How Epigenomics Contributes to the Understanding of Gene Regulation in *Toxoplasma gondii*

MATHIEU GISSOT and KAMI KIM

Departments of Medicine (Infectious Diseases) and Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York 10461

ABSTRACT. How apicomplexan parasites regulate their gene expression is poorly understood. The complex life cycle of these parasites implies tight control of gene expression to orchestrate the appropriate expression pattern at the right moment. Recently, several studies have demonstrated the role of epigenetic mechanisms for control of coordinated expression of genes. In this review, we discuss the contribution of epigenomics to the understanding of gene regulation in *Toxoplasma gondii*. Studying the distribution of modified histones on the genome links chromatin modifications to gene expression or gene repression. In particular, coincident trimethylated lysine 4 on histone H3 (H3K4me3), acetylated lysine 9 on histone H3 (H3K9ac), and acetylated histone H4 (H4ac) mark promoters of actively transcribed genes. However, the presence of these modified histones at some non-expressed genes and other histone modifications at only a subset of active promoters implies the presence of other layers of regulation of chromatin structure in *T. gondii*. Epigenomics analysis provides a powerful tool to characterize the activation state of genomic loci of *T. gondii* and possibly of other Apicomplexa including *Plasmodium* or *Cryptosporidium*. Further, integration of epigenetic data with expression data and other genome-wide datasets facilitates refinement of genome annotation based upon experimental data.

**Key Words.** Apicomplexa, epigenetics, *Plasmodium*, *Toxoplasma*.

The life cycle of *Toxoplasma gondii* is complex with multiple differentiation steps that are critical to survival of the parasite in its human and feline hosts (Kim and Weiss 2004). As for other parasitic Apicomplexa, the *Toxoplasma* life cycle can be divided into a sexual cycle and an asexual cycle occurring in a feline definitive host and in warm-blooded animal intermediate hosts, respectively (Dubey, Miller, and Frenkel 1970). Sexual stages are characterized by the production of macro- and microgametes that will form a zygote and lead to the development of oocysts, which ultimately produce sporozoites. The sporozoites are ingested and transform into tachyzoites in the intermediate host. The rapidly growing tachyzoites convert into the latent bradyzoite, which is thought to persist for the life of the host. As clinical disease is frequently due to reactivation of latent infection, tachyzoite–bradyzoite interconversion is a key transition during the infection of the intermediate host. Ingestion by the cat of tissue cysts containing bradyzoites completes the life cycle.

Transcriptome studies of the various stages of the life cycle of the parasite revealed a complex pattern of expression associated with each form of the parasite. Large scale expression sequence tag (EST) studies demonstrated that some transcripts were specifically expressed at different developmental stages of the life cycle (Ajioka et al. 1998; Li et al. 2003, 2004; Manger et al. 1998). Further studies focused on the transcriptome analysis of the tachyzoite to bradyzoite transition, using a library of mutants deficient for this conversion step. cDNA microarray analysis of these bradyzoite mutants suggested that a hierarchical gene expression program may control the bradyzoite differentiation (Singh, Brewer, and Boothroyd 2002).

The most comprehensive transcriptome study of the life cycle of the parasite was performed using the sequential analysis of gene expression technique and followed the development of the VEG (type III) in vitro from sporozoites to tachyzoites and finally to bradyzoites. The investigators also studied the tachyzoites of three laboratory-adapted strains (RH, type I; ME49, type II; VEGmsg, type III) (Radke et al. 2005). Approximately 18% of genes were stage specific with coregulated genes specific to each form of the life cycle. These coregulated genes were scattered across the genome indicating the importance of the promoter and trans-acting factors for the regulation of each particular gene (Radke et al. 2005).

Unlike the kinetoplastid protozoan parasites, apicomplexan parasites including *T. gondii* seem to greatly rely on transcriptional regulation to control gene expression. Supporting this idea, most of the basal transcription machinery required for the transcriptional control of gene expression appears to be encoded by the *T. gondii* genome (Callebaut et al. 2005; Meissner and Soldati 2005). This basal transcriptional complex is present in most eukaryotes and is composed of RNA polymerase II and its associated general transcription factors. The activity of this complex is modulated by sequence-specific transcription factors (STTF), cis-promoting elements present in promoters, as well the chromatin context. In addition, coregulators interact with the chromatin state and the STTFs connecting those actors to the basal transcription machinery (Thomas and Chiang 2006).

Until recently, STTFs were regarded as poorly represented in apicomplexan genomes (Coulson, Hall, and Ouzounis 2004), leading to speculation that epigenetic or post-transcriptional mechanisms were the primary mode of gene regulation. However, the presence of unique cis-promoting elements in the intergenic sequences is more consistent with the presence of apicomplexan-specific STTFs with unrecognized structural features: the cis elements mapped to date appear unrelated to other eukaryotic transcription binding sites (Behnke et al. 2008; Bohne, Wirsing, and Gross 1997; Kibe et al. 2005; Ma et al. 2004; Van Poppel et al. 2006). Reanalysis of apicomplexan genomes led to the discovery of a plant-like family of transcription factors with an Arabidopsis APETALA2 (AP2)-like DNA-binding domain (Balaji et al. 2005). Approximately 50 putative AP2 transcription factors have been found in the *T. gondii* genome (Iyer et al. 2008), and most appear to be expressed in *T. gondii* tachyzoites (MG, and KK, unpubl. data). Further studies are needed to decipher their role in gene regulation, but the AP2 DNA-binding domains in *Plasmodium* and *Cryptosporidium* were recently reported to bind specific DNA motifs (De Silva et al. 2008).

The conservation of the structural features of chromatin modifiers and remodelers facilitated the identification of a wide array of enzymes able to modify or remodel chromatin (Sullivan and...
Hakimi 2006). Similarly, a large number of predicted proteins encompass domains able to mediate the interaction with modified histones. Surprisingly, the *T. gondii* genome encodes histone modifier enzymes that are absent in yeast (Sullivan and Hakimi 2006), suggesting an important role for chromatin structure in *T. gondii* gene regulation.

**Epigenetics in Toxoplasma gondii.** Cytosine methylation at CpG dinucleotides mediates silencing of promoter activity in higher eukaryotes (Fazzari and Greally 2004). The presence of methylated cytosines plays a role in gene silencing in unicellular eukaryotes, such as *Dictyostelium* (Katoh et al. 2006) and *Entamoeba* (Fisher, Siman-Tov, and Ankri 2004), but is absent in the *Saccharomyces cerevisiae* genome. The cytosine-5 DNA methyltransferases (DNMT) are enzymes able to catalyze the methylation of cytosine in the DNA context. The DNMT2 family is conserved in lower eukaryotes and catalyzes DNA cytosine methylation in *Dictyostelium* (Katoh et al. 2006) and *Entamoeba* (Fisher et al. 2004), mainly at repetitive elements and retroelements. We discovered a homolog of the DNMT2 family of DNMTs in the *T. gondii* genome, but were unable to find methylated cytosine in the genomic DNA (Gissot et al. 2008). Although, we cannot exclude the presence of methylated cytosine in the genomic DNA of other stages of the *T. gondii* life cycle, the absence of this modified base in *Cryptosporidium* genome (Gissot et al. 2007). This custom oligonucleotide microarray ChIP and hybridization on a tiling microarray (ChIP on chip) was found in *Toxoplasma gondii*: this mark failed to correlate with promoters of repressed genes but instead localized to repetitive DNA and telomeres (Sautel et al. 2007). Another repressive mark, the trimethylation of lysine 9 on H3 (i.e. H3K9me3), a heterochromatin marker in other eukaryotes, was found in *Plasmodium* in the promoters of inactive var genes (Chookajorn et al. 2007; Lopez-Rubio et al. 2007), and was also reported in *T. gondii* (Sautel et al. 2007). However, its distribution does not seem to correlate with promoters in *T. gondii* (Sautel et al. 2007).

Although the repertoire of chromatin remodeling enzymes is similar in *Plasmodium* and *T. gondii* (Sullivan and Hakimi 2006), it is not yet clear how conserved individual histone modifications are among the Apicomplexa or whether these modifications correspond to similar states of gene activation. Mass spectrometry analysis of the *T. gondii* histones would be invaluable for precise mapping of histone modifications that are present in the epigenetic arsenal of the parasite. Further, the biochemical activities and specificity of candidate chromatin remodeling enzymes have been experimentally verified for only a subset of these factors (Bhatti et al. 2006; Saksouk et al. 2005; Sautel et al. 2007; Smith et al. 2005).

**Epigenetics and modified histones in gene regulation in Toxoplasma gondii.** Initial studies investigating the role of modified histones in *T. gondii* focused upon a small number of developmentally regulated genes (Saksouk et al. 2005; Sautel et al. 2007) using a combination of chromatin immunoprecipitation (ChIP) and single locus PCR. The low resolution and sensitivity of PCR on a small number of loci limits the conclusions of those studies.

We investigated the epigenetic organization and the transcription patterns of a contiguous 1% of the *T. gondii* genome using ChIP and hybridization on a tiling microarray (ChIP on chip) (Gissot et al. 2007). This custom oligonucleotide microarray

<table>
<thead>
<tr>
<th>Modified histone</th>
<th>ChIP</th>
<th>ChIP-chip</th>
<th>ChIP-chip peaks</th>
<th>Localization of the peaks</th>
<th>Correlation to gene expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>H3K4me1</td>
<td>+</td>
<td>+</td>
<td></td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>+</td>
<td>+</td>
<td></td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>+</td>
<td>+</td>
<td></td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>H3K9ac</td>
<td>+</td>
<td>+</td>
<td></td>
<td>Promoters&lt;sup&gt;a&lt;/sup&gt;</td>
<td>+</td>
</tr>
<tr>
<td>H3K9me3</td>
<td>+</td>
<td>+</td>
<td></td>
<td>Promoters&lt;sup&gt;a&lt;/sup&gt;</td>
<td>+</td>
</tr>
<tr>
<td>H3R17me2</td>
<td>+</td>
<td>+</td>
<td></td>
<td>Telomeres and repetitive sequences&lt;sup&gt;b&lt;/sup&gt;</td>
<td>–</td>
</tr>
<tr>
<td>H3K18ac</td>
<td>+</td>
<td>+</td>
<td></td>
<td>Promoters&lt;sup&gt;a&lt;/sup&gt;</td>
<td>+</td>
</tr>
<tr>
<td>H4ac</td>
<td>+</td>
<td>+</td>
<td></td>
<td>Promoters&lt;sup&gt;a&lt;/sup&gt;</td>
<td>+</td>
</tr>
<tr>
<td>H4K20me3</td>
<td>+</td>
<td>+</td>
<td></td>
<td>Telomeres and repetitive sequences&lt;sup&gt;b&lt;/sup&gt;</td>
<td>–</td>
</tr>
<tr>
<td>H4K20me1</td>
<td>+</td>
<td>+</td>
<td></td>
<td>Promoters&lt;sup&gt;a&lt;/sup&gt;</td>
<td>+</td>
</tr>
</tbody>
</table>

Histone modifications reported in the literature for *T. gondii* include methylation and acetylation of histone H3 and histone H4 with studies reported using chromatin immunoprecipitation followed by PCR amplification of genomic loci of interest (ChIP) or chromatin immunoprecipitation followed by hybridization to a tiled genomic array (ChIP-chip).

Studies reporting the histone modification are

<sup>a</sup>Gissot et al. (2007).
<sup>b</sup>Saksouk et al. (2005).
<sup>c</sup>Sautel et al. (2007), localization of peaks inferred from results with SET8-myc.
<sup>d</sup>MG, and KK. (unpubl. data).
<sup>e</sup>NA, not applicable; ND, not determined.
contained 12,995 50 mer features tiling a 650-kb region of chromosome 1b of the RH strain, with an average resolution of one oligonucleotide every 50 bp. Using commercial antibodies specific for modified histones, we observed 52 clear, discrete, and coincident peaks of H4ac, H3K9ac, and H3K4me3 within the 650-kb region tiled on the microarray. Interestingly, a fourth modified histone (H3R17me2) was only enriched in a small subset of promoters. Overall, 51 of the 52 regions with a cluster of H3K9ac, H4ac, and H3K4me3 peaks had significant cDNA hybridization signals adjacent to them, demonstrating a strong correlation between the presence of modified histone peaks and gene expression. In contrast, two regions of clustered genes, with stage-specific expression based on EST data and no expression at the tachyzoite stage, were shown to lack these modified histone peaks (Gissot et al. 2007).

One region with modified histone peaks was associated with a gene encoding a putative UV-induced double-strand break-repair protein that had no detectable mRNA expression. This gene might be poised for rapid expression as demonstrated in human T cells (Roh et al. 2006). Gene expression can be rapidly regulated at certain loci with the inactive PolII associated with the promoter, ready to induce transcription (Margaritis and Holstege 2008). Up to 20% of the regions with modified histones were not associated with a corresponding expressed gene in a survey of human promoters, suggesting that this may be a common mechanism of gene regulation (Kim et al. 2005).

In yeast, phosphorylation of the serine 5 (S5) of the PolII is a marker of elongation because this post-translational modification occurs upon activation and initiation of elongation (Cho 2007). A survey of the distribution of the PolII S5 phosphorylation across the yeast genome revealed that PolII is poised at the promoters of rapidly expressed genes, where the peaks of H3K9ac and H4ac are, but not in the body of the gene (Radonjic et al. 2005). In eukaryotes, the pausing of PolII is dependant upon negative factors (5,6-dichloro-1-β-ribofuransosylbenzimidazole sensitivity-inducing factor and the negative elongation factor), and factors able to relieve the inhibition of the PolII activity as the positive transcription–elongation factor-b complex (Margaritis and Holstege 2008). Predicted proteins with low but significant similarity with those factors are present in T. gondii and therefore may participate in this regulatory mechanism. However, understanding the extent or usage of this type of regulation in T. gondii awaits the results of a genome-wide survey of the distribution of modified histones, the localization of different phosphorylated forms of PolII, and the correlation of both of these with gene expression.

We have now expanded our ChIP-chip studies to the whole genome of T. gondii using each of the three archetypic strains (RH, type I; PLK/ME49, type II; CEP/CTG, type III) hybridized to a genome-wide tiled microarray based on the genomic sequence of the ME49 strain (http://www.toxodb.org). H3K9ac and H3K4me3 ChIP-chip of tachyzoite DNA from each of these strains revealed that PolII is poised at the promoters of modified histone peaks, the peaks of H3K9ac and H4ac are, but not in the body of the gene (Roh et al. 2006). Gene expression can be rapidly regulated at certain loci with the inactive PolII associated with the promoter, ready to induce transcription (Margaritis and Holstege 2008). Up to 20% of the regions with modified histones were not associated with a corresponding expressed gene in a survey of human promoters, suggesting that this may be a common mechanism of gene regulation (Kim et al. 2005).

Epigenomics allows the mapping of active promoters. The regions encompassing the peaks of modified histones were located at the 5’-end of predicted genes and strongly correlated to gene expression. To determine if these regions had promoter function, we cloned genomic regions encompassing the H3K9ac and H4ac peaks, which are undistinguishable one from another, and tested their promoter activity using a luciferase reporter. These regions were able to directionally drive the expression of the reporter (Gissot et al. 2007), illustrating the utility of modified histones to define the location of the active promoters on this portion of the genome.

The size of the peaks is on average 1,550 bp, much longer than the previously well-described promoters driving the expression of genes like GRA1, DHRF, α-tubulin in T. gondii (558, 467, and 530 bp, respectively). The ChIP technique relies on the hybridization of DNA fragments of 500–1,000 bp generated by sonication of the genomic DNA. The presence of one modified histone at the extremity of the fragment is theoretically enough to yield a positive signal, artificially increasing the size of regions with positive signal. An alternative ChIP-chip technique, based on the hybridization of micrococcal nuclease-digested DNA, permits a higher resolution, selecting only the histone-bound DNA, which is protected from digestion (Clark and Shen 2006). The regions defined by ChIP-chip probably do not represent the minimal promoters and encompass more sequence than is strictly needed to drive the expression of the genes.

Our working model envisions the chromatin and the SSTF as coregulators of the transcriptional activity of promoters in T. gondii (Fig. 1). Whether the SSTF recruits the chromatin modifier and remodelers (Fig. 1A) or the opposite (Fig. 1B), both types of factors act in a coordinated fashion to stabilize the PolII complex at the transcriptional start site.
genes. We found two mispredicted genes in the portion of the ChrIb we studied (Gissot et al. 2007). Therefore, the mapping of active promoters can inform an experimentally based refinement of the annotation of the genome. Expansion of our analysis to multiple strains and different stages of the life cycle will also enhance accurate gene annotation on other less explored forms of the parasite.

We have also shown that promoters have directional activity and that the distribution of the H3K4me3 is predictive of the orientation of transcription. Using this feature, we discovered two regions where no gene was predicted for which we detected modified histones and mRNA hybridization (Gissot et al. 2007). One of the regions encoded a putative cytochrome oxidase subunit III gene and the other had no obvious open reading frame and could represent a non-coding RNA. Because the 5'-end of the gene is the most difficult to predict (Bajic et al. 2006), locating the promoter of genes represents valuable experimental data that can be incorporated into algorithms used for gene prediction and genome annotation.

Conclusion. The use of epigenomics will allow a better understanding of gene regulation in Apicomplexa and provide invaluable data for experimentally based refinement of genome annotation. Although the general picture has been investigated with the first characterization of the T. gondii tachyzoite epigenome, preliminary data suggest that the modified histone code is more complex than envisioned previously: there are H3K9ac peaks unlinked to a H3K4me3 peak and only a subset of promoters have associated H3R17me2. Deciphering the T. gondii epigenome during developmental transitions, as well as in different strains, is likely to provide further insights into the mechanisms of gene activation and repression in T. gondii.

ACKNOWLEDGMENTS

M. G. was supported by a Philippine Foundation Fellowship. This work was supported by NIH NIAID RO1 AI060496 to K. K.

LITERATURE CITED


