P. falciparum* genome projects revealed various candidate nuclear-encoded apicoplast proteins, based on sequence similarity to known plastid genes in other systems (Cline & Henry 1996; Haucke & Schatz 1997; Martin & Hermann 1998; Martin *et al.* 1998; Soll & Tien 1998). Several of these proteins have been localized within the apicoplast by antibody staining (Waller *et al.* 1998; Jomaa *et al.* 1999; Jelenska *et al.* 2001). Putative nuclear-encoded apicoplast proteins derived from either *T. gondii* or *P. falciparum* have also been expressed in *T. gondii* tachyzoites (Roos *et al.* 1999*b*). Recombinant fusions with enzymatically active reporter molecules should facilitate organelle purification (He *et al.* 2001*a*), while fluorescent protein fusions permit analysis of apicoplast dynamics in living cells, such as studies which investigate how this organelle divides (Striepen *et al.* 2000).

(e) *Targeting to the apicoplast*

In mitochondria and chloroplasts, considerable effort has been invested in identifying the mechanisms whereby nuclear-encoded proteins are imported across the one, two, or three membranes separating distinct organellar compartments from the cytosol (Pilon *et al.* 1995; Cline & Henry 1996; Haucke & Schatz 1997; Martin & Hermann 1998; Martin *et al.* 1998; Soll & Tien 1998). Import signals are not well conserved, typically consisting of weakly basic amphipathic alpha helices; these N-terminal domains are cleaved upon import into the organelle. A neural net approach successfully predicts many chloro-

plast-targeting sequences in higher plants (von Heijne & Nishikawa 1991), but is far less effective than the identification of secretory signal sequences using similar strategies (Nielsen *et al.* 1997). The presence of four membranes surrounding the apicoplast (Köhler *et al.* 1997; McFadden & Roos 1999) poses an unusually complicated barrier to protein import. How are proteins, encoded in the parasite nucleus, targeted into the organelle?

The predicted polypeptide for several nuclear-encoded apicoplast proteins reveals no conserved amino-acid sequence, but these proteins are characterized by long N-terminal extensions relative to the predicted mature protein, and Western blotting shows two bands, of a size consistent with the unprocessed pre-protein and a mature protein (Waller *et al.* 1998). The ability to express recombinant proteins in *T. gondii* confirms that these N-terminal extensions—derived from either *T. gondii* or *P. falciparum* genes—are both necessary, and sufficient, for targeting of proteins (such as GFP) into the apicoplast (Waller *et al.* 1998).

The amino terminus of nuclear-encoded apicoplast proteins looks nothing like a chloroplast- (or mitochondrial-) targeting signal, however. Rather, the N-terminal 15–40 amino acids are usually hydrophobic in nature, and look remarkably similar to a eukaryotic secretory signal sequence. Considered in isolation, these domains also function as secretory signals, targeting reporter molecules through the secretory pathway in transfected parasites (Roos *et al.* 1999*b*; DeRocher *et al.* 2000; Waller *et al.* 2000; Yung *et al.* 2001). Moreover, while removal of the N-terminal hydrophobic domain from nuclear-encoded apicoplast proteins abolishes organellar targeting, targeting to the apicoplast can be restored by replacing this domain with any generic signal sequence—such as signals derived from parasite surface proteins, or even mammalian viral glycoproteins. The subterminal domain of nuclear-encoded apicoplast proteins (extending from *ca.* 40 to 200 amino acids downstream of the secretory signal sequence) looks more like a chloroplast-targeting signal, and is also necessary for targeting to the apicoplast. Remarkably, this domain can be replaced with a known chloroplast-targeting signal (derived from the *ftsZ* gene of the higher plant (*Arabidopsis thaliana*; M. J. Crawford, unpublished data).

In sum, targeting to the apicoplast is accomplished using a bipartite leader, as shown in figure 2 (Roos *et al.* 1999*b*). A secretory signal sequence is responsible for translocation across one membrane into the ER, followed (presumably) by signal peptidase cleavage to expose a plastid-targeting domain. This second domain then mediates translocation across the remaining membranes into the apicoplast lumen (where processing to the mature protein occurs (He *et al.* 2001*b*). Similar pathways have been proposed for targeting to other secondary endosymbiotic plastids (Apt *et al.* 1995; Bodyl 1997; Lang *et al.* 1998; Schwartzbach *et al.* 1998), although the lack of transfection systems for these algae precludes direct experimental confirmation. Further cut-and-paste molecular genetic experiments indicate that it may be possible to distinguish subdomains within the plastid transit peptide, dissociating microdomains involved in targeting to the apicoplast, membrane translocation, and protein processing. These experiments also suggest the possibility of,

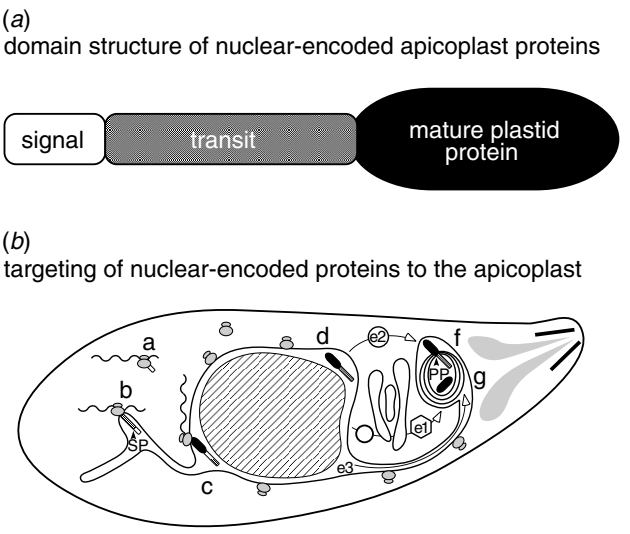

Figure 2. Protein import into the apicoplast. The vast majority of apicoplast proteins—including all proteins involved in metabolic activities—are encoded in the nuclear genome and post-translationally imported into the apicoplast. All nuclear-encoded apicoplast proteins identified to date include a bipartite leader sequence that is both necessary and sufficient for import into the organelle (Waller *et al.* 1998). (*a*) The structure of the acyl carrier protein (drawn to scale). Translocation is believed to occur as follows (Roos *et al.* 1999*b*). (*b*) Synthesis of an N-terminal signal sequence (a) mediates SRP-dependent targeting to the rough ER, and cotranslational sec61p-dependent translocation across the ER membrane (b). The signal sequence is then cleaved by SP to expose a plastid transit peptide (c). This pro-protein then wends its way through the ER to the apical end of the nucleus (d). The precise mechanism of targeting to the apicoplast is unknown, but may proceed via the classical ER–Golgi secretory pathway (e1), via direct vesicle transport from the ER (e2), or via connections between the ER and Golgi (e3) (see § 3 for further discussion). Once proteins reach the apicoplast, the transit peptide mediates translocation across the remaining membranes (f). Cleavage by a PP releases the mature protein into the apicoplast lumen (g).

as yet, undefined commonalities between targeting to the apicoplast and targeting to the rhoptries (O. S. Harb, unpublished data).

This 'Rube Goldberg' scheme provides a clever strategy for combining two independent translocation domains (secretory signal sequence + plastid transit peptide) to produce a composite signal capable of targeting nuclear-encoded apicoplast proteins across four membranes. How nuclear-encoded plastid proteins in the ER reach the apicoplast remains uncertain, however. Traffic via the Golgi seems logical ('e1' in figure 2), and would explain the close apposition of the apicoplast with the Golgi apparatus, but two lines of evidence argue against this hypothesis. Treatment with Brefeldin A disrupts the Golgi and inhibits targeting and processing of rhoptry proteins in *T. gondii*, but fails to inhibit the targeting or processing of apicoplast proteins (K. M. Hager, unpublished data). Similarly, appending a C-terminal ER-retrieval signal (HDEL) (Hager *et al.* 1999) to recombinant nuclear-encoded apicoplast proteins fails to inhibit apicoplast targeting, while efficiently retaining other secretory proteins within the ER. It is possible that specialized vesicles target

proteins directly from the ER to the apicoplast ('e2' in figure 2). Alternatively, the apicoplast might physically reside within the ER ('e2' in figure 2); consistent with this hypothesis, ER is frequently observed in close association with the apicoplast (M. K. Shaw, unpublished data).

(f) *But what does the apicoplast do?*

The above studies identified several nuclear-encoded apicoplast proteins, most of which are probably involved in housekeeping functions (e.g. ribosomal proteins), and a few probable enzymes (based on sequence similarity to proteins of known function) (Waller *et al.* 1998). This information provided anecdotal hints suggesting a role for the apicoplast in lipid metabolism, but one would prefer a complete metabolic pathway map for the organelle. With the genome sequence for *P. falciparum* now virtually complete (Gardner *et al.* 1998; Bowman *et al.* 1999; the *Plasmodium* Genome Database Collaborative 2001), all nuclear-encoded apicoplast genes are now available ... if only we could recognize them!

Unfortunately, as noted above, there is no reliable method for recognizing nuclear-encoded apicoplast proteins (as is the case for proteins destined for mitochondria or chloroplasts): apicoplast-targeting signals are highly degenerate. Probable homologues of chloroplast proteins, known to be encoded in the nuclear genome of plants or algae (Cline & Henry 1996; Haucke & Schatz 1997; Soll & Tien 1998; Martin & Hermann 1998; Martin *et al.* 1998) include hundreds of candidate genes, including, for example, cytoplasmic and mitochondrial ribosomal proteins, in addition to their counterparts destined for the apicoplast. Similarly, predicted signal sequences may be found on proteins destined for the rhoptries, micronemes, dense granules, parasite plasma membrane, parasitophorous vacuole, or host cell ... in addition to nuclear-encoded plastid proteins. Conducting such searches on a genome-wide scale and integrating the results that emerge from these searches is greatly facilitated using computational databases, particularly where queries can be integrated into the framework of a relational database.

4. PlasmoDB: DESIGNING AND MINING THE *PLASMODIUM* GENOME DATABASE

The rapid progress of the *P. falciparum* Genome Sequencing Project (Fletcher 1998; Gardner *et al.* 1998; Bowman *et al.* 1999), and the scale of this undertaking, requires a centralized database providing access to primary sequences emerging in the three sequencing centres involved in this project (The Sanger Centre, Hinxton, UK; the Institute for Genome Research, Rockville, MD, USA; and Stanford University, Stanford, CA, USA), along with ancillary analyses of gene and protein predictions, expression data, etc. The *Plasmodium* Genome Database, designated 'PlasmoDB' was established in June 2000, and is accessible online at http://PlasmoDB.org (The *Plasmodium* Genome Database Collaborative 2001) (figure 3). A CD-based version of PlasmoDB (*Plasmodium* GenePlot) is also available for researchers lacking reliable high-speed web-access. By the time this report appears in print, the current release is on track to reach *ca.* 10 000 hits per day, around the clock, from *ca.* 100 countries. PlasmoDB seeks to provide 'one-stop shopping' for *P. falciparum* genomics research.

Both finished, annotated genomic sequence and unfinished data (raw sequence reads and draft contig assemblies) are available in PlasmoDB, in recognition of the fact that for large eukaryotic genomes, the time between initial sequence production and the release of annotated chromosomes or contigs may stretch to years. Unfinished data (such as provisional assemblies of fivefold shotgun coverage) are also likely to be of increasing importance, given the economics of genome sequencing. Shotgun sequencing to an average depth of fivefold is expected to yield more than 90% of the complete DNA sequence, and identify virtually all genes in a genome the size of *Plasmodium* or related species—at a small fraction of the total cost required for sequencing to completion (because gap closure and chromosome 'finishing' is highly labour-intensive and therefore costly).

Of course, sequence data alone are of limited value, and careful annotation is impractical for unfinished sequences, given the constantly changing nature of the underlying data. PlasmoDB provides a variety of automated analyses (see the Acknowledgements page in PlasmoDB for proper citation of relevant data sources and software tools), including

(i) calculations of A/T bias;
(ii) identification of simple sequence repeats and conserved sequence motifs;
(iii) gene predictions based on a variety of algorithms;
(iv) predicted protein coding sequence and secondary structure;
(v) protein sequence motif identification;
(vi) results of BLAST searches against GenBank;
(vii) searches for near-exact matches among the unfinished sequence dataset (permitting identification of overlapping sequences resulting from mis-assembly of unfinished sequence data);
(viii) comparisons between *P. falciparum* and *P. yoelii*.

These analyses are also provided for finished, annotated sequences, enabling researchers to examine competing models whenever the 'official' annotation conflicts with experimental results in the laboratory. Additional projects aim to

(i) integrate the physical, genetic and optical sequence maps;
(ii) incorporate further sequence annotation and curation, including metabolic pathway analyses and references to published research;
(iii) incorporate results from RNA expression analysis (EST sequencing, SAGE, glass slide and Affymetrix oligonucleotide arrays) and proteomics projects being conducted in various laboratories around the world;
(iv) provide cross-comparisons with sequencing data from related species; genome and/or EST sequencing projects are currently underway for several species of *Plasmodium* parasites (*P. vivax*, *P. knowlesi*, *P. yoelii*, *P. berghei*, *P. chabaudi*, *P. gallinaceum*, and a field isolate of *P. falciparum*), and for various other apicomplexan species (including *Babesia bovis*, *Theileria annulata*, *Theileria parva*, *Sarcocystis neu-*

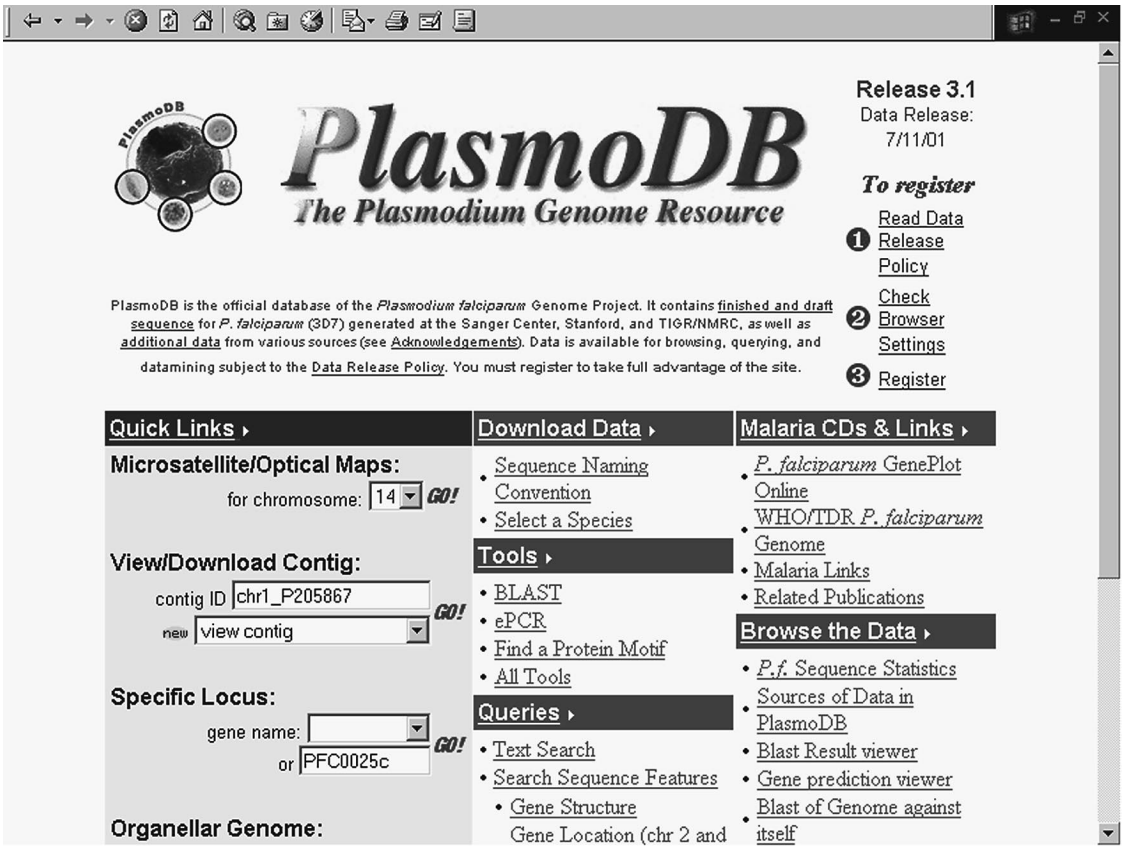

Figure 3. The *Plasmodium* Genome Database 'PlasmoDB'. Home page of the *Plasmodium* Genome Database (v. 3.0), noting URL http://PlasmoDB.org, data sources, and the data release and registration policy (The *Plasmodium* Genome Database Collaborative 2001). Links provide rapid access to a variety tools for graphical viewing of the data and user-based queries.

rona, *Neospora caninum*, *T. gondii*, *E. tenella*, and *C. parvum*) (Tarleton & Kissinger 2001).

All of the above data are incorporated into a large relational database using a highly structured format to accommodate diverse datasets related to genomic sequence and gene expression (The *Plasmodium* Genome Database Collaborative 2001; Davidson *et al.* 2001). Tools have been designed to facilitate complex biological queries, including many that are specific to *Plasmodium* parasites and malaria as a disease. For example, investigators might wish to identify

(i) multigene families;
(ii) predicted proteins exhibiting a particular user-defined motif;
(iii) predicted proteins exhibiting similarity to known proteases in other species;
(iv) highly polymorphic genes encoding proteins destined for the parasite surface;
(v) genes on chromosome 7 that are expressed only during gametocytogenesis;
(vi) *var* genes with an exon–intron structure that falls outside the normal range for this variant surface antigen gene family.

5. COMPUTATIONAL DATABASE MINING: A METABOLIC PATHWAY MAP FOR THE APICOPLAST

The ability to formulate and integrate complex queries such as those suggested above (not all of which are yet possible in PlasmoDB) permits vastly more efficient identification of nuclear-encoded apicoplast genes than is possible by examining individual ESTs one at a time, or by running individual BLAST searches for specific chloroplast proteins in the *P. falciparum* database. A schema for the identification and validation of nuclear-encoded apicoplast proteins is provided in figure 4 (Roos *et al.* 1999*c*).

Data to be queried include all available apicomplexan parasite sequences. In practice, the *P. falciparum* genome database provides the most complete resource at present, but the *T. gondii* EST database (Ajioka *et al.* 1998) has also been shown to include several nuclear-encoded apicoplast proteins. Several large-scale genome and/or EST sequencing projects are currently underway for various other apicomplexans, as noted above (Tarleton & Kissinger 2001). *Cryptosporidium* provides a particularly interesting case, as circumstantial evidence suggests that this parasite lacks an apicoplast (Zhu *et al.* 2000), and preliminary analysis of the *C. parvum* genome supports this hypothesis.

These data are then subjected to a variety of computational queries, many of which can be conducted within PlasmoDB (although others require separate analysis, usually based on straightforward PERL scripts; http://www.bioperl.org). As noted above, virtually all of these queries, considered in isolation, suffer from an unacceptably high false positive rate (identification of all secretory proteins, for example, yields primarily non-apicoplast proteins). However, combining queries to identify genes with multiple 'hits' (e.g. proteins containing a signal sequence and a long N-terminal extension that might possibly serve as

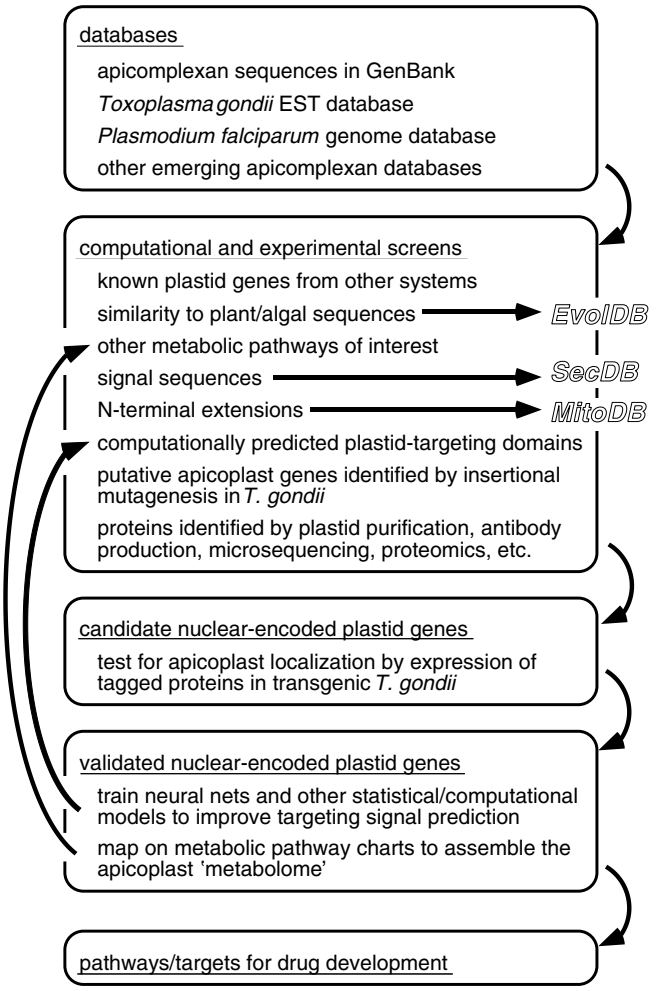

Figure 4. Schema for identifying nuclear-encoded apicoplast proteins. Data from various sources are subjected to a variety of computational or experimental screens to identify candidate nuclear-encoded plastid genes (Jelenska *et al.* 2001). These screens also reveal numerous genes of potential interest for other studies: phylogenetically-conserved sequences that may be useful for broad spectrum vaccine or drug design (Beckers *et al.* 1995); genes that may have been derived from the alga that donated the original apicoplast (EvolDB) (Palmer & Delwiche 1996; Martin & Hermann 1998; Martin *et al.* 1998); potential secretory pathway proteins associated with the parasite surface, the parasitophorous vacuole, modification of the host cell, attachment or invasion (SecDB) (Roos *et al.* 1999*a*; Lingelback & Joiner 1998); and candidate mitochondrial proteins (MitoDB) (Haucke & Schatz 1997; Neupert 1997). Validation in the *T. gondii* system distinguishes true and false positives (Waller *et al.* 1998; Jomaa *et al.* 1999; Jelenska *et al.* 2001), both of which are useful for devising new computational or statistical screens (e.g. training neural nets; von Heijne & Nishikawa (1991); Nielsen et al. (1997)) to improve the identification of nuclear-encoded apicoplast proteins. Validated hits provide further clues for database searches, candidate targets for drug development (McFadden & Roos 1999), and clues to the characterization of apicoplast metabolic function.

a plastid-targeting domain and show significant similarity to proteins known to be targeted into chloroplasts), makes it possible to assemble a database of candidate nuclear-encoded apicoplast genes. For example, while *ca.* 7% of the entire *P. falciparum* genome (*ca.* 600 predicted genes) contain probable signal sequences, only about 60 of these also exhibit probable N-terminal extensions. N-terminal extensions are difficult to recognize, resulting in a high rate of both false positives and false negatives, but many of the above genes are clear homologues of known chloroplast proteins. This list undoubtedly probably contains several false positives, and certainly fails to include many proteins destined for the apicoplast. It is important to recognize, however, that the goal in computational database mining is not to provide the correct answer without resorting to experimental analysis, but to obtain an answer that can feasibly be tested experimentally.

Candidate nuclear-encoded apicoplast proteins are easily tested in *T. gondii,* as noted above (Waller *et al.* 1998; Roos *et al.* 1999*b*). For example, recombinant fusions between the gene in question and a yellow fluorescent protein reporter can be transfected into transgenic parasites in which the apicoplast is already labelled with a cyan fluorescent protein. Dozens, or even hundreds, of candidates can readily be tested in this manner to distinguish true versus false positives. The availability of apicoplast-deficient mutants also facilitates analysis of candidate genes (He *et al.* 2001*a*). Validated nuclear-encoded apicoplast genes are potential targets for drug development (Siddall 1992), and can also be used to build up a picture of apicoplast metabolic pathways, as outlined in figure 5.

Both housekeeping and metabolic functions appear to be associated with the apicoplast. The former category includes the A and B subunits of a type II topoisomerase that is the probable target of ciprofloxacin (Fichera & Roos 1997; Weissing *et al.* 1997), the *rpoA* subunit of RNA polymerase that is 'missing' from the apicoplast genome (Wilson & Williamson 1997; http://www.sas.upenn.edu/~jkissing/toxomap.html), several tRNA synthetases, numerous ribosomal proteins and other transcription and translation factors, putative organellar import proteins (tic and toc homologues) (Cline & Henry 1996; Haucke & Schatz 1997; Soll & Tien 1998), and proteins involved in maintaining redox balance (Vollmer *et al.* 2001). Biosynthetic functions include pathways associated with the formation of amino acids, haem, isoprenoids (Jomaa *et al.* 1999), and fatty acids (Waller *et al.* 1998; Jelenska *et al.* 2001). The import of triose phosphates and/or PEP probably serves as a carbon source for these reactions, although hexose transport cannot be excluded. Direct transport of metabolites such as amino acids, pyruvate and ACoA is also a possibility, but appropriate transporters have not been identified to date. Conversion of PEP to pyruvate through the action of pyruvate kinase probably provides a key metabolite, required for the synthesis of amino acids involved in haem biosynthesis. Pyruvate is also required for the synthesis of isoprenoids from DOXP (Jomaa *et al.* 1999; Eisenreich *et al.* 2001), and for production of ACoA for fatty acid synthesis/elongation (Somerville 1991; Jelenska *et al.* 2001).

Several (but not yet all) of the enzymes necessary for haem biosynthesis have been identified in the *P. falciparum* database. Preliminary analysis suggests that at least the early components of the haem biosynthetic pathway are present within the apicoplast (as in chloroplasts), but these pathways also share similarities with mitochondrial haem biosynthesis. The discovery of DOXP pathway enzymes in *P. falciparum* caused considerable excitement (Jomaa

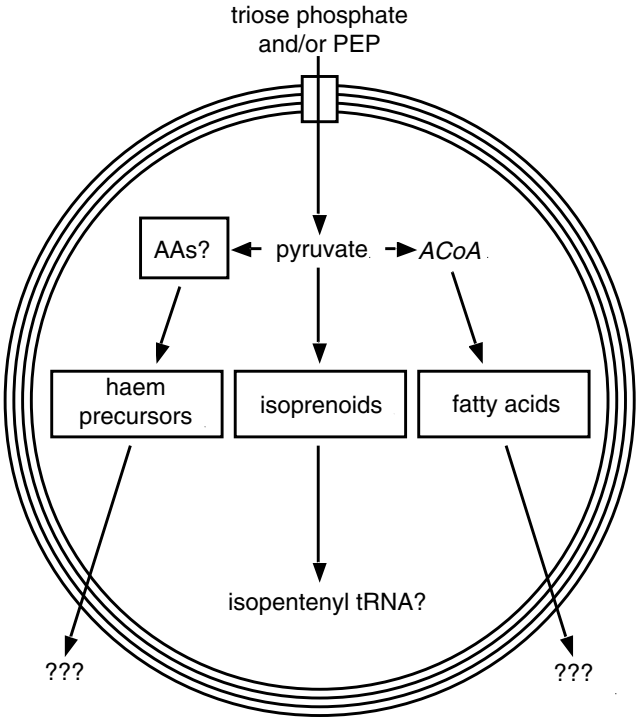

Figure 5. A (nearly) complete metabolic pathway map for the apicoplast? More than 150 probable nuclear-encoded apicoplast proteins have been identified to date. Most of these proteins are associated with housekeeping functions (plastid DNA replication, RNA transcription, protein synthesis, protein import, maintaining redox balance, etc.). Biosynthetic activities of the apicoplast depend on pyruvate, which is probably acquired through the import of sugars and/or PEP and subsequent action of pyruvate kinase. Isoprenoids are synthesized via the DOXP pathway, and are probably required for tRNA isopentenylation. Amino acids are required for haem biosynthesis, the initial stages of which (amino-levulinic acid production) appear most similar to mitochondrial pathways, while later stages may be more akin to chloroplast pathways. ACoA is produced by a pyruvate dehydrogenase complex, and used for the synthesis of fatty acids via ACoA carboxylase and a type II FAS complex. The structure and functions of metabolites produced by these pathways are not yet known in detail, but many of the enzymes identified provide potential targets for drug development. See §5 for further details.

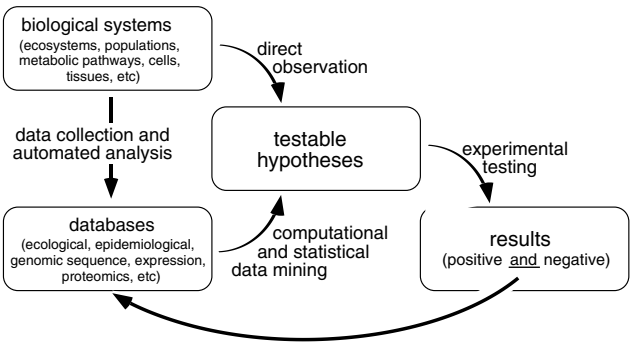

Figure 6. Integrating computational and experimental biology. Biologists traditionally look for patterns within complex systems (tropical rainforests, cells, populations) by applying whatever tools are available (microscopes, gel electrophoresis, statistical analyses), and using the resulting observations to develop hypotheses that can be tested experimentally. The growth of large-scale datasets increasingly mandates the use of computational databases and machine-learning algorithms. Integrating computational and experimental approaches is challenging (Roos 2001), but particularly rewarding, as discussed in § 6.

et al. 1999), given the lack of an HMGCoA pathway in apicomplexan parasites. However it is unlikely that bulk parasite cholesterol is synthesized in the apicoplast; most cholesterol is probably salvaged from the host cell (Coppens *et al.* 2000). The DOXP pathway is probably required for isoprenylation of tRNAs, as in plant chloroplasts (Somerville 1991), but additional activities are also possible. An apicoplast-specific ACoA carboxylase (Jelenska *et al.* 2001) and all subunits of a type II fatty acyl synthase complex (Waller *et al.* 1998) have been identified in the *P. falciparum* genome (and several have also been identified in *T. gondii*), but as for the haem and DOXP pathways, the precise nature and function of the metabolites produced remain undefined.

It is also important to note that there may be significant differences in the function of the apicoplast in different parasites. The apicoplast genome itself is highly conserved between *P. falciparum* and *T. gondii,* but uses a different genetic code (UGA-encoded tryptophan) (http://www.sas.edu/~jkissing/toxomap.html). Interestingly, there is, as yet, no evidence for the presence of a DOXP pathway for isoprenoid biosynthesis in *T. gondii,* and inhibitors of this pathway fail to kill *T. gondii* tachyzoites (Jomaa *et al.* 1999).

At this stage, it is difficult to enumerate precisely how many nuclear-encoded apicoplast genes have been identified, because this depends on which criteria are considered acceptable for validation. Possible criteria include:

(i) an unequivocal bipartite targeting sequence;
(ii) localization to the apicoplast by immunostaining;
(iii) targeting of fluorescent proteins to the organelle;
(iv) failure to accumulate in plastid-deficient parasites;
(v) components of a plausible plastid metabolic pathway;
(vi) pharmacologically inhibitable using apicoplast-specific drugs.

Only a handful of proteins are positive on all counts, but based on a combination of computational, experimental and theoretical criteria, over 150 proteins can probably be claimed as destined for the apicoplast. This catalogue of apicoplast proteins is still incomplete, but it is doubtful if any major pathways have been overlooked.

6. FINAL COMMENTS: INTEGRATING COMPUTATIONAL AND EXPERIMENTAL BIOLOGY

Traditional biological research differs from computational research in that, in the former, experiments are carried out at the bench or in the field, rather than at the computer. The descending staircase shown in figure 6 might reasonably be described as the 'central dogma' of biological research: observation of poorly-defined biological systems with whatever tools are at hand, to produce hypotheses that can be tested *in vivo.* Computational

biology adds large-scale biological databases to this mix, providing another class of systems for analysis.

Biological databases differ from primary biological systems (e.g. a tropical rainforest) in being more rigorously defined, and are therefore amenable to computational analysis. An additional feature of genome-scale databases is that their completeness makes negative data more informative than has usually been the case in biological studies (Roos 2001). For example, it is now possible to state categorically that *P. falciparum* lacks the genes encoding HMGCoA reductase, or thymidine kinase—rather than making the much weaker statement that these genes have not yet been found. Keeping track of both positive and negative data makes it possible to develop new computational and statistical models that better model biological observations (feedback loops in figures 4 and 6), thereby enlisting computational researchers as intellectual colleagues in biological investigation, rather than merely as technical support employed to write code.

As scientific colleagues, our goal is neither to ignore computational tools, nor to solve biological problems through purely computational means, but to take advantage of whatever each approach does best. In the example presented above, our incomplete understanding of the biological algorithms underlying apicoplast function means that no computational tool is yet able to recognize nuclear-encoded apicoplast genes with acceptable accuracy. Similarly, the available technology is, as yet, inadequate to the task of identifying all nuclear-encoded apicoplast genes by purely experimental means. In combination, however, the time required to generate the pathway map outlined in figure 5 (approximately two years) represents surprisingly rapid progress, especially when compared with the effort required to produce a similar understanding of chloroplasts (for example).

Successful integration of computational and experimental approaches (Roos 2001) requires that computer scientists and statisticians come to recognize that the constantly changing nature of biological understanding means that simple queries are often more valuable than complex algorithms that must be rewritten frequently as the biological data and/or questions change. Biologists must come to terms with computational and statistical analyses as a powerful new class of tools. As generating data becomes increasingly easier, making sense of this deluge becomes more difficult. The day is not far off when more biology will be done at the computer than at the bench!

Genome projects are large-scale undertakings, and none of the above work would have been possible without the impressive achievements of the *T. gondii* EST project (Ajioka *et al*. 1998), the *P. falciparum* Genome Project (Fletcher 1998; Gardner *et al*. 1998; Bowman *et al.* 1999), and their associated databases (Ajioka *et al*. 1998; The *Plasmodium* Genome Database Collaborative 2001). The *T. gondii* EST project has been a collaborative undertaking between investigators at Washington University, Cambridge University, The Sanger Centre, Stanford University, Montana State University, and the University of Pennsylvania, with support from Merck Research Laboratories, the Wellcome Trust, and the US National Institutes of Health. The *T. gondii* Database of Clustered ESTs (http://www.cbil.upenn.edu/ParaDBs/Toxoplasma/index.html) was developed in conjunction with Brian Brunk, Chris Overton, and Li Li (University of Pennsylvania). *Plasmoduim falciparum* genome sequencing, assembly and annotation was conducted by The Sanger Centre (chromosomes 1, 3–9, and 13), a consortium composed of The Institute for Genome Research, along with the Naval Medical Research Center (chromosomes 2, 10, 11, and 14) and The Stanford Genome Technology Center (chromosome 12), with support from the Wellcome Trust, US National Institutes of Health (NIAID), US Department of Defense, and the Burroughs Wellcome Fund. The *Plasmodium* Genome Database PlasmoDB (http://PlasmoDB.org) was developed in collaboration with colleagues at Melbourne University, and the Computational Biology and Informatics Laboratory at the University of Pennsylvania, with support from the Burroughs Wellcome Fund. The authors thank the scientists and funding agencies supporting these projects for making sequence data publicly available prior to publication of the completed sequence. The authors also thank the relevant research communities for having supported these projects. Finally, the authors thank colleagues in the Computational Biology Program at the University of Pennsylvania, and the many researchers whose insight into the evolution and biology of the apicoplast has contributed substantially to this review, particularly present and former members of the Haselkorn laboratory (Chicago, USA), Keeling laboratory (Vancouver, Canada), McFadden laboratory (Melbourne, Australia), Palmer laboratory (Bloomington, USA), Roos laboratory (Philadelphia, USA), Vaidya laboratory (Philadelphia, USA) and Wilson laboratory (Mill Hill, UK). D.S.R. is the recipient of a Scholar Award in Molecular Parasitology from the Burroughs Wellcome Fund.

REFERENCES

Aikawa, M. 1971 *Plasmodium*: the fine structure of malarial parasites. *Exp. Parasitol* **30**, 284–320.

Ajioka, J. W. (and 12 others) 1998 Gene discovery by EST sequencing in *Toxoplasma gondii* reveals sequences restricted to the apicomplexa. *Genome Res*. **8**, 18–28.

Apt, K. E., Clendennen, S. K., Powers, D. A. & Grossman, A. R. 1995 The gene family encoding the fucoxanthin chlorophyll proteins from the brown alga *Macrocystis pyrifera*. *Mol. Gen. Genet.* **246**, 455–464.

Beckers, C. J. M., Roos, D. S., Donald, R. G. K., Luft, B. J., Schwab, J. C. & Joiner, K. A. 1995 Inhibition of cytoplasmic and organellar protein synthesis in *Toxoplasma gondii*: implications for the target of macrolide antibiotics. *J. Clin. Invest.* **95**, 367–376.

Bhattacharya, D. (ed.) 1997 *Origins of algae and their plastids*. Vienna: Springer.

Black, M. W. & Boothroyd, J. C. 1998 Development of a stable episomal shuttle vector for *Toxoplasma gondii*. *J. Biol. Chem.* **273**, 3972–3979.

Blanchard, J. L. & Hicks, J. S. 1999 The non-photosynthetic plastid in malaria parasites and other apicomplexans is derived from outside the green plastid lineage. *J. Eukaryotic Microbiol.* **46**, 367–375.

Bodyl, A. 1997 Mechanism of protein targeting to the chlorarachniophyte plastids and the evolution of complex plastids with four membranes: a hypothesis. *Botan. Acta* **110**, 395–400.

Boothroyd, J. C., Kim, K., Pfefferkorn, E. R., Sibley, L. D. & Soldati, D. 1994 Forward and reverse genetics in the study of the obligate intracellular parasite *Toxoplasma gondii*. *Meth. Mol. Genet.* **3**, 1–29.

Bowman, S. (and 35 others) 1999 The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **400**, 532–538.

Carruthers, V. & Sibley, L. D. 1997 Sequential protein secretion from three distinct organelles of *Toxoplasma gondii* accompanies invasion of human fibroblasts. *Eur. J. Cell Biol.* **73**, 114–123.

Cavalier-Smith, T. 1993 Kingdom protozoa and its 18 phyla. *Microbiol. Rev.* **57**, 953–994.

Chobotar, W. & Scholtyseck, E. 1982 Ultrastructure. In *The biology of the Coccidia* (ed. D. M. Hammond), pp. 101–165. Baltimore, MD: University Park Press.

Cline, K. & Henry, R. 1996 Import and routing of nucleus-encoded chloroplast proteins. *A. Rev. Cell Devl Biol.* **12**, 1–26.

Coppens, I., Sinai, A. P. & Joiner, K. A. 2000 *Toxoplasma gondii* exploits host low-density lipoprotein receptor-mediated endocytosis for cholesterol acquisition. *J. Cell Biol.* **149**, 167–180.

Cundliffe, E. 1990 Recognition sites for antibiotics within rRNA. In *The ribosome: structure, function, and evolution* (ed. W. H. Hill, A. Dahlberg, R. A. Garrett, P. B. Moore, D. Schlessinger & J. R. Warner), pp. 479–490. Washington, DC: American Society of Microbiology.

Davidson, S., Crabtree, J., Brunk, B. P., Schug, J., Tannen, V., Overton, G. C. & Stoeckert Jr, C. J. 2001 K2/Klesli and GUS: experiments in integrated access to genomic data sources. *IBM Syst. J.* **40**, 512–531.

Delwiche, C. F., Kuhsel, M. & Palmer, J. D. 1995 Phylogenetic analysis of *tufA* sequences indicates a cyanobacterial origin of all plastids. *Mol. Phylogenet. Evol.* **4**, 110–128.

DeRocher, A., Hagen, C. B., Froehlich, J. E., Feagin, J. E. & Parsons, M. 2000 Analysis of targeting sequences demonstrates that trafficking to the *Toxoplasma gondii* plastid branches off the secretory system. *J. Cell Sci.* **113**, 3969–3977.

Di Cristina, M., Spaccapelo, R., Soldati, D., Bistoni, F. & Crisanti, A. 2000 Two conserved amino acid motifs mediate protein targeting to the micronemes of the apicomplexan parasite *Toxoplasma gondii*. *Mol. Cell Biol.* **20**, 7332–7341.

Donald, R. G. K. & Roos, D. S. 1994 Homologous recombination and gene replacement at the dihydrofolate reductase/thymidylate synthase locus in *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **63**, 243–253.

Donald, R. G. K. & Roos, D. S. 1998 Gene knock-outs and allelic replacements in *Toxoplasma gondii*: HXGPRT as a selectable marker for 'hit-and-run' mutagenesis. *Mol. Biochem. Parasitol.* **91**, 295–305.

Eisenreich, W., Rohdich, F. & Bacher, A. 2001 Deoxyxylulose phosphate pathway to terpenoids. *Trends Plant Sci.* **6**, 78–84.

Fast, N. M., Kissinger, J. C., Roos, D. S. & Keeling, P. J. 2001 Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol. Biol. Evol.* **18**, 418–426.

Feagin, J. E. 1994 The extrachromosomal DNAs of apicomplexan parasites. *A. Rev. Microbiol.* **48**, 81–104.

Fichera, M. E. & Roos, D. S. 1997 A plastid organelle as a drug target in apicomplexan parasites. *Nature* **390**, 407–409.

Fichera, M. E., Bhopale, M. K. & Roos, D. S. 1995 *In vitro* assays elucidate peculiar kinetics of clindamycin action against *Toxoplasma gondii*. *Antimicrob. Agents Chemother.* **39**, 1530–1537.

Fletcher, C. 1998 The *Plasmodium falciparum* Genome Project. *Parasitol. Today* **14**, 342–344.

Frenkel, J. K. 1973 Toxoplasmosis. A parasite life cycle, pathology, and immunology. In *The Coccidia* (ed. D. M. Hammond), pp. 343–410. Baltimore, MD: University Park Press.

Gajadhar, A. A., Marquardt, W. C., Hall, R., Gunderson, J., Ariztia-Carmona, E. V. & Sogin, M. L. 1991 Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Crypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Mol. Biochem. Parasitol.* **45**, 147–154.

Gardner, M. J. (and 26 others) 1998 Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**, 1126–1132.

Gauthier, J. D., Feig, B. & Vasta, G. R. 1995 Effect of fetal bovine serum glycoproteins on the *in vitro* proliferation of the oyster parasite *Perkinsus marinus*: development of a fully defined medium. *J. Eukaryotic Microbiol.* **42**, 307–313.

Geary, T. G., Divo, A. A. & Jensen, J. B. 1989 Stage specific actions of antimalarial drugs on *Plasmodium falciparum* in culture. *Am. J. Trop. Med. Hyg.* **40**, 240–244.

Hager, K. M., Striepen, B., Tilney, L. G. & Roos, D. S. 1999 The nuclear envelope serves as an intermediary between the ER and Golgi complex in the intracellular parasite *Toxoplasma gondii*. *J. Cell Sci.* **112**, 2631–2638.

Haucke, V. & Schatz, G. 1997 Import of proteins into mitochondria and chloroplasts. *Trends Cell Biol.* **7**, 103–106.

He, C. Y., Shaw, M. K., Pletcher, C. H., Striepen, B., Tilney, L. G. & Roos, D. S. 2001*a* A plastid segregation defect in the protozoan parasite *Toxoplasma gondii*. *EMBO J.* **20**, 330–339.

He, C. Y., Striepen, B., Pletcher, C. H., Murray, J. M. & Roos D. S. 2001*b* Targeting and processing of nuclear-encoded apicoplast proteins in plastid segregation mutants of *Toxoplasma gondii*. *J. Biol. Chem.* **276**, 28 436–28 442.

Jelenska, J., Crawford, M. J., Harb, O. S., Zuther, E., Haselkorn, R., Roos, D. S. & Gornicki, P. 2001 Subcellular localization of acetyl-CoA carboxylase in the apicomplexan parasite *Toxoplasma gondii*. *Proc. Natl Acad. Sci. USA* **98**, 2723–2728.

Jomaa, H. (and 12 others) 1999 Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* **285**, 1573–1576.

Köhler, S., Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, P., Wilson, R. J. M., Palmer, J. D. & Roos, D. S. 1997 A plastid of probable green algal origin in apicomplexan parasites. *Science* **275**, 1485–1489.

Lang, M., Apt, K. E. & Kroth, P. G. 1998 Protein transport into 'complex' diatom plastids utilizes two different targeting signals. *J. Biol. Chem.* **273**, 30 973–30 978.

Laughon, B. E. (and 20 others) 1991 Summary of the workshop on future directions in discovery and development of therapeutic agents for opportunistic infections associated with AIDS. *J. Infect. Dis.* **164**, 244–251.

Levine, N. D. 1988 Progress in taxonomy of the apicomplexan protozoa. *J. Protozool.* **35**, 518–520.

Lingelbach, K. & Joiner, K. A. 1998 The parasitophorous vacuole membrane surrounding *Plasmodium* and *Toxoplasma*: an unusual compartment in infected cells. *J. Cell Sci.* **111**, 1467–1475.

Luft, B. J. & Remington, J. S. 1992 Toxoplasmic encephalitis in AIDS patients. *Clin. Infect. Dis.* **15**, 211–222.

McAuley, J. (and 20 others) 1994 Early and longitudinal evaluations of treated infants and children and untreated historical patients with congenital toxoplasmosis: the Chicago Collaborative Treatment Trial. *Clin. Infect. Dis.* **18**, 38–72.

McFadden, G. I. 1999 Plastids and protein targeting. *J. Eukaryotic Microbiol.* **46**, 339–346.

McFadden, G. I. & Roos, D. S. 1999 Apicomplexan plastids as drug targets. *Trends Microbiol.* **7**, 328–333.

McFadden, G. I., Reith, M. E., Munholland, J. & LangUnnasch, N. 1996 Plastid in human parasites. *Nature* **381**, 482.

McFadden, G. I., Waller, R. F., Reith, M. E. & Lang-Unnasch, N. 1997 Plastids in apicomplexan parasites. In *Origins of algae and their plastids* (ed. M. Bhattacharya), pp. 261–287. Vienna: Springer.

Martin, W. & Herrmann, R. G. 1998 Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.* **118**, 9–17.

Martin, W., Stoebe, B., Goremykin, V., Hansmann, S., Hasegawa, M. & Kowallik, K. V. 1998 Gene transfer to the

nucleus and the evolution of chloroplasts. *Nature* **393**, 162–165.

Neupert, W. 1997 Protein import into mitochondria. *A. Rev. Biochem.* **66**, 863–917.

Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. 1997 Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engng* **10**, 1–6.

Palmer, J. D. & Delwiche, C. F. 1996 Second-hand chloroplasts and the case of the disappearing nucleus. *Proc. Natl Acad. Sci. USA* **93**, 7432–7435.

Pfefferkorn, E. R. 1988 *Toxoplasma gondii* as viewed from a virological perspective. In *The biology of parasitism* (ed. P. T. Eglund & A. Sher), pp. 479–501 (*MBL Lect. Biol.* **9**). New York: Alan R. Liss.

Pfefferkorn, E. R. & Pfefferkorn, L. C. 1980 *Toxoplasma gondii*: genetic recombination between drug resistant mutants. *Exp. Parasitol.* **50**, 305–316.

Pilon, M., Wienk, H., Sips, W., de Swaaf, M., Talboom, I., van't Hof, R., de Korte-Kool, G., Demel, R., Weisbeek, P. & de Kruijff, B. 1995 Functional domains of the ferredoxin transit sequence involved in chloroplast import. *J. Biol. Chem.* **270**, 3882–3893.

The *Plasmodium* Genome Database Collaborative 2001 PlasmoDB: an integrative database of the *Plasmodium falciparum* genome. Tools for accessing and analyzing finished and unfinished sequence data. *Nucleic Acids Res.* **29**, 66–69.

Remington, J. S. & Desmonts, G. 1989 Toxoplasmosis. In *Infectious diseases of the fetus and newborn infant* (ed. J. S. Remington & J. O. Klein III), pp. 89–195. Philadelphia: W. B. Saunders.

Roos, D. S. 2001 Bioinformatics—trying to swim in a sea of data. *Science* **291**, 1260–1261.

Roos, D. S., Donald, R. G. K., Morrissette, N. S. & Moulton, A. L. C. 1994 Molecular tools for genetic dissection of the protozoan parasite *Toxoplasma gondii*. *Meth. Cell Biol.* **45**, 27–63.

Roos, D. S., Sullivan Jr, W. J., Striepen, B., Bohne, W. & Donald, R. G. K. 1997 Tagging genes and trapping promoters in *Toxoplasma gondii* by insertional mutagenesis. *Methods* **13**, 112–122.

Roos, D. S., Crawford, M. J., Donald, R. G. K., Fohl, L. M., Hager, K. M., Kissinger, J. C., Reynolds, M. G., Striepen, B. & Sullivan Jr, W. J. 1999*a* Transport and trafficking: *Toxoplasma* as a model for *Plasmodium*. *Novartis Fdn Symp.* **226**, 176–198.

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

Roos, D. S., Darling, J. A., Reynolds, M. G., Hager, K. M., Striepen, B. & Kissinger, J. C. 1999*c* *Toxoplasma* as a model parasite: apicomplexan biochemsitry, cell biology, molecular genetics…and beyond. In *Biology of parasitism* (ed. C. Tschudi & E. Pearce), pp. 143–167. Boston: Kluwer.

Schlichterle, M., Wahlgren, M., Scherf, A. & Perlmann, H. 2000 Methods in malaria research. See http://www.malaria.mr4.org/mr4pages/MR4_Protocols.html.

Schmatz, D. M. 1997 The mannitol cycle in *Eimeria*. *Parasitology* **114**, S81–S89.

Schwartzbach, S. D., Osafune, T. & Löffelhardt, W. 1998 Protein import into cyanelles and complex chloroplasts. *Plant Mol. Biol.* **38**, 247–263.

Shaw, M. K. & Tilney, L. G. 1992 How individual cells develop from a syncytium: merogony in *Theileria parva* (Apicomplexa). *J. Cell Sci.* **101**, 109–123.

Sheffield, H. G. & Melton, M. J. 1968 The fine structure and reproduction of *Toxoplasma gondii*. *J. Parasitol.* **54**, 209–226.

Siddall, M. 1992 Hohlzylinders. *Parasitol. Today* **8**, 90–91.

Sinden, R. 1983 The cell biology of sexual development in *Plasmodium*. *Parasitology* **86**, 7–28.

Soll, J. & Tien, R. 1998 Protein translocation into and across the chloroplastic envelope membranes. *Plant Mol. Biol.* **38**, 191–207.

Somerville, C. B. J. 1991 Plant lipids: metabolism, mutants, and membranes. *Science* **252**, 80–87.

Striepen, B., He, C. Y., Matrajt, M., Soldati, D. & Roos, D. S. 1998 Expression, selection, and organellar targeting of the Green Fluorescent Protein in *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **92**, 328–338.

Striepen, B., Crawford, M. J., Shaw, M. K., Tilney, L. G., Seeber, F. & Roos, D. S. 2000 The plastid of *Toxoplasma gondii* is divided by association with the centrosomes. *J. Cell Biol.* **151**, 1–12.

Striepen, B., Soldati, D., Garcia-Reguet, N., Dubremetz, J.-F. & Roos, D. S. 2001 Targeting of soluble proteins to the rhoptries and micronemes in *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **113**, 45–53.

Tarleton, R. L. & Kissinger, J. C. 2001 Parasite genomics: current status and future prospects. *Curr. Opin. Immunol.* **13**, 395–402.

Ullman, B. & Carter, D. 1995 Hypoxanthine-guanine phosphoribosyltransferase as a therapeutic target in protozoal infections. *Infect. Agents Dis.* **4**, 29–40.

Vollmer, M., Thomsen, N., Wiek, S. & Seeber, F. 2001 Apicomplexan parasites possess distinct nuclear-encoded, but apicoplast-localized, plant-type ferredoxin-NADP(+) reductase and ferredoxin. *J. Biol. Chem.* **276**, 5483–5490.

von Heijne, G. & Nishikawa, K. 1991 Chloroplast transit peptides: the perfect random coil? *FEBS Lett.* **278**, 1–3.

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.

Waller, R. F., Reed, M. B., Cowman, A. F. & McFadden, G. I. 2000 Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. *EMBO J.* **19**, 1794–1802.

Weissig, V., Vetro-Widenhouse, T. S. & Rowe, T. C. 1997 Topoisomerase II inhibitors induce cleavage of nuclear and 35-kb plastid DNAs in the malarial parasite *Plasmodium falciparum*. *DNA Cell Biol.* **16**, 1483–1492.

Wellems, T. E. 1991 Molecular genetics of drug resistance in *Plasmodium falciparum* malaria. *Parasitol. Today* **7**, 110–116.

Wilson, R. J. M. & Wiliamson, D. H. 1997 Extrachromosomal DNA in the Apicomplexa. *Microbiol. Mol. Biol. Rev.* **61**, 1–16.

World Health Organization 1997 World malaria situation in 1994. *Wkly Epidemiol. Res.* **72**, 269–274.

Yung, S., Unnasch, T. R. & Lang-Unnasch, N. 2001 Analysis of apicoplast targeting and transit peptide processing in *Toxoplasma gondii* by deletional and insertional mutagenesis. *Mol. Biochem. Parasitol.* **118**, 11–21.

Zhu, G., Marchewka, M. J. & Keithly, J. S. 2000 *Cryptosporidium parvum* appears to lack a plastid genome. *Microbiology* **146**, 315–321.