Published in final edited form as:
Exp Parasitol. 2010 January ; 124(1): 2. doi:10.1016/j.exppara.2008.12.014.

Cryptosporidium: Genomic and Biochemical Features

Stanley Dean Rider Jr. and **Guan Zhu**
Department of Pathobiology, College of Veterinary Medicine & Biomedical Sciences, Texas A&M University, 4467 TAMU, College Station, TX 77843, USA

Stanley Dean Rider: drider@cvm.tamu.edu; Guan Zhu: gzhu@cvm.tamu.edu

Abstract

Recent progress in understanding the unique biochemistry of the two closely related human enteric pathogens *Cryptosporidium parvum* and *C. hominis* has been stimulated by the elucidation of the complete genome sequences for both pathogens. Much of the work that has occurred since that time has been focused on understanding the metabolic pathways encoded by the genome in hopes of providing increased understanding of the parasite biology, and in the identification of novel targets for pharmacological interventions. However, despite identifying the genes encoding enzymes that participate in many of the major metabolic pathways, only a hand full of proteins have actually been the subjects of detailed scrutiny. Thus, much of the biochemistry of these parasites remains a true mystery.

Keywords

Cryptosporidium; genomes; biochemistry; metabolism

1. Introduction

Cryptosporidium parvum is an obligate intracellular intestinal parasite of mammals that causes cryptosporidiosis, one of the most important causes of diarrhea worldwide. Cryptosporidiosis is typically a self-limiting from of intense diarrhea, but among individuals with compromised immune systems, the illness may be terminal. Early efforts to identify drugs to fight cryptosporidiosis were hampered, in part, due to the extreme difficulty of working with this parasite. *Cryptosporidium parvum* is refractory to continuous *in vitro* cultivation and is available is only small quantities, thus making many traditional biochemical methods, and high throughput drug screening nearly impossible. Inexplicably, efforts to utilize drugs that were effective on related pathogens nearly always failed. To overcome this difficulty, individual genes involved in various pathways have been cloned and analyzed using heterologous expression systems. *In vitro* biochemical testing of inhibitors against recombinant proteins, followed by small-scale testing against parasites, has provided interesting leads for future therapies. This process has been simplified by elucidating the *C. parvum* genome sequence. Hence, the major metabolic pathways, and many novel drug targets for the parasite have also been identified.

A very recent review on the Cryptosporidium biochemistry has been published as a chapter in the 2nd edition of *Cryptosporidium* and Cryptosporidiosis (Zhu, 2008). In this review, we will

provide a snapshot on the most current knowledge and advances on the areas of *Cryptosporidium* molecular biology and biochemistry, ranging from the genome features and core metabolic pathways, to the unique structural and surface proteins in this parasite.

2. Genome analyses

The *C. parvum* karyotype was established using rare-cutting restriction enzymes and pulsed-field gel electrophoresis (Caccio et al., 1998). The parasite has eight chromosomes ranging in size from 0.9 to 1.4 megabases in size. A HAPPY map of the genome was also constructed using markers based on publicly available DNA sequences (Piper et al., 1998). This map established ten linkage groups for the eight chromosomes and, along with a genome sequence survey, served as a prelude to elucidating the nearly complete *C. parvum* genome sequence (Liu et al., 1999). It became abundantly clear after the genome sequence was completed (Abrahamsen, et al., 2004) that a common reason for drug failures was that this parasite lacks many of the pharmacological targets that other apicomplexan parasites (such as *Plasmodium falciparum* and *Toxoplasma gondii*) possess. No evidence of a plastid is present (Zhu et al., 2000), thus pharmacologigal inhibition of plastid metabolic pathways is not an option in *Cryptosporidium* as it is in many other apicomplexan species. Further, no mitochondrial genome was present, although a relict mitochondrion may exist in this apicomplexan. Aside from an unidentified RNA virus (Khramtsov, et al., 2000), no extranuclear genomes exist in *Cryptosporidium*. The *C. parvum* nuclear genome, which is under 10 megabases in size, encodes a predicted set of ~3800 loci, roughly 4% of which have been assigned a putative enzymatic function (Abrahamsen, et al., 2004).

Overall, the genome of *C. parvum* is quite small in comparison to *P. falciparum* and *T. gondii*. The reduced size is partly due to a lack of genes for plastid maintenance and variant surface antigen gene families, as well as a minimal metabolism that is dependent largely on nutrient acquisition from its host. The genome has reduced intron numbers with about 5% of genes possessing introns and the parasite has smaller intergenic spaces than related organisms (Abrahamsen, et al., 2004). The reduced number of introns in *Cryptosporidium* has been proposed to be the result of ancient retrotransposon activity where mRNAs reverse-transcribed into cDNA have undergone homologous recombination with genomic regions (Roy and Penny, 2007).

Phylogenomics analyses of *C. parvum* genes provided a preliminary estimate of the number of loci that are likely to have originated through lateral gene transfer (Huang et al., 2004). Many are of probable streptophyte (green plant) origin (194 loci) with nearly an equal number of genes (108 loci) that are of eubacterial origin. Loci with phylogentic affinities to Achaea (30 loci), chlorophytes (green algae, 10 loci), rhodophytes (red algae, 8 loci) and cyanobacteria (7 loci) were also identified. Comparisons with other apicomplexan species (*P. falciparum*, *Theileria annulata* and *T. gondii*) revealed that although similar numbers of genes from each species show affinity to each of the non-apicomplexan taxonomic groups mentioned above, only a small fraction of the putatively transferred loci are shared among all four lineages. Thus, considerable gene loss, and/or gene gain appears to have occurred independently among related parasites. These results also imply that the ancestor to *Cryptosporidium* possessed a plastid that was lost secondarily (Huang et al., 2004).

The genome sequence for *C. hominis* (Xu, et al., 2004) and sequences for *C. muris* (unpublished, but deposited into public databases) have also been produced and provide a valuable comparison to that of *C. parvum*. The genomes of both *C. parvum* and *C. hominis* are homologous, and have essentially the same gene content (Abrahamsen, et al., 2004, Xu, et al., 2004). They are between 90–95% identical at the nucleotide sequence level, with the variable regions primarily due to the presence of micro- or mini-satellites (Tanriverdi and Widmer,

2006). Because of the highly identical nature of the two parasite genomes, results of studies on one parasite are likely to be reflected in the other in their general biochemical features.

3. Genome maintenance and gene regulation

The *C. parvum* DNA replication machinery includes two unique replication protein A (RPA) heterotrimeric complexes that are expressed differentially during the parasite life cycle and in response to DNA damage (Rider and Zhu, 2008a). The RPA1 subunits provide the distinguishing characteristics of the two complexes that utilize the same RPA2 and RPA3 subunits (Rider, et al., 2005). Recombinant versions of the RPA1 subunits, generated with bacterial expression systems, also display unique DNA binding preferences, further indicating divergence of function (Millership and Zhu, 2002). It appears that the RPA heterotrimer possessing a short type RPA1 subunit (CpRPA1A) is the main RPA complex for early replication of the parasite's genome, while the other complex possessing a typical long-type RPA1 subunit (CpRPA1B) is probably involved in recombination and DNA repair. Expression of the CpRPA1A subunit precedes and coincides with DNA replication, as indicated by the immunofluorescence of CpRPA1 and bromodeoxyuracil incorporated into replicating parasite DNA. Induction of the CpRPA1B subunit after damage by UV irradiation, along with its ability to bind tightly to very short (≥5bp) single stranded DNAs provides compelling evidence for its role in repair (Rider and Zhu, 2008b). The remaining replication and repair machinery are typical of eukaryotic organisms, with a reduced set of components. The ORC and MCM proteins, along with RF-A, RF-C, CDC45, and PCNA work with DNA polymerases for replication while a complete excision repair pathway is available in the event of DNA damage. DMC1 and other recombination-related enzymes are available for recombination, which has been demonstrated to take place through genetic crosses of *C. parvum* parasites (Tanriverdi et al., 2007). An intriguing note is that the parasite is dependent upon a single DNA ligase for all replication and repair pathways. However, other than the RPA1 subunits, the replication and repair machinery has not been studied in detail.

Cryptosporidium is dependent on a single TATA-binding protein (TBP) to help with transcription at a basal level and whose DNA binding preferences remain unknown. For some genes, transcriptional co-activators such as the multiprotein bridging factor (MBF) may be involved in recruiting TBP to sites of transcription through interactions with other transcriptional regulators. In *C. parvum*, the MBF has been tested using yeast as a heterologous system in functional complementation assays and MBF was found to interact with TBP (Millership, et al., 2004, Zhu, et al., 2000). The basal transcriptional machinery along with a small number of transcription factors was originally thought to be all that was present in the *C. parvum* genome, and many families of transcriptional regulators are absent from this and other apicomplexan genomes (Templeton, et al., 2004). However, *C. parvum* (and related apicomplexans) was found to have an extended family of transcription factors that possess from one to four AP2 domains (Balaji et al., 2005). AP2 transcription factors are named after the abbreviation for the *Arabidopsis thaliania* gene *APETELA2*, a founding member of the gene family (Bowman, et al., 1989, Sakuma, et al., 2002). This family of transcription factors is common among the green plants, but also among diatoms, and some ciliates (Iyer et al., 2008). Two independent studies provide evidence that these transcription factors regulate important processes in *C. parvum*. By examining the upstream regions of co-regulated genes, a few putative cis-regulatory elements in *C. parvum* were identified. In a subset of glycolysis pathway genes, the sequence 5′-TGCATGCGA-3′ was identified as a possible candidate regulatory element (Mullapudi et al., 2007). Interestingly, a recent study of AP2 domain transcription factors from *P. falciparum* and *C. parvum* revealed that recombinant *C. parvum* AP2 transcription factor (cgd2_3490) binds the consensus element 5′-TGCATGCA-3′, a sequence that is very similar to the consensus of the putative cis-acting elements previously identified upstream of the glycolysis pathway loci (De Silva et al., 2008). Thus, it is likely that

the plant-like AP2 transcription factors represent major regulatory proteins for *C. parvum* transcription. *Cryptosporidium* is interesting in that it possesses two E2F/DP winged-helix DNA-binding domain transcription factors not found in *P. falciparum* (Templeton, et al., 2004). However the specific roles these transcription factors play in *C. parvum* have yet to be elucidated.

Beyond transcription factors, gene-silencing mechanisms dependent on RNA appear to be lacking in *C. parvum* and no orthologs of the machinery necessary to carry out those processes have been identified. Thus, RNA interference is unlikely to be successful in helping determine gene function in this parasite. Chromatin modifications, however, represent a fundamental aspect of the gene regulatory cascade. At a basic level, chromatin is DNA wrapped around a nucleosome octamer made of two each of the structural histone proteins H2A, H2B, H3 and H4. Linker histones (H1) may also be present in some organisms that bind the internucleosomal space, and certain histone variants may replace one or more of the core histones under special circumstances. There do not appear to be any linker histones (H1) encoded by the *C. parvum* genome, which is consistent with other apicomplexan genomes. *Cryptosporidium* encodes only one H4 histone, two highly divergent H2B proteins, and two H2A variants. A centromeric H3 and two H3 variants that represent the replication-dependent and the replication-independent "replacement" H3 are also encoded by the genome. Post-translational modifications of histones, including acetylation and methylation, represent a combinatorial code that has not been completely deciphered. This code regulates processes such as chromatin condensation, gene expression, DNA replication and recombination. In terms of post-translational modifications of histones, *C. parvum* possesses a number of interesting proteins involved in histone methylation that have unusual domain architectures, indicating some specialization of functions in this aspect of chromatin regulation. However, there is no obvious mechanism for histone demethylation because neither the amine oxidase type or jumonji domain type histone demethylases could be clearly identified. Histone acetylation is likely performed by two MYST-type, one GCN5, and one HAT1 type of histone acetyl transferases, while histone deacetylation is under control of five different histone deacetylases (HDACs), including a single sirtuin. Again, there is some similarity to plants in that one of the putative histone deacetylases is a nucleolar HD2-type similar to that observed originally in maize, and another ankyrin repeat containing HDAC has orthologs in *Ostreococcus*, diatoms and other apicomplexans but not animals. The movement and exchange of nucleosomes, which is just as essential as histone modifications, is probably carried out by the actions of about a dozen members of the Swi/Snf superfamily of ATPases, including CpSRCAP and others, some of which have unique domain architectures. Thus far, HDAC inhibitors have been shown to be effective at disrupting *C. parvum* growth *in vitro*, indicating an essential role of chromatin remodeling enzymes in *C. parvum* development (Darkin-Rattray, et al., 1996). A high throughput screen to identify pharmacological inhibitors specific for the ankyrin-repeat histone deacetylase was recently developed (Rider and Zhu, 2008a).

4. Structural proteins

The oocysts of *Cryptosporidium* are highly resistant to many chemicals, including common disinfecting agents. The resilience of the parasite to chemical disinfectants may be mediated in part through an outer wall that is a common feature of apicomplexan parasites that produce environmentally stable oocysts. The exact make-up of the oocyst wall and the mechanisms regulating its construction remain elusive. However, *Cryptosporidium* oocyst wall proteins (COWPs) are an integral part of this structure, and are encoded by at least nine genes. *Plasmodium falciparum*, which has oocysts that remain in side the mosquito host, apparently lacks such proteins while *T. gondii* and the cryptosporidial species (all of which produce a stable oocyst) share this feature. The proteins themselves each contain a signal peptide thought to help direct them to the parasite's exterior. COWPs contain multiple copies of two types of

recognizable cysteine rich motifs (designated as type I and type II) with type I being present in all COWPs (Lally, et al., 1992, Spano, et al., 1997). Of the nine COWPs, only three possess a type II cysteine rich motif. Although COWPs make up a part of the structure of the oocyst wall, recent evidence indicates the presence of N-acetyl-galactosamine-containing molecules on the surface of the oocyst (Stein, et al., 2006). The presence of these molecules may indicate post-translational modifications of the oocyst wall proteins, or the presence of yet unidentified materials that make up the oocyst wall. Consistent with these observations, but in contrast to *P. falciparum*, *C. parvum* appears to encode enzymes capable of a multitude of sugar modifications in both the N-linked and O-linked glycosylation pathways, as well as GPI anchor biosynthesis (Templeton, et al., 2004). The N-linked glycosylation is initiated with specific asparagine residues in poteins and occurs through the use of a lipid-linked precursor (dolichol-PP-GlcNAc2 Man9 Glc3) containing 14 sugar molecules (Samuelson et al., 2005). The production of the lipid-linked precursor requires the presence of multiple glycosyltransferases which may differ among different organisms (due to loss of specific enzymes). The O-linked pathway involves the addition of GlcNAc to the Serine or Threonine residues of proteins through O-linked N-acetylglucosaminyltransferase (Banerjee et al., 2008). One interesting variation in the GPI anchor biosynthesis pathway of *C. parvum* is the presence of a bacterial type oligosaccharide deacetylase that may replace the function of a typical eukaryotic type N-acetylglucosaminyl phosphatidylinositol deacetylase during GPI anchor biosynthesis.

More than 30 extracellular mucin-like protein-coding genes have been identified in the *C. parvum* genome (Templeton, et al., 2004). These extracellular domain-containing proteins are modified by various pathways, including the N-linked and O-linked glycosylation pathways. The mucin-like proteins that possess these sugar modifications have been characterized in some detail. Gp900 (its mature form has a mass near 900 kDa) was the first mucin-like protein to be described in *C. parvum* and was found to be highly immunogenic and to contain abundant N-glycosylation (Petersen et al., 1992). Additional glycoproteins have been examined experimentally, with the gp40/15 protein(s) being among the better-characterized proteins. Recombinant gp40/15 has been expressed in *Escherichia coli* and *T. gondii*. In *C. parvum*, gp40/15 is expressed as a precursor protein that is cleaved into two proteins of 40 and 15 kDa (hence the name gp40/15). Cleavage is facilitated by a furin-like protease activity that is calcium dependent (Wanyiri, et al., 2007). Administration of furin inhibitors reduces *C. parvum* infection of host cells indicating cleavage of the protein may be important to host cell invasion. Following cleavage, both proteins associate with one another on the cell surface, consistent with their role in host cell invasion (O'Connor et al., 2007). The same protein was expressed in *T. gondii*, and although not efficiently processed through cleavage, proper glycosylation was demonstrated and the recombinant protein was able to bind specific lectins similar to the native protein (O'Connor, et al., 2003). Recently, a protein called p30 was discovered as a lectin from *C. parvum* that is capable of binding to the host cell surface (Bhat, et al., 2007). Specifically, p30 is a sporozoite protein required for host cell invasion that binds to *N*-acetyl-galactosamine residues, confirming that these types of proteins help target the parasite to host cells. Interestingly, p30 was found to associate with gp40 as well as other mucin-like glycoproteins (e.g. gp900) and this may indicate some cooperation among these proteins for mediating host cell invasion (Bhat, et al., 2007). Another extracellular protein, p23, is a sporozoite surface antigen that was expressed in *T. gondii*. The recombinant protein from *Toxoplasma* lysates was used to immunize mice, which generated protective antibodies that helped against subsequent *C. parvum* challenge (Shirafuji, et al., 2005). Thus, methods aimed at vaccine development are being pursued in parallel with those aimed towards chemotherapies.

More than 50 other surface protein genes have been identified in *C. parvum* (Abrahamsen, et al., 2004). A group of thrombospondin-related anonymous proteins (TRAP) represents a family of surface proteins in *C. parvum*, while there appears to be only a single TRAP present in *P.*

falciparum (Templeton et al., 2004). In *P. falciparum*, the TRAP protein is required for host cell invasion. This accomplished through gliding motility whereby the extracellular portion of the TRAP protein is connected to the inner parasite membrane through an actin-myosin motor complex (Baum, et al., 2006). In *C. parvum*, gliding has also been demonstrated, and is comparable to other apicomplexans. Structural proteins involved in gliding motility include the TRAP proteins, as well as cytoskeletal proteins. *Cryptosporidium* possesses a single actin protein, but also seven other actin-like proteins (Gordon and Sibley, 2005). An actin-myosin system is required for gliding motility and host cell invasion in *C. parvum* and inhibitors of this process disrupt both gliding and host cell invasion (Chen et al., 2004).

The overall structure of the parasite, including its apical complex, is dependent upon microtubules constructed from α- and β-tubulins. A single β-tubulin gene is encoded by the *C. parvum* genome, and is one of the few genes possessing an intron (Caccio et al., 1997). Cytochalisin D, an inhibitor of microtubule function, was recently shown to disrupt apical organelle discharge, as well as parasite invasion, but not attachment of the parasite to host cells (Chen et al., 2004). Thus, microtubules are an important factor in host cell invasion independent of host cell attachment by the parasite. This is consistent with attachment being a separate function mediated by extracellular glycoproteins like gp40/15.

5. Carbohydrate and lipid metabolism

There does not appear to be any fatty acid beta-oxidation pathway in *C. parvum*, suggesting that fatty acids are not an energy source. Despite the presence of an organelle that may represent a relict mitochondrion, there is no Krebs cycle in *C. parvum*. A few proteins potentially involved in mitochondrial function remain (17 in total), including HSP65, HSP70, TIM17 and TOM40 (Abrahamsen, et al., 2004). Some Fe-S cluster proteins, a cyanine-resistant alternative oxidase and pyridine nucleotide transhydrogenase add support for the presence of a mitochondrial remnant and indicate a possible function in iron-sulfur cluster biogenesis or respiration (Abrahamsen, et al., 2004, Slapeta and Keithly, 2004, Suzuki, et al., 2004).

However, the main energy pathway in *C. parvum* is probably glycolysis, and the parasite may be dependent on polysaccharides like amylopectin for basic energy storage (Harris, et al., 2004, Thompson, et al., 2005). Enzymes for the production and degradation of polysaccharides are present, including phosphorylase and 1,6-glucosidase along with a putative starch branching enzyme. Sucrose can be broken down into fructose and glucose while maltose can be broken down into glucose through the action of glucoside glucohydrolase. Both glucose and fructose can be phosphorylated by hexokinase and then fed into other pathways. Fructose-6-phosphate serves as the substrate for the production of mannose-6-phosphate via mannose-6-phosphate isomerase. Mannose-6-phosphate then is converted through two additional enzymatic steps (facilitated by phosphomanomutase and manose-1-phosphate guanyltransferase) to produce GDP-mannose, which feeds into the pathways for N-glycan (e.g. glycoprotein) and complex lipid synthesis. The enzymes involved in *Cryptosporidium* sugar metabolism have not been studied biochemically. To enable conservation of ATP consumption during glycolysis, *C. parvum* utilizes a pyrophosphate-dependent type of phosphofructokinase, rather than a typical eukaryotic ATP-dependent phosphofructokinase. Some reductive carboxylate cycle enzymes are available to produce acetyl-CoA, pyruvate, oxaloacetate and malate from acetate. Acetyl-CoA can be converted to malonyl CoA by acetyl CoA carboxylase and malonyl CoA may serve as a major substrate for the production of fatty acids or polyketides.

Cryptosporidium parvum fatty acid metabolism has been studied in some detail. Fatty acids are generated with fatty acid synthetases, and in *C. parvum* a giant (1500 kDa) modular Type I fatty acid synthetase (FAS) likely serves this function (Zhu, 2004, Zhu, et al., 2000). This enzyme can utilize malonyl CoA to extend the fatty acid chain of an activated acyl carrier

protein. Acyl carrier protein in *Cryptosporidium parvum* is activated from an apo-form to the holo-form through a phosphopantetheinyl transferase (PPT), which is an SFP-type PPT, as would be expected for a Type I FAS (Cai et al., 2005). (Note: In those organisms with a Type II system, another type of PPT activates Type II FAS.) Polyketides are also likely to be synthesized in a similar manner by virtue of a giant (900 kDa) polyketide synthetase (PKS). Both enzymes in *C. parvum* have a clear substrate preference for long chain fatty acids (e.g., C16:0), but the final products for either of these proteins remain unknown (Zhu, et al., 2002). Cerulenin has been identified as an inhibitor of FAS ketoacyl synthase activity *in vitro* (Zhu et al., 2004) and triacsin C has been shown to inhibit the acyl ligase activity of the modular PKS (Fritzler and Zhu, 2007). A single long chain fatty acid elongase (LCE)-mediated system also exists in the *C. parvum* genome that may work in a fashion similar to FAS by producing longer chain fatty acids from medium-chain fatty acyl CoAs (Fritzler, et al., 2007). However, cerulenin is not an effective inhibitor of LCE activity.

It is hypothesized that because there appear to be only enzymes for chain elongation, de-novo biosynthesis of fatty acids probably does not occur in *C. parvum.* If fatty acids are not synthesized de-novo, but are the result of chain elongation of medium or long-chain fatty acids, the question remains, how does the parasite acquire the initial substrate fatty acids? Thus far, there has not been any transporter demonstrated to be responsible for the import of fatty acids into *C. parvum*. However, *C. parvum* possesses three fatty acyl synthases (ACS) that could perform this function. These enzymes activate fatty acids by converting them into fatty acyl CoAs and in some systems, these same proteins can function in lipid trafficking and import (Zhu, 2004). Additional lipid trafficking proteins include a fatty acid binding protein (FABP) and an ankyrin repeat-containing acyl-CoA binding protein (ACBP) similar to those found in higher plants (Zeng et al., 2006). Two oxysterol binding protein related proteins (ORP) have also been characterized that show a preference for various phosphatidylinositol phosphates and sulfatide (Zeng and Zhu, 2006). One ORP has been demonstrated to reside in the parasitophorous vacuole membrane, suggesting it may play a role in lipid trafficking between the host cell and the parasite. Other proteins may play a role in lipid uptake (see transporters section below), but this has not yet been demonstrated.

6. Nucleotide metabolism

Purine biosynthesis does not take place in *C. parvum*, and this is a common feature of apicomplexans. A purine salvage pathway is present that relies on the presence of adenosine provided by the host that is likely imported through an adenosine transporter. Adenosine kinase is used to convert adenosine into adenosine monophosphate (AMP) (Galazka, et al., 2006). AMP is converted to guanosine monophosphate (GMP) and inosine monophosphate (IMP) through GMP synthase and by AMP deaminase, respectively. AMP is also converted into xanthosine monophosphate (XMP) through the use of a bacterial type IMP dehodrogenase (CpIMPDH) that was probably acquired through lateral gene transfer in the *Cryptosporidium* lineage (Striepen et al., 2002). CpIMPDH was expressed in both *T. gondii* and *E. coli*. In *T. gondii*, CpIMPDH unexpectedly complemented a ΔHXGPRT strain (carrying a deletion in hypoxanthine-xanthine-guanine phosphoribosyltransferase) because the *Cryptosporidium parvum* CpIMPDH protein is resistant to mycophenolic acid, a compound that was used in the assay to prevent purine salvage by the endogenous *Toxoplasma* enzymes (Striepen et al., 2002). *Cryptosporidium*, unlike many other apicomplexans, was subsequently found to lack an HXGPRT, indicating that only a single salvage pathway exists for generating GMP. In *E. coli* CpIMPDH expression was found to be able to complement the function in an IMPDH-knockout strain, enabling the creation of a novel drug screen for CpIMPDH (Umejiego, et al., 2008, Umejiego, et al., 2004).

Cryptosporidium lacks a *de novo* pathway for purine biosynthesis from glutamine and relies on nucleotide import for scavenging pyrimidines. Two different uracil phosphoribosyltransferases (UPRTs) are utilized to make uracil monophosphate (UMP). The two UPRTs differ in that one is fused with a uridine kinase domain that is capable of synthesizing cytidine monophosphate (CMP) from cytidine. A bacterial-type thymidine kinase is used to synthesize dTMP. Finally, the thymidylate synthase (TS) fused with dihydrofolate reductase (DHFR) is used for the methylation of dUMP to produce dTMP. This well-characterized enzyme has recently had its crystal structure solved, revealing the spatial orientation of the DHFR relative to the TS domain, as well as providing detailed molecular structures of both domains (Anderson, 2005). More recent studies on the structure and function of the DHFR-TS protein have revealed insight into the catalytic mechanisms and resulted in the identification selective inhibitors (Doan et al., 2007; Bolstad et al., 2008; Martucci et al., 2008).

7. Amino acid and polyamine metabolism

Cryptosporidium is incapable of amino acid biosynthesis de novo and relies on various transporters for the import of amino acids from the host and intestinal lumen. The parasite possesses at least 11 amino acid transporters for scavenging amino acids, which is quite high in contrast to the single amino acid transporter present in *P. falciparum* (Abrahamsen et al., 2004). However, a number of amino acid interconversions can be performed. For example, serine (Ser) and glycine (Gly) may be interconverted by serine hydroxymethyl transferase, asparagine (Asn) can be made from aspartate (Asp) by asparagine synthetase, and Glutamate (Glu) can be converted back to glutamine (Gln) by glutamine synthetase. Tryptophan synthase may synthesize tryptophan (Trp) from indole or indoleglycerol phosphate and tryptophan synthesis, which may help the parasite evade the host defense response designed to starve parasites of tryptophan. Methionine (Met) and S-adenosylmethionine (SAM) may be interconverted through SAM synthetase. Homocysteine and S-adenosyl homocysteine (SAH) are interconverted by S-adenosyl homocysteine hydrolase (SAHH). The *Cryptosporidium parvum* SAHH has been expressed as a recombinant enzyme in *E. coli* and was found to be inhibited by d-eritadenine, S-DHPA and Ara-A (Ctrnacta, et al., 2007). There is evidence that *P. falciparum* may utilize selenocysteine because several annotated genes that possess selenocysteine insertion (SECIS) elements also contain unexpected stop codons (Lobanov, et al., 2006). The *C. parvum* genome lacks SECIS elements and the machinery to recognize them, indicating that unlike *P. falciparum*, *Cryptosporidium* can not utilize selenocysteine. The remaining amino acids are probably acquired directly from the host.

Biochemical studies have demonstrated that a set of enzymes exists to convert arginine into polyamines (Keithly, et al., 1997). *Cryptosporidium parvum* utilizes a plant-like pathway for the generation of polyamines. During polyamine synthesis, arginine is first converted to agmatine by arginine decarboxylase (ADC). Agmatine is further converted to putrescine *via* agmatine iminohydrolase (AIH). Longer polyamines such as spermidine and spermine, can be made from putrescine by spermidine synthase and spermine synthase, respectively, with the latter using spermidine as a substrate. Spermine can also be converted back into spermidine, and then to putrescine by a spermidine:spermine N-acetyltransferase (SSAT) (Cook, et al., 2007, Yarlett, et al., 2007). Various inhibitors have been shown to inhibit polyamine biosynthesis activities in *C. parvum*, including difluoromethylarginine (DFMA), an inhibitor of ADC activity (Keithly, et al., 1997). Although ADC activity was consistently detected in *Cryptosporidium parvum* parasites, a gene encoding an ADC homolog has not been identified in the completed genome sequence. This may indicate that the ADC gene is highly unusual/divergent, or that an unidentified pathway is responsible for the observed ADC activity. SSAT activities have also been detected in this parasite indicating that the interconversion between

different polyamines may make it difficult to disrupt the polyamine pathway downstream of putrescine biosynthesis, as this may require disrupting multiple enzymes.

8. Transporters

As eluded to previously, many metabolic pathways for *de novo* synthesis for some essential compounds is lacking in *C. parvum*, that some pathways are non existent or highly reduced, and that the parasite is dependent on the host for many of the basic building blocks of nucleic acid, carbohydrate, lipid, and protein biosynthesis. As a result, this parasite must acquire these compounds from the host or the surrounding intestinal lumen. To facilitate this process, the *C. parvum* genome encodes an expanded repertoire of transporters (Abrahamsen, et al., 2004). Homology based searches are able to identify transporters involved in the movement of nucleotides, sugars and amino acids. However most of these transporters have yet to be examined experimentally to determine their true substrates. A large number (~19) of ABC-type transporters exists that may facilitate the transport of lipids, sterols, metabolites and drugs, and includes one ABC transporter related to the chloroquine resistance gene *PfCRT* from *P. falciparum* (Zapata et al., 2002). Another ABC transporter has been localized to the presumptive feeder organelle that exists at the parasite-host cell boundary, implying a role in nutrient uptake (Perkins, et al., 1999). Other members have been identified that belong to the multidrug resistance gene family and may be involved in the transport of peptides. There is also a precedent for ATP-binding cassette (ABC) transporters to transport lipids, so further examinations of the ABC transporters in *C. parvum* will be an interesting avenue for future study. However, the functions of most of these transporters remain to be determined.

At least seven P-type ATPases that are typically involved in transporting cations are encoded by the *C. parvum* genome. One of the *Cryptosporidium* P-type ATPases has phospholipid as its predicted substrate. This suggests that phospholipids, or other lipid molecules may be transported via one or more of the *C. parvum* ATPases. As there has not yet been a lipid importer demonstrated for *Cryptosporidium*, further examination of this protein may prove fruitful. At least three of the *C. parvum* ATPases have already been examined, including one that belongs to a group whose substrates remains unknown (Zhu and Keithly, 1997, Zhu, et al., 2002) and it was localized to the parasite apical complex. A second putative Ca^{2+}-ATPase that was found to be localized near the nucleus and apical regions of parasites (Zhu and Keithly, 1997). Finally, Another transporter of the ATPase type, predicted to be a heavy metal transporter was examined and found to specifically bind copper and was associated with plasma membranes (LaGier, et al., 2001).

One other ion pump represented in the genome is the multisubunit vacuolar proton translocating ATPase (V-type ATPase). All necessary subunits of the V-type ATPase, including the seven transmembrane proteins, are present in the parasite genome (Abrahamsen, et al., 2004). This transporter has not yet been subjected to experimental investigation.

9. Proteomics analyses

It is now possible to attempt to clone and express all the genes of an organism for structural studies. This has been attempted for 6 apicomplexan genomes: 5 plasmodium species and *C. parvum* (Vedadi, et al., 2007). More than 60 protein-coding genes from *C. parvum* were cloned and roughly half could be expressed as recombinant proteins. This resulted ultimately in the generation of 7 crystal structures being solved for different *C. parvum* proteins. These structural studies could provide a basis for molecular modeling studies aimed at designing novel inhibitors for understanding *C. parvum* biology and for developing new drugs. Other high-throughput protein-based analyses have been performed recently to identify expressed peptides in *C. parvum*. Data from several studies have confirmed the presence of several protein coding genes for which no other data (e.g. expressed sequence tags) has indicated their expression

(Snelling et al., 2007; Sanderson et al., 2008). These proteomics resources are accessible to the research community through CryptoDB (see below). At the present time, only a few proteomics analyses have been performed, but with the complete genome sequence, and improvements and interest in proteomics technologies, additional studies are likely to occur in the future.

10. Genomic and biochemical resources for cryptosporidial species

The genomes of all three *Cryptosporidium* species sequenced to date are available in a public database called CryptoDB (Heiges, et al., 2006, Puiu, et al., 2004). CryptoDB is available at http://www.cryptodb.org and has recently been incorporated into a more comprehensive assembly of parasite databases under EuPathDB (Aurrecoechea, et al., 2008) that was formerly ApiDB (Aurrecoechea et al., 2007) and is available at http://www.eupathdb.org. These databases offer a number of bioinformatics tools, including BLAST and text searches, and fully annotated versions of the *C. parvum* and *C. hominis* genomes. Highly expressed genes with full-length cDNA sequences for *C. parvum* and related parasites are overlayed onto their respective genomic sequences at Comparasite (Watanabe et al., 2007) available at http://comparasite.hgc.jp. An excellent resource for metabolic pathways in *C. parvum* and many other sequenced genomes is the Kyoto Encyclopedia of Genes and Genomes (KEGG)'s pathway database (Kanehisa and Goto, 2000) present at http://www.genome.ad.jp/kegg/pathway.html, which maintains a collection of manually curated major metabolic pathways (based on genome annotations). Enzyme sequences are annotated and cross-linked to a number of other databases, including CryptoDB. These same pathways are also presented in EuPathDB with color-coded images that compare several apicomplexan genomes. An additional metabolic resource for Cryptosporidium which is also accessible through CryptoDB, is the CryptoCyc database (http://apicyc.apidb.org/CPARVUM/server.html) and is useful for comparison with predicted metabolism pathways in other organisms (Caspi et al., 2008).

11. Concluding remarks

The molecular biology and biochemistry of the enteric human pathogen *C. parvum* have benefited substantially in the last few years from the presence of an existing genome sequence and some technologies have been applied to *C. parvum* that are wholly dependent upon the genome sequence for their interpretation (e.g. proteomics analyses). In conjunction with the widely available genomic DNA sequence for this parasite, and due to the continued intractability of the parasite itself, numerous genes have been cloned, expressed and examined as potential therapeutic targets using heterologous expression systems. The potential targets range from replicating chromatin to polyamine and nucleotide metabolism to cell surface proteins and transporters, just to name just a few. Each of the studies on the biochemistry and molecular genetics of *C. parvum* in the post-genomic era is providing optimistic new leads for future understanding of the interesting biology of this important parasite.

Acknowledgments

Studies derived from the authors' laboratory have been mainly supported by grants from the National Institute of Allergy and Infectious Diseases (NIAID) at the National Institutes of Health (NIH) (AI068461 to SDR; AI44594 and AI055278 to GZ).

References

Aurrecoechea C, Heiges M, Wang H, Wang Z, Fischer S, Rhodes P, Miller J, Kraemer E, Stoeckert CJ Jr, Roos DS, Kissinger JC. ApiDB: integrated resources for the apicomplexan bioinformatics resource center. Nucleic Acids Research 2007;35:D427–430. [PubMed: 17098930]

Balaji S, Babu MM, Iyer LM, Aravind L. Discovery of the principal specific transcription factors of Apicomplexa and their implication for the evolution of the AP2-integrase DNA binding domains. Nucleic Acids Research 2005;33:3994–4006. [PubMed: 16040597]

Banerjee S, Robbins PW, Samuelson J. Molecular characterization of nucleocytosolic O-GlcNAc transferases of Giardia lamblia and Cryptosporidium parvum. Glycobiology. 2008

Bolstad DB, Bolstad ES, Frey KM, Wright DL, Anderson AC. Structure-based approach to the development of potent and selective inhibitors of dihydrofolate reductase from cryptosporidium. Journal of Medicinal Chemistry 2008;51:6839–6852. [PubMed: 18834108]

Caccio S, La Rosa G, Pozio E. The beta-tubulin gene of *Cryptosporidium parvum*. Molecular and Biochemical Parasitology 1997;89:307–311. [PubMed: 9364975]

Caccio S, Camilli R, La Rosa G, Pozio E. Establishing the Cryptosporidium parvum karyotype by NotI and SfiI restriction analysis and Southern hybridization. Gene 1998;219:73–79. [PubMed: 9757000]

Cai X, Herschap D, Zhu G. Functional characterization of an evolutionarily distinct phosphopantetheinyl transferase in the apicomplexan *Cryptosporidium parvum*. Eukaryotic Cell 2005;4:1211–1220. [PubMed: 16002647]

Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, Paley S, Rhee SY, Shearer AG, Tissier C, Walk TC, Zhang P, Karp PD. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Research 2008;36:D623–631. [PubMed: 17965431]

Chen XM, O'Hara SP, Huang BQ, Nelson JB, Lin JJ, Zhu G, Ward HD, LaRusso NF. Apical organelle discharge by *Cryptosporidium parvum* is temperature, cytoskeleton, and intracellular calcium dependent and required for host cell invasion. Infection and Immunity 2004;72:6806–6816. [PubMed: 15557601]

De Silva EK, Gehrke AR, Olszewski K, León I, Chahal JS, Bulyk ML, Llinás M. Specific DNA-binding by apicomplexan AP2 transcription factors. Proceedings of the National Academy of Sciences U S A 2008;105:8393–8398.

Doan LT, Martucci WE, Vargo MA, Atreya CE, Anderson KS. Nonconserved residues Ala287 and Ser290 of the Cryptosporidium hominis thymidylate synthase domain facilitate its rapid rate of catalysis. Biochemistry 2007;46:8379–8391. [PubMed: 17580969]

Fritzler JM, Zhu G. Functional characterization of the acyl-[acyl carrier protein] ligase in the *Cryptosporidium parvum* giant polyketide synthase. International Journal for Parasitology 2007;37:307–316. [PubMed: 17161840]

Gordon JL, Sibley LD. Comparative genome analysis reveals a conserved family of actin-like proteins in apicomplexan parasites. BMC Genomics 2005;6:179. [PubMed: 16343347]

Huang J, Mullapudi N, Lancto CA, Scott M, Abrahamsen MS, Kissinger JC. Phylogenomic evidence supports past endosymbiosis, intracellular and horizontal gene transfer in Cryptosporidium parvum. Genome Biology 2004;5:R88. [PubMed: 15535864]

Iyer LM, Anantharaman V, Wolf MY, Aravind L. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. International Journal for Parasitology 2008;38:1–31. [PubMed: 17949725]

Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research 2000;28:27–30. [PubMed: 10592173]

Liu C, Vigdorovich V, Kapur V, Abrahamsen MS. A random survey of the Cryptosporidium parvum genome. Infection and Immunity 1999;67:3960–3969. [PubMed: 10417162]

Martucci WE, Vargo MA, Anderson KS. Explaining an unusually fast parasitic enzyme: folate tail-binding residues dictate substrate positioning and catalysis in Cryptosporidium hominis thymidylate synthase. Biochemistry 2008;47:8902–8911. [PubMed: 18672899]

Mullapudi N, Lancto CA, Abrahamsen MS, Kissinger JC. Identification of putative cis-regulatory elements in *Cryptosporidium parvum* by de novo pattern finding. BMC Genomics 2007;8:13. [PubMed: 17212834]

O'Connor RM, Wanyiri JW, Cevallos AM, Priest JW, Ward HD. *Cryptosporidium parvum* glycoprotein gp40 localizes to the sporozoite surface by association with gp15. Molecular and Biochemical Parasitology 2007;156:80–83. [PubMed: 17719100]

Petersen C, Gut J, Doyle PS, Crabb JH, Nelson RG, Leech JH. Characterization of a > 900,000-M(r) Cryptosporidium parvum sporozoite glycoprotein recognized by protective hyperimmune bovine colostral immunoglobulin. Infection and Immunity 1992;60:5132–5138. [PubMed: 1452347]

Piper MB, Bankier AT, Dear PH. A HAPPY map of Cryptosporidium parvum. Genome Research 1998;8:1299–1307. [PubMed: 9872984]

Rider SD Jr, Zhu G. An improved histone deacetylase (HDAC) assay and its application for high-throughput screening of inhibitors against *Cryptosporidium parvum* enzymes. FASEB Journal 2008a; 22 [meeting abstract].

Rider SD Jr, Zhu G. Differential expression of the two distinct replication protein A subunits from *Cryptosporidium parvum*. Journal of Cellular Biochemistry 2008b;104:2207–2216. [PubMed: 18452165]

Roy SW, Penny D. Widespread intron loss suggests retrotransposon activity in ancient apicomplexans. Molecular Biology and Evolution 2007;24:1926–1933. [PubMed: 17522085]

Samuelson J, Banerjee S, Magnelli P, Cui J, Kelleher DJ, Gilmore R, Robbins PW. The diversity of dolichol-linked precursors to Asn-linked glycans likely results from secondary loss of sets of glycosyltransferases. Proceedings of the National Academy of Sciences U S A 2005;102:1548–1553.

Sanderson SJ, Xia D, Prieto H, Yates J, Heiges M, Kissinger JC, Bromley E, Lal K, Sinden RE, Tomley F, Wastling JM. Determining the protein repertoire of Cryptosporidium parvum sporozoites. Proteomics 2008;8:1398–1414. [PubMed: 18306179]

Snelling WJ, Lin Q, Moore JE, Millar BC, Tosini F, Pozio E, Dooley JS, Lowery CJ. Proteomics analysis and protein expression during sporozoite excystation of *Cryptosporidium parvum* (Coccidia, Apicomplexa). Molecular & Cellular Proteomics 2007;6:346–355. [PubMed: 17124246]

Striepen B, White MW, Li C, Guerini MN, Malik SB, Logsdon JM Jr, Liu C, Abrahamsen MS. Genetic complementation in apicomplexan parasites. Proceedings of the National Academy of Sciences U S A 2002;99:6304–6309.

Tanriverdi S, Blain JC, Deng B, Ferdig MT, Widmer G. Genetic crosses in the apicomplexan parasite *Cryptosporidium parvum* define recombination parameters. Molecular Microbiology 2007;63:1432–1439. [PubMed: 17302818]

Templeton TJ, Iyer LM, Anantharaman V, Enomoto S, Abrahante JE, Subramanian GM, Hoffman SL, Abrahamsen MS, Aravind L. Comparative analysis of apicomplexa and genomic diversity in eukaryotes. Genome Research 2004;14:1686–1695. [PubMed: 15342554]

Watanabe J, Wakaguri H, Sasaki M, Suzuki Y, Sugano S. Comparasite: a database for comparative study of transcriptomes of parasites defined by full-length cDNAs. Nucleic Acids Research 2007;35:D431–438. [PubMed: 17151081]

Zapata F, Perkins ME, Riojas YA, Wu TW, Le Blancq SM. The *Cryptosporidium parvum* ABC protein family. Molecular and Biochemical Parasitology 2002;120:157–161. [PubMed: 11849715]

Zeng B, Cai X, Zhu G. Functional characterization of a fatty acyl-CoA-binding protein (ACBP) from the apicomplexan *Cryptosporidium parvum*. Microbiology 2006;152:2355–2363. [PubMed: 16849800]

Zeng B, Zhu G. Two distinct oxysterol binding protein-related proteins in the parasitic protist *Cryptosporidium parvum* (Apicomplexa). Biochemical and Biophysical Research Communications 2006;346:591–599. [PubMed: 16765916]

Zhu G, Keithly JS. Molecular analysis of a P-type ATPase from *Cryptosporidium parvum*. Molecular and Biochemical Parasitology 1997;90:307–316. [PubMed: 9497052]

Zhu G, Marchewka MJ, Keithly JS. *Cryptosporidium parvum* appears to lack a plastid genome. Microbiology 2000;146(Pt 2):315–321. [PubMed: 10708370]

Zhu G. Current progress in the fatty acid metabolism in *Cryptosporidium parvum*. J Eukaryotic Microbiology 2004;51:381–388. [PubMed: 15352319]

Zhu G, Li Y, Cai X, Millership JJ, Marchewka MJ, Keithly JS. Expression and functional characterization of a giant Type I fatty acid synthase (CpFAS1) gene from *Cryptosporidium parvum*. Molecular and Biochemical Parasitology 2004;134:127–135. [PubMed: 14747150]

Zhu, G. Biochemistry. In: Fayer, R.; Xiao, L., editors. Cryptosporidium and Cryptosporidiosis. CRC Press; Boca Raton, FL: 2008. p. 57-77.