

Stefan A. Juranek (219), *Howard Hughes Medical Institute, Laboratory of RNA Molecular Biology, Rockefeller University, New York, New York 10021*

Hans Joachim Lipps (219), *Institute of Cell Biology, University of Witten/Herdecke, 58448 Witten, Germany*

Hilton H. Mollenhauer (191), *Texas A&M University, College Station, Texas 77845*

D. James Morr  (191), *Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, Indiana 47907*

James B. Phillips (75), *Biological Sciences Department, The Open University, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom*

Dan J. Sillence (151), *Leicester School of Pharmacy, Hawthorne Building, De Montfort University, Leicester, LE1 9BH, United Kingdom*

## **Adhesion Molecules and Other Secreted Host-Interaction Determinants in Apicomplexa: Insights from Comparative Genomics**

Vivek Anantharaman, Lakshminarayan M. Iyer, S. Balaji, and L. Aravind  
National Center for Biotechnology Information, National Library of Medicine,  
National Institutes of Health, Bethesda, Maryland 20894

Apicomplexa have developed distinctive adaptations for invading and surviving within animal cells. Here a synthetic overview of the diversity and evolutionary history of cell membrane-associated, -secreted, and -exported proteins related to apicomplexan parasitism is presented. A notable feature in this regard was the early acquisition of adhesion protein domains and glycosylation systems through lateral transfer from animals. These were utilized in multiple contexts, including invasion of host cells and parasite-specific developmental processes. Apicomplexans possess a specialized version of the ancestral alveolate extrusion machinery, the rhoptries and micronemes, which are deployed in invasion and delivery of proteins into host cells. Each apicomplexan lineage has evolved a unique spectrum of extruded proteins that modify host molecules in diverse ways. Hematozoans, in particular, appear to have evolved novel systems for export of proteins into the host organelles and cell membrane during intracellular development. These exported proteins are an important aspect of the pathogenesis of *Plasmodium* and *Theileria*, being involved in response to fever and in leukocyte proliferation respectively. The complement of apicomplexan surface proteins has primarily diversified via massive lineage-specific expansions of certain protein families, which are often coded by subtelomeric gene arrays. Many of these families have been found to be central to immune evasion. Domain shuffling and accretion have resulted in adhesins with new domain architectures. In terms of individual genes, constant selective pressures from the host immune response has resulted in extensive protein polymorphisms and gene losses. Apicomplexans have also evolved complex regulatory mechanisms controlling expression and

maturation of surface proteins at the chromatin, transcriptional, posttranscriptional, and posttranslational levels. Evolutionary reconstruction suggests that the ancestral apicomplexan had thrombospondin and EGF domain adhesins, which were linked to the parasite cytoskeleton, and played a central role in invasion through formation of the moving junction. It also suggests that the ancestral parasite had O-linked glycosylation of surface proteins which was partially or entirely lost in hematozoan lineages.

**KEY WORDS:** Apicomplexa, Apicomplexan parasitism, Adhesions, Alveolate extrusion, Hematozoans, Protein export. © 2007 Elsevier Inc.

## I. Introduction

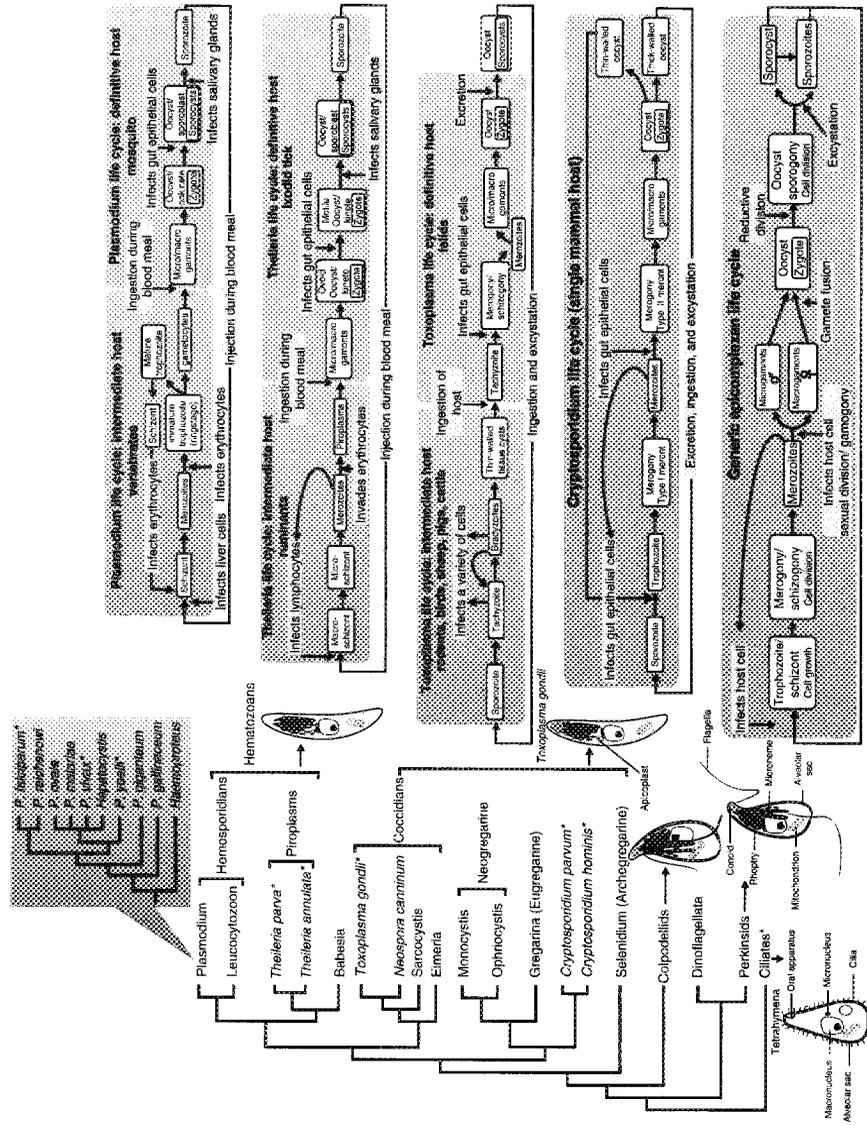
Apicomplexa are an extraordinary lineage of eukaryotic microorganisms that have been recognized since the earliest microscopic observations—Leeuwenhoek described spores of *Eimeria* from infected rodents. To date approximately 2400 species have been described using microscopic observations (Levine, 1988; Vivier and Desportes, 1990). More recently, sampling of eukaryotic diversity via environmental sequencing of ribosomal DNA (rDNA) and other phylogenetic markers suggests that traditional observations might have considerably underappreciated their actual diversity (Moon-van der Staay *et al.*, 2001). All well-studied apicomplexans are parasites of animals and infect most major invertebrate and vertebrate lineages. Apicomplexans are causative agents of major human and veterinary diseases: (1) Malaria is caused by several *Plasmodium* species in various mammals, including humans, reptiles and birds (Kreier, 1977). (2) Two forms of theileriosis, namely East Coast fever and tropical theileriosis in cattle and water buffaloes are caused respectively by *Theileria parva* and *T. annulata* (Dobbelaere and McKeever, 2002). (3) Babesiosis or Red-water fever is caused by *Babesia* species in cattle, horses, dogs, and rarely humans (Levine, 1988). (4) *Toxoplasma gondii* causes toxoplasmosis in several mammals including rodents, felids, and humans. In humans, especially, fetal infections can result in serious eye and brain damage (Dubey and Beattie, 1988). (5) *Neospora caninum* infects dogs, cattle, and other mammals and has been implicated as a major cause of abortion in cattle. (6) *Sarcocystis* species infects muscles of various wild herbivores, such as hare and deer, thereby making them easier prey for their predators (Levine, 1988). (7) *Eimeria* species infect birds and are a major cause of morbidity in poultry. (8) *Isospora* is the causative agent of an acute diarrhea in humans and other mammals and is particularly common in immunosuppressed or young individuals. (9) *Cryptosporidium* species have

been implicated in gastrointestinal disease that is a major cause of morbidity in HIV-infected patients (Fayer, 1997). Several poorly studied apicomplexans are also known to infect all major invertebrate lineages; most common among these are gregarines which infect various organs of arthropods, annelids, and mollusks (Levine, 1988; Vivier and Desportes, 1990). As yet uncharacterized apicomplexans are also believed to infect unicellular foraminiferans (Koepecna *et al.*, 2006).

Despite a wealth of traditional morphological investigations using light and electron microscopy (EM), molecular studies on apicomplexa were slowed by difficulties posed by complex life cycles, problems in maintaining growth or perpetuating development in cell cultures, and in several cases absence of effective genetic tools. However, the avalanche of information coming from genome sequences and associated high-throughput proteomics and gene expression are rapidly improving our understanding of every aspect of apicomplexan biology and providing new tools to complement traditional approaches (Aravind *et al.*, 2003b; Roos, 2005). An area of tremendous interest in light of the quest for antiparasite treatments is the parasite cell surface and extracellular events related to host-parasite interactions (Miller and Hoffman, 1998). Also relevant are developments in understanding regulatory processes that allow stage-specific expression of particular surface proteins, as well as the differential gene-expression mechanisms that allow immune evasion by parasites (Ralph and Scherf, 2005). In this article we synthesize these data in an evolutionary context to present an overview of proteins involved in interactions of apicomplexan parasites with their hosts. In particular, we concentrate on transmembrane and secreted or extruded proteins at the focus of the parasite-host interface. We begin by providing a background on apicomplexan biology, phylogenetics, and the current state of genomics and use this as a scaffold to explore developments relating to the core topics.

### A. Apicomplexa: Overview of Morphology, Life Cycles, Host, and Tissue Range

The motile form (the zoite) of the apicomplexan cell that initiates infection has a highly polar cell structure. The apical end of the zoite contains a polar ring from which emerge microtubules going back all the way into the cell. Also associated with the apex are paired pedunculate organelles termed rhoptries, which are unique extrusive organelles. Another distinctive organelle is the apicoplast, which is a degenerate photosynthetic organelle descending from a eukaryotic secondary endosymbiont, most probably of rhodophyte ancestry (Foth and McFadden, 2003). The apical end of the cell is also enriched in dense tubular bodies termed micronemes, which contain molecules required



for interacting with the host cell (Fig. 1). During invasion they appear to fuse to the neck of the rhoptry to discharge their contents. Together, these features form the distinctive apical complex that lends apicomplexa its name. Apicomplexa are also characterized by a life cycle along with certain hallmark phases that appear to be present in most studied members of the clade (Fig. 1). The zoite initiates infection by entering the host cell or associating intimately with it. It then grows by deriving nutrients from the host (the trophozoite stage) and often undergoes a phase of repeated mitotic divisions termed schizogony or merogony. Daughter zoites of this phase are generally termed merozoites and go on to infect a new set of cells. There is

FIG. 1 Alveolate phylogeny and life cycles of apicomplexan species with completely sequenced genomes. Cartoon representations of various species are shown to highlight conserved cellular features shared by apicomplexa with other closely related alveolate species. A generic life cycle shows stages and events conserved in all apicomplexans, and specific life cycles are shown for the main species with sequenced genomes. The sporozoite of *Cryptosporidium* emerges from excreted oocysts and infects vertebrate gut epithelial cells. A distinct class of merozoites eventually gives rise to intracellular micro- and macro-gamonts. These develop into micro- and macro-gametes that fuse to form the intracellular zygote enclosed within either thin-walled or thick-walled oocysts. The former maintain auto-infection, whereas the latter are excreted out and perpetuate infection to new hosts upon ingestion. *T. gondii* infects gut epithelial cells of cats (definite host). In course of infection some merozoites eventually lead to micro- and macro-gametes which fuse to form an oocyst-enclosed zygote inside an epithelial cell. Excreted oocysts sporulate to form two sporocysts within them, which in turn contain four sporozoites each. These resistant sporulated oocysts, when ingested by intermediate hosts, typically mice, release sporozoites and initiate asexual development. Upon invading host cells, sporozoites transform into tachyzoites that rapidly multiply within cells through binary fission and rupture them to infect new cells. Tachyzoites are disseminated via macrophages, lymphocytes, and in plasma and spread throughout the host body. This marks the acute phase of infection, which is followed by the chronic phase. Here, tachyzoites transform to morphologically distinct bradyzoites that multiply slowly and form thin-walled tissue cysts. These enter the definitive host upon predation of the intermediate host. In *Theileria*, the sporozoite delivered by blood-sucking ticks invades mammalian host lymphocytes and undergoes schizogony to give rise to macroschizonts. Such lymphocytes are transformed resulting in a leukemia-like phenotype. Some macroschizonts metamorphose into microschizonts that then spawn merozoites, which invade erythrocytes. Return to the definitive host occurs when such infected erythrocytes are ingested by the tick. In the tick gut, they emerge from erythrocytes and spawn gametes that fuse to form a zygote. The zygote invades gut epithelial cells where it develops into a kinete that traverses the gut wall and via the coelom localizes to salivary glands. In the glandular epithelium they undergo massive division termed sporogony to form thousands of sporozoites that escape into salivary ducts. In *Plasmodium* sporozoites first invade vertebrate host hepatocytes and undergo rounds of schizogony. Resultant merozoites infect erythrocytes, where they grow as trophozoites, passing through a distinctive morphology termed the ring stage, and give rise to merozoites. Some merozoites transform into gametocytes that are taken up by the mosquito during feeding and develop into flagellated micro- and macro-gametes in the mosquito gut. Upon fusion of gametes, a kinete forms that traverses the gut and on the external side of the gut wall forms an oocyst. Within it are generated hundreds of sporozoites, which reach the salivary glands for a new round of transmission. Genomes that are completely sequenced or nearly complete are marked with an asterisk in the tree. (See also color insert.)

also a sexual phase that typically results in sexually dimorphic gametes. Fusion of gametes results in a zygote that is enclosed within an oocyst. Within the oocyst there is a reductive division, which might be followed by additional rounds of mitotic division, finally resulting in a new zoite termed the sporozoite that is ready to infect new hosts. Divisions within the oocyst may result in formation of resistant sporocysts that are shed to the exterior of the host and transmitted to new hosts via the environment. Here, sporozoites emerge from sporocysts to infect cells of the new host. Although this pattern is shared by all apicomplexans, in actual life cycles of various members of this clade, these phases play out in a bewildering diversity of contexts with respect to the number of hosts, type of hosts, and tissue types they infect (Levine, 1988; Vivier and Desportes, 1990) (Fig. 1).

The simplest life cycles involve a single host (monoxenous) and are typified by that of *Cryptosporidium* (Fig. 1) and members of the gregarine grade of apicomplexa. However, in many of the latter organisms, trophozoites are not located entirely within host cells, but partially invade them, and are attached via structures known as the epimerite or mucron. In addition to invading the gut epithelium, several of these parasites enter the coelom and invade a range of other tissues (e.g., *Monocystis agilis* invades seminal vesicles of earthworms, whereas *Ophriocystis* invades insect Malpighian tubules) (Levine, 1988; Vivier and Desportes, 1990). Selective pressures from the necessity of reaching new hosts appear to have favored inclusion of additional hosts, resulting in heteroxenous life cycles (Lafferty, 1999) (Fig. 1). These additional hosts or *intermediate hosts* typically serve as carriers that act as vehicles in delivering the parasite to the *definitive host* in which the parasite's sexual cycle is completed. There have been multiple convergent innovations of heteroxenous cycles even within apicomplexa that parallel similar life cycles in a variety of eukaryotic parasites, such as kinetoplastids, various platyhelminth lineages, nematodes, and acanthocephalans (Lafferty, 1999). One such cycle, seen in *Toxoplasma gondii* and related coccidian apicomplexans involves reaching a predatory mammalian definitive host via prey which serves as the intermediate host (Fig. 1). *T. gondii* and several related apicomplexans are also known or suspected to modify the behavior or health of intermediate hosts such that they become more susceptible to predation by definitive hosts and thereby perpetuate their life cycle. Similar predation-dependent heteroxenous cycles are encountered in invertebrate-parasitic apicomplexans such as *Aggregata eberthi*, which uses crabs as intermediate hosts and their molluscan predators, cuttlefish, as definitive hosts. Hematozoans (hemosporidians and piroplasms) appear to have evolved from precursors that infected arthropods by adding vertebrate intermediate hosts (Fig. 1). Emergence of vertebrate hemophagy in invertebrates appears to have been a driving force that led to using vertebrate tissues, especially blood cells as the

site for asexual development, thereby enabling reentry into the definitive host (Levine, 1988; Vivier and Desportes, 1990).

## B. Evolutionary History of Apicomplexa and Implications for Origins of Parasitism

Like many other specialized parasites, apicomplexa exhibit drastic morphological modifications with respect to other eukaryotes but retain a degree of structural conservatism within the clade (Leander *et al.*, 2003). This resulted in considerable confusion in traditional taxonomical schemes of relationships within apicomplexa. However, several molecular phylogenies point to an emerging consensus on intra-apicomplexan relationships, as well as some clarity regarding their affinities to other major eukaryotic lineages (Fig. 1). Within apicomplexa, the basal-most lineage recognized in most studies is the archigregarine clade, which is typified by *Selmidium*, a parasite of marine worms. The next clade comprises a large assemblage that links *Cryptosporidium* with two major gregarine clades typified respectively by *Monocystis* + *Ophriocystis* and *Gregarina*. The two correspond approximately to neogregarines and eugregarines of classical taxonomy. Although the bootstrap support for this assemblage is weak, it is consistently recovered in most studies supporting its reality (Kopečna *et al.*, 2006; Leander *et al.*, 2006). This is followed by a crown group formed by classical coccidians including, among others, *Toxoplasma*, *Sarcocystis*, and *Eimeria* and hematozoans, which are specialized vertebrate blood parasites. The hematozoan clade in turn contains piroplasms such as *Theileria* and hemosporidians containing various species of the genus *Plasmodium* and related genera such as *Hemoproteus* and *Leucocytozoon* (Yotoko and Elisei, 2006). Genome comparisons of *Theileria* and *Plasmodium* show regions of microsynteny and also reveal several shared protein sets to the exclusion of other apicomplexans, thereby strongly supporting the monophyly of hematozoans (Pañ *et al.*, 2005). Monophyly of the coccidian + hematozoan crown group is similarly strongly supported by several uniquely shared protein sets that have become available from genome sequencing efforts.

The closest sister clade of apicomplexa is a group of environmentally prevalent, predatory, biflagellate protists, the colpodellids (typified by *Colpodella*) (Kuvardina *et al.*, 2002; Leander *et al.*, 2003). Together these are related to an assemblage formed by dinoflagellates and their sister-group, the parasitic perkinsids. All these groups further unify with ciliates to form the alveolate clade (Fig. 1). The alveolate lineage is supported by several morphological synapomorphies, most of which appear in some form in apicomplexans. These include presence of a system of inner membranous sacs or alveoli, micropores of diverse morphologies that communicate with the cell exterior

and specific "extrusome" organelles (Leander *et al.*, 2003). Additionally, it is believed that the common ancestor of alveolates or an even earlier ancestor shared by alveolates and the entire stramenopile (chromist) lineage possessed an originally photosynthetic secondary endosymbiont of possible rhodophyte provenance (Bhattacharya *et al.*, 2004; Foth *et al.*, 2003). This endosymbiont, precursor of the apicomplexan apicoplast, appears to have been repeatedly lost or at least become incapable of photosynthesis on multiple occasions in alveolates. In apicomplexans, colpodellids, perkinsids, and other poorly characterized forms such as *Acrocoelus* (a hemichordate gut parasite) and *Cryptophagus* (a cryptomonad algal parasite), the primary extrusome is the rhoptry. In ciliates several morphologically distinct forms of extrusomes (e.g., ejectives and trichocysts) are seen. These observations suggest that the common ancestor of alveolates already possessed a specialized extrusion system that might have played a role in interaction with other organisms and the environment and a micropore that possibly played a role in feeding (Leander *et al.*, 2003). These ancestral preadaptations aided repeated emergence of parasitism in several alveolate lineages.

Several dinoflagellates and their closest sister group perkinsids are parasites of diverse eukaryotes. Both perkinsids, dinoflagellates like *Paulsenella* (an algal parasite), and predatory colpodellids deploy their rhoptry secretions to respectively invade host cells or penetrate them and directly draw up cytoplasmic contents. Close morphological similarities of colpodellids and perkinsids implies the common ancestor of this entire subgroup of the alveolate clade was probably a similarly modeled, actively motile predator that used rhoptry secretions to attack target cells, penetrate them, and draw their contents (Kuvardina *et al.*, 2002; Leander *et al.*, 2003). Consistent with this proposal, trophozoites of basal apicomplexans, such as *Selenidium*, directly imbibe nutrients by penetrating the host cell rather than being an entirely intracellular parasite like crown group apicomplexans (Vivier *et al.*, 1990). Aspects of the ancestral predatory/parasitic life style continue to show up in invasion strategies of extant forms such as *Plasmodium*. Imaging studies on sporozoites indicate they actively glide against the blood flow to reach the hepatic sinusoidal cell layer. They then tightly adhere to it and pass through Kupffer cells to reach hepatocytes. Sporozoites reach their target hepatocyte only after leaving a trail of dead hepatocytes, which they penetrate en route to their destination (Frevort *et al.*, 2005). Thus, one could conceive evolution of apicomplexan parasites as an adaptive series from colpodellid-like predators to archigregarine-like parasites to completely intracellular forms. On one hand, this was accompanied by gain of adaptations for intracellular parasitism, evasion of host defenses, and modification of host cell and host behavior. On the other hand, there was degeneration in the form of loss of flagella and decreased metabolic capabilities. Monophyly of alveolates is firmly established on molecular and morphological

grounds. The higher-order relationship of alveolates and stramenopiles (the chromoalveolate clade), which include diverse photosynthetic taxa such as diatoms and brown algae, bicoseoid flagellates, and oomycetes, while requiring more investigation, has been supported by certain morphological and molecular features (Bhattacharya *et al.*, 2004). Comparative genome analysis and trees built from concatenated multiple alignments of all ribosomal proteins suggest the chromoalveolate assemblage form a sister group to the crown group that includes animal, fungi, amebozoans such as *Dictyostelium*, and plants. Outside of these are kinetoplastids and diplomonads such as *Giardia* forming successively earlier branching eukaryotic lineages (Bhattacharya *et al.*, 2004; Templeton *et al.*, 2004a).

## II. Comparative Genomics and Apicomplexan Biology

### A. Current State of Apicomplexan Genomics

Apicomplexan genomics was initiated with sequencing of the complete genome of *Plasmodium falciparum*. Genomes of the human benign tertian malaria parasite *P. vivax* and the murine parasite *P. yoelii* were also sequenced in parallel, although their assembly and annotation is in a considerably poorer condition than that of *P. falciparum* (Carlton *et al.*, 2002; Gardner *et al.*, 2002, 2005). Lower coverage sequencing and limited annotation have also become available for genomes of *P. berghei* and *P. chabaudi*, which are rodent parasites widely used as models for human malaria (Hall *et al.*, 2005). The genome of early-branching *Cryptosporidium parvum* was sequenced next and more recently complemented with a second species, *C. hominis* (Abrahamsen *et al.*, 2004; Xu *et al.*, 2004). This was followed by sequencing of two piroplasm species, *T. parva* and *T. annulata* (Gardner *et al.*, 2005; Pain *et al.*, 2005). All these genome sequences have become available through the Genbank database of NCBI and the apicomplexan database ApiDB. The genome of *T. gondii* has also been sequenced and has been made available via ToxoDB/ApiDB, but to date there has not been a publication on the genome sequence (Roos, 2005). These genomic efforts have also facilitated a variety of high-throughput analyses, data sets from which are publicly available to varying degrees. Of these, most detailed studies have been on *P. falciparum*, where multiple groups have published high-throughput transcriptome and proteome analysis of the intraerythrocytic developmental cycle (IDC) (Bozdech *et al.*, 2003; Hall *et al.*, 2005; Le Roch *et al.*, 2003). Transcriptome analysis of febrile response of *P. falciparum* is underway, and that of intrahepatocytic development of *P. yoelii* has been published (Oakley *et al.*, 2007; Wang *et al.*, 2004). Likewise, the

transcriptome analysis of the blood stages of *T. parva* has also been published (Bishop *et al.*, 2005). Transcriptome analysis has been complemented by in-depth proteomic analysis, at least in *P. falciparum*, and is promising to shed light on parasite proteins targeted to the merozoite cell surface as well as those routed to the infected erythrocyte surface (Florens *et al.*, 2002, 2004; Hall *et al.*, 2005). Additionally, a substantial subsection of the protein-protein interaction map of *P. falciparum* has been determined through two-hybrid interaction mapping (LaCount *et al.*, 2005). However, this data is riddled with several false positives due to presence of large low-complexity stretches in *Plasmodium* proteins.

Apicomplexan genomes are marked by striking base-compositional differences: *T. gondii* has an even GC content (~50%), whereas those of *Theileria* (~34%), *Cryptosporidium* (30%), and *Plasmodium* (19.5%) have low GC content (Gardner *et al.*, 2005). The strikingly AT-rich genome of *Plasmodium* is also reflected in its proteins in the form of enrichment for asparagine as well as numerous low-complexity inserts of asparagine-rich segments even within globular domains of proteins (Aravind *et al.*, 2003b). General results of apicomplexan comparative genomics point in the direction of considerable molecular specialization of each parasite, beyond a core set of uniquely apicomplexan features. Predicted apicomplexan proteomes range from approximately 4000 to over 7000 annotated proteins, which is comparable to what is seen in free-living and pathogenic fungi with yeastlike morphologies, *Entamoeba*, *Giardia*, and some kinetoplastids. Apicomplexan genomes are clearly larger than microsporidian genomes, but distinctly smaller than those of free-living alveolates, namely the ciliate *Tetrahymena*, and multicellular crown-group eukaryotes (Eisen *et al.*, 2006; Gardner *et al.*, 2005; Templeton *et al.*, 2004a). Apicoplast and mitochondrial genomes of all apicomplexans show considerable reduction, with the extreme case being *Cryptosporidium*. In this organism the apicoplast and its genome have been completely lost, whereas the mitochondrion is degenerate and has lost its genome (Abrahamsen *et al.*, 2004). The mitochondrial genome of *Theileria* is reduced to a 7.1-Kb DNA element with limited coding capacity (Kairo *et al.*, 1994). Even though metabolic capabilities of all studied apicomplexans are highly reduced relative to free-living eukaryotes with comparable genome sizes, they display noticeable lineage-specific diversity. In *T. gondii* and *Plasmodium*, there is evidence for several distinct functionally linked metabolic processes, such as fatty acid synthesis, amino acid interconversions, and carbohydrate metabolism, happening in the cytoplasm, apicoplast, and mitochondrion, with the apicoplast playing a particularly major anabolic role (Ralph, 2005). At the other extreme, practically all of these anabolic pathways appear to have been lost in *Cryptosporidium*, and it possesses very limited metabolic capabilities, implying heavy dependence on the host (Abrahamsen *et al.*, 2004). *Theileria* represents an intermediate situation,

having lost several enzymes seen in *Plasmodium* and *T. gondii*, related to shikimate and type-II fatty acid metabolism (Gardner *et al.*, 2005).

Henceforth, when referring to apicomplexans we mean the aforementioned taxa with completely sequenced genomes. Functional partitioning of apicomplexans proteomes show similarities as well as differences from free-living eukaryotes with similar genome sizes. Numbers of protein kinases and phosphatases, GTPases, and basal chromatin and transcription complex components are similar, suggesting that they are likely to possess comparable signal transduction and global transcription regulatory capabilities. Except for *Cryptosporidium*, which has lost a large number of introns and spliceosomal components, apicomplexans possess a robust complement of introns and a well-developed splicing apparatus. Apicomplexans markedly differ from free-living eukaryotes in allocating a notable section of their proteomes for parasitic adaptations (Templeton *et al.*, 2004a). Most significant among these proteins are the secreted and cell-surface transmembrane or membrane-anchored proteins. These proteins shall form the focus of this review and will be discussed in detail in the remainder of the article.

## B. Genomic Perspective on Transmembrane (TM) and Secreted Proteins in Apicomplexa and Other Eukaryotes

Eukaryotic TM and secreted proteins (collectively termed "surface proteins") are usually characterized by presence of signals that determine their localization or topology (Nilsson *et al.*, 2005). The situation is more complex in apicomplexans because they spend a significant part of their life cycle within host cells. Except piroplasms, others reside within a parasitophorous vacuole (PV) membrane within host cells. Hence, after secretion from the parasite membrane (PM), a protein might technically localize to the PV lumen or exit it to reach the host cytoplasm or cell membrane (Lingelbach and Joiner, 1998). Additionally, in most apicomplexans, there is a subset of proteins with hydrophobic signal peptides followed by a substantially sized leader sequence that serves as a localization signal for apicoplast targeting rather than secretion (Foth *et al.*, 2003). Localization to the membrane is determined by hydrophobic TM helices that anchor proteins to cell membrane, and they are broadly classified as type I (N-terminus extracellular or luminal), type II (C-terminus extracellular or luminal), multipass, lipid-anchored, or GPI-anchored TM proteins. In apicomplexans, a TM protein could potentially localize to the nuclear, ER, mitochondrial, apicoplast, or plasma membranes (Nilsson *et al.*, 2005). Additionally, in the course of intracellular development, they could localize to the PV membrane (PVM), if present, or the host cell membrane (Lingelbach and Joiner, 1998). However, a number of lines of evidence, such as immunological/proteomics

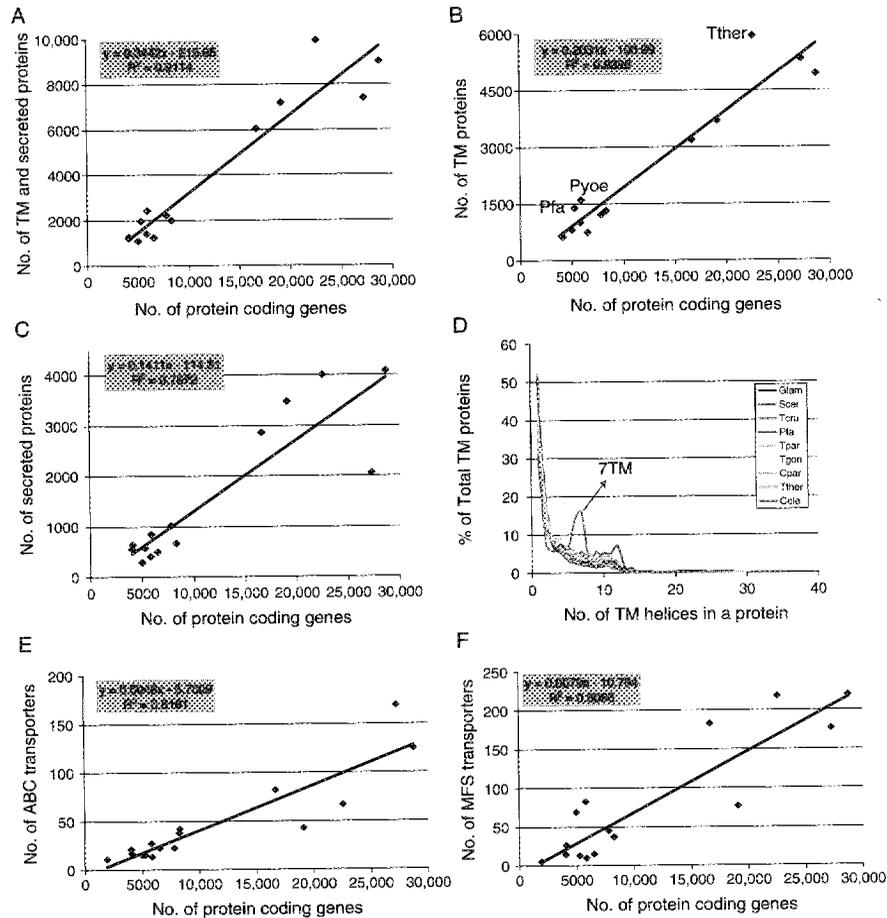


FIG. 2 Distribution of surface proteins in apicomplexa and other eukaryotes. (A) Scaling of all TM and secreted (surface) proteins with genome size. Organisms shown in the graph are *Giardia lamblia*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Trypanosoma cruzi*, *Plasmodium falciparum* (*Pfa*), *Plasmodium yoelii* (*Pyoe*), *Theileria parva* (*Tpar*), *Toxoplasma gondii* (*Tgon*), *Cryptosporidium parvum* (*Cpar*), *Tetrahymena thermophila* (*Tther*), *Leishmania major*, *Arabidopsis thaliana* (*At*), *Caenorhabditis elegans*, and *Drosophila melanogaster*. (B) Scaling of just TM proteins with genome size. The organisms studied are as shown in A. The organisms with higher fraction of TM proteins than expected are marked. (C) Scaling of just secreted proteins with genome size. The organisms studied are as shown in A. (D) Fraction of the total number of TM proteins with a given number of TM segments (1, 2, 3...n). The number of TM segments in each TM-containing protein of each organism were counted and binned, and the results are shown in the graph. The organisms are shown in the legend on the side. The spike of 7 TM proteins (odorant/chemoreceptors) in *C. elegans* is shown (E) Scaling of ABC transporters with genome size. In addition to the organism studied in A, *Encephalitozoon cuniculi* and *Ostreococcus tauri* are shown. (F) Scaling of Major Facilitator Superfamily proteins with genome size. The organisms studied are as shown in (E). Transmembrane regions and signal peptides were

studies, conserved localization signals, and sequence conservation and phylogenetic patterns, suggest that the majority of membrane proteins coded by apicomplexan genomes appear to localize to the external membranes rather than organellar or internal membranes (Florens *et al.*, 2002; Hall *et al.*, 2005; Sam-Yellowe *et al.*, 2004a; Tonkin *et al.*, 2006).

External membrane and secreted proteins, which are major determinants of survival and pathogenesis, mediate a range of biological functions. General functional groups of these proteins that will be considered in this article include: (1) TM proteins that localize either to the parasite cell membrane or to the host-cell membrane (if the parasite is residing in the host cell) and play a role in adhesion, immune-evasion, host-cell remodeling, and mobilizing nutrients. (2) Secreted proteins that might perform a variety of functions such as formation of protective cysts, degradation or modification of host molecules, or modification of host behavior. (3) Proteins that are exported from the intracellular parasite into the host cell cytoplasm and influence host physiology by interacting with components of different host-cell compartments. There will be relatively limited discussion of membrane proteins localizing to internal and organellar membranes. We also do not cover all aspects of transporter evolution, reviewing just those aspects relevant to specific facets of parasite-host interactions.

Systematic analysis of complements of predicted membrane and secreted proteins in diverse eukaryotes ranging from early branching diplomonads to large genomes of ciliates, animals, and plants shows that their counts scale more or less linearly ( $R^2 = 0.91$ ) with proteome size (Fig. 2). Thus, most eukaryotes devote approximately 30% of their protein-coding capacity to secreted or TM proteins. TM and membrane-anchored proteins by themselves show a strong linear trend ( $R^2 = 0.92$ ) and represent approximately 15 to 20% of eukaryotic proteomes. Secreted proteins by themselves also show a linear trend with respect to proteome size, but the correlation is lower ( $R^2 = 0.8$ ). Linear scaling of protein counts is also discernible for some of the largest conserved superfamilies of membrane proteins observed in all eukaryotic genomes, such as the major facilitator superfamily (MFS) and ABC transporters ( $R^2 = 0.8-0.81$ ); these two transporters constitute around 6% of membrane proteins, on average in a eukaryotic proteome. Most noticeable instances of overrepresentation in terms of overall counts of TM proteins are seen in *Plasmodium* and the nematode *Caenorhabditis elegans*, suggesting that particular selective pressures of lifestyle might indeed produce

identified using the prediction algorithms of TMHMM and SignalP and the TASS package (V. A., S. B., L. A., unpublished). In graphs A, B, C, E, and F, regression analysis was conducted, and the linear best fit is shown. The p-value was  $<10^{-5}$  in all the cases, demonstrating a linear relationship between the proteins studied and genome size at 99% or greater significance level. The corresponding  $R^2$  values and the equation are shown on the graph. (See also color insert.)

deviations from the general trend. In all eukaryotes the most frequent group of membrane proteins are single pass and GPI-anchored membrane proteins and constitute about 45% of all the membrane proteins. Counts of multipass proteins with two or more TM segments steadily decrease in number following a roughly exponential decay, with the exception of animals that shown anomalous counts for 7 TM proteins (resulting from expansions of 7 TM chemoreceptors) (Nilsson *et al.*, 2005) (Fig. 2).

A simple explanation for the striking linear scaling of surface proteins is that with increasing genome size their numbers increased in direct proportion through duplications. However, a more careful examination of these proteins suggests this is not the case. Most secreted proteins of eukaryotes are not conserved across kingdoms or even within monophyletic lineages such as apicomplexa. Although some membrane proteins such as ABC transporters and MFS transporters, or proteins of inner membrane trafficking systems contain lineages conserved throughout eukaryotes (Igarashi *et al.*, 2004; Saier *et al.*, 2001), remaining classes of membrane proteins show no evidence for widespread conservation. Even within widespread conserved superfamilies such as ABC and MFS, there is no evidence that the majority of members show orthologous relationships across eukaryotes. This implies that despite a relatively strong constraint on the total fraction of membrane and secreted proteins coded by eukaryotes throughout their evolution, actual evolutionary affinities and type of families of membrane and secreted proteins contributing to the total fraction can widely vary across organisms. This also suggests that the selective pressure imposing linear scaling of numbers of membrane proteins probably arises from a fundamental cellular constraint. A possible constraint is a strong limitation on the fraction of proteins being synthesized in a cell that can be secreted or routed to membrane proteins at a given point. This idea is also supported by the observation that organisms such as *Plasmodium* or *C. elegans* showing major deviations do not express their surplus of membrane proteins at the same time. Rather, the surplus membrane proteins, namely rifins/stevors in *Plasmodium* (see Section III.B.2) and 7 TM odorant receptors in *C. elegans* are expressed a few at a time, respectively in different cell cycles or different neurons (Florens *et al.*, 2002, 2004; Spehr and Leinders-Zufall, 2005).

Thus, parasitic adaptations of apicomplexa in general are not strongly reflected in differences in the fraction of membrane and secreted proteins, but qualitatively in terms of the distinctness of the types of such proteins in the proteome. Thus, the key to investigating parasitism-related adaptations among surface proteins is identification of innovations shared by apicomplexa as a whole and innovations specific to internal branches of the clade. These are best understood in terms of lineage-specific expansions and innovations in membrane and secreted protein families, which supply molecular determinants of parasite-host interactions.

### C. Lineage-Specific Expansions and Diversification of Apicomplexan Secreted and Membrane Proteins

Lineage-specific expansions (LSEs) are defined as the proliferation in the number of a particular protein family (protein domain) in a given lineage relative to a sister lineage (Lespinet *et al.*, 2002). For example, there is a single or few proteins with Duffy binding-like (DBL) domains in various *Plasmodium* species such as *P. vivax*, but *P. falciparum* codes for numerous PfEMP-1 proteins with multiple DBL domains (Singh *et al.*, 2006; Tolia *et al.*, 2005) (Table I). Thus, there is a LSE of the DBL domain proteins in the latter lineage. Previous studies have shown that LSEs are a prevalent feature of all eukaryotic genomes including *Plasmodium* species. The number of lineage-specifically expanded clusters of a given member count have been shown to follow a power-law scaling in various eukaryotes (Fig. 3), indicating that there are a few massively expanded families of proteins in each lineage. Sequence analysis shows a major fraction of the expanded proteins might have no close homologs in other lineages. Analysis, across diverse eukaryotes, of proteins that showed massive LSEs suggested they usually belong to a few general functional categories: (1) transcription factors; (2) secreted/TM proteins that are required in abundant quantity (e.g., for generation of extracellular protein matrices or certain secreted enzymes) or in diverse forms. This category includes both immune evasion proteins of parasites and pathogen recognition proteins of hosts that exhibit diversity or antigenic variation or (3) proteins involved in a detoxification or recognition of xenobiotics (Lespinet *et al.*, 2002).

Most large LSEs in apicomplexans, which have been functionally characterized, encode secreted or membrane proteins with several distinct roles related to host interaction, pathogenesis, and immune evasion (Aravind *et al.*, 2003b; Templeton *et al.*, 2004a). All apicomplexan genomes encode at least one major LSE of proteins that might play a major role in cytoadhesion and antigenic variation (Table I). Examples of these include unrelated PfEMP1 (var genes) of the Dbl superfamily and rifins/stevors of the rifin superfamily in *P. falciparum*, vir family in *P. vivax*, and related expansions such as the yir and bir families in *P. yoelii* and *P. berghei* (del Portillo *et al.*, 2001; Gardner *et al.*, 1998). Comparable expansions are seen in other apicomplexan lineages, namely the FAINT domain family in *Theileria*, the SRS/SAG1 domain family in *T. gondii*, and mucin proteins in *Cryptosporidium* (Abrahamsen *et al.*, 2004; He *et al.*, 2002; Pain *et al.*, 2005) (Table I). It is likely that all of these LSEs have general functional properties comparable to var, rifin, and vir families of *Plasmodium*. Evidence suggests in *Plasmodium* they exploit their high copy number for differential expression in different cell cycles or for diversification through recombination. Thus, one driving force for LSE in these cases is probably the need to have sufficient antigen

TABLE 1

Examples of characterized lineage specific expansions in apicomplexan proteomes

Name of protein/domain	Counts <sup>a</sup>	Comments/architecture
<b>A. <i>Plasmodium falciparum</i> expansions</b>		
Rifin (stevens include) superfamily	Pfal(173)	Predominantly subtelomeric gene family with PEXEL motif; localized to erythrocyte membrane, believed to adopt 2TM topology
erythrocyte membrane protein 1 (PFEM/P1)	Pfal(60-70)	Predominantly subtelomeric gene family. Typical versions have 4-7 DBL domains and C terminal Pfemp1-C domains. The first DBL domain plays a major role in adhesion by binding heparan sulfates
EBA175-like proteins	Pfal(4)	Homologs of <i>P. vivax</i> DBP. 2 DBL + MAEBL-C domain. Bind sialates on erythrocyte protein glycophorin A
PFL2555w-like	Pfal(20)	Subtelomeric gene family with PEXEL. SP + PRESAN domain
RESA-type DNAJ proteins	Pfa(12)	Subtelomeric gene family with PEXEL. SP + PRESAN + DnaJ domains
PFB0995w-like 2TM family	Pfal(12)	Subtelomeric gene family with PEXEL. Integral membrane protein typified by PFB0995w-like (SP + 2TM)
Pfmc-2TM (MAL7P1.5-like)	Pfal(11)	Subtelomeric gene family with PEXEL. Maurer's cleft localized proteins (SP + 2TM). Analogous expansion found in <i>P. yoelii</i> .
PFACS (PFD0085c)	Pfal(9)	Subtelomeric gene family with PEXEL. Fatty acyl-CoA synthetase targeted to erythrocyte cytoplasm
Merozoite surface protein 7 family	Pfal(7)	Have a signal peptide followed by a variable low-complexity charged region. Only one gene is subtelomeric. Has a C-terminal PEXEL-like element. Not expanded in other <i>Plasmodium</i> species
ETR-AMP family (PF14_0016-like)	Pfal(7)	Subtelomeric gene family. SP + charged region + TM. Localize to the PVM
PFB0075c-like	Pfal(5)	Subtelomeric gene family with PEXEL. SIG + TM
CLAG	Pfal(4)	Rhoptry neck protein apparently required for deployment of extruded adhesins
Surfins (PFA0725w-like)	Pfal(10)	Subtelomeric gene family with PEXEL-like motif. SM + VGID + Pfemp1-C repeats. Expressed on both merozoite and erythrocyte surfaces.
<b>B. <i>Theileria parva</i> expansions</b>		
lysophospholipase PF10_0018-like	Pfal(4)	Subtelomeric gene family. SP + phospholipase catalytic domain. Might be involved in remodeling host membranes or metabolizing lipids.
Plasmepsin	Pfal(4)	Subtelomeric gene family. SP + aspartyl protease domain
PfST-2TM family	Pfal(3)	SP + 2TM
PFB1045w	Pfal(2)	PFB1045w Subtelomeric with Signal peptide PEXEL and C term Pfemp1-C
SERA proteases	Pfal(8)	A family of papain fold thiol proteases (some members have serine instead of cysteine in active site)
<b>C. <i>Theileria parva</i> expansions</b>		
FAINT Repeat containing proteins	Tpar(190)	Predominantly subtelomeric gene family. SP + FAINT domains. This group contains both SFI repeat and TASH families. Some versions may have multiple repeats of the FAINT domain.
Theileria-specific repeat protein (TSRP) family	Tpar(40)	Predominantly subtelomeric gene family. TM proteins with a conserved 7TM domain
TP04_0914-like ABC Transporter	Tpar(12)	Subtelomeric gene family. TMs + ABC ATPase + TMs + ABC ATPase
TP03_0463-like family	Tpar(7)	Subtelomeric SP + conserved domain with conserved D and H
TP03_0339-like MFS transporters	Tpar(8)	Major Facilitator Superfamily; homologs found in other eukaryotes
TP03_0543-like MFS transporters	Tpar(5)	Major Facilitator Superfamily
TP04_0455-like transporters related to bacteria RnfA	Tpar(6)	Predicted transporters with a conserved TM domain related to the N terminal domain of bacterial RnfA
TP05_0009-like 2TM proteins	Tpar(6)	Might be similar to the Plasmodium 2TM proteins
TP03_0633-like		Sp + conserved domain with D and H residues
<b>C. <i>Cryptosporidium parvum</i> expansions</b>		
Crypto Mucins	Cpar(98)	Surface proteins with mucin segments. Likely to be O-glycosylated
Oocyst wall protein	Cpar(20)	SP + Cysteine-rich module. Found in 1-3 copies in Toxoplasma
SKSR	Cpar(7)	Subtelomeric gene family. SP + domain with conserved SKSR signature

(continued)

TABLE 1 (continued)

Name of protein/domain	Counts <sup>a</sup>	Comments/architecture
FLGN	Cpar(6)	Subtelomeric gene family. SP + domain with conserved FLGN signature
MEDLE	Cpar(6)	Subtelomeric gene family. SP + domain with conserved MEDLE signature
GGC	Cpar(3)	Subtelomeric gene family. SP + domain with conserved GGC signature
DP-fucose transporter	Cpar(5)	Predicted GDP-fucose transporter. Probably required for uptake of fucose
<b>D. <i>Toxoplasma gondii</i> expansions</b>		
SGA-1 / SRS domain	Tgon(70)	Predominantly subtelomeric gene family. SP + SAG1-related SRS + GPI-anchor
Toxo Mucins	Tgon(28)	SP + $\beta$ -sandwich fold globular domain + Mucin segments + GPI-anchor
2TM protein family 42.m03640 (Chr X)	Tgon(9)	<i>Toxoplasma</i> -specific 2TM proteins. Some genes are subtelomeric
Cysteine-rich family 583.m05769 (Chr XI)	Tgon(7)	<i>Toxoplasma</i> -specific family SP + Cys-rich domain
Charged ITM family 76.m02667 (Chr V)	Tgon(6)	Signal peptide + charged region + ITM resemble <i>P. falciparum</i> ETRAMP proteins
Secreted protein family 80.m02161 (Chr IX)	Tgon(5)	<i>Toxoplasma</i> -specific family of 4TM proteins with peculiar localize to rhoptries repeats. Both genes are next to each other on Chr IV.
4TM family 641.m01582 (Chr IV)	Tgon(2)	
MFS amino acid transporters	Tgon(6)	Major Facilitator Superfamily possibly involved in amino acid transport
NTP4-type nucleoside triphosphatases	Tgon(5)	SP + nucleoside-triphosphatase domain

<sup>a</sup>The counts provided here are approximate counts. Some members could be pseudogenes, and there could be count variation between strains. Pfal, *Plasmodium falciparum*; Tgon, *Toxoplasma gondii*; Cpar, *Cryptosporidium parvum*; Tpar, *Theileria*.

Other abbreviations: SP, signal peptide; ITM, transmembrane segment. The *T. gondii* gene names are as currently provided by ToxoDB (<http://www.toxodb.org/toxo/home.jsp>). The Roman numerals are used to designate chromosome location as in ToxoDB.

## APICOMPLEXAN SURFACE PROTEINS

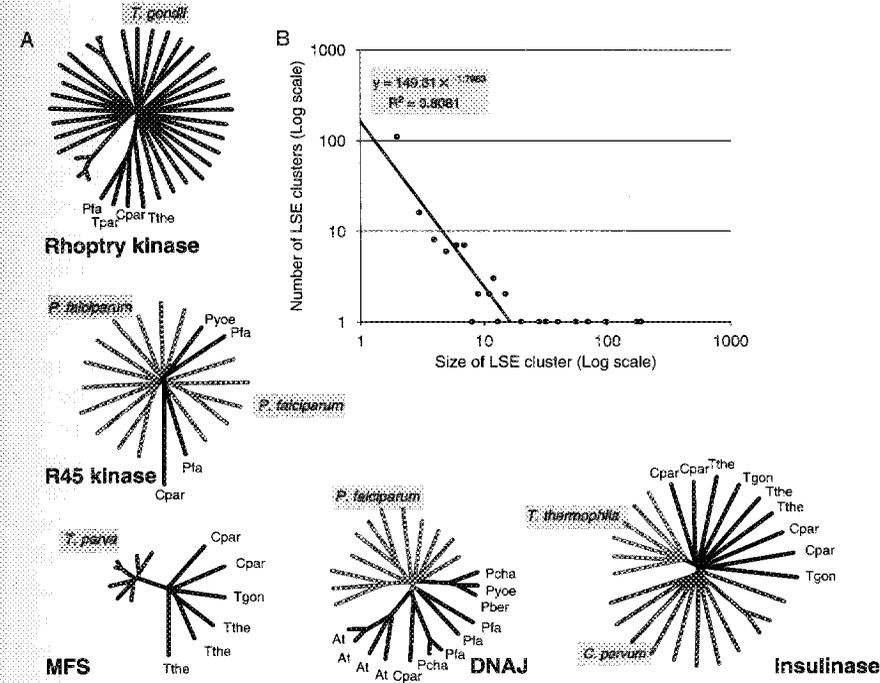


FIG. 3 Lineage-specific expansions in apicomplexa. (A) Phylogenetic analysis of selected alveolate lineage-specific expansions. Groups supported by a bootstrap value  $>70\%$  are colored violet for *Toxoplasma gondii*, orange for *Plasmodium falciparum*, red for *Theileria parva*, blue for *Cryptosporidium parvum*, and brown for *Tetrahymena thermophila*. Trees for Rhopty (Ron2) kinase expansion in *T. gondii* (with closest related kinases from other organisms as outgroups), R45 kinase expansion of *P. falciparum*, TP03\_0339-like Major Facilitator Superfamily transporter proteins of *T. parva*, RESA-like DnaJ proteins of *P. falciparum*, and Insulinase expansion of *C. parvum* are shown. Organism abbreviations are as shown in Fig. 2. *P. berghei* (Pber) and *P. chabaudi* (Pcha) are additional abbreviations. (B) Size distribution of the lineage-specific clusters in all four apicomplexans. Cluster size (X-axis) is plotted against the number of lineage specific clusters of each size in double logarithmic coordinates. The equation of the power law distribution fitting the linear part of the data is shown on the graph. (See also color insert.)

variation because these exposed proteins, which might also be required for cytoadherence, are likely to come under attack from the host immune system (Freitas-Junior *et al.*, 2000).

Examples of smaller LSEs among membrane proteins include different types of solute transporters. For example, both *Cryptosporidium* and *Theileria* show an expansion of a specific class of ABC transporters, whereas both *Theileria* and *T. gondii* show an expansion of MFS transporters (Fig. 3). Like the aforementioned variant surface proteins, many of the expanded transporters are also encoded in subtelomeric regions, suggesting the

possibility that they are expressed differentially in different cell cycles to provide antigenic variation for immune evasion (Figuciredo *et al.*, 2002; Freitas-Junior *et al.*, 2000) (Table I). However, as noted, the overall fraction of transporters encoded by apicomplexa is on average comparable to the other eukaryotes, and subtelomeric transporter gene expansions are also seen in free-living eukaryotes such as *S. cerevisiae* (Mortimer *et al.*, 1992) (Fig. 2). Hence, these transporters might instead represent specially adapted versions that expanded due to efficient nutrient uptake properties in unique intrahost locations. Another functional theme associated with LSEs of apicomplexan surface proteins is exemplified by the oocyst wall protein (OWP) family in *Cryptosporidium* (Templeton *et al.*, 2004b). In this case, the multiple OWP genes probably play a role similar to gene amplifications in providing sufficient templates for synthesizing large amounts of protein, especially for production of thick-walled cysts required for transmission to new hosts.

Another major group of proteins that show lineage-specific expansions are those exported into the host cytoplasm. Several examples are seen in *P. falciparum* (Fig. 3), such as the RESA-like specialized DNAJ-domain proteins, SERA proteases, plasmepsin proteases, and fatty acyl CoA synthetases (PfACS). Plasmepsins are involved in degradation of hemoglobin, and the PfACSs localize to the peripheral erythrocyte skeleton and mobilize fatty acyl-CoA for the parasite, which is unable to carry out *de novo* fatty acid synthesis (Liu *et al.*, 2006; Matesanz *et al.*, 2003; Miller *et al.*, 2002b). Lineage-specific TASH family of proteins from *T. annulata* includes several proteins with DNA-binding AT-hook motifs, which localize to the host-cell nucleus and appear to alter its chromatin structure and potentially transcription of host genes (Pain *et al.*, 2005; Shiels *et al.*, 2004). A similar theme is suggested by two independent groups of kinases respectively expanded in some *Plasmodium* species such as *P. falciparum* (the R45-kinase family) and *T. gondii* (ROP2-like rhoptyr kinases), at least some of which might directly phosphorylate host proteins (El Hajj *et al.*, 2006; Schneider and Mercereau-Puijalon, 2005) (Fig. 3, Table II). In these cases, two major selective pressures could have led to LSE: (1) LSEs might allow targeting a diverse range of host proteins by means of a common biochemical mechanism—similar catalytic (e.g., kinases) or protein-ligand interactions (e.g., AT-hook-DNA interactions). Members of the R45 kinase expansion in *P. falciparum* show differential expression patterns, whereas *T. gondii* ROP2-like kinases show major differences in the active site, implying that they might target different host proteins at different spatial or temporal points. (2) LSEs might also allow more efficient channeling of host resources for the parasite. Thus, expansions of plasmepsins and PfACS might respectively increase efficiency of hemoglobin digestion and fatty acid mobilization.

Several striking examples of differential LSEs of surface proteins are seen even between species of the genus *Plasmodium* with otherwise congruent life

TABLE II  
Examples of Surface Proteins with Ancient Conserved Domains

Extracellular molecules			Phyletic distribution	Architectures/comments
Domain family	Specific proteins/ gene names			
Adhesins	TSPI-domain	TRAP	Pl, Th	SP + vWA + TSP1; links moving junction to parasite cytoskeleton
		MTRAP	Pl	SP + 2xTSP1 + TM; links moving junction to parasite cytoskeleton in merozoites
		CTRP	Pl	SP + 5xvWA + 6xTSP1; Essential for ookinete penetration of mosquito midgut epithelium
	Apple domain	PFF0800w	Pl	SP + TSP1 + vWA + TM; Possible merozoite adhesin
		PTRAMP	Pl	SP + TSP1 + TM; Expressed in merozoites and colocalizes with AMA-1 to the micronemes and cell surface
		TgMIC2	Tgon	SP + vWA + 6xTSP1; Microneme protein also present on parasite surface and binds target cells
Apple domain	CSP	Pl	SP + TSP1 + TM; Major surface protein of sporozoites. Believed to be required for invasion of mosquito salivary glands and known to bind heparan sulfate proteoglycans during hepatocyte invasion. Shows population polymorphism	
	AMA-1	Pl, Tgon, Th	SP + 2xAPPLE + TM; Stored in micronemes and is probably involved in merozoite attachment to via binding of erythrocyte protein Kx. Shows population polymorphism	
	MAEBL	Pl, Th	SP + 4xAPPLE + MAEBL-C + TM; Essential for sporozoite invasion of salivary glands. Is also expressed by merozoites and shows alternative splicing to generate a soluble version	
	SAG1	Tgon	SP + SAG1; Cupredoxin fold domain present on tachyzoite and bradyzoite cell surfaces	

(continued)

TABLE II (continued)

Extracellular molecules		Architectures/comments	
Domain family	Specific proteins/ gene names	Phyletic distribution	
MAC-perforin	MAOP, SPECT (I)	Pl, Th, Tgon	SP + Macperforin, Essential for ookinete midgut invasion in Pl. SPECT (I) is required for liver cell invasion
EGF-repeats containing proteins	TgMIC6	Tgon	SP + 3xEGF + TM; Escorts TgMIC1 and TgMIC4 to the micronemes in tachyzoites
	TgMIC8	Tgon	SP + 14xEGF + TM; Escorts TgMIC3 to the micronemes tachyzoites
Laminin-like coiled-coils	MSP1, MSP4, MSP5, MSP8, MSP10	Pl	SP + 2xEGF + TM; Merozoites surface proteins. MSP1 is believed to be required for invasion by binding erythrocyte band 3. Is proteolytically processed before invasion
	Pfs25	Pl	SP + 4xEGF + TM; Expressed on zygotes and ookinets of Pfall
Sushi	SPECT2	Pl	SP + $\alpha$ -helical extracellular domain + TM, required for liver invasion in Pl
	TgRON1, PfASP	Cpar, Pl, Tgon	SP + RONI-N + Sushi + TM (TM only in Pl); Expressed in the late schizont stage of Pl merozoites in erythrocytes, Localizes to the rhoptry neck in Tgon; May be essential for invasion
Secreted ookinete adhesive protein	SOAP	Pl	SP + cysteine-rich domain; Expressed in ookinets and young oocysts; targeted to micronemes; involved in mosquito midgut invasion by perhaps binding to mosquito laminin. Also expressed in Pl blood stages
<b>Peptidases</b>			
Papain fold Peptidase domain	Falcpain	Pl	SP + papain; Hemoglobins expressed in the food vacuole of Pl trophozoites in the blood stages
Pepsin-like aspartyl protease	Toxopain-1	Tgon	SP + papain; Localizes to the rhoptries and is involved in the processing of rhoptry proteins. In PV, it may be involved in degrading host proteins
	Cryptopain	Cpar	TM + SERA protease; Sporozoite surface protein
	Plasmeprin-I, -II, -IV	Pl	TM + Aspartyl protease; Hemoglobins expressed in the blood stages of Pl
Subtilisin	TgSUB2	Tgon	SP + subtilisin + TM; Essential gene expressed in tachyzoites that associates with rhoptries and may be involved in processing rhoptry proteins
<b>Signaling proteins</b>			
PP2C-like phosphatases	PISUB2	Pl	SP + subtilisin + TM; Merozoite protein stored in micronemes; involved in the shedding of MSP1 and AMA1
Kinases	PP2C-hn and homologs	Pl, Tgon	SP + PP2C; Rhoptry protein that is secreted into the host nucleus and may target phosphorylated host nuclear proteins. Other secreted PP2C proteins of <i>Toxoplasma</i> may similarly target host proteins.
	R45 kinases	Pl, Tgon, Cpar	SP + kinase; The Pl R45 is expressed in trophozoites and exported to the erythrocyte membrane. These proteins may be generally involved in host protein phosphorylation
Potassium channels	ROP2 family of kinases	Tgon	SP + kinase; ROP2 is probably inactive and mediates binding of host mitochondria to parasitophorous vacuole
	PF14_0043	Pl, Tgon, Cpar	K-channel + cyclase + TPR repeats; Ancient alveolate specific protein. May regulate the extrusosome using ion flux and cyclic nucleotides
	PFL1315w (PFkchl)	Pl, Tgon, Th, Cpar	SP + K-channel + K-channel-C; expressed in the intra-erythrocytic life cycle. K-channel-C is a eukaryote-specific domain fused to K-channels
<b>Other secreted and membrane anchored enzymes and proteins that are poorly characterized</b>			
PPX1-like phosphatases	MAL13P1.121	Pl, Tgon	SP + PPX1; Might catalyze the hydrolysis of extracellular nucleotides or phosphorylated small molecules and may affect host signaling
MAM-copper amine oxidase	Cgd3_3430	Cpar, Tgon	SP + MAM + Cu-amine-oxidase; May oxidize the amine side chains of basic amino acids such as lysine in extracellular proteins. Lysyl oxidases are important for biogenesis of fibrillar extracellular matrices
Insulinase	Tgon VIII#59, m00083	Cpar, Pl, Tgon, Th	SP + Insulinase; May be involved in processing rhoptry and micronemal proteins. In Tgon, an insulinase is expressed in the rhoptry
EF-hand	PF14_0607	Cpar, Pl, Tgon, Th	8xTM + 4xEF-hand. A component of the calcium-dependent signaling pathway. Might be involved in calcium-dependent extrusosome activation
	cgd2_2010	Cpar, Pl, Tgon, Th	SP + 2xTM + EF-hand; May be involved in calcium-dependent signaling.

Species abbreviations: Pl, *Plasmodium*; Pfall, *Plasmodium falciparum*; Tgon, *Toxoplasma gondii*; Cpar, *Cryptosporidium parvum*; Th, *Theileria*.  
Other abbreviations: PV, parasitophorous vacuole; SP, signal peptide; TM, transmembrane segment.

cycles, hosts, and tissue specificity (Carlton *et al.*, 2002; del Portillo *et al.*, 2001; Gardner *et al.*, 1998). Further, members of several of these LSEs show evidence for rapid sequence divergence or high nonsynonymous versus synonymous codon substitution rates (Plotkin *et al.*, 2004). Hence, different members of expanded clusters appear to be under strong selection to diversify, either to acquire distinct roles or evade host attack. One notable example of this diversification within LSEs is illustrated by the subset of DBL domain proteins of *P. falciparum* involved in erythrocyte invasion (i.e., EBA175 and its paralogs). Members of this LSE have acquired capabilities to bind multiple erythrocyte receptors, thereby providing *P. falciparum* with multiple means of invading erythrocytes, in contrast to *P. vivax* which is entirely dependent on binding the Duffy antigen (Mayer *et al.*, 2004). Thus, LSEs are not only of notable adaptive significance at the level of diverse apicomplexan genera, but also in closely related species within a genus. In conjunction with sequence analysis, these observations also provide contextual precedence for deciphering functions of uncharacterized proteins that show LSEs. For instance, two major uncharacterized LSEs unique to the *P. yoelii*/*P. berghei* lineage, respectively typified by membrane proteins PY00238 and PY07566, with rapidly diverging sequences, are likely to be novel variable antigens that might have a role in cytoadherence. On the other hand, a unique lineage-specific family of 11 TM proteins from *Cryptosporidium*, encoded in subtelomeric regions, might potentially function as a novel type of transporter.

### III. Conserved Domains in Apicomplexan Adhesion and Host-Interaction Proteins

The diversity of apicomplexan surface proteins involved in pathogenesis is astonishing. Yet, previous studies have suggested they can be effectively organized based on the evolutionary histories of conserved protein domains found in them (Aravind *et al.*, 2003b; Templeton *et al.*, 2004a). Accordingly, we use this as a framework for the further discussion of various aspects of apicomplexan parasitism. The main categories of domains as per their evolutionary categories in apicomplexan surface proteins include: (1) conserved globular domains that are otherwise found primarily in animal or bacterial cell-surface proteins, which mainly function in cytoadherence; (2) conserved domains, as well as distinctive low complexity segments in proteins, found exclusively in apicomplexa or only certain lineages within apicomplexa; and (3) another prominent category, distinct from the former two, composed of diverse enzymatic domains that catalyze a range of extracellular or cell surface reactions or their catalytically inactive derivatives.

Not surprisingly, proteins containing these domains perform diverse biological roles, but their functions show several unifying themes. A summary of these functional themes and structural information for representative protein-containing domains from all these categories is provided in Tables I and II.

#### A. Conserved Domains of Animal, Bacterial, and Ancient Eukaryotic Provenance

The major role of a distinctive set of protein domains in cytoadherence first came to light in the course of early molecular studies on the animal connective tissue, immune cytoadherence, and blood-clotting systems. Examples of these protein domains include EGF repeats, thrombospondin-1 (TSP1), APPLE, kringle, von Willebrand factor A (vWA), fibronectin-type II and type III domains (FNII and FNIII), and scavenger receptor (SR) domains (Patthy, 1999). These domains span the entire spectrum of structural diversity, including entirely disulfide-supported forms such as the EGF-repeat, ancient  $\alpha/\beta$  globular folds such as the vWA domain and the FNIII domain containing  $\beta$ -sandwich folds with immunoglobulin (Ig)-like topology. Some of these domains exclusively occur in extracellular contexts (e.g., kringle or APPLE), whereas others occur in both intracellular and extracellular contexts (e.g., vWA), with the latter versions typically forming evolutionarily distinct groups. Yet, most of these domains have similar functions, mediating cell-cell or cell-connective tissue interactions by means of protein-protein or protein-polysaccharide interactions of varying specificity (Patthy, 1999). Early molecular studies on circumsporozoite protein (CSP) and TRAP from *Plasmodium* (Robson *et al.*, 1988) showed that they contained the TSP1 domain. Likewise, the merozoite surface protein, MSP1, and sexual stage antigen Pfs25, also from *Plasmodium*, were revealed to contain EGF repeats (Blackman *et al.*, 1991; Kaslow *et al.*, 1988) (Fig. 4A). These studies suggested that apicomplexa possess surface proteins that share conserved domains with animal cytoadhesion proteins. These commonalities in animal and parasite proteins sparked considerable interest because these molecules were seen as potential vaccine targets and they hinted that both parasite and host might exploit similar mechanisms for cytoadherence. Subsequently accumulating genomic sequence made it increasingly clear that not just *Plasmodium* but also other apicomplexans possessed several proteins with a diverse range of domains typical of animal adhesion molecules (Fig. 4A). This suggested that animal-like adhesion domains were probably acquired by the apicomplexans in the course of their long parasitic evolution through lateral gene transfer from the animal host (Aravind *et al.*, 2003b). However, representatives of a subset of these domains also turned up in plants, fungi, other eukaryotes, and bacteria (Aravind *et al.*, 2003a), raising the possibility

that they are ancient vertically inherited domains that were similarly deployed in animal and apicomplexan cytoadherence.

Availability of genomes from the major branches of apicomplexa, as well as free-living alveolate outgroups such as *Tetrahymena* (Eisen *et al.*, 2006) and several other eukaryotes has enabled us to more or less settle the issue of lateral acquisition of animal domains versus vertical inheritance from more ancient ancestors. The emerging picture for adhesion domains is rather complex (Templeton *et al.*, 2004a). Some are indeed of unequivocal animal provenance acquired through lateral transfer, but others were vertically inherited from earlier eukaryotic ancestors and yet others through lateral transfer from bacteria (Fig. 4B). A total of 18 types of noncatalytic extracellular adhesion domains from apicomplexan proteins could be established as being of ultimately animal origin. Several domains in this list, namely TSP1, Sushi/CCP, Notch/Lin1 (NL1), NEC (Neurexin-Collagen domain), Fibronectin type 2 (FN2), MAM, and the Scavenger receptor domain, are only found in animals and apicomplexans. The remaining domains (e.g., vWA and SCP/PR1, kringle, and the newly identified F09F7.1-like domains) are found in other eukaryotes or prokaryotes, but apicomplexan versions are closer to animal versions to exclusion of all others. In these cases, apicomplexan versions of the domain showed either a significant relationship in phylogenetic trees or uniquely shared patterns of disulfide-bonding cysteines or arrangement of cysteines with their animal counterparts (Templeton *et al.*, 2004a). Some domains, such as EGF repeats, are found in a number of other eukaryotic lineages including free-living alveolates (Eisen *et al.*, 2006). The small size and poor sequence conservation, beyond the disulfide-bonding cysteines precludes statistically well-supported resolution of phylogenetic relationships of all apicomplexan EGF repeats. However, presence of some distinctive versions of this domain (e.g., the ephrin-receptor-type EGF domain), uniquely shared with animals, suggests that at least certain apicomplexan representatives have an ultimately animal provenance. *T. gondii*

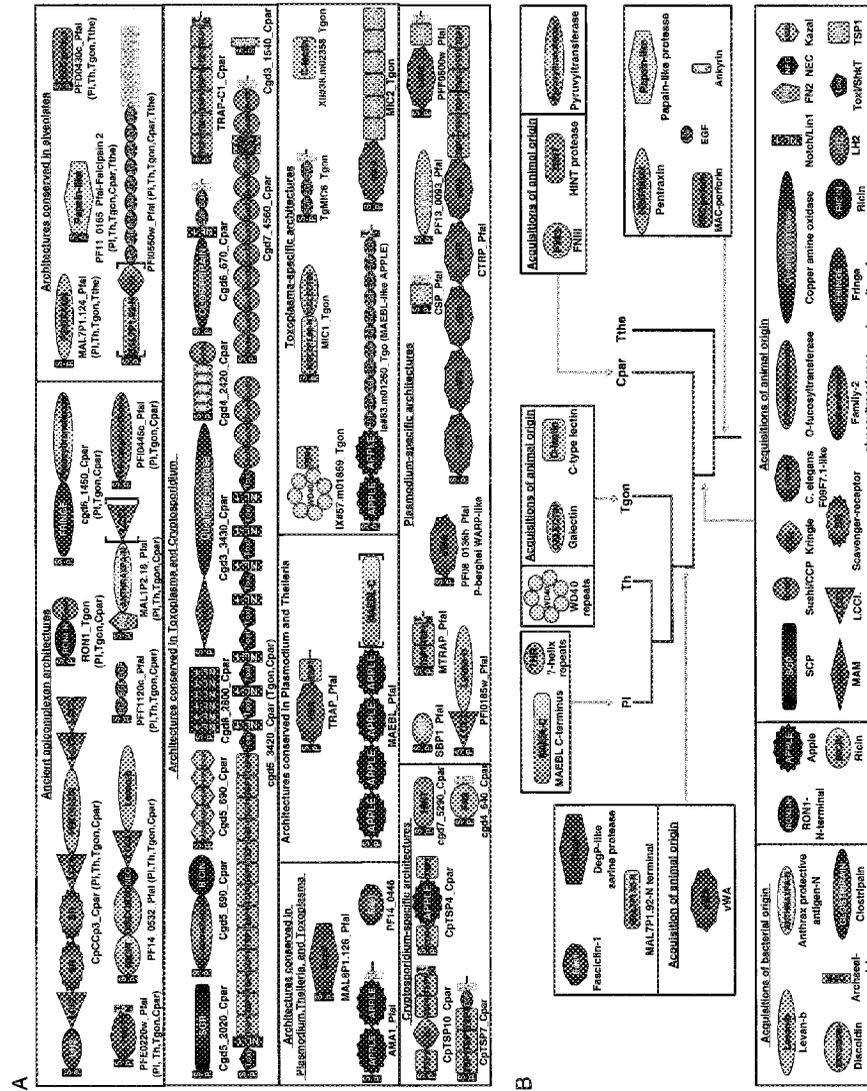


FIG. 4 Domain architectures and evolution of conserved adhesion and invasion proteins. The figure particularly emphasizes domains that were acquired by apicomplexa from bacteria and animals. (A) Domain architectures are grouped in boxes based on their conservation pattern within alveolates. Proteins are denoted by their gene/common names and their species abbreviations separated by an underscore, shown below the architecture. For conserved domain architectures found in more than one alveolate lineage, phyletic patterns are shown next to protein names in brackets. Domains enclosed in square brackets are not present in all orthologs of the displayed protein. Domain names are given in Fig. 3B. SP represents a signal peptide, and TM represents a transmembrane region. (B) Domains are shown around an alveolate tree and grouped according to their inferred point of origin or acquisition as depicted by blue arrows. Within each group, domains predicted to have been laterally acquired from bacteria or animals are further segregated. Boxes without a title depict domains acquired in apicomplexa by vertical inheritance or of unclear derivation. Cpar, *Cryptosporidium parvum*; Pl, *Plasmodium*; Pfa, *Plasmodium falciparum*; Tgon, *Toxoplasma gondii*; Tthe, *Tetrahymena thermophila*; Th, *Theileria*. (See also color insert.)

contains two proteins respectively, with a C-type lectin domain and galectin domain (Saouros *et al.*, 2005), both of which are found commonly in animals and infrequently in plants or fungi. In phylogenetic analysis, the *T. gondii* versions show weakly supported, but consistent, grouping with animal forms. Taken together with their sporadic distribution in apicomplexa, it is likely that the C-type lectin and galectin domains in coccidians are probably of animal origin.

Systematic examination of apicomplexan genomes revealed that domains showing animal affinities are more frequently found in cytoadherence proteins rather than uniformly across all functional categories of proteins (Templeton *et al.*, 2004a). In addition to the noncatalytic adhesion domains, apicomplexans contain several secreted or TM proteins with adhesion-related enzymatic domains showing evolutionary affinities to animal protein domains (see later for discussion, Fig. 4B). These observations suggest that domains acquired at some point in evolution from animal hosts were predominantly recruited in functional contexts pertaining to cytoadherence. Nevertheless, current comparative genomic analysis indicates a subset of adhesion domains, previously believed to be of animal provenance, appears to have been vertically inherited by apicomplexans from a more ancient eukaryotic common ancestor. Chief among these are the pentraxin domain (also found in vertebrate serum proteins such as the C-reactive protein and the serum-amyloid protein) and the Mac-perforin domain (found in vertebrate membrane attack complex/perforin-like proteins) (Aravind *et al.*, 2003b; Templeton *et al.*, 2004a). Apicomplexan proteins with both these domains contain orthologs in ciliates suggesting they were inherited from the alveolate common ancestor. Apicomplexans and ciliates also share a membrane protein (e.g., *P. falciparum* MAL7P1.92) with 10 TM segments and a core conserved N-terminal extracellular module with at least two EGF repeats (Fig. 4A). This protein family has undergone a huge lineage-specific expansion in ciliates, and orthologs are also sporadically found in other unicellular eukaryotes such as the chlorophyte *Ostreococcus tauri* and *Dictyostelium discoideum*. Hence, this protein was probably vertically inherited from the alveolate common ancestor, rather than laterally transferred from animals.

A third group of domains in apicomplexan adhesion molecules show clear affinities to domains found in bacterial surface molecules involved in cell-cell interactions, S-layer biogenesis, and film formation (Fig. 4B). These include the Anthrax-protective antigen  $\beta$ -strand rich domain, the levanase-associated (levan-b), the Cys-arch (archaeal-cysteine rich), and discoidin domains (Templeton *et al.*, 2004a). The majority of these domains might participate in carbohydrate-protein interactions as in the case of their bacterial counterparts (Pradel *et al.*, 2004). A few other adhesion domains in apicomplexan proteins, such as the APPLE domain, fascilin1 domain,  $\beta$ -helix repeats and

some versions of the ricin domain, are widely distributed in both eukaryotic and bacterial adhesion proteins. This, taken with their absence in currently available ciliate genomes makes their provenance unclear. They could have potentially been acquired through lateral transfer from either an animal or bacterial source, or inherited from an earlier ancestor, but lost in ciliates. SRS/SAG1 surface antigen proteins of *T. gondii* (He *et al.*, 2002) contain a conserved globular domain found in animal ephrins that are ligands involved in neural development and also extracellular copper-chelating proteins (plastocyanins) from cyanobacteria and plants. The extreme sequence divergence of *T. gondii* SRS/Sag1 domain from all these forms makes it equally probable it was acquired through lateral transfer from animals or evolved from a plastocyanin-like precursor inherited from the apicoplast. Likewise, the  $\beta$ -propeller domain found in *P. falciparum* GBP and GBPH2 (Nolte *et al.*, 1991) proteins have homologs in both animals and bacteria. Their unique presence in just one apicomplexan taxon supports an origin through lateral transfer, but their comparable similarity to both bacterial and animal proteins makes it impossible to identify the original source of these domains. Likewise, origins of the divergent  $\beta$ -helix repeats in the N-terminal module of skeleton-binding protein 1 (PfSBP1) of *P. falciparum* (Maier *et al.*, 2006) and *Cryptosporidium* cgd5\_4480 remain unclear, although their sporadic occurrence is suggestive of potential involvement of lateral gene transfer. Further genomic sequences and protein-structure data might resolve some of these uncertain evolutionary affinities.

Phyletic distribution of conserved adhesion protein domains in apicomplexa sheds light on the temporal sequence of their evolution (Fig. 4B). Most domains of clear-cut bacterial provenance appear to have been acquired prior to the divergence of apicomplexans from their common ancestor. A few inherited from the common alveolate ancestor, such as the mac-perforin and pentraxin domains also appear to be of ultimately bacterial origin. This suggests that the major period of lateral transfer from bacteria of such domains was in the early phase of alveolate and apicomplexan evolution. Although some such domains could ultimately be derived from the apicoplast chloroplast, which is of cyanobacterial origin, it is more likely they were derived from environmentally co-occurring bacteria, as was previously suggested for the mac-perforin and discoidin domains through phylogenetic analysis. Interestingly, the majority of domains of animal origin are also traceable to the ancestral apicomplexan, suggesting that the major acquisition of animal-like adhesion domains occurred early in their evolution. However, there are examples such as the C-type lectin, FNIII, and vWA domains that show a more sporadic phyletic distribution, suggesting that there have been occasional lineage-specific acquisitions of adhesion domains from the animal hosts in the course of apicomplexan evolution (Templeton *et al.*, 2004a) (Fig. 4B).

Apicomplexan surface proteins generally do not share identical or closely related architectures with counterparts from animals or bacteria containing homologous domains. In apicomplexan proteins, domains of different origins may be combined together to result in novel architectures. For example, in the *Plasmodium* protein, MALIP2.18, LCCL, and FNII domains of animal origin are combined with the Anthrax-protective antigen domain of bacterial origin (Claudianos *et al.*, 2002; Delrieu *et al.*, 2002; Pradel *et al.*, 2004). Another such case is the member of the aforementioned family with EGF repeats and 10 TM segments. In one of the paralogs of this family in hematozoans (e.g., PFI0550w), ephrin-receptor type EGF repeats and a kringle domain of animal origin were added to the conserved core inherited from the ancestral alveolate (Fig. 4A). Thus, although particular adhesion domains were acquired through lateral transfer from different sources, they were uniquely combined in apicomplexans to spawn novel architectures. In animals there is a strong correspondence between the protein domain and the exon coding it in adhesion proteins (Patthy, 1999). However, no significant correlation was observed in multidomain apicomplexan adhesion proteins. Hence, it appears that architectural diversity in apicomplexan surface proteins, in contrast to their animal counterparts, did not arise via conventional exon-shuffling (Templeton *et al.*, 2004a). Domain architectures of adhesion proteins are fluid within the apicomplexa. They diversify mainly due to domain accretion, in which new domains are added to the pre-existing architectural core (e.g., in PFI0550w as mentioned previously). However, a few architectures such as a protein with FNII and Anthrax-protective antigen domain; a protein with multiple LCCL domains, SR, LH2, and PTX; and a domain and a protein with the F09F7.1-like domain are conserved throughout apicomplexa, suggesting these architectures had already emerged early in evolution and were maintained due to strong functional constraints (Fig. 4A). The maximum architectural innovation (17 distinct architectures) is observed in the *Cryptosporidium* lineage, suggesting that it had a long independent history after separating from the crown group of coccidian and hematozoans. Surprisingly, hematozoans do not share many unique architectures compared to those traceable to the ancestor of the crown group, suggesting that architectural innovation in these adhesion proteins did not play a major role in adapting to vertebrate blood parasitism. In addition to architectural diversification, some conserved adhesion domains might show rapid sequence evolution. One such case is the APPLE domain, found in proteins throughout apicomplexa. However, there is a specific family of them in the crown group apicomplexa, prototyped by *Plasmodium* AMA1 and MAEBL proteins in which the APPLE domains have greatly diverged in sequence while retaining their structure (Bai *et al.*, 2005; Nair *et al.*, 2002). This suggests rapid and extreme sequence divergence might have also been important in adaptation of conserved adhesion modules for specific new functions.

## B. Globular Domains, Transmembrane, and Low-Complexity Segments of Exclusively Apicomplexan Provenance

Apicomplexa also display a wide range of unique protein domains that are either found throughout the clade or more often only in particular taxa. These domains figure prominently among the larger lineage-specific expansions shown by apicomplexans (Table I). At least a subset of them appears to include surface antigens that show antigenic variation enabling immune evasion. There are three broad categories of such domains in surface proteins: (1) Distinct globular domains of diverse structural categories; (2) multi-TM domains with variable solvent-exposed loops; and (3) low-complexity segments enriched in particular amino acids that sometimes undergo covalent modification (Fig. 5).

### 1. Apicomplexa-Specific Globular Domains

This category of domains first came to light in classical studies on *Plasmodium* variant surface antigens. Typical examples are globular domains in products of lineage-specifically expanded *vir* and *var* gene families, respectively, from *P. vivax* and *P. falciparum*. Best studied of these are products of the *var* gene family, PfEMP1 proteins, which adhere to endothelial cell receptors and thereby enable the parasite to avoid splenic clearance (Smith *et al.*, 1995; Su *et al.*, 1995). PfEMP1 contains multiple copies of the DBL domain (*Duffy-binding-like*), and they are also found in 1 to 2 copies in the erythrocyte glycoprotein A-binding protein, EBA-175, and related erythrocyte-invasion proteins such as BAEBL, EBL-1, and JESEBL from *P. falciparum* (Mayer *et al.*, 2001; Singh *et al.*, 2006; Tolia *et al.*, 2005) (Fig. 5). Orthologs of this second group of proteins are found in other *Plasmodium* species in single or small copy number and include the Duffy-antigen binding protein DBP from *P. vivax* (Singh *et al.*, 2006). The DBL domain fold is versatile with different versions specializing in binding polysaccharide chains (sialates on glycoprotein A bound by EBA-175 and heparan sulfate bound by PfEMP1 DBL domains) or sulfated tyrosines (those on Duffy antigen bound by DBP) (Choe *et al.*, 2005; Singh *et al.*, 2006; Vogt *et al.*, 2003). The EBL-1 and DBP proteins share another *Plasmodium*-specific cysteine-rich globular domain (the MAEBL-C domain) present C-terminal to DBL domains in these proteins and also C-terminal to APPLE domains in MAEBL (Kappe *et al.*, 1998). Most PfEMP1 proteins instead contain a novel C-terminal  $\alpha$ -helical globular domain (PfEMP1C) of approximately 90 amino acids. The PfEMP1-C domain is also independently present in multiple copies combined with an N-terminal *vir*-type globular domain (VGD) in *P. falciparum* surfactin proteins or as stand-alone copies in PFB1045w-like proteins, the giant protein Pf332 and *P. knowlesi* SICAVAR



interactions represent “easier” pathways for innovation of globular domains, which bypass the more intensive selection needed to generate elaborate hydrogen-bonding interactions of complex  $\alpha + \beta$  domains. Such an evolutionary pathway is strikingly illustrated by solved structures of DBL domains (Singh *et al.*, 2006; Tolia *et al.*, 2005). Structural analysis of a single DBL domain shows that it contains three subdomains: an N-terminal flap predominantly stabilized by disulfide bonds; the first helical subdomain with four conserved helices, of which three form a central bundle packed by hydrophobic interactions and are further stabilized by disulfide bonds; and a second helical subdomain that is similarly structured. Structural comparison of the two helical subdomains reveals that they are homologous 4-helical units that have duplicated from a common ancestor. Thus, the DBL domain appears to have emerged in part by serial duplication: long helical segments initially duplicated to form the ancestral version of the two helical subdomains. This whole unit further duplicated resulting in a two subdomain structure, which appears to have contained an exposed cleft. This cleft then was filled in by innovation of a cysteine-supported N-terminal flap (the first subdomain), resulting in a stable globular DBL domain. Thus, the principal stabilizing forces in emergence of this domain were the hydrophobic interactions between helices and the disulfide bonds. VGD and related PRESAN (*Plasmodium RESA* N-terminal domain) domains and PfEMP1-C domains, are other comparable examples of innovations of entirely  $\alpha$ -helical domains. On the other hand, the MAEBL-C domain and cysteine-rich repeats in the OWP domain appear to represent examples of innovations of cysteine-supported structures.

In contrast, FAINT domains and the N-terminal domain of the *T. gondii* mucin-like proteins are different in being all  $\beta$ -strand domains. They have about 7 to 8 predicted  $\beta$  strands and are of comparable length to  $\beta$ -sandwich folds with Ig domain-like topologies (Patthy, 1999). Hence, similar to the SRS antigen extracellular domain from *T. gondii* (He *et al.*, 2002), they might be divergent representatives of pre-existing  $\beta$ -sandwich domains. This implies these  $\beta$ -rich domains might not be real *de novo* lineage-specific innovations, but divergent versions of ancient Ig-like folds.

## 2. Apicomplexa-Specific Multi-TM Domains

All apicomplexans contain several lineage-specific multi-TM proteins unrelated to ancient conserved multi-TM proteins such as transporters and ion channels. These proteins have two or more TM segments along with soluble extracellular and intracellular loops of variable lengths. One prominent group of such multi-TM proteins are the 2 TM proteins, which were initially identified in studies on Maurer's cleft proteins in *P. falciparum* (Khattab and Klinkert, 2006;

Sam-Yellowe *et al.*, 2004a). This study noted the PfMC-2 TM family was structurally and functionally related to the PfST-2TM family and the large rifin superfamily (including stevors) from *P. falciparum* and the PyST-B-2 TM family from *P. yoelii*, *P. chabaudi*, and *P. berghei*. Further analysis of *Plasmodium* genomes reveals there might be several other such proteins, some of which form moderate-sized families such as the PFB0995w-like 2 TM family in *P. falciparum*. Together, these proteins might form a large assemblage of at least 205 proteins in *P. falciparum* (including  $\sim 180$  of the rifin superfamily, 11 of PfMC-2 TM, 3 to 4 of Pf-ST-2 TM, and 12 PFB0995w-like 2 TM) and 32 proteins in *P. yoelii* (all of the PyST-B-2 TM family, Table I). These proteins share a characteristic structure with an N-terminal signal-like sequence followed by a cytoplasmic loop, in turn followed by two transmembrane regions with a highly variable externally located loop between them and a charged cytoplasmic C-terminal tail (Fig. 5) (Khattab and Klinkert, 2006; Sam-Yellowe *et al.*, 2004a). The N-terminal signal-like sequence is often associated with a conserved cysteine that could potentially be lipid-modified while the protein is being localized to the membrane. The signal is typically followed by PEXEL motifs (see Section IV.E) that target the protein for erythrocyte export. The N-terminal cytoplasmic tail also often contains cysteines (Sam-Yellowe *et al.*, 2004) that might stabilize this structure via disulfide bonding. In rifins alone, the equivalent of the N-terminal TM segment has an unusual sequence composition, suggesting it might have functional interactions different from the rest of the family. Analysis of genomes of other apicomplexans revealed smaller expansions of comparable 2 TM domains in *Theileria* (e.g., TP05\_0009 family) and *T. gondii* (Table I). One of these 2 TM families typified by PFL0745c is interestingly conserved in all completely sequenced apicomplexans. Although there is no detectable sequence similarity between these proteins and the *Plasmodium* proteins, it is likely that they all adopt similar structures. Thus, 2 TM proteins appear to have a much wider presence in apicomplexa and possibly emerged from a common ancestor earlier in evolution.

*Theileria* contains a unique LSE of TM proteins with a conserved 7 TM domain (Fig. 5), the *Theileria*-specific repeat protein family (TSRP; also known as *Theileria parva* repeat, TPR family). We recommend the former nomenclature to avoid conflation with the well-known but unrelated tetratricopeptide or TPR repeats) (Gardner *et al.*, 2005; Pain *et al.*, 2005). The family contains around 40 proteins, most of which are closely related, suggesting a proliferation. This domain appears to have a distinct conservation pattern in both the TM and loop segments including one GXG signature embedded in a TM segment, suggesting it might form channel-like structures. There are several other smaller families of multi-TM proteins conserved in one or few apicomplexan lineages, but most are of unclear functional significance (Table I).

### 3. Low-Complexity Protein Families Among Apicomplexan Surface Proteins

Extracellular domains of numerous apicomplexan surface proteins are composed almost entirely of low-complexity segments or regions that show a biased/repetitive amino acid composition with overrepresentation of certain types of residues. Such low complexity segments may also be combined with previously described types of structured domains in the same polypeptide (Fig. 5). These low-complexity segments do not necessarily have similar functions and might have convergently evolved on multiple occasions. Yet, in most apicomplexans several genuine families of lineage-specifically expanded low-complexity proteins can be identified (Table I). For example, the lineage-specific ETRAMP family from *Plasmodium*, whose members localize to the PVM, contains proteins with an external-facing, highly positively charged, lysine-rich, low-complexity region (Spielmann *et al.*, 2003). A comparable family of proteins with external low-complexity segments was detected in *T. gondii*. *T. gondii* also codes for two proteins each with a C-terminal 4 TM domain and an N-terminal extracellular region with 3 to 6 repeats of a specific highly polar low-complexity sequence (Fig. 5). Thus, such surface proteins with charged or polar low-complexity external domains appear to be widely utilized in the apicomplexan crown group. Alcoholic side chains of low-complexity segments enriched in serine/threonine serve as sites for polysaccharide attachment and are termed mucins. These are found in a large number of surface proteins from *Cryptosporidium* and *T. gondii*, which possess an O-linked glycosylation pathway (Stwora-Wojczyk *et al.*, 2004; Templeton *et al.*, 2004a) (see Section IV.A; Table I). In *Cryptosporidium* there is an expansion of such proteins, the glycoprotein 40/15 family, whose extracellular regions are largely comprised of mucin segments. *Cryptosporidium* also contains a remarkable 11,696 residue-long secreted protein (cgd3\_720) with 17 copies of a *Cryptosporidium*-specific all- $\beta$  strand cysteine-rich globular domain, most copies of which contain an insert of a mucin segment. In both *T. gondii* and *Cryptosporidium*, there are several mucin segments that form low-complexity stalks for globular domains (Fig. 5). Glycosylated S/T-rich stretches or mucins thus appear to be another common theme between animal and apicomplexan adhesion proteins (Templeton *et al.*, 2004a).

#### C. Secreted and Cell-Surface Enzymatic Domains in Apicomplexans

Apicomplexans possess several secreted or membrane-anchored enzymes, some of which might be extruded via rhoptries or exported into the host cell during intracellular parasitism. Some of the most common among these extracellular

enzymes are several types of peptidases, protein kinases (Table II), a diverse group of hydrolases, and some metabolic enzymes that utilize host substrates for synthesis of metabolites needed by the intracellular parasite.

#### 1. Peptidases

Secretion or deployment of membrane-anchored peptidases is a common feature of several parasites (Rosenthal, 2004). Importance of these enzymes in apicomplexan physiology was first realized in *P. falciparum*, where a number of peptidases were found to be used in degrading the host cell proteins (Greenbaum *et al.*, 2002; Liu *et al.*, 2006; Miller *et al.*, 2002b; Rosenthal, 2004). The falcipain-/cryptopain-like family of thiol proteases with peptidase domains of the papain-like fold and the toxomepsin2-like family with pepsin-like aspartyl peptidase domains are found in all apicomplexans and also in ciliates (Eisen *et al.*, 2006; Rosenthal, 2004). In each alveolate lineage, they show small or large LSEs and potential substrate diversity (Tables I and II). Their phyletic patterns suggest they are likely to be secreted digestive proteases inherited from the ancestral alveolate. *Cryptosporidium*, *T. gondii*, and *Plasmodium* encode a subtilisin-like protease (typified by TgSUB2 and PfSUB2) secreted into rhoptries (Bradley *et al.*, 2005; Miller *et al.*, 2003). Analysis of rhoptry proteins has also uncovered a DegP-like trypsin-type serine protease conserved in the apicomplexan crown group and secreted into rhoptries. Evolutionary affinities of both these proteases suggest they were probably acquired by the apicomplexan through lateral transfer from a bacterial source. Another interesting protease found in *Cryptosporidium* and *T. gondii* (e.g., cgd6\_670), but lost in other lineages, is a membrane-anchored clostripain-type peptidase (thiol peptidase of the caspase fold) The N-terminal peptidase domain of this protein is bacterial in origin and is combined with C-terminal NLI and EGF domains of animal origin (Templeton *et al.*, 2004a) (Fig. 4).

There are also LSEs of other secreted protease families restricted to particular taxa (Table I): *P. falciparum* shows expansions of pepsin-like aspartyl proteases and papain-like thiol proteases, respectively, forming the plasmepsin and SERA protease families (Liu *et al.*, 2006; Miller *et al.*, 2002b). Interestingly, some members of the SERA family have an active serine in place of the usual catalytic cysteine seen in other members of this fold. *Cryptosporidium* has a LSE of over 11 insulinase-like metalloproteases, whereas *T. gondii* has a single homolog of this protein localizing to the rhoptries (Abrahamsen *et al.*, 2004; Bradley *et al.*, 2005). Some other membrane-associated proteases are present more sporadically in certain apicomplexans. One notable example is a large secreted zincin-like metalloprotease found in hematozoans (e.g., PFD0425w) and is related to a family of such enzymes found in the predatory bacterium *Bdellovibrio*. Interestingly,

*Cryptosporidium* also possesses a stand-alone secreted version of the HINT-type peptidase domain (Templeton *et al.*, 2004a) that is closely related to the equivalent domain found in animal developmental regulators, the Hedgehog proteins (Dellovade *et al.*, 2006) (Fig. 4). This appears to be a clear case of lateral transfer from animals, and as in the latter might function as a signaling molecule.

## 2. Secreted Protein Kinases and Phosphatases

Members of the R45 kinase family have signal peptides, a peculiar variant of the ATP-binding site in the N-terminal kinase subdomain, a conserved N-terminal extension with a characteristic tryptophan, and cysteines forming potential disulfide bonds in the C-terminal kinase subdomain. They are present in *Plasmodium*, *T. gondii*, and *Cryptosporidium*, suggesting they were present in the ancestral apicomplexan. The *P. falciparum* versions contain a PEXEL motif (see Section IV.E.1) and are imported into the host erythrocyte where they are believed to target host proteins (Oakley *et al.*, 2007; Schneider *et al.*, 2005). Presence of signal peptides in other apicomplexans suggests they were ancestrally secreted kinases that probably mediated host-protein phosphorylation from early in their evolution. Expansion of this family in a few *Plasmodium* species is mirrored by the expansion of a very different kinase family, namely the Rop kinases in *T. gondii* (El Hajj *et al.*, 2006). The remarkable LSE of Rop kinases in *T. gondii* includes 34 distinct kinases, at least 15 of which localize to tachyzoite rhoptries and are extruded into the host cell during invasion (Bradley *et al.*, 2005) (Table II). These kinases are characterized by the presence of a conserved pair of cysteines in the C-terminal kinase subdomain that are likely to form a disulfide bridge. Some of these proteins appear to have disruptions of catalytic residues in the kinase domain, suggesting they might be inactive and merely act as peptide-binding molecules. *Plasmodium* and *T. gondii* also contain a family of secreted protein phosphatases of the PP2C superfamily. Although a single copy is seen in all *Plasmodium* species, it shows an LSE in *T. gondii* with around eight members, of which at least one is delivered by rhoptries and localizes to the host nucleus (Gilbert *et al.*, 2006). It is quite possible that, like secreted kinases, these phosphatases are also extruded into the host cell via rhoptries and target specific phosphorylated serines and threonines on them (Table II).

## 3. Other Secreted Enzymes

Of the several other extracellular enzymes hydrolases, especially phosphohydrolases seem to be dominant. At least six different types of such phosphohydrolases are seen in apicomplexans, namely pyrophosphatase (PPX),

apyrase, glycerophosphoryldiesterase, a membrane-anchored alkaline phosphatase-like enzyme, an AP-endonuclease-like phosphodiesterase, and rhoptry protein Rop9, which is a phosphatase of the HAD superfamily. All apicomplexans also encode two uncharacterized secreted  $\alpha/\beta$  hydrolases, which might possibly function as lipoesterases or even proteases. *T. gondii* and *Cryptosporidium* also code for a peculiar copper-dependent amine oxidase, whose N-terminal catalytic domain and C-terminal adhesion-related MAM domain both appear to have been acquired through lateral transfer from the animal lineage (Templeton *et al.*, 2004a) (Fig. 4). However, all of these enzymes, irrespective of their ultimate provenance, show phyletic patterns indicating their presence in the common ancestor of apicomplexa, with subsequent loss in certain lineages. This suggests the ancestral apicomplexan already deployed a sizable battery of cell-surface enzymes. At least a subset of these, like one of the  $\alpha/\beta$  hydrolases, are also present in ciliates (Eisen *et al.*, 2006), suggesting they might belong to a more ancient alveolate set of secreted enzymes. In contrast, secreted versions of the PfACs involved in fatty acyl CoA synthesis appear to be a *Plasmodium*-lineage-specific innovation, derived from related enzymes that normally function in intracellular contexts (Matesanz *et al.*, 2003; Norimine *et al.*, 2006). *Plasmodium* also exports a paralogous group of lysophospholipases that might degrade host phospholipids to provide it with lipid intermediates or to modify the host membrane (Oakley *et al.*, 2007). Another lineage-specific enzyme expressed specifically in insect stages of *Plasmodium* development is the chitinase, belonging to the TIM barrel fold. This enzyme might have been acquired through later transfer from insects themselves and is required for degrading chitin while penetrating the peritrophic membrane of the mosquito gut (Vinetz *et al.*, 2000).

## 4. Membrane-Embedded Enzymes with Possible Role in Survival, Signaling, or Pathogenesis

Like other eukaryotes, apicomplexans possess a range of membrane-embedded enzymes that might drive ionic and molecular transport (e.g., classical P-type ATPases and ABC transporter ATPases), lipid and glycolipid biosynthesis, or membrane protein modification and processing (Abrahamsen *et al.*, 2004; Gardner *et al.*, 2005). However, apicomplexans do not appear to have membrane-anchored receptor kinases and possess only a few other receptor enzymes compared to crown-group eukaryotes. As they spend most of their existence as intracellular parasites or need to shield their cell surfaces from host immune responses, they might not extensively depend on enzymatic receptor molecules. Yet, there are few notable exceptions, which might play a potential role in the parasite's interactions with the host. Two adenylyl cyclases with highly distinctive domain architectures

are found exclusively in apicomplexans and ciliates (Linder *et al.*, 1999; Weber *et al.*, 2004). The first of these represented in all four sequenced apicomplexan taxa is a gigantic protein with 22 TM segments, a P-type ATPase domain, and two cyclase catalytic domains (PF11\_0395). The second, which is lost in *Theileria*, combines a distinctive N-terminal potassium channel domain with a C-terminal cyclase domain followed by TPR repeats (PF14\_0043). Their remarkable conservation suggests they were part of the ancestral signaling apparatus shared by alveolates. Localization of the first version to outer alveolar membranes suggests it might transmit signals in response to ionic changes in the alveolar sac. The second version might be a dual regulator that uses both ion flux and cyclic nucleotide generation to control alveolate extrusomes. This suggests cAMP signaling induced by these enzymes is likely to play a major role in the parasite response to extracellular stimuli or environmental ionic changes.

Other membrane-embedded enzymes with possible signaling roles include an enigmatic pan-apicomplexan protein (e.g., PF14\_0350), with an N-terminal multi-TM domain fused to a C-terminal intracellular GCN5-related N-acetyltransferase domain. This implies apicomplexans might possibly deploy an unprecedented signaling mechanism involving acetylation of downstream proteins by this membrane-embedded enzyme. Another enigmatic enzyme is a highly distinctive S2P-superfamily membrane-embedded Zn-metalloprotease of apicoplast endosymbiont origin, which is found in *T. gondii* and *Plasmodium* (e.g., PF13\_0260). S2P proteases have been implicated in both bacteria and eukaryotes in signaling via intramembrane cleavage of substrate proteins (Makinoshima *et al.*, 2006). This phenomenon plays a major role in regulating expression of virulence factors in *Vibrio cholerae* and *Mycobacterium tuberculosis* (Makinoshima *et al.*, 2006), pointing to the possibility that similar membrane-associated cleavages may be critical for apicomplexan host-pathogen interaction.

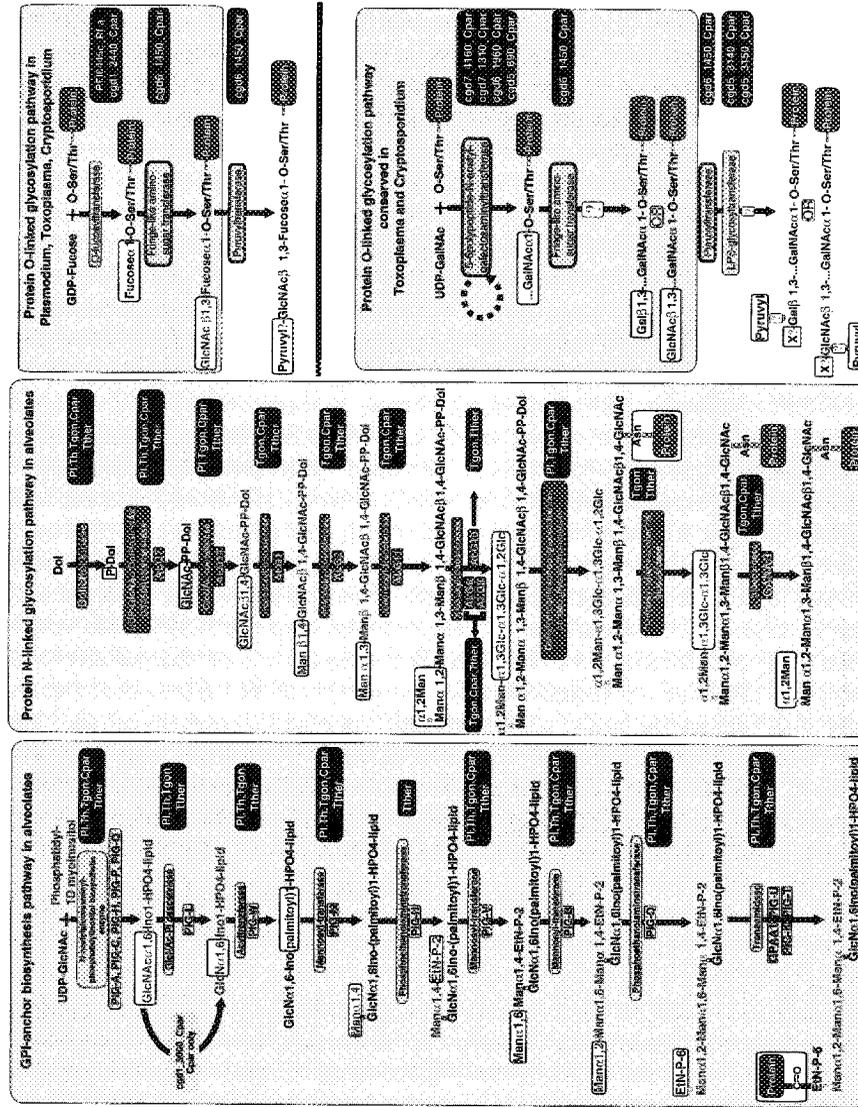
#### IV. Maturation, Deployment, and Export of Apicomplexan Surface Proteins

Apicomplexans share with other eukaryotes an elaborate apparatus, composed of vesicular transport systems and the Golgi network, which is required for maturation of surface proteins (Lingelbach and Przyborski, 2006). However, they also have the unique extrusion system of the ancestral alveolate in the form of the apical complex and the system for exporting proteins into the host cell. Here, we consider some key molecular aspects of surface protein maturation and modification relevant for parasitic interactions. Then we explore new insights regarding the specialized export and extrusion systems.

#### A. Surface Protein Maturation: Glycosylation Pathways

Apicomplexans possess two major glycosylation pathways in common with other eukaryotes—the GPI anchor pathway that synthesizes glycosylphospholipid anchor of membrane-anchored proteins and the N-linked glycosylation pathway that carries out oligosaccharide modification of amide groups of asparagines (Templeton *et al.*, 2004a; Varki *et al.*, 1999) (Fig. 6). However, reconstructed apicomplexan versions of these two pathways show several unique features. *Cryptosporidium*, *T. gondii*, and *Plasmodium* also possess one or both of two O-linked glycosylation pathways that modify hydroxyl groups of alcoholic amino acids (Stwora-Wojczyk *et al.*, 2004; Templeton *et al.*, 2004a). Comparable O-linked glycosylation pathways occur primarily in animals (Varki *et al.*, 1999). GPI anchor biosynthesis in apicomplexans is similar to other eukaryotes except for the absence of enzymes catalyzing adducts of multiple ethanolamine units and the fourth mannosyl unit. This latter feature is shared with ciliates, suggesting the mannose chain in alveolate GPI anchors is shorter (Fig. 6). Ciliates possess only two homologs of ethanolamine transfer enzymes (GPI7 and PIG-O), whereas apicomplexans only possess one (PIG-O). This might mean that either PIG-O catalyzes multiple ethanolamine transfers by itself, or alveolates have fewer ethanolamine adducts, with apicomplexans specifically having only the one adduct to which the polypeptide is covalently linked. *Cryptosporidium*, interestingly, lacks the classical deacetylase catalyzing the second step of the GPI pathway. However, it encodes an unrelated bacterial-type sugar deacetylase that might effectively perform the same function (Templeton *et al.*, 2004a). All alveolates lack the GPI anchor deacetylase which removes the palmitoyl chain in the terminal step of the pathway. *Cryptosporidium* and *Tetrahymena* also lack the palmitoyltransferase (PIG-W). This raises the intriguing possibility that at least in some alveolates the pathway might entirely bypass the palmitoylation step (Fig. 6). In crown-group apicomplexans, which possess a palmitoyltransferase, a different distantly related enzyme, such as the membrane-anchored esterase prototyped by PF11\_0168, could perform deacetylase function.

Alveolates possess enzymes to synthesize the primer for polysaccharide elongation in N-glycosylation, dolichyl phosphate, from dolichol. However, analysis of reconstructed N-glycosylation pathways suggests alveolates are unlikely to possess elaborately branched oligosaccharide side chains present in crown-group eukaryotes (Fig. 6). *Plasmodium* possesses only the first two enzymes for addition of two N-acetylglucosamine units and the final enzyme that transfers the polysaccharide chain from dolichyl phosphate to an asparagine side chain in a protein (Templeton *et al.*, 2004a). In contrast, remaining alveolates possess enzymes for addition of a longer linear chain of mannose and glucose to core N-acetylglucosamine (NAG) units, as well as enzymes that trim the terminal glucose residues to produce pure mannan-modified



proteins. Close similarity of ciliate and apicomplexan N-glycosylation pathways suggests that the latter more or less retain the essential core of the ancestral alveolate pathway, without any currently known elaborations. Loss of much of the pathway in malarial parasites suggests selective pressures from host immune response against heavily glycosylated proteins have possibly played a role in its abbreviation or alteration (Gowda *et al.*, 1997).

The shared O-linked glycosylation pathway of apicomplexans contains two core enzymes (Fig. 6). First is the protein O-fucosyltransferase (e.g., PFI0445c) that transfers fucose to the hydroxyl group in target proteins. The second step is catalyzed by the fringe-type glycosyltransferase that serially elongates a carbohydrate chain on the fucose primer by adding  $\beta$ 1,3 NAG units. In *Cryptosporidium* alone, the fringe-type glycosyltransferase is fused to a C-terminal polysaccharide pyruvyl transferase (Andreishcheva *et al.*, 2004), suggesting the O-linked carbohydrate is further decorated by pyruvylation. The first apicomplexan enzyme is closely related to the animal fucosyltransferases (POFUT1/2) and the second enzyme to animal fringe proteins. In animals both enzymes are known to cooperate in producing oligosaccharide modifications specifically on serines or threonines found in EGF and TSPI domains (Luo *et al.*, 2006); hence, it is possible that apicomplexan versions similarly carry out O-linked glycosylation of the equivalent domains found in various parasite adhesion molecules. Apicomplexans also encode a sugar transporter related to the animal fucose-GDP transporter involved in fucose uptake for surface adhesion protein modification in animal cells (Lubke *et al.*, 2001; Luhn *et al.*, 2001). It is possible they use this transporter for obtaining fucose-GDP from their hosts. The second O-linked glycosylation pathway is found only in *Cryptosporidium* and *T. gondii* and is rudimentary compared to that of animals (Fig. 6). Presence of numerous surface proteins with mucin segments in both these taxa indicates this system is mainly geared toward modification of these segments on endogenous

FIG. 6 Protein glycosylation pathways in apicomplexa. Four distinct pathways are depicted: GPI anchor biosynthesis, N-linked glycosylation, and two types of O-linked glycosylation. Conserved sections of pathways are enclosed in colored boxes, and lineage-specific or speculative parts are shown outside boxes. Yeast homologs for proteins involved in GPI-anchor biosynthesis and N-linked glycosylation pathways are shown below the enzyme name. Intra-alveolate phyletic distributions of enzymes involved in N-linked glycosylation and GPI anchor biosynthesis are shown in black boxes next to enzymes. Genes in *Plasmodium* or *Cryptosporidium* are shown in blue boxes next to the enzyme names in the O-linked glycosylation pathways. The reconstructed oligosaccharide chain is represented using abbreviations for various sugars and functional groups. Sugar residues or functional groups added or modified during a particular step of the pathway are enclosed in white boxes. Speculative parts of the pathway are marked with a "?". Dol, dolichol; EtN-P, Phosphoethanolamine; Gal, Galactose; GalNAc, N-acetylgalactosamine; Glc, Glucose; GlcNAc, N-acetylglucosamine; Man, Mannose; Ino, Inositol. "X?" indicates the uncharacterized sugar added by the LPS-glycosyltransferases. (See also color insert.)

proteins (Stwora-Wojczyk *et al.*, 2004; Templeton *et al.*, 2004a). This pathway is initiated by the 5 to 6 polypeptide N-acetyl-galactosaminyltransferase proteins that belong to three distinct orthologous groups, each having an animal representative. These enzymes catalyze transfer of NGlc from UDP-N-acetylgalactosamine to the hydroxyl group in the protein. It is possible that in apicomplexans subsequent elongation of the carbohydrate chain by addition of  $\beta$  1,3-N-acetylglucosamine or  $\beta$  1,3 galactose is carried out by the fringe-like enzymes from the preceding pathway or by iterative action of the first family of enzymes. *Cryptosporidium* also contains two genes (cgd5\_3140, cgd5\_3150) for enzymes related to the bacterial LPS glycosyltransferases, which might allow further organism-specific modification of either O- or N-linked polysaccharides (Templeton *et al.*, 2004a) (Fig. 6).

Phyletic patterns of these O-linked pathways, clear-cut evolutionary affinities of core enzymes and the fucose transporter to animal versions, and presence of the target adhesion protein domains such as EGF and TSP1 shared with animals suggest that this pathway was acquired early in apicomplexan evolution through lateral transfer of corresponding genes from animal hosts. The pathway related to the modification of mucin segments is retained in *Cryptosporidium* and *T. gondii*, suggesting that it was originally selected for an important role in adhesion and survival in conditions specific to animal guts. Subsequently, with emergence of predominantly hemal parasitism in hematozoans, there appears to have been intense selection against O-glycosylated proteins by the vertebrate immune system. As a result, one (the mucin modification) or both (in piroplasms) pathways were eliminated in different hematozoan lineages. Thus, the O-glycosylation, especially the fucose-primed pathway, represents a probable example of acquisition of both target protein domains and multiple enzymatic domains through lateral gene transfer from animals.

## B. Surface Protein Maturation: Protein Processing

Apicomplexans retain much of the conserved eukaryotic machinery for processing of surface proteins through proteolytic cleavage, disulfide bond formation, and lipid modifications. This includes core catalytic and structural components of the signal peptidase complex, CAAX proteases, lipid transferases, diverse disulfide bond isomerases, and membrane-associated proteases such as rhomboids and presenilins. We mainly discuss apicomplexa-specific peculiarities in evolution of these components and their role in regulating the dynamics of localization and exposure of diverse cytoadherence and other host-interaction proteins. All apicomplexans have multiple systems for modifying proteins with lipids, and thereby allow them to be anchored the membrane. N-myristoyltransferases that modify N-terminal

glycines of proteins with myristoyl groups are conserved in all alveolates (Rajala *et al.*, 2000). The ancestral alveolate also appears to have had at least 2  $\alpha$ -subunits and 3  $\beta$ -subunits of the prenyltransferase (Zhang *et al.*, 1996), which potentially combine to form a farnesyltransferase that adds farnesyl groups to CAAX signals and two distinct geranylgeranyltransferases that modify CAAL or CC signal in proteins. The apicomplexan Rab11 protein with a CC signal, which might be critical for trafficking of proteins to rhoptries (Bradley *et al.*, 2005), is one potential conserved substrate for type II geranylgeranylation. Strangely, *T. gondii* does not seem to encode any homologs of the  $\alpha$ -subunit. In contrast the CAAX metallopeptidase that cleaves proteins for farnesylation appears to be missing in *Plasmodium*; however, *Theileria* has three different versions. It would be of tremendous interest to determine whether the *Plasmodium* protein has been displaced in evolution by another peptidase because these enzymes could potentially be exploited as drug targets. Protein palmitoylation system (Mitchell *et al.*, 2006) is extremely well-developed in all apicomplexans with 7 to 18 palmitoyltransferases of the DHHC family found in different taxa. One of these, typified by PfAnkDHHC from *Plasmodium*, is conserved in all apicomplexans and has a unique architecture with N-terminal ankyrin repeats. This enzyme localizes to the Golgi and is expressed in late schizont, just before merozoites escape (Seydel *et al.*, 2005). This suggests stage-specific palmitoylation of proteins might be conserved across apicomplexans and serve to target proteins to rhoptries and micronemes by means of lipid rafts. In particular, presence of conserved cysteines in N-termini of several *Plasmodium* 2<sup>TM</sup> proteins (see earlier) and the pan-apicomplexan family of lipid-anchored proteins typified by *Cryptosporidium* antigen CP15/60 (PF10\_0107) are excellent candidates for modification by palmitoylation.

Two major classes of membrane-embedded peptidases catalyzing intramembrane protein cleavage in eukaryotes are the presenilin and rhomboid proteases (Dowse *et al.*, 2005; Kopan *et al.*, 2004). Rhomboids are found in three to eight copies in different apicomplexans, and in both *T. gondii* and *Plasmodium* they participate in cleavage of diverse surface proteins required for cytoadherence (Baker *et al.*, 2006; Baum *et al.*, 2006; Dowse *et al.*, 2005; O'Donnell *et al.*, 2006). It is known that parasites shed different adhesion molecules repeatedly in particular stages of their invasion. This allows them to sequentially form initial transient contacts with the host cell, then reorient the apical end for invasion, form the moving junction with invagination of the host membrane, and finally complete internalization by dissolving the moving junction. In *Plasmodium*, two of eight rhomboid proteases (PfROM1 and PfROM4) have been shown to be capable of cleaving a diverse range of adhesins from all stages of the parasite's life cycle, including the TRAP protein (Baker *et al.*, 2006). Similarly, at least three of six *T. gondii* rhomboids have been implicated in cleavage of micronemal proteins, such as

TgMic2 (Dowse *et al.*, 2005). This suggests the action of rhomboid proteases might be a conserved feature critical for shedding adhesins extruded by the apical complex during completion of invasion. PfROM3 is expressed in gametocytes (Baker *et al.*, 2006), and it would be of interest to investigate its role in processing stage-specific surface proteins required for sexual union. A single presenilin is found in all apicomplexans except *Theileria* and might be required for as yet unexplored aspects of surface proteins dynamics.

Another group of enzymes required for surface protein maturation in eukaryotes are protein disulfide bond isomerases, and apicomplexans possess several paralogs of both versions belonging to the thioredoxin fold and the unrelated ERV1-like superfamily (Mahajan *et al.*, 2006). Investigation has shown that one thioredoxin fold protein prototyped by PfPDI-8 (MAL8P1.17), which was ancestrally present in apicomplexans, facilitates folding of the cysteine-rich adhesin EBA-175 in *P. falciparum* (Mahajan *et al.*, 2006). Given abundant cysteine-rich surface proteins in all apicomplexans, one might expect other PDIs to also be involved in proper folding of these proteins during secretion or membrane targeting. In this context, it is of considerable interest to note that PF13\_0272 typifies a predicted PDI with two membrane-spanning segments unique to apicomplexans and highly conserved in them. This protein could possibly mediate a specialized folding pathway for membrane-anchored, cysteine-rich proteins in apicomplexa.

### C. Surface Protein Deployment and Connection with Cytoskeleton

Studies indicate apicomplexan gliding movement and motoring of moving junctions during invasion might share cytoskeletal components. TSP1-domain adhesins such as TRAP, TgMIC2, and MTRAP (PF10\_0281), which are central to invasion, might be linked by aldolase to the glideosome complex consisting of F-actin, Myosin A, and myosin light chain (Baum *et al.*, 2006; Kappe *et al.*, 1999). Glideosome components also include (Gaskins *et al.*, 2004): (1) GAP50, a TM protein containing a divergent calcineurin-like phosphatase domain that might be catalytically inactive; (2) GAP45 that contains a C-terminal Zn-chelating domain similar to the knuckle-type Zn-clusters. Both proteins are conserved throughout apicomplexa and are unique ancestral innovations complementing the pan-apicomplexan TSP1-domain adhesins. Localization of these adhesins to the motor complex appears to depend on distinctive cytoplasmic tails, which are enriched in charged residues and bear an aromatic residue just prior to the C-terminus. GAP50 is homologous to another membrane-associated protein, IMC2 conserved in

*T. gondii* (two copies) and *Plasmodium*. IMC2 and articulins, which are conserved across alveolates, form the subpellicular protein-network (Mann *et al.*, 2001; Templeton *et al.*, 2004a). The two related membrane-associated calcineurin-like proteins might possibly link the two distinct cytoskeletal complexes with the pellicle. Thus, phosphopeptide-binding or phosphatase activity of these proteins might be an unexplored aspect of cytoskeletal dynamics. It will also be of interest to investigate whether other components of the ARP complex, such as p34 or p20, which are highly divergent in apicomplexa, are also part of the motor-adhesin linkage.

### D. Rhoptries, Micronemes, and Protein Extrusion During Invasion

Release of rhoptry and microneme contents, following apical reorientation of the parasite, is the key step in commencement of invasion. Rhoptry and microneme contents appear to be required for forming the host-parasite interface during invasion, as well as the PV upon completion of invasion. Primary rhoptry proteins themselves seem to be packed in a crystalline array in the rhoptry's bulb. Detection of specific proteins, such as CLAG (Kaneko *et al.*, 2001, 2005), localizing only to the neck of the rhoptry, suggests it forms a specialized duct comparable to unicellular glands in animals. Insights regarding the operation and function of rhoptries have emerged from large-scale proteomic studies on *T. gondii* and *P. falciparum* rhoptries (Bradley *et al.*, 2005; Sam-Yellowe *et al.*, 2004b). As a result, it appears that rhoptry components might be distinguished into several distinct functional categories:

1. *Structural and operational components of rhoptries* include proteins that actually contribute to the structure of organelle and its working. This category contains ancient eukaryotic proteins related to protein trafficking that appear to have been recruited for rhoptry-specific functions. The GTPase Rab11 is one such conserved component potentially critical for protein trafficking into rhoptries. A sodium-hydrogen exchanger typified by TgNHE2 appears to be critical for maintenance of osmolarity in rhoptries. Another such multi-TM protein (PFF0645c), which is also found in ciliates and animals, might form a critical structural component of the rhoptry membrane. Interestingly, an uncharacterized protein rhoptry membrane protein conserved in all apicomplexans contains 8 TM domains followed by four calcium-binding EF-hand domains at the C-terminus (PF14\_0607). This might point to involvement of calcium signaling in the operation of rhoptries. CLAG/Ron2 proteins of the crown group apicomplexa appear to be a key structural component of the neck.

2. *Processing components of rhoptries* appear to be needed to further process rhoptry proteins that initially pass through the classical secretory pathway by means of their signal peptides. Although a short YXXh (where h is a hydrophobic residue) motif has been implicated in localization of proteins to rhoptries in *T. gondii* (Hoppe *et al.*, 2000), there is no evidence for such a motif in other apicomplexans. However, all apicomplexans possess a group of conserved papain-like proteases that might be critical for processing rhoptry proteins. Likewise, the rhoptry localized insulinase-like metalloproteinase, subtilisin-2 (TgSUB2), DegP-like serine proteases, and their orthologs in other apicomplexans might be critical for processing of rhoptry-localized proteins.

3. Some surface proteins extruded by rhoptries are *conserved to differing degrees in apicomplexa* and include various catalytic and noncatalytic proteins that are potentially released into the host cell. Conserved enzymes include diverse phosphatases such as PPX1, Rop9-like HAD phosphatases, and the PP2C-like protein phosphatases, whereas noncatalytic proteins include sushi domain proteins which are either secreted or have a GPI anchor (e.g., TgRon1) and an ancient apicomplexan protein with eight conserved cysteines (e.g., cgd5\_170).

4. Analysis of both *T. gondii* and *Plasmodium* rhoptry proteins suggests there are *lineage-specific rhoptry surface proteins*, which might have conserved domains, like the kinase domain, or no known conserved domain. Other than several kinase domain proteins, *T. gondii* also extrudes eight lineage-specific secreted or membrane-anchored proteins with no homologs found even in other apicomplexans.

The last two categories of proteins are likely to be principal purveyors of host-parasite interactions. Kinase domain proteins appear to have acquired diverse roles—ROP2 seems to be inactive and is involved in recruiting host mitochondria to the PV, presumably to fuel the growing parasite (Sinai *et al.*, 2001). Some extruded kinases and phosphatases might be targeted to host-cell organelles (e.g., the nucleus) and play a critical role in altering the behavior of the host (Gilbert *et al.*, 2006). PPX1-like phosphatases localize to the PV and might be needed to supply phosphate to the parasite. Evidence from *Plasmodium* suggests the conserved sushi domain localizing to the rhoptry neck (TgRon1) in *T. gondii* might play a role in invasion and is subsequently shed by proteolytic cleavage (O'Keeffe *et al.*, 2005). The lineage-specifically expanded Py235 family of *P. yoelii* rhoptry proteins with coiled-coil regions and their homologs, *P. vivax*, RBP-2, and PfRBP-2Ha/b are required to target erythrocytes, and in *P. vivax*, they appear to be critical for preferentially targeting reticulocytes (Galinski *et al.*, 2000; Keen *et al.*, 1994; Rayner *et al.*, 2000).

Throughout apicomplexa, micronemes deliver adhesins, with EGF, vWA, and APPLE domains (e.g., TgMIC2, TgMIC6, TgMIC8, TRAP, AMA1, CTRP), invasion proteins (e.g., MAC-perforin proteins), and low-complexity proteins (e.g., PfSPECT1 and TgMIC12) (Baum *et al.*, 2006; Ishino *et al.*, 2004, 2005). Many microneme TM proteins possess a C-terminal tail with charged residues and a conserved aromatic residue close to the C-terminus and are targeted for cleavage by rhomboid proteases during adhesin-shedding upon completion of invasion (Opitz *et al.*, 2002). Different stages of the parasite might express different microneme contents; for example, *Plasmodium* sporozoite micronemes deliver PbSPECT(1), a secreted protein with extracellular coiled-coil segments similar to animal extracellular matrix protein laminin domain, and SPECT2, with a MAC-perforin domain for initial penetration of Kupffer cells (Ishino *et al.*, 2004). Interestingly, analysis of *T. gondii* rhoptries did not recover microneme-targeted proteins, such as adhesins (e.g., TgMIC12). Thus, it seems likely adhesins and other components are separately trafficked into micronemes and probably delivered to the neck of rhoptries, just prior to invasion. This is supported by the association of rhoptry neck protein PfRon4 with the AMA1 adhesin delivered by micronemes during *P. falciparum* invasion (Alexander *et al.*, 2006).

In evolutionary terms, very few rhoptry proteins are also found in ciliates and mainly include operational components such as Rab11, TgNHE2, or PFF0645c. Processing proteases of the papain-like fold and perhaps the insulinase-like peptidase also show comparable phyletic patterns (Table II). Few other conserved, uncharacterized membrane proteins, as well as a large armadillo-repeat protein (e.g., MAL13P1.308), and some cytoskeletal proteins which are shared by ciliates and apicomplexans might also be as-yet-identified components of rhoptries or alveolar sacs. The ancestral extrusion system of alveolates might have in large part been an adaptation of pre-existing cytoskeletal, protein-processing and secretory elements, with a few lineage-specific innovations. At least four proteins are suggested to be ancestral components of the apicomplexan rhoptry. These include a secreted enzyme (Rop9 HAD phosphatase), potential adhesion molecules (the sushi domain protein [O'Keeffe *et al.*, 2005; Templeton *et al.*, 2004a]), and a potential signaling molecule (the 8-TM protein with EF-hand domains). The apicomplexan crown group appears to have acquired DegP-like proteases and Clag/Ron2- and Ron4-like rhoptry neck proteins. Most other proteins appear to be lineage-specific adaptations. Thus, although rhoptries are morphologically conservative structures, their major protein contents appear to have undergone notable lineage-specific diversification in apicomplexa, probably providing the fine-tuning to suit each parasite's requirements. Likewise, microneme products also display lineage-specific diversity, but at least one adhesin each with TSP1 and EGF domains, and a MAC-perforin protein was potentially present in ancestral micronemes.

## E. Protein Export by Apicomplexans During Intracellular Life Cycle

### 1. The PV and the PEXEL/VTS System in Plasmodium

Phyletic distribution of the PV indicates it is an ancestral feature of apicomplexa, which probably emerged as a result of the ancestral invasion strategy used by the progenitor of the dinoflagellate-perkinsid-colpodellid-apicomplexan clade. The PV is derived from invagination of the host cell formed during invasion (Lingelbach and Joiner, 1998). Consistent with this, experiments in different apicomplexans have shown the PV is largely composed of host lipids, but initiation of its formation depends on discharge of rhoptry contents. It is also likely the PVM, while excluding host proteins, incorporates proteins initially discharged from rhoptries and those exported by the internalized parasite. Lack of PVs in piroplasms is a derived condition resulting from their escape from incipient invaginations formed during invasion (Lingelbach and Joiner, 1998; Ward *et al.*, 1993). The PV is believed to participate in protein export as well as nutrient concentration and uptake. The latter function has several facets such as providing diffusion pores for nutrients or recruitment of host mitochondria around the PV in the case of *T. gondii*. Both parasites residing within PVs or directly in the host cytoplasm are known to export proteins that enter different organelles of the host cell or localize to the host cell membrane. This is one of the most important areas of parasite physiology that is poorly understood. Several advances, especially in *Plasmodium*, have provided leads regarding protein export into the host, but at the same time have raised multiple unanswered puzzles. In the course of IDC, *Plasmodium* faces challenges very different from most other parasites. Mammalian erythrocytes, in particular, have limited metabolic activity, being more or less just hemoglobin packets. As a result, findings from *Plasmodium* IDC might not necessarily generalize to other stages, such as the hepatic cycle, or across other apicomplexans, which often infect nucleated cells with an active physiology. Newer ultrastructural studies have suggested the PV membrane itself is continuous with the membranous network termed the tubulovesicular network (TVN) that pervades the erythrocyte during the ring stage (Wickert *et al.*, 2003a). However, there is considerable debate among researchers as to whether structures known as Maurer's clefts (MC) are also directly connected to the TV network or whether the MCs bud off from the TVN (Spycher *et al.*, 2006; Wickert *et al.*, 2003a). Several studies seem to indicate that parasite-encoded trafficking proteins such as COPII, SAR1-like GTPases, Sec31, and Sec23 localize to MCs, suggesting they might play a major role in the transport of molecules to the host cell or the host membrane (Sam-Yellowe *et al.*, 2004a; Wickert *et al.*, 2003b). Details of the process by which membrane proteins are trafficked to

the PVM and beyond are also debated and not well understood (Lingelbach and Przyborski, 2006).

The main advance in understanding protein export was the discovery of a short peptide motif termed PEXEL (*Plasmodium* export element) or VTS (vacuolar translocation signal) required for targeting of both TM and soluble secreted proteins to the host cell or host membrane, respectively (Hiller *et al.*, 2004; Marti *et al.*, 2004). This motif is of the form [RK]X [LFI]XE, where X is any amino acid, and is likely to form a short-charged helical segment. It is often encoded by a separate exon and usually occurs at the N-terminus of the protein, downstream of the conventional signal peptide, and in proteins with other globular domains, it typically precedes the globular domain. Independent studies have confirmed the requirement of the PEXEL motif and shown its ability to drive trafficking of heterologous proteins, such as GFP, into the host cytoplasm (Hiller *et al.*, 2004; Marti *et al.*, 2004; Spielmann *et al.*, 2006). Yet, several issues question the wider significance of the PEXEL motif: (1) There are several proteins without PEXEL motifs (e.g., MAHRP1, REX1, and REX2) that are also exported into the host cell (Spielmann *et al.*, 2006; Spycher *et al.*, 2006). (2) The PEXEL motif is implicated in transport of soluble proteins, TM proteins, as well as the more mysterious PfEMP1 which lacks any obvious TM or signal peptide (Hiller *et al.*, 2004; Marti *et al.*, 2004; Spielmann *et al.*, 2006). Conventional knowledge of protein transport has indicated vesicle or membrane fusion mechanisms for export of TM proteins and use of generic or specific pores or translocators for soluble proteins (Spycher *et al.*, 2006). Hence, use of a common targeting motif by different types of proteins is puzzling. (3) The rise to prominence of the PEXEL motif appears to have happened relatively recently in species such as *P. falciparum*. Proliferation of the exon encoding the PEXEL motif region and combination with portions of genes encoding remainders of exported proteins, followed by lineage-specific expansion of several of these genes, has been the main cause for the explosion of PEXEL-containing proteins. (4) The PEXEL motif is not conserved in other apicomplexan genera, despite at least some of them being known to export proteins into hosts. These points suggest the PEXEL motif is an important adaptation in *Plasmodium*, but it is unlikely it is the ancestral apicomplexan host-targeting signal or the only signal for protein export. Nevertheless, it has considerable predictive value in delineating the complement of proteins exported by *Plasmodium* into the host cytoplasm, especially during the IDC (Hiller *et al.*, 2004; Marti *et al.*, 2004). One study has also proposed the phylogenetically distant oomycete intracellular parasite, *Phytophthora*, might have an export motif similar to the PEXEL (Bhattacharjee *et al.*, 2006). However, even if confirmed, given the previous points, it is likely to be a case of chance resemblance or convergence.



Many host-cell-targeted enzymes in *P. falciparum* may or may not contain canonical PEXEL motifs. These include the plasmepsin and SERA peptidases, the PfACS family of acyl CoA synthetases and lysophospholipases involved in lipid metabolism, and R45 family of kinases. Of these, plasmepsins and PfACSs appear to be involved in obtaining nutrients for the parasite by mobilizing fatty acyl CoA and amino acids from digestion of hemoglobin, respectively. A subset of R45 kinases appears to be overexpressed in response to conditions such as febrile temperature, suggesting they might modify target host proteins for phosphorylation in specific conditions to allow parasite survival (Oakley *et al.*, 2007). It would also be of interest to investigate whether host-targeted lipid metabolism enzymes are required for modifying internal membranes or forming the PV and TVN. The RESA family of DNAJ domain proteins is another major set of exported proteins with PEXEL motifs (Aravind *et al.*, 2003b). The majority of RESA-like genes show elevated expression under febrile temperatures (Oakley *et al.*, 2007), and disruption of the RESA gene results in decreased parasite survival under these conditions. RESA appears to interact with the spectrin cytoskeleton and possibly stabilizes it from disruption under high temperatures by recruiting components of heat shock machinery such as proteins of the HSP70 family (Silva *et al.*, 2005). RESA proteins contain an N-terminal PRESAN domain that occurs in over 70 distinct copies in *P. falciparum*, including numerous stand-alone subtelomerically encoded versions with N-terminal signal peptides and PEXEL motifs (Figs. 5, 7B, Table I). They are present in fewer copies in other species such as *P. vivax*. The PRESAN domain is distantly related to the  $\alpha$ -helical VGD but lacks the cysteines present in them (Oakley *et al.*, 2007), suggesting that in all cases it is likely to function within the host cytoplasm. There is also evidence from genome-scale expression analysis that most genes encoding PRESAN domain proteins are expressed during the IDC (Bozdech *et al.*, 2003; Daily *et al.*, 2005) (Fig. 7A). Given the precedence of RESA's interaction with spectrin (Silva *et al.*, 2005), it is possible that it is required for interacting and remodeling the host cytoskeleton, which might be especially significant for formation of the parasite-induced structures such as MCs and knobs. PRESAN domains, which were probably derived from the VGD, along with the spread of the PEXEL motif among subtelomerically encoded protein families appear to represent major new adaptations of the *Plasmodium* lineage.

A single ortholog of three paralogous *P. falciparum* proteins REX3/REX4/MAL3P8.16 is seen in other *Plasmodium* species. These proteins contain a conserved all  $\alpha$ -helical domain of about 150 residues and are likely to form aggregates in the host cytoplasm through interhelical interactions (Spielmann *et al.*, 2006). They represent a set of poorly characterized conserved proteins with PEXEL motifs present across *Plasmodium* and might be early adaptations of the genus involved in modifying the host cytoplasm.

They also indicate that in *P. yoelii* and *P. vivax* the PEXEL motif might assume more divergent forms like RXLX[AS]. It is possible that PEXEL motif-based protein export emerged as a parasitic adaptation to survive in the metabolically limited erythrocytes. This was possibly followed by an evolutionary arms race resulting from development of different host defense mechanisms, including possibly the febrile response. As a result, certain families of PEXEL-containing proteins appear to have undergone expansion as countermeasures in species such as *P. falciparum*. In contrast, apicomplexans such as *T. gondii*, which reside in host cells with more active metabolism than erythrocytes, possibly do not require as extensive an export system as *Plasmodium* during intracellular development.

Macroschizonts of *Theileria* undergoing nuclear division within leukocytes cause their proliferation and dedifferentiation and protect them against apoptosis. As they continue development into merozoites, proliferation of leukocytes slows down and ultimately stops. These unique aspects of *Theileria* biology have been attributed to secretion of proteins into the host cell by the parasite. Two related types of DNA-binding proteins, TashATs and SuATs, with AT-hook motifs that translocate to the host nucleus might play a major role in this process in *T. annulata* (Pain *et al.*, 2005; Shiels *et al.*, 2004). These proteins are encoded as a part of a large tandem cluster of chromosomally internal genes and might use their AT-hook proteins to alter host chromatin structure similar to the HMG-I/Y proteins (Aravind and Landsman, 1998). TashATs might induce transcription changes that favor proliferation, and SuAT1 might have a role in regulating host cytoskeletal gene expression. Interestingly, related *T. parva* TpHN genes do not encode canonical AT-hook proteins (Pain *et al.*, 2005). Some of them have motifs such as "PGRP" C-terminal to their FAINT domains that might bind DNA analogous to AT-hooks of the *T. annulata* proteins. Thus, proliferation of proteins with AT-hook motifs C-terminal to FAINT domains is a development that occurred only in the *T. annulata* lineage. Both the *Theileria* species and *Babesia* encode an array of related proteins with FAINT domains (Fig. 5) likely to be secreted into the host cytoplasm and act as mediators of physiological changes associated with intracellular parasitism by piroplasms. The possible Ig-like  $\beta$ -sandwich structure predicted for FAINT domain provides an excellent scaffold that might have diverged to bind a variety of target molecules.

Comparison of functions of various extruded or exported proteins point to certain analogous adaptations that appear to have evolved convergently. Ancestrally apicomplexans might have used extruded kinases as a means of altering host physiology through phosphorylation. Proliferation of exported or extruded kinases in both *Plasmodium* and *T. gondii* suggest that this theme was repeatedly expanded, probably broadening the range of targeted proteins. Within hematozoans, *Plasmodium* and *Theileria* have relied upon LSE

and diversification of a single novel domain (PRESAN or FAINT) as an interface for different aspects of host interaction. However, *Plasmodium* appears to be unique in deploying a panoply of proteins ranging from metabolic enzymes and proteases to host-cell stabilizing components such as RESA. Another salient point is that, as a rule, there is immense lineage-specific diversity in host-interaction proteins even in closely related species with similar hosts and life cycles. This is best exemplified by the numerous AT-hook proteins within FAINT domain expansion in *T. annulata* and RESA-like DNAJ domains expanded in *P. falciparum* (Fig. 5). These observations imply that proteins at the interface of host-parasite association are under intense selective pressures and are constantly modified via fresh innovations.

## V. Regulation of Surface Protein Gene Expression

Regulation of genes encoding surface proteins is an important aspect of parasitic life cycles with each stage requiring a different set of proteins to be secreted or expressed on the membrane. Additionally, antigenic variation in apicomplexans appears to be mainly due to specific regulation of gene expression. Though these processes were poorly understood, data suggest that like other eukaryotes, apicomplexans show evidence for chromatin level, transcriptional, and posttranscriptional RNA-level regulation. However, there is evidence apicomplexans differ considerably in their repertoire of regulatory factors, especially transcription factors and RNA-level regulators (Ralph *et al.*, 2005).

### A. Chromatin-Level Controls, Telomere Effect, and Transcription Regulation

Apicomplexans are seen to have a robust complement of chromatin proteins, including certain lineage-specific versions, suggesting they effectively deploy chromatin remodeling and epigenetic mechanisms to regulate gene expression. Studies on antigenic variation of *P. falciparum* have yielded considerable details regarding chromatin level regulation (Ralph *et al.*, 2005). At a given time parasites overexpress one var gene while silencing most others. Promoters of both active and silenced var genes can drive expression of plasmid-borne reporters, suggesting they are all competent to drive transcription. This pointed to epigenetic factors playing a major role in regulation var gene expression. The highly conserved intron in var genes can silence expression of var promoters, and this intron itself contains a promoter which can generate "sterile" transcripts that do not yield proteins. Establishment of silencing by activity of the intron requires passage through S-phase. Taken

together, these observations implied that a RNA transcribed from the intron promoter could be involved in establishing repressive chromatin structure that shuts off a var gene (Deutsch *et al.*, 2001; Frank *et al.*, 2006). Other studies have shown subtelomeric location of var genes also plays an important role (Duraisingh *et al.*, 2005; Figueiredo *et al.*, 2002; Ralph *et al.*, 2005). First, it provides condensing chromatin (heterochromatin) that propagates inward into subtelomeric regions from telomeres. Second, rep20 DNA repeats found in *P. falciparum* chromosomes are often close to var promoters and allow stable maintenance of transcription state by shielding it from surrounding influences. Also, PfSir2, a histone deacetylase from telomeres, deacetylates histones associated with var promoters and thereby favors their repression (Duraisingh *et al.*, 2005). Furthermore, it was shown that silenced var genes cluster together in condensed heterochromatin in the nuclear periphery, whereas the single active var gene leaves such a cluster and probably localizes close to the nuclear pore, where it might be in a euchromatic state. Here the var promoter is hyperacetylated and drives active transcription. Shutoff is believed to occur when the competing weaker intron promoter stochastically generates a noncoding RNA that facilitates assembly of heterochromatin. Possible involvement of noncoding RNA points to close coupling between posttranscriptional silencing with chromatin-level silencing (Ralph *et al.*, 2005).

In all apicomplexans, most large lineage-specifically expanded families are encoded by subtelomeric gene arrays (Fig. 7B, Table I). In *P. falciparum*, there are var genes, various families of PEXEL-containing proteins, and other membrane proteins such as ETRAMPS. In *P. yoelii* and *P. vivax* there are respectively PyST-B-2 TM and vir gene families. In *Theileria*, the FAINT domain, ABC transporter, and TSRP families are predominantly subtelomeric, whereas in subtelomeric regions of *T. gondii* genes for SAG1/SRS proteins and a group of mucins are found (Table I). In *Cryptosporidium*, there are several smaller gene families coding such proteins as MEDLE, SKSR, FGLN, and *Cryptosporidium*-specific 11 TM in subtelomeric regions. Such expanded subtelomeric families are also encountered in both free-living eukaryotes (e.g., *S. cerevisiae*) and other distantly related eukaryotic parasites (e.g., *Giardia*) (Mortimer *et al.*, 1992; Upcroft *et al.*, 1997). This suggests that the primary role of subtelomeric location of expansions is in regulating transcription of genes, which is likely to be a general phenomenon in eukaryotes. However, a unique feature observed in apicomplexa was the much-greater-than-average enrichment of TM or secreted proteins in subtelomeric regions (Fig. 7B). This suggests subtelomeric location is a major aspect in regulation of genes for surface proteins. As with *P. falciparum* var genes, it is possible in some of these cases the subtelomeric gene clusters are used for exclusive expression of one or a few members of the family. However, studies on gene expression in IDC development and febrile conditions in

*P. falciparum* and during asexual infection in *C. parvum* indicate that some of these subtelomeric surface protein gene clusters might actually show similar coexpression patterns in similar stages or conditions (Abrahamsen *et al.*, 2004; Bozdech *et al.*, 2003; Hall *et al.*, 2005; Oakley *et al.*, 2007). Thus, subtelomeric regions might actually have a more complex chromatin structure, with certain subsections displaying condition-specific extended zones of open chromatin that allow concerted transcription of whole groups of genes. Subtelomeric location also provides an added advantage in evolutionary terms—clustering of telomeres within the nucleus could accentuate gene conversion and recombination between paralogous members of multigene families from heterologous chromosomes, thereby resulting in increased adhesive and antigenic diversity (Freitas-Junior *et al.*, 2000).

In the past, all studies indicated an astonishing paucity of transcription factors compared to eukaryotes with similar number of genes (Aravind *et al.*, 2003b). However, our investigations suggest apicomplexans do possess a unique expanded group of transcription factors that contain DNA-binding domains of the AP2 family (ApiAP2 proteins), which is also found in plant transcription factors (Balaji *et al.*, 2005). Analysis of expression patterns of these proteins in *Plasmodium* suggest these are likely to be the dominant transcription factors that set up the transcriptional program in the course of parasite development. Expression analysis in *P. falciparum* points to a cascading program of gene expression, starting with genes involved in housekeeping and generalized and culminating in lineage-specific genes including those involved in host-response and adhesion (Bozdech *et al.*, 2003; Daily *et al.*, 2005; Hall *et al.*, 2005) (Fig. 7A). This observation coupled with expression patterns of ApiAP2 proteins hints that specific guilds of these transcription factors might regulate specific sets of membrane proteins associated with each developmental stage (Balaji *et al.*, 2005). *T. gondii* has a particularly well-developed complement (around 50 paralogs) of ApiAP2 transcription factors, suggesting it might have elaborately developed transcriptional regulation. Future investigation of roles of apicomplexan transcription factors is likely to be an important factor in understanding regulation of gene expression.

#### B. Posttranscriptional or RNA-Based Regulation

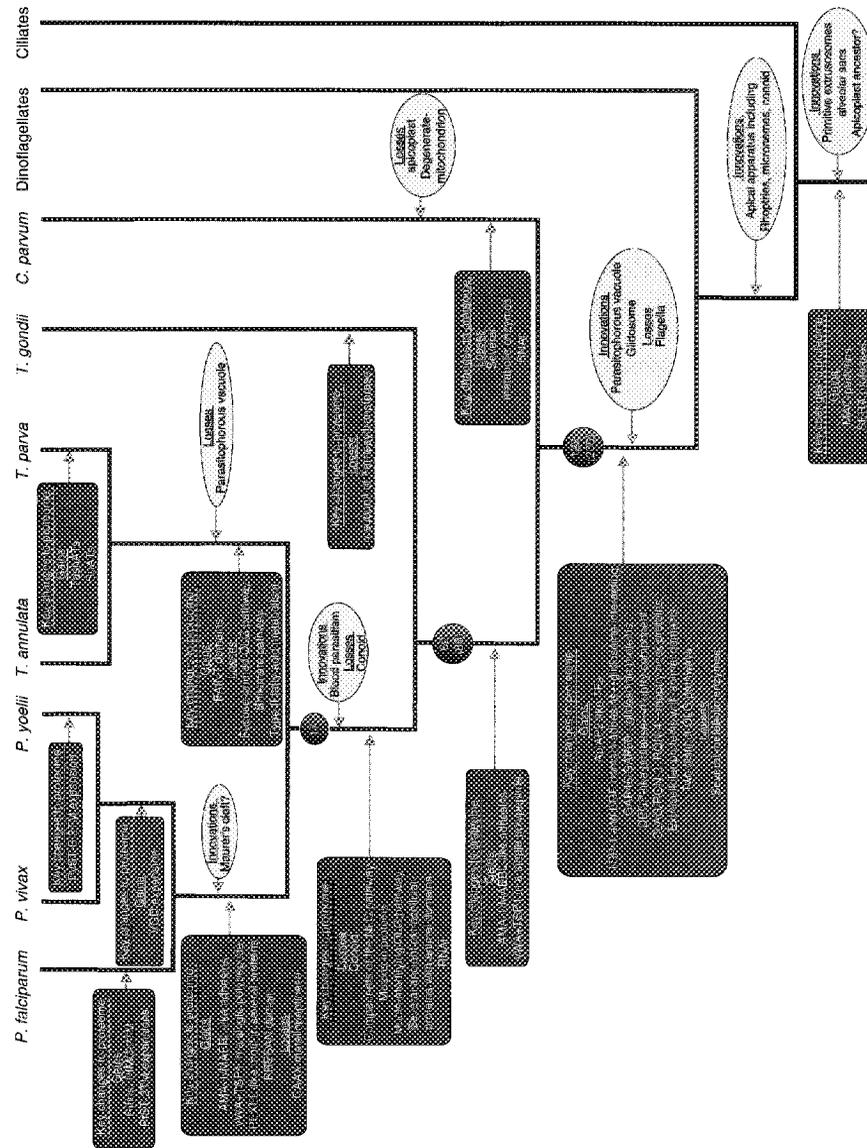
The classical RNAi pathway, including key proteins such as the Dicer helicase-nuclease involved in maturation of miRNAs/siRNAs, the RNA-dependent RNA polymerase involved in amplification and propagation, and argonaute proteins involved in degradation of the targeted mRNA, was ancestrally present in eukaryotes. Yet this system appears to have been

lost in all apicomplexans studied to date except *T. gondii* (Aravind *et al.*, 2003b; Ullu *et al.*, 2004). Nevertheless, involvement of possible noncoding transcripts in var gene regulation suggests other RNA-based post-transcriptional regulatory mechanisms comparable to those found in other eukaryotes might be in action. For example, apicomplexans encode orthologs of the *S. pombe* Mei2p that bind a noncoding meiRNA and regulate cell cycle progression. Apicomplexans also uniquely share with the plant lineage FCA-type RNA-binding proteins that have been implicated in mRNA processing in flowering (Razem *et al.*, 2006). Furthermore, *Plasmodium* and *Cryptosporidium* contain expansions of two distinct but related types of proteins, also related to certain plant proteins, with the RNA-binding CCCH domain (PFE1245w and cgd6\_4910, respectively). These proteins could mediate as yet unexplored posttranscriptional regulatory mechanisms. Thus, there are several hints that novel RNA-level regulation might emerge as a major factor in expression of adhesion and proteins complement of apicomplexan genomes.

## VI. General Evolutionary Considerations and Conclusions

### A. Evolution of Parasitism and Comparisons with Other Parasites

Despite technical difficulties in studying apicomplexans, our understanding of this lineage has immensely increased in lieu of information from comparative genomics. At present, apicomplexa represents one of the best models to characterize evolutionary trajectories of surface proteins of an exclusively parasitic lineage of organisms. Evolution of surface protein complements is very dynamic at practically every organizational level. At the highest level, it is characterized by massive lineage-specific innovations and expansions of adhesins and other surface proteins across the entire apicomplexan phylogeny. These expanded gene families often entirely alter the “genomic landscape,” especially of subtelomeric gene-arrays (Fig. 7). At the middle level, we find extensive architectural innovation through domain shuffling, novel deployments of enzymatic domains to target host substrates, and drastic functional shifts in various protein families (e.g., DBL domains). At the microscopic level, we find evidence for constant adaptation through subtle sequence variations. In each category of surface proteins (Section III), several parasite proteins tend to show amino acid polymorphisms in population, suggesting that new variants are constantly selected to evade any host counteradaptation (Mayer *et al.*, 2004).



We are also now in a position to reasonably reconstruct the temporal picture of apicomplexan innovations (Fig. 8). Although apicomplexan genomes are highly streamlined and their parasitic adaptations are extremely ancient, some distinct evolutionary steps and preadaptations can be clearly discerned. First, the ancestral alveolate had robust signaling systems including several calcium-dependent kinases, ion channels, and adenylyl cyclases, which appear to have been retained in part as the conserved core of apicomplexan signaling systems. Apicomplexans inherited several protein domains from the ancestral apicoplast endosymbiont even though it appears to have vastly degenerated in many extant lineages (Aravind *et al.*, 2003b; Foth *et al.*, 2003; Templeton *et al.*, 2004a). Apicoplast contributions were mainly incorporated in regulatory and metabolic systems, but a small subset of these domains were possibly recruited for more direct pathogenic roles. The ancestral unique core of the alveolate extrusion system was probably relatively simple as implied by the relatively small number of genes with potential involvement in this system. It appears to have developed in different alveolates to greater or lesser degrees, with the apical complex emerging well before the origin of apicomplexans *sensu strictu*. The ancestral extrusion apparatus of apicomplexans was already extruding a distinct set of enzymes with potential to modify host molecules. Subsequently, a diverse range of proteins with catalytic, adhesive, and structural properties came to be extruded in different apicomplexan lineages. This is a unique feature of apicomplexan parasites that distinguishes them from other well-studied eukaryotic or bacterial parasites. Hematozoans, in the least, also developed a system of targeting proteins to the host cell during intracellular development. This feature, although possibly having parallels in other intracellular parasites like oomycetes, has undergone extraordinary development in these apicomplexans and is thus far unprecedented in the parasite world.

An exciting feature of apicomplexan parasites is that early in their evolution they acquired several conserved adhesion protein domains and glycosylation pathways which modify these domains from animal hosts through lateral gene transfer. These were combined with similar adhesion domains acquired from ultimately bacterial sources or ancient alveolate inheritances to generate an array of novel adhesion molecules (Fig. 4). Some of these ancient pan-apicomplexan adhesion proteins, such as those with LCCL domains, were also recruited for parasite sexual development and maturation

FIG. 8 Reconstructed evolutionary history of apicomplexa showing some major transitions. The major morphological or operational transitions are enclosed in yellow ovals, and key changes to the proteome are enclosed in blue boxes. Numbers on nodes indicate total number of surface proteins (excluding plastid and mitochondrial membrane or imported proteins) present in all members of the species unified at a particular node. The number of surface proteins reconstructed to have been uniquely present or acquired at a node is indicated in brackets. NLG, N-linked glycosylation; OLG, O-linked glycosylation. (See also color insert.)

(Claudianos *et al.*, 2002; Deirieu *et al.*, 2002; Pradel *et al.*, 2004). Presence of conserved adhesins with TSP1 domains and a glideosome with unique GAP45/50 components suggests some “animal-like” adhesins were combined with the gliding motor for invasion in the ancestral apicomplexan. Large-scale acquisition of components of pathogenesis-related molecules through lateral transfer has not been confirmed for other eukaryotic parasites, though there is some evidence for comparable phenomena, in bacterial parasites of animals and plants. However, even in the latter instances, there is little evidence for full-scale recruitment of adhesion domains of host provenance (Koonin *et al.*, 2001). Several large eukaryotic DNA viruses have a wide range of protein domains acquired through lateral transfer from hosts. These include a few surface proteins related to adhesion and several anti-apoptotic proteins (Iyer *et al.*, 2006). Apicomplexans differ from these viruses in largely lacking antiapoptotic protein domains acquired through lateral transfer from the host. Secondly, unlike these viruses, apicomplexans display extensive shuffling of the adhesion protein domains, which they laterally acquired early in evolution. Like several bacterial and eukaryotic pathogens, apicomplexans display LSEs of distinct antigenically variant adhesins and virulence factors. In most cases, these variant surface proteins contain  $\alpha$ -helical- or cystine-supported cores, suggesting this is a general evolutionary trend among parasites (Aravind *et al.*, 2003b).

Several bacterial intracellular pathogens and comparable eukaryotic parasites such as microsporidians show massive genomic degeneration, with gene loss across most functional categories (Katinka *et al.*, 2001; Sallstrom *et al.*, 2005). Apicomplexans also show gene reduction, particularly genes coding metabolic enzymes, but their overall genome degeneration is not comparable to that observed in the aforementioned parasites. They retain reasonably robust regulatory systems, which are closer in complexity to free-living forms with similar gene numbers. Thus, apicomplexans, by virtue of their complex life cycles and parasitic adaptations are probably constrained at an intermediate level of genome degeneration in the parasite world. Comparative genomics strengthens the molecular phylogenetic view of intra-apicomplexa relationships and clarifies how gene loss and innovation occurred in different apicomplexans. Within the crown group, there appears to have been innovation of several adhesion-related features, such as MAEBL-like and AMA1-like proteins with divergent APPLE domains and proteins combining vWA and TSP1 domains in a single polypeptide. Analysis of conserved genes across apicomplexa indicates there has been notable gene loss in *Theileria* relative to other lineages.

Extensive sharing of unique gene sets by hematozoans indicates an extended shared history. This suggests that blood parasitism emerged on a single occasion in the ancestral hematozoan as it evolved from a gut parasite of a hemophagous invertebrate by acquiring a vertebrate intermediate host to

increase its chances of transmission. Several proteins, which are used to invade both arthropod and vertebrate cells, contain the same set of “animal-like” adhesion domains (Table II). Hence, acquisition of vertebrate intermediate hosts might have depended on reusing pre-existing adhesins. Subsequently, the diversification within hematozoa appears to have occurred via a more complex pattern of host switching. Convergent evolution of hemophagy in different invertebrates, such as leeches, ticks, and mosquitoes, appears to have allowed hematozoans to gain new definitive hosts, whereas broadening of vertebrate targets by hemophagous invertebrates might have transmitted the parasite to new intermediate hosts. For instance, studies on *Plasmodium* and related genera indicate that these parasites might have switched between avian and reptilian hosts on more than one occasion, and at least once from reptiles to mammals (Yotoko *et al.*, 2006). More recently, host switching among Old World monkeys, humans, and New World monkeys might have played a critical role in the spread of *P. vivax* (Mu *et al.*, 2005). Another interesting aspect revealed by comparative genomics is a group of molecular adaptations shared by *Cryptosporidium* and *T. gondii*, despite them being paraphyletic branches of the tree. These include a suite of extracellular proteins with animal-like domains, OWPs, mucins, and the O-linked glycosylation pathway to modify them. Their presence in the two successively basal branches of apicomplexa, together with their life cycle similarities, implies that they were part of the complex of adaptations for animal gut parasitism in the ancestral apicomplexan.

## B. General Conclusions

The biggest revelation from apicomplexan comparative genomics is the degree of lineage-specific diversification of adhesion and host-interaction molecules, even between closely related species. Conversely, it has also shown that there are several conserved elements of host-parasite interaction that appear to be present throughout animal-infecting apicomplexa. More generally, comparative genomics also provides a platform to fill lacunae in key areas of apicomplexan pathogenesis through future investigations. Thus, we are now on a much firmer basis to decide between investment in pan-apicomplexan models and taxon-specific investigations. Some examples of major problems that might be addressed on a more general basis across apicomplexans are: (1) connections between adhesion and deployment of parasite cytoskeletal motors early in invasion; (2) processing of precursor polypeptides in rhoptries; (3) mechanism of extrusion by rhoptries and micronemes; (4) role of conserved extruded rhoptry and microneme enzymes and adhesion molecules in interaction with the host; (5) involvement of conserved signaling systems such as potassium channels and adenylyl cyclases; and

(6) implications of folding, covalent modifications, proteolytic processing, and shedding of adhesins. For these studies investigators could resort to more tractable apicomplexan systems such as *T. gondii*. In contrast, major lineage-specific expansions that mediate key aspects of pathogenesis necessarily require case-by-case taxon-specific investigations. Furthermore, given their redundancy, conventional knockout-based approaches might not provide sufficient functional information, necessitating more elaborate approaches via high-throughput methods. Each parasite also has several lineage-specific secreted proteins with no detectable relationship to other proteins (e.g., those highlighted in analysis of *T. gondii* rhoptry components [Bradley *et al.*, 2005]), which need new directed approaches to elucidate their functions. However, caution is always warranted while using analogies rather than homologies when comparing features across parasites. Because high-throughput methods are becoming increasingly standard and affordable, an explosion of new data addressing these problems is likely to occur. One considerably neglected area of investigation is establishment of activities of uncharacterized secreted enzymes. Here recently developed methods of biochemical genomics might provide new leads (Phizicky *et al.*, 2003).

Finally, it should be noted that the new insights could have enormous implication for development of vaccines and therapeutics against apicomplexan parasites, which have proven to be considerably refractory to various treatments (Miller *et al.*, 1998, 2002a). Much effort has been expended on development of vaccines against *P. falciparum*, and surface proteins expressed in every stage of its life cycle have been targeted. Adhesion proteins with conserved domains of animal, bacterial, and ancient eukaryotic origin are obvious candidates because some of them possess indispensable functions and they do not show much antigenic variation (Miller *et al.*, 1998, 2002). Genomics has played a central role in increasing the sheer number of potential targets, but the "meta-information" critical for prioritizing candidates is only now beginning to build up. For instance, expression patterns during development, posttranslational maturation of molecules through covalent modifications and proteolytic cleavage, and positional distribution on the cell surface (e.g., all over the cell or only in apices) are critical constraints in potential success of a vaccine. It is hoped that future improvements in our understanding of the biochemistry and molecular biology of adhesins translate into improvements for vaccine development and diagnostics. Discovery of several secreted enzymes that might be critical for parasite invasion and survival within cells provides new targets for a different approach to anti-apicomplexan therapies—namely drug development. Given the molecular quirks of many apicomplexan enzymes as well as adaptors such as the DNAJ domain in RESA, it is hoped specific small molecule inhibitors might be useful in functionally disrupting these proteins (Bentley, 2006; Winstanley *et al.*, 2006). Major advances in host and vector genomics (ape,

mouse, and mosquito genome projects) have considerably aided in understanding host immune responses to parasites. Several ongoing genome-scale studies promise to deliver new results in this direction. These taken with advances in population genetics of hosts and parasites are hoped to contribute a future layer of information regarding parasite biology and development of anti-apicomplexan therapies.

It may be objectively said that we are in the midst of a revolution in understanding the complexities of apicomplexa—its implications for fundamental cell biology as well as therapeutics are simply staggering.

### Acknowledgments

The authors gratefully acknowledge the support from the Intramural program of the National Library of Medicine, National Institutes of Health. We apologize for not being able to cite numerous primary works due to the sheer enormity of the literature on the topic and the need to keep the article within reasonable limits. Supplementary material in the form of comprehensive collection of surface proteins and alignments is available at <ftp://ftp.ncbi.nih.gov/pub/aravind>.

### References

- Abrahamsen, M. S., Templeton, T. J., Enomoto, S., Abrahante, J. E., Zhu, G., Lancto, C. A., Deng, M., Liu, C., Widmer, G., Tzipori, S., Buck, G. A., Xu, P., *et al.* (2004). Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science* **304**, 441–445.
- Alexander, D. L., Arastu-Kapur, S., Dubremetz, J. F., and Boothroyd, J. C. (2006). *Plasmodium falciparum* AMA1 binds a rhoptry neck protein homologous to TgRON4, a component of the moving junction in *Toxoplasma gondii*. *Eukaryot. Cell* **5**, 1169–1173.
- Andreishcheva, E. N., Kunkel, J. P., Gemmill, T. R., and Trimble, R. B. (2004). Five genes involved in biosynthesis of the pyruvylated Galbeta1,3-epitope in *Schizosaccharomyces pombe* N-linked glycans. *J. Biol. Chem.* **279**, 35644–35655.
- Aravind, L., Anantharaman, V., and Iyer, L. M. (2003a). Evolutionary connections between bacterial and eukaryotic signaling systems: A genomic perspective. *Curr. Opin. Microbiol.* **6**, 490–497.
- Aravind, L., Iyer, L. M., Wellem, T. E., and Miller, L. H. (2003b). *Plasmodium* biology: Genomic gleanings. *Cell* **115**, 771–785.
- Aravind, L., and Landsman, D. (1998). AT-hook motifs identified in a wide variety of DNA-binding proteins. *Nucleic Acids Res.* **26**, 4413–4421.
- Bai, T., Becker, M., Gupta, A., Strike, P., Murphy, V. J., Anders, R. F., and Batchelor, A. H. (2005). Structure of AMA1 from *Plasmodium falciparum* reveals a clustering of polymorphisms that surround a conserved hydrophobic pocket. *Proc. Natl. Acad. Sci. USA* **102**, 12736–12741.
- Baker, R. P., Wijetilaka, R., and Urban, S. (2006). Two plasmodium rhomboid proteases preferentially cleave different adhesins implicated in all invasive stages of Malaria. *PLoS Pathog.* **2**, e113.
- Balaji, S., Babu, M. M., Iyer, L. M., and Aravind, L. (2005). Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. *Nucleic Acids Res.* **33**, 3994–4006.

- Baum, J., Richard, D., Healer, J., Rug, M., Krnjanski, Z., Gilberger, T. W., Green, J. L., Holder, A. A., and Cowman, A. F. (2006). A conserved molecular motor drives cell invasion and gliding motility across malaria life cycle stages and other apicomplexan parasites. *J. Biol. Chem.* **281**, 5197–5208.
- Bentley, G. A. (2006). Functional and immunological insights from the three-dimensional structures of Plasmodium surface proteins. *Curr. Opin. Microbiol.* **9**, 395–400.
- Bhattacharjee, S., Hiller, N. L., Liolios, K., Win, J., Kanneganti, T. D., Young, C., Kamoun, S., and Haldar, K. (2006). The malarial host-targeting signal is conserved in the Irish potato famine pathogen. *PLoS Pathog.* **2**, e50.
- Bhattacharya, D., Yoon, H. S., and Hackett, J. D. (2004). Photosynthetic eukaryotes unite: Endosymbiosis connects the dots. *Bioessays* **26**, 50–60.
- Bishop, R., Shah, T., Pelle, R., Hoyle, D., Pearson, T., Haines, L., Brass, A., Hulme, H., Graham, S. P., Taracha, E. L., Kanga, S., Lu, C., et al. (2005). Analysis of the transcriptome of the protozoan *Theileria parva* using MPSS reveals that the majority of genes are transcriptionally active in the schizont stage. *Nucleic Acids Res.* **33**, 5503–5511.
- Blackman, M. J., Ling, I. T., Nicholls, S. C., and Holder, A. A. (1991). Proteolytic processing of the *Plasmodium falciparum* merozoite surface protein-1 produces a membrane-bound fragment containing two epidermal growth factor-like domains. *Mol. Biochem. Parasitol.* **49**, 29–33.
- Bozdech, Z., Liinas, M., Pulliam, B. L., Wong, E. D., Zhu, J., and DeRisi, J. L. (2003). The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol.* **1**, E5.
- Bradley, P. J., Ward, C., Cheng, S. J., Alexander, D. L., Collier, S., Coombs, G. H., Dunn, J. D., Ferguson, D. J., Sanderson, S. J., Wastling, J. M., and Boothroyd, J. C. (2005). Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in *Toxoplasma gondii*. *J. Biol. Chem.* **280**, 34245–34258.
- Carlton, J. M., Angiuoli, S. V., Suh, B. B., Kooij, T. W., Perte, M., Silva, J. C., Ermolaeva, M. D., Allen, J. E., Selengut, J. D., Koo, H. L., Peterson, J. D., Pop, M., et al. (2002). Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* **419**, 512–519.
- Choe, H., Moore, M. J., Owens, C. M., Wright, P. L., Vasilieva, N., Li, W., Singh, A. P., Shakri, R., Chitnis, C. E., and Farzan, M. (2005). Sulphated tyrosines mediate association of chemokines and *Plasmodium vivax* Duffy binding protein with the Duffy antigen/receptor for chemokines (DARC). *Mol. Microbiol.* **55**, 1413–1422.
- Claudianos, C., Dessens, J. T., Trueman, H. E., Arai, M., Mendoza, J., Butcher, G. A., Crompton, T., and Sinden, R. E. (2002). A malaria scavenger receptor-like protein essential for parasite development. *Mol. Microbiol.* **45**, 1473–1484.
- Daily, J. P., Le Roch, K. G., Sarr, O., Ndiaye, D., Lukens, A., Zhou, Y., Ndir, O., Mboup, S., Sultan, A., Winzler, E. A., and Wirth, D. F. (2005). *In vivo* transcriptome of *Plasmodium falciparum* reveals overexpression of transcripts that encode surface proteins. *J. Infect. Dis.* **191**, 1196–1203.
- Deitsch, K. W., Calderwood, M. S., and Wellem, T. E. (2001). Malaria. Cooperative silencing elements in var genes. *Nature* **412**, 875–876.
- del Portillo, H. A., Fernandez-Becerra, C., Bowman, S., Oliver, K., Preuss, M., Sanchez, C. P., Schneider, N. K., Villalobos, J. M., Rajandream, M. A., Harris, D., Pereira da Silva, L. H., et al. (2001). A superfamily of variant genes encoded in the subtelomeric region of *Plasmodium vivax*. *Nature* **410**, 839–842.
- Dellova, T., Romer, J. T., Curran, T., and Rubin, L. L. (2006). The hedgehog pathway and neurological disorders. *Annu. Rev. Neurosci.* **29**, 539–556.
- Delrieu, I., Waller, C. C., Mota, M. M., Grainger, M., Langhorne, J., and Holder, A. A. (2002). PSLAP, a protein with multiple adhesive motifs, is expressed in *Plasmodium falciparum* gametocytes. *Mol. Biochem. Parasitol.* **121**, 11–20.

- Dobbelaere, D., and McKeever, D. (2002). "Theileria." Springer, Cambridge.
- Dowse, T. J., Pascali, J. C., Brown, K. D., and Soldati, D. (2005). Apicomplexan rhomboids have a potential role in microneme protein cleavage during host cell invasion. *Int. J. Parasitol.* **35**, 747–756.
- Dowse, T. J., and Soldati, D. (2005). Rhomboid-like proteins in Apicomplexa: Phylogeny and nomenclature. *Trends Parasitol.* **21**, 254–258.
- Dubey, J. P., and Beattie, C. P. (1988). "Toxoplasmosis of Animals and Man." CRC Press, Boca Raton, FL.
- Duraisingh, M. T., Voss, T. S., Marty, A. J., Duffy, M. F., Good, R. T., Thompson, J. K., Freitas-Junior, L. H., Scherf, A., Crabb, B. S., and Cowman, A. F. (2005). Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell* **121**, 13–24.
- Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J. R., Badger, J. H., Ren, Q., Amedeo, P., Jones, K. M., Tallon, L. J., Decher, A. L., et al. (2006). Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote. *PLoS Biol.* **4**, e286.
- El Hajj, H., Demey, E., Poncet, J., Lebrun, M., Wu, B., Galeotti, N., Fourmaux, M. N., Mercereau-Pujalon, O., Vial, H., Labesse, G., and Dubremetz, J. F. (2006). The ROP2 family of *Toxoplasma gondii* rhoptry proteins: Proteomic and genomic characterization and molecular modeling. *Proteomics* **6**, 5773–5784.
- Fayer, R. (1997). "Cryptosporidium and Cryptosporidiosis." CRC Press, Boca Raton, FL.
- Figureido, L. M., Freitas-Junior, L. H., Bottius, E., Olivo-Marin, J. C., and Scherf, A. (2002). A central role for *Plasmodium falciparum* subtelomeric regions in spatial positioning and telomere length regulation. *EMBO J.* **21**, 815–824.
- Fiorens, L., Liu, X., Wang, Y., Yang, S., Schwartz, O., Peglar, M., Carucci, D. J., Yates, J. R., 3rd, and Wub, Y. (2004). Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. *Mol. Biochem. Parasitol.* **135**, 1–11.
- Fiorens, L., Washburn, M. P., Raine, J. D., Anthony, R. M., Grainger, M., Haynes, J. D., Moch, J. K., Muster, N., Sacci, J. B., Tabb, D. L., Witney, A. A., Wolters, D., et al. (2002). A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419**, 520–526.
- Foth, B. J., and McFadden, G. I. (2003). The apicoplast: A plastid in *Plasmodium falciparum* and other Apicomplexan parasites. *Int. Rev. Cytol.* **224**, 57–110.
- Frank, M., Dzikowski, R., Costantini, D., Amulic, B., Berdugo, E., and Deitsch, K. (2006). Strict pairing of var promoters and introns is required for var gene silencing in the malaria parasite *Plasmodium falciparum*. *J. Biol. Chem.* **281**, 9942–9952.
- Freitas-Junior, L. H., Bottius, E., Pirit, L. A., Deitsch, K. W., Scheidig, C., Guinet, F., Nehrbass, U., Wellem, T. E., and Scherf, A. (2000). Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* **407**, 1018–1022.
- Frevert, U., Engelmann, S., Zougbede, S., Stange, J., Ng, B., Matuschewski, K., Liebes, L., and Yee, H. (2005). Intravital observation of *Plasmodium berghei* sporozoite infection of the liver. *PLoS Biol.* **3**, e192.
- Galinski, M. R., Xu, M., and Barnwell, J. W. (2000). *Plasmodium vivax* reticulocyte binding protein-2 (PvRBP-2) shares structural features with PvRBP-1 and the *Plasmodium yoelii* 235 kDa rhoptry protein family. *Mol. Biochem. Parasitol.* **108**, 257–262.
- Gardner, M. J., Bishop, R., Shah, T., de Villiers, E. P., Carlton, J. M., Hall, N., Ren, Q., Paulsen, I. T., Pain, A., Berriman, M., Wilson, R. J., Sato, S., et al. (2005). Genome sequence of *Theileria parva*, a bovine pathogen that transforms lymphocytes. *Science* **309**, 134–137.
- Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., et al. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511.

- Gardner, M. J., Tettelin, H., Carucci, D. J., Cummings, L. M., Aravind, L., Koonin, E. V., Shallom, S., Mason, T., Yu, K., Fujii, C., Pederson, J., Shen, K., et al. (1998). Chromosome 2 sequence of the human malaria parasite *Plasmodium falciparum*. *Science* **282**, 1126–1132.
- Gaskins, E., Giik, S., DeVore, N., Mann, T., Ward, G., and Beckers, C. (2004). Identification of the membrane receptor of a class XIV myosin in *Toxoplasma gondii*. *J. Cell Biol.* **165**, 383–393.
- Gilbert, L. A., Ravindran, S., Turetzky, J. M., Boothroyd, J. C., and Bradley, P. J. (2006). *Toxoplasma gondii* targets a protein Phosphatase 2C to the nucleus of infected host cells. *Eukaryot. Cell* **6**, 73–83.
- Gowda, D. C., Gupta, P., and Davidson, E. A. (1997). Glycosylphosphatidylinositol anchors represent the major carbohydrate modification in proteins of intraerythrocytic stage *Plasmodium falciparum*. *J. Biol. Chem.* **272**, 6428–6439.
- Greenbaum, D. C., Baruch, A., Grainger, M., Bozdech, Z., Medzihradsky, K. F., Engel, J., DeRisi, J., Holder, A. A., and Bogoy, M. (2002). A role for the protease falcipain 1 in host cell invasion by the human malaria parasite. *Science* **298**, 2002–2006.
- Hall, N., Karras, M., Raine, J. D., Carlton, J. M., Kooij, T. W., Berriman, M., Florens, L., Janssen, C. S., Pain, A., Christophides, G. K., James, K., Rutherford, K., et al. (2005). A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. *Science* **307**, 82–86.
- He, X. L., Grigg, M. E., Boothroyd, J. C., and Garcia, K. C. (2002). Structure of the immunodominant surface antigen from the *Toxoplasma gondii* SRS superfamily. *Nat. Struct. Biol.* **9**, 606–611.
- Hiller, N. L., Bhattacharjee, S., van Ooij, C., Liolios, K., Harrison, T., Lopez-Estrano, C., and Haldar, K. (2004). A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* **306**, 1934–1937.
- Hoppe, H. C., Ngo, H. M., Yang, M., and Joiner, K. A. (2000). Targeting to rhoptry organelles of *Toxoplasma gondii* involves evolutionarily conserved mechanisms. *Nat. Cell Biol.* **2**, 449–456.
- Igarashi, Y., Aoki, K. F., Mamitsuka, H., Kuma, K., and Kanehisa, M. (2004). The evolutionary repertoires of the eukaryotic-type ABC transporters in terms of the phylogeny of ATP-binding domains in eukaryotes and prokaryotes. *Mol. Biol. Evol.* **21**, 2149–2160.
- Ishino, T., Chinzei, Y., and Yuda, M. (2005). A *Plasmodium sporozoite* protein with a membrane attack complex domain is required for breaching the liver sinusoidal cell layer prior to hepatocyte infection. *Cell Microbiol.* **7**, 199–208.
- Ishino, T., Yano, K., Chinzei, Y., and Yuda, M. (2004). Cell-passage activity is required for the malarial parasite to cross the liver sinusoidal cell layer. *PLoS Biol.* **2**, E4.
- Iyer, L. M., Balaji, S., Koonin, E. V., and Aravind, L. (2006). Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Res.* **117**, 156–184.
- Janssen, C. S., Phillips, R. S., Turner, C. M., and Barrett, M. P. (2004). *Plasmodium* interspersed repeats: The major multigene superfamily of malaria parasites. *Nucleic Acids Res.* **32**, 5712–5720.
- Kairo, A., Fairlamb, A. H., Gobright, E., and Nene, V. (1994). A 7.1 kb linear DNA molecule of *Theileria parva* has scrambled rDNA sequences and open reading frames for mitochondrially encoded proteins. *EMBO J.* **13**, 898–905.
- Kaneko, O., Tsuboi, T., Ling, I. T., Howell, S., Shirano, M., Tachibana, M., Cao, Y. M., Holder, A. A., and Torii, M. (2001). The high molecular mass rhoptry protein, RhopH1, is encoded by members of the clag multigene family in *Plasmodium falciparum* and *Plasmodium yoelii*. *Mol. Biochem. Parasitol.* **118**, 223–231.
- Kaneko, O., Yim Lim, B. Y., Iriko, H., Ling, I. T., Otsuki, H., Grainger, M., Tsuboi, T., Adams, J. H., Mattei, D., Holder, A. A., and Torii, M. (2005). Apical expression of three RhopH1/Clag proteins as components of the *Plasmodium falciparum* RhopH complex. *Mol. Biochem. Parasitol.* **143**, 20–28.
- APICOMPLEXAN SURFACE PROTEINS
- Kappe, S., Bruderer, T., Gantt, S., Fujioka, H., Nussenzweig, V., and Menard, R. (1999). Conservation of a gliding motility and cell invasion machinery in Apicomplexan parasites. *J. Cell Biol.* **147**, 937–944.
- Kappe, S. H., Noe, A. R., Fraser, T. S., Blair, P. L., and Adams, J. H. (1998). A family of chimeric erythrocyte binding proteins of malaria parasites. *Proc. Natl. Acad. Sci. USA* **95**, 1230–1235.
- Kaslow, D. C., Quakyi, I. A., Syin, C., Raum, M. G., Keister, D. B., Coligan, J. E., McCutchan, T. F., and Miller, L. H. (1988). A vaccine candidate from the sexual stage of human malaria that contains EGF-like domains. *Nature* **333**, 74–76.
- Katinka, M. D., Duprat, S., Cornillot, E., Metenier, G., Thomarat, F., Prensier, G., Barbe, V., Peyretailade, E., Brottier, P., Wincker, P., Delbac, F., El Alaoui, H., et al. (2001). Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature* **414**, 450–453.
- Keen, J. K., Sinha, K. A., Brown, K. N., and Holder, A. A. (1994). A gene coding for a high-molecular mass rhoptry protein of *Plasmodium yoelii*. *Mol. Biochem. Parasitol.* **65**, 171–177.
- Khattab, A., and Klinkert, M. Q. (2006). Maurer's clefts-restricted localization, orientation and export of a *Plasmodium falciparum* RIFIN. *Traffic* **7**, 1654–1665.
- Koonin, E. V., Makarova, K. S., and Aravind, L. (2001). Horizontal gene transfer in prokaryotes: Quantification and classification. *Annu. Rev. Microbiol.* **55**, 709–742.
- Kopan, R., and Ilagan, M. X. (2004). Gamma-secretase: Proteasome of the membrane? *Nat. Rev. Mol. Cell Biol.* **5**, 499–504.
- Kopecna, J., Jirku, M., Obornik, M., Tokarev, Y. S., Lukes, J., and Modry, D. (2006). Phylogenetic analysis of coccidian parasites from invertebrates: Search for missing links. *Protist* **157**, 173–183.
- Kreier, J. (1977). "Parasitic Protozoa." Academic Press, New York.
- Kuvarina, O. N., Leander, B. S., Aleshin, V. V., Mylnikov, A. P., Keeling, P. J., and Simdyanov, T. G. (2002). The phylogeny of colpodellids (Alveolata) using small subunit rRNA gene sequences suggests they are the free-living sister group to apicomplexans. *J. Eukaryot. Microbiol.* **49**, 498–504.
- LaCount, D. J., Vignali, M., Chettier, R., Phansalkar, A., Bell, R., Hesselberth, J. R., Schoenfeld, L. W., Ota, I., Sahasrabudhe, S., Kurschner, C., Fields, S., and Hughes, R. E. (2005). A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature* **438**, 103–107.
- Lafferty, K. D. (1999). The evolution of trophic transmission. *Parasitol. Today* **15**, 111–115.
- Le Roch, K. G., Zhou, Y., Blair, P. L., Grainger, M., Moch, J. K., Haynes, J. D., De La Vega, P., Holder, A. A., Batalov, S., Carucci, D. J., and Winzeler, E. A. (2003). Discovery of gene function by expression profiling of the malaria parasite life cycle. *Science* **301**, 1503–1508.
- Leander, B. S., and Keeling, P. J. (2003). Morphostasis in alveolate evolution. *Trends Ecol. Evol.* **18**, 395–402.
- Leander, B. S., Kuvarina, O. N., Aleshin, V. V., Mylnikov, A. P., and Keeling, P. J. (2003). Molecular phylogeny and surface morphology of *Colpodella edax* (Alveolata): Insights into the phagotrophic ancestry of apicomplexans. *J. Eukaryot. Microbiol.* **50**, 334–340.
- Leander, B. S., Lloyd, S. A., Marshall, W., and Landers, S. C. (2006). Phylogeny of marine Gregarines (Apicomplexa)—Pterospira, Lithocystis and Lankesteria—and the origin(s) of coelomic parasitism. *Protist* **157**, 45–60.
- Lespinet, O., Wolf, Y. I., Koonin, E. V., and Aravind, L. (2002). The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Res.* **12**, 1048–1059.
- Levine, N. D. (1988). "The Protozoan Phylum Apicomplexa." CRC Press, Boca Raton, FL.

- Linder, J. U., Engel, P., Reimer, A., Kruger, T., Plattner, H., Schultz, A., and Schultz, J. E. (1999). Guanylyl cyclases with the topology of mammalian adenylyl cyclases and an N-terminal P-type ATPase-like domain in *Paramecium*, *Tetrahymena* and *Plasmodium*. *EMBO J.* **18**, 4222–4232.
- Lingelbach, K., and Joiner, K. A. (1998). The parasitophorous vacuole membrane surrounding *Plasmodium* and *Toxoplasma*: An unusual compartment in infected cells. *J. Cell Sci.* **111** (Pt 11), 1467–1475.
- Lingelbach, K., and Przyborski, J. M. (2006). The long and winding road: Protein trafficking mechanisms in the *Plasmodium falciparum* infected erythrocyte. *Mol. Biochem. Parasitol.* **147**, 1–8.
- Liu, J., Istvan, E. S., Gluzman, I. Y., Gross, J., and Goldberg, D. E. (2006). *Plasmodium falciparum* ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. *Proc. Natl. Acad. Sci. USA* **103**, 8840–8845.
- Lubke, T., Marquardt, T., Etzioni, A., Hartmann, E., von Figura, K., and Korner, C. (2001). Complementation cloning identifies CDG-IIc, a new type of congenital disorders of glycosylation, as a GDP-fucose transporter deficiency. *Nat. Genet.* **28**, 73–76.
- Luhn, K., Wild, M. K., Eckhardt, M., Gerardy-Schahn, R., and Vestweber, D. (2001). The gene defective in leukocyte adhesion deficiency II encodes a putative GDP-fucose transporter. *Nat. Genet.* **28**, 69–72.
- Luo, Y., Nita-Lazar, A., and Haltiwanger, R. S. (2006). Two distinct pathways for O-fucosylation of epidermal growth factor-like or thrombospondin type 1 repeats. *J. Biol. Chem.* **281**, 9385–9392.
- Mahajan, B., Noiva, R., Yadava, A., Zheng, H., Majam, V., Mohan, K. V., Moch, J. K., Haynes, J. D., Nakhasi, H., and Kumar, S. (2006). Protein disulfide isomerase assisted protein folding in malaria parasites. *Int. J. Parasitol.* **36**, 1037–1048.
- Maier, A. G., Rug, M., O'Neill, M. T., Beeson, J. G., Marti, M., Reeder, J., and Cowman, A. (2006). Skeleton Binding Protein 1 functions at the parasitophorous vacuole membrane to traffic PfEMP1 to the *Plasmodium falciparum*-infected erythrocyte surface. *Blood* **109**, 1289–1297.
- Makinoshima, H., and Glickman, M. S. (2006). Site-2 proteases in prokaryotes: Regulated intramembrane proteolysis expands to microbial pathogenesis. *Microbes Infect.* **8**, 1882–1888.
- Mann, T., and Beckers, C. (2001). Characterization of the subpellicular network, a filamentous membrane skeletal component in the parasite *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **115**, 257–268.
- Marti, M., Good, R. T., Rug, M., Knuepfer, E., and Cowman, A. F. (2004). Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* **306**, 1930–1933.
- Matesanz, F., Teitez, M. M., and Alcina, A. (2003). The *Plasmodium falciparum* fatty acyl-CoA synthetase family (PfACS) and differential stage-specific expression in infected erythrocytes. *Mol. Biochem. Parasitol.* **126**, 109–112.
- Mayer, D. C., Kaneko, O., Hudson-Taylor, D. E., Reid, M. E., and Miller, L. H. (2001). Characterization of a *Plasmodium falciparum* erythrocyte-binding protein paralogous to EBA-175. *Proc. Natl. Acad. Sci. USA* **98**, 5222–5227.
- Mayer, D. C., Mu, J. B., Kaneko, O., Duan, J., Su, X. Z., and Miller, L. H. (2004). Polymorphism in the *Plasmodium falciparum* erythrocyte-binding ligand JESEBL/EBA-181 alters its receptor specificity. *Proc. Natl. Acad. Sci. USA* **101**, 2518–2523.
- Miller, L. H., Baruch, D. I., Marsh, K., and Doumbo, O. K. (2002a). The pathogenic basis of malaria. *Nature* **415**, 673–679.
- Miller, L. H., and Hoffman, S. L. (1998). Research toward vaccines against malaria. *Nat. Med.* **4**, 520–524.
- Miller, S. A., Thathy, V., Ajioka, J. W., Blackman, M. J., and Kim, K. (2003). TgSUB2 is a *Toxoplasma gondii* rhoptry organelle processing proteinase. *Mol. Microbiol.* **49**, 883–894.

- Miller, S. K., Good, R. T., Drew, D. R., Delorenzi, M., Sanders, P. R., Hodder, A. N., Speed, T. P., Cowman, A. F., de Koning-Ward, T. F., and Crabb, B. S. (2002b). A subset of *Plasmodium falciparum* SERA genes are expressed and appear to play an important role in the erythrocytic cycle. *J. Biol. Chem.* **277**, 47524–47532.
- Mitchell, D. A., Vasudevan, A., Linder, M. E., and Deschenes, R. J. (2006). Protein palmitoylation by a family of DHHC protein S-acyltransferases. *J. Lipid Res.* **47**, 1118–1127.
- Moon-van der Staay, S. Y., De Wachter, R., and Vaultot, D. (2001). Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409**, 607–610.
- Mortimer, R. K., Contopoulou, C. R., and King, J. S. (1992). Genetic and physical maps of *Saccharomyces cerevisiae*, Edition 11. *Yeast* **8**, 817–902.
- Mu, J., Joy, D. A., Duan, J., Huang, Y., Carlton, J., Walker, J., Barnwell, J., Beerli, P., Charleston, M. A., Pybus, O. G., and Su, X. Z. (2005). Host switch leads to emergence of *Plasmodium vivax* malaria in humans. *Mol. Biol. Evol.* **22**, 1686–1693.
- Nair, M., Hinds, M. G., Coley, A. M., Hodder, A. N., Foley, M., Anders, R. F., and Norton, R. S. (2002). Structure of domain III of the blood-stage malaria vaccine candidate, *Plasmodium falciparum* apical membrane antigen 1 (AMA1). *J. Mol. Biol.* **322**, 741–753.
- Nilsson, J., Persson, B., and von Heijne, G. (2005). Comparative analysis of amino acid distributions in integral membrane proteins from 107 genomes. *Proteins* **60**, 606–616.
- Nohte, D., Hundt, E., Langsley, G., and Knapp, B. (1991). A *Plasmodium falciparum* blood stage antigen highly homologous to the glycophorin binding protein GBP. *Mol. Biochem. Parasitol.* **49**, 253–264.
- Norimine, J., Ruef, B. J., Palmer, G. H., Knowles, D. P., Herndon, D. R., Rice-Ficht, A. C., and Brown, W. C. (2006). A novel 78-kDa fatty acyl-CoA synthetase (ACS1) of *Babesia bovis* stimulates memory CD4+ T lymphocyte responses in *B. bovis*-immune cattle. *Mol. Biochem. Parasitol.* **147**, 20–29.
- Oakley, M., Kumar, S., Anantharaman, V., Zheng, H., Haynes, J., Moch, K., Fairhurst, R., McCutchan, T. F., and Aravind, L. (2007). Molecular factors and biochemical pathways induced by febrile temperature in *plasmodium falciparum* parasites. *Infect Immun.* In press.
- O'Donnell, R. A., Hackett, F., Howell, S. A., Treeck, M., Struck, N., Krnajska, Z., Withers-Martinez, C., Gilberger, T. W., and Blackman, M. J. (2006). Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite. *J. Cell Biol.* **174**, 1023–1033.
- O'Keefe, A. H., Green, J. L., Grainger, M., and Holder, A. A. (2005). A novel Sushi domain-containing protein of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **140**, 61–68.
- Opitz, C., Di Cristina, M., Reiss, M., Ruppert, T., Crisanti, A., and Soldati, D. (2002). Intramembrane cleavage of microneme proteins at the surface of the apicomplexan parasite *Toxoplasma gondii*. *EMBO J.* **21**, 1577–1585.
- Pain, A., Renaud, H., Berriman, M., Murphy, L., Yeats, C. A., Weir, W., Kerhornou, A., Aslett, M., Bishop, R., Bouchier, C., Cochet, M., Coulson, R. M., et al. (2005). Genome of the host-cell transforming parasite *Theileria annulata* compared with *T. parva*. *Science* **309**, 131–133.
- Patthy, L. (1999). "Protein Evolution." 1st ed. Blackwell Publishing, London.
- Phizicky, E., Bastiaens, P. I., Zhu, H., Snyder, M., and Fields, S. (2003). Protein analysis on a proteomic scale. *Nature* **422**, 208–215.
- Plotkin, J. B., Dushoff, J., and Fraser, H. B. (2004). Detecting selection using a single genome sequence of *M. tuberculosis* and *P. falciparum*. *Nature* **428**, 942–945.
- Pradel, G., Hayton, K., Aravind, L., Iyer, L. M., Abrahamsen, M. S., Bonawitz, A., Mejia, C., and Templeton, T. J. (2004). A multidomain adhesion protein family expressed in *Plasmodium falciparum* is essential for transmission to the mosquito. *J. Exp. Med.* **199**, 1533–1544.

- Rajala, R. V., Datla, R. S., Moyana, T. N., Kakkar, R., Carlsen, S. A., and Sharma, R. K. (2000). N-myristoyltransferase. *Mol. Cell Biochem.* **204**, 135–155.
- Ralph, S. A. (2005). Strange organelles—*Plasmodium mitochondria* lack a pyruvate dehydrogenase complex. *Mol. Microbiol.* **55**, 1–4.
- Ralph, S. A., and Scherf, A. (2005). The epigenetic control of antigenic variation in *Plasmodium falciparum*. *Curr Opin Microbiol.* **8**, 434–440.
- Rayner, J. C., Galinski, M. R., Ingravallo, P., and Barnwell, J. W. (2000). Two *Plasmodium falciparum* genes express merozoite proteins that are related to *Plasmodium vivax* and *Plasmodium yoelii* adhesive proteins involved in host cell selection and invasion. *Proc. Natl. Acad. Sci. USA* **97**, 9648–9653.
- Razem, F. A., El-Kereamy, A., Abrams, S. R., and Hill, R. D. (2006). The RNA-binding protein FCA is an abscisic acid receptor. *Nature* **439**, 290–294.
- Robson, K. J., Hall, J. R., Jennings, M. W., Harris, T. J., Marsh, K., Newbold, C. I., Tate, V. E., and Weatherall, D. J. (1988). A highly conserved amino-acid sequence in thrombospondin, properdin and in proteins from sporozoites and blood stages of a human malaria parasite. *Nature* **335**, 79–82.
- Roos, D. S. (2005). Genetics. Themes and variations in apicomplexan parasite biology. *Science* **309**, 72–73.
- Rosenthal, P. J. (2004). Cysteine proteases of malaria parasites. *Int. J. Parasitol.* **34**, 1489–1499.
- Saier, M. H., Jr., and Paulsen, I. T. (2001). Phylogeny of multidrug transporters. *Semin. Cell Dev. Biol.* **12**, 205–213.
- Sallstrom, B., and Andersson, S. G. (2005). Genome reduction in the alpha-Proteobacteria. *Curr. Opin. Microbiol.* **8**, 579–585.
- Sam-Yellowe, T. Y., Florens, L., Johnson, J. R., Wang, T., Drazba, J. A., Le Roch, K. G., Zhou, Y., Batalov, S., Carucci, D. J., Winzeler, E. A., and Yates, J. R., 3rd (2004a). A *Plasmodium* gene family encoding Maurer's cleft membrane proteins: Structural properties and expression profiling. *Genome Res.* **14**, 1052–1059.
- Sam-Yellowe, T. Y., Florens, L., Wang, T., Raine, J. D., Carucci, D. J., Sinden, R., and Yates, J. R., 3rd. (2004b). Proteome analysis of rho-tryptophan-enriched fractions isolated from *Plasmodium merozoites*. *J. Proteome Res.* **3**, 995–1001.
- Saouros, S., Edwards-Jones, B., Reiss, M., Sawmynaden, K., Cota, E., Simpson, P., Dowse, T. J., Jakle, U., Ramboarina, S., Shivarattan, T., Matthews, S., and Soldati-Favre, D. (2004b). A novel galectin-like domain from *Toxoplasma gondii* micronemal protein 1 assists the folding, assembly, and transport of a cell adhesion complex. *J. Biol. Chem.* **280**, 38583–38591.
- Schneider, A. G., and Mercereau-Puijalon, O. (2005). A new Apicomplexa-specific protein kinase family: Multiple members in *Plasmodium falciparum*, all with an export signature. *BMC Genom.* **6**, 30.
- Seydel, K. B., Gaur, D., Aravind, L., Subramanian, G., and Miller, L. H. (2005). *Plasmodium falciparum*: Characterization of a late asexual stage golgi protein containing both ankyrin and DHHC domains. *Exp. Parasitol.* **110**, 389–393.
- Shiels, B. R., McKellar, S., Katzer, F., Lyons, K., Kinnaird, J., Ward, C., Wastling, J. M., and Swan, D. (2004). A *Theileria annulata* DNA binding protein localized to the host cell nucleus alters the phenotype of a bovine macrophage cell line. *Eukaryot. Cell* **3**, 495–505.
- Silva, M. D., Cooke, B. M., Guillotte, M., Buckingham, D. W., Sauzet, J. P., Le Scanf, C., Contamin, H., David, P., Mercereau-Puijalon, O., and Bonnefoy, S. (2005). A role for the *Plasmodium falciparum* RESA protein in resistance against heat shock demonstrated using gene disruption. *Mol. Microbiol.* **56**, 990–1003.
- Sinai, A. P., and Joiner, K. A. (2001). The *Toxoplasma gondii* protein ROP2 mediates host organelle association with the parasitophorous vacuole membrane. *J. Cell Biol.* **154**, 95–108.

- Singh, S. K., Hora, R., Belrhali, H., Chitnis, C. E., and Sharma, A. (2006). Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. *Nature* **439**, 741–744.
- Smith, J. D., Chitnis, C. E., Craig, A. G., Roberts, D. J., Hudson-Taylor, D. E., Peterson, D. S., Pinches, R., Newbold, C. I., and Miller, L. H. (1995). Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82**, 101–110.
- Spehr, M., and Leinders-Zufall, T. (2005). One neuron—multiple receptors: Increased complexity in olfactory coding? *Sci. STKE* **2005**, pe25.
- Spielmann, T., Ferguson, D. J., and Beck, H. P. (2003). etramps, a new *Plasmodium falciparum* gene family coding for developmentally regulated and highly charged membrane proteins located at the parasite-host cell interface. *Mol. Biol. Cell* **14**, 1529–1544.
- Spielmann, T., Hawthorne, P. L., Dixon, M. W., Hannemann, M., Klotz, K., Kemp, D. J., Klonis, N., Tilley, L., Trenholme, K. R., and Gardiner, D. L. (2006). A cluster of ring stage-specific genes linked to a locus implicated in cytoadherence in *Plasmodium falciparum* codes for PEXEL-negative and PEXEL-positive proteins exported into the host cell. *Mol. Biol. Cell* **17**, 3613–3624.
- Spycher, C., Rug, M., Klonis, N., Ferguson, D. J., Cowman, A. F., Beck, H. P., and Tilley, L. (2006). Genesis of and trafficking to the Maurer's clefts of *Plasmodium falciparum*-infected erythrocytes. *Mol. Cell Biol.* **26**, 4074–4085.
- Stwora-Wojczyk, M. M., Kissinger, J. C., Spitalnik, S. L., and Wojczyk, B. S. (2004). O-glycosylation in *Toxoplasma gondii*: Identification and analysis of a family of UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferases. *Int. J. Parasitol.* **34**, 309–322.
- Su, X. Z., Heatwole, V. M., Wertheimer, S. P., Guinet, F., Herrfeldt, J. A., Peterson, D. S., Ravetch, J. A., and Wellems, T. E. (1995). The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* **82**, 89–100.
- Templeton, T. J., Iyer, L. M., Anantharaman, V., Enomoto, S., Abrahante, J. E., Subramanian, G. M., Hoffman, S. L., Abrahamsen, M. S., and Aravind, L. (2004a). Comparative analysis of apicomplexa and genomic diversity in eukaryotes. *Genome Res.* **14**, 1686–1695.
- Templeton, T. J., Lancto, C. A., Vigdorovich, V., Liu, C., London, N. R., Hadsall, K. Z., and Abrahamsen, M. S. (2004b). The *Cryptosporidium oocyst* wall protein is a member of a multigene family and has a homolog in *Toxoplasma*. *Infect. Immun.* **72**, 980–987.
- Tolia, N. H., Enemark, E. J., Sim, B. K., and Joshua-Tor, L. (2005). Structural basis for the EBA-175 erythrocyte invasion pathway of the malaria parasite *Plasmodium falciparum*. *Cell* **122**, 183–193.
- Tonkin, C. J., Pearce, J. A., McFadden, G. I., and Cowman, A. F. (2006). Protein targeting to destinations of the secretory pathway in the malaria parasite *Plasmodium falciparum*. *Curr. Opin. Microbiol.* **9**, 381–387.
- Üllu, E., Tschudi, C., and Chakraborty, T. (2004). RNA interference in protozoan parasites. *Cell Microbiol.* **6**, 509–519.
- Uproft, P., Chen, N., and Uproft, J. A. (1997). Telomeric organization of a variable and inducible toxin gene family in the ancient eukaryote *Giardia duodenalis*. *Genome Res.* **7**, 37–46.
- Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., and Marth, J. (1999). Essentials of glycobiology cold spring harbor laboratory press, New York.
- Vinetz, J. M., Valenzuela, J. G., Specht, C. A., Aravind, L., Langer, R. C., Ribeiro, J. M., and Kaslow, D. C. (2000). Chitinases of the avian malaria parasite *Plasmodium gallinaceum*, a class of enzymes necessary for parasite invasion of the mosquito midgut. *J. Biol. Chem.* **275**, 10331–10341.

- Vivier, E., and Desportes, I. (1990). Phylum Apicomplexa. In "Handbook of Protozoology" (L. Margulis, J. O. Corliss, M. Melkonian, and D. J. Chapman, Eds.), pp. 549-573. Jones and Bartlett Publishers, Boston.
- Vogt, A. M., Barragan, A., Chen, Q., Kironde, F., Spillmann, D., and Wahlgren, M. (2003). Heparan sulfate on endothelial cells mediates the binding of *Plasmodium falciparum*-infected erythrocytes via the DBL1 alpha domain of PfEMP1. *Blood* **101**, 2405-2411.
- Wang, Q., Brown, S., Roos, D. S., Nussenzweig, V., and Bhanot, P. (2004). Transcriptome of axenic liver stages of *Plasmodium yoelii*. *Mol. Biochem. Parasitol.* **137**, 161-168.
- Ward, G. E., Miller, L. H., and Dvorak, J. A. (1993). The origin of parasitophorous vacuole membrane lipids in malaria-infected erythrocytes. *J. Cell Sci.* **106**(Pt 1), 237-248.
- Weber, J. H., Vishnyakov, A., Hambach, K., Schultz, A., Schultz, J. E., and Linder, J. U. (2004). Adenylyl cyclases from *Plasmodium*, *Paramecium* and *Tetrahymena* are novel ion channel/enzyme fusion proteins. *Cell Signal* **16**, 115-125.
- Wickert, H., Rohrbach, P., Scherer, S. J., Krohne, G., and Lanzer, M. (2003b). A putative Sec23 homologue of *Plasmodium falciparum* is located in Maurer's clefts. *Mol. Biochem. Parasitol.* **129**, 209-213.
- Wickert, H., Wissing, F., Andrews, K. T., Stich, A., Krohne, G., and Lanzer, M. (2003a). Evidence for trafficking of PfEMP1 to the surface of *P. falciparum*-infected erythrocytes via a complex membrane network. *Eur. J. Cell Biol.* **82**, 271-284.
- Winstanley, P., and Ward, S. (2006). Malaria chemotherapy. *Adv. Parasitol.* **61**, 47-76.
- Winter, G., Kawai, S., Haeggstrom, M., Kaneko, O., von Euler, A., Kawazu, S., Palm, D., Fernandez, V., and Wahlgren, M. (2005). SURFIN is a polymorphic antigen expressed on *Plasmodium falciparum* merozoites and infected erythrocytes. *J. Exp. Med.* **201**, 1853-1863.
- Xu, P., Widmer, G., Wang, Y., Ozaki, L. S., Alves, J. M., Serrano, M. G., Puiu, D., Manque, P., Akiyoshi, D., Mackey, A. J., Pearson, W. R., Dear, P. H., et al. (2004). The genome of *Cryptosporidium hominis*. *Nature* **431**, 1107-1112.
- Yotoko, K. S. C., and Elisei, C. (2006). Malaria parasites (Apicomplexa, Haematozoa) and their relationships with their hosts: Is there an evolutionary cost for the specialization? *J. Zool. System. Evol. Res.* **44**, 265-270.
- Zhang, F. L., and Casey, P. J. (1996). Protein prenylation: Molecular mechanisms and functional consequences. *Annu. Rev. Biochem.* **65**, 241-269.

## Cell Responses to Biomimetic Protein Scaffolds Used in Tissue Repair and Engineering

Robert A. Brown\* and James B. Phillips†

\*Tissue Regeneration & Engineering Center, Institute of Orthopedics, University College London, Stanmore Campus, London, HA7 4LP, United Kingdom  
†Biological Sciences Department, The Open University, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom

Basic science research in tissue engineering and regenerative medicine aims to investigate and understand the deposition, growth, and remodeling of tissues by drawing together approaches from a range of disciplines. This review discusses approaches that use biomimetic proteins and cellular therapies, both in the development of clinical products and of model platforms for scientific investigation. Current clinical approaches to repairing skin, bone, nerve, heart valves, blood vessels, ligaments, and tendons are described and their limitations identified. Opportunities and key questions for achieving clinical goals are discussed through commonly used examples of biomimetic scaffolds: collagen, fibrin, fibronectin, and silk. The key questions addressed by three-dimensional culture models, biomimetic materials, surface chemistry, topography, and their interaction with cells in terms of durotaxis, mechano-regulation, and complex spatial cueing are reviewed to give context to future strategies for biomimetic technology.

**KEY WORDS:** Tissue engineering, Biomimetic proteins, Regenerative medicine, Collagen fibronectin, Fibrin, Silk, 3D modeling. © 2007 Elsevier Inc.

### I. Introduction

The field of research covered by this review represents and feeds into the basic science platform for a wide range of biotechnical activities commonly grouped under the umbrellas of tissue engineering and regenerative medicine.