The transcriptional response to encystation stimuli in *Giardia lamblia* is restricted to a small set of genes

Laura Morf[1], Cornelia Spycher[1], Hubert Rehrauer[2], Catharine Aquino Fournier[2], Hilary G. Morrison[3] and Adrian B. Hehl[1]*

[1]Institute of Parasitology, University of Zurich, 8057 Zurich, Switzerland; [2] Functional Genomics Center Zurich, 8057 Zurich, Switzerland, [3]Marine Biological Laboratory, Woods Hole, Massachusetts, 02543, USA

*Address correspondence to:

Adrian B. Hehl, E-mail: adrian.hehl@access.uzh.ch
Institute of Parasitology, University of Zurich.
Winterthurerstrasse 266a , 8057 Zurich, Switzerland.

ABSTRACT

The protozoan parasite *Giardia lamblia* undergoes stage-differentiation in the small intestine of the host to an environmentally resistant and infectious cyst. Encystation involves secretion of an extracellular matrix comprised of cyst wall proteins (CWPs) and a β(1-3)-GalNAc homopolymer. Upon induction of encystation, genes coding for CWPs are switched on, and mRNAs coding for a transcription factor Myb and enzymes involved in cyst wall glycan synthesis are upregulated. Encystation in vitro is triggered by several protocols, which call for changes in bile concentrations or availability of lipids, and elevated pH. However, the conditions for induction are not standardized and we predicted significant protocol-specific side effects. This makes reliable identification of encystation factors difficult. Here, we exploited the possibility to induce encystation with two different protocols, which we show to be equally effective, for a comparative mRNA profile analysis. The standard encystation protocol induced a bipartite transcriptional response with surprisingly minor involvement of stress genes. A comparative analysis revealed a core set of only 18 encystation genes and showed that a majority of genes was indeed upregulated as a side effect of inducing conditions. We also established a Myb binding sequence as a signature motif in encystation promoters, suggesting coordinated regulation of these factors.

INTRODUCTION

Differentiation of *Giardia lamblia* to cysts is a key step in the simple life cycle of this ubiquitous intestinal parasite. Induced trophozoites, the flagellated, motile, and attachment competent stage, exit the cell cycle at the G2-M transition (Bernander *et al.*, 2001) and begin to synthesize the components of an extracellular matrix, the cyst wall (CW). Synthesis and export of the CWPs to specialized organelles termed encystation-specific vesicles (ESVs) is completed between 8 and 10 h post induction (p.i.). After proteolytic cleavage of the CWP2 C-terminus the CWPs are sorted into two fractions which are deposited sequentially and polymerize on the surface of the cyst (Konrad *et al.*, 2010). The first layer of the CW which eventually completely encloses the parasite (Lauwaet *et al.*, 2007; Marti and Hehl, 2003) is secreted in the last minutes of differentiation simultaneously with nuclear division and morphological transformation of the cell (cyst formation). In vitro, the entire encystation process typically takes 20-24 h in our hands and can be induced by modifying medium components and/or concentrations of bile, fatty acids (cholesterol) or lactic acid (Boucher and Gillin, 1990; Gillin *et al.*, 1989; Kane *et al.*, 1991; Lujan *et al.*, 1996a). Because the condition(s) which induce encystation in vivo are not known, all factors and components, mentioned above, are implicated since their concentrations vary in the small intestine, as does the pH which increases from ~6 in the duodenum to 7.5 - 8.0 in the distal ileum. Although all available data indicate that differentiation is induced by one or several environmental signals, the mechanism(s) for reception and transduction have not been characterized. Upregulation of the genes coding for the three major structural proteins and the enzymes involved in the synthesis of the glycan component of the cyst wall has been documented by various small- and large scale analysis methods (Kim *et al.*, 2009; Palm *et al.*, 2005; Que *et al.*, 1996; Stefanic *et al.*, 2006; Sun *et al.*, 2003; Yang *et al.*, 2003). Considering how effectively the cyst wall protects the parasite in the environment, the exported CWM appears to have a surprisingly low complexity: The three cyst wall proteins (CWP1-3) have been identified as the major protein components of the structure (Lujan *et al.*, 1995; Sun *et al.*, 2003). In addition, a simple β (1-3) - GalNAc homopolymer glycan (Gerwig *et al.*, 2002; Jarroll *et al.*, 1989) contributes ~60% of the CWM, but its manner of integration with CWPs in this extracellular matrix is unknown. A membrane-anchored, cysteine-rich protein (HCNCp) which localizes to the endoplasmatic reticulum and the periphery of cysts (Davids *et al.*, 2006) could function as a possible link between the CW proper and the cell surface. The three CWPs are paralogous members of a protein family (Sun *et al.*, 2003) and the corresponding genes are not transcribed in trophozoites. Transcript levels during in vitro encystation peak around 7 h p.i. and appear to be sharply downregulated thereafter (Davis-Hayman *et al.*, 2003; Hehl *et al.*, 2000). Inducible expression of a reporter under the control of a CWP1 promoter showed that down-regulation of the reporter mRNA after peak induction was dependent on a short downstream region flanking the CWP1 ORF (Hehl *et al.*, 2000). Another category of upregulated genes, in addition to those coding for structural proteins, are factors needed for the synthesis of the cyst wall glycan. Transcription of five enzymes involved in synthesis of UDP-GalNAc from glucose was found to be induced relatively late during

encystation (Lopez *et al.*, 2003). At least one of those enzymes, UDP-N-acetylglucosamine phosphorylase, is also allosterically regulated (Bulik *et al.*, 2000).

Several investigations have addressed how expression of CWPs and the enzymes involved in cyst wall biosynthesis may be stage-regulated. A Myb2-like protein (*Giardia* Myb) is one of the few transcription factors identified in *Giardia* (Huang *et al.*, 2008; Sun *et al.*, 2002; Yang *et al.*, 2003). Recently, a plant-like WRKY protein was also shown to bind to promoter sequences and influence transcription levels (Pan *et al.*, 2009). Both trans-acting factors have binding sites within 100 bp upstream of the open reading frames (ORFs) of CWP1, CWP2, and genes coding for several other factors involved in biosynthesis of the cyst wall. While it seems clear that the *Giardia* Myb plays a role in regulating transcription of CWPs and of itself by binding to short sequence motifs upstream of these ORFs, the literature does not quite agree on the nature of these motifs. Random site selection experiments suggested that a 5'-GTTT(G/T)(G/T)-3' motif and Myb was involved which binds to both CWP1 and Myb promoters (Yang *et al.*, 2003). Gillin and coworkers described a binding motif 5'-C(T/A)ACAG-3' in the upstream regions of the Myb, CWP1-3 and G6PI-B genes to which this putative transcription factor binds (Sun *et al.*, 2002). Interestingly, three Myb binding sequences (MBS) inserted upstream of a constitutively expressed gene confer only weak stage regulation (2.6 fold), but insertion upstream of the regulated G6PI-B gene promoter boosts induction >10 fold in encysting cells. This suggests that *Giardia* Myb is able to enhance transcriptional upregulation. A binding motif 5'- (c/t)TGAC(c/t) -3' for a giardial WRKY protein has been identified in the upstream region of the stage-regulated WRKY, CWP1, CWP2, and Myb genes, but also in a number of non-regulated genes (Pan *et al.*, 2009). The WRKY binding motif appears to act as a positive regulatory element of transcription in vegetative and in differentiating cells. An interesting aspect of encystation-specific promoter function was revealed by a deletion study of the CWP2 upstream sequence. Davis-Hayman and coworkers (Davis-Hayman *et al.*, 2003) showed that deletions in a 64 nt region upstream of the CWP2 gene led to expression of a reporter also in proliferating trophozoites, effectively abolishing stage-regulation. Thus, while Myb and other transcription factors may play a role in modulating expression of encystation genes, this evidence suggests that stage-specific control of CWP transcription is mediated by gene silencing in trophozoites and de-repressed upon induction of encystation.

Taken together, the current in vitro data indicate that rapid induction of gene expression during encystation is controlled on multiple levels, which encompass basic transcription activity, enhancement and de-repression. However, with the exception of a putative repressor for CWP genes, no single factor or cis-acting motif appears to have a dominant role. A recent study shows that ultimately, trophozoite to cyst differentiation is controlled on an epigenetic level and depends on histone acetylation (Sonda *et al.*, 2010). The protocol for induction of encystation in vitro is key. None of the current experimental approaches to identify encystation genes, including those used to generate the serial analysis of gene expression (SAGE) data (Birkeland *et al.*, 2010; Palm *et al.*, 2005) available on the GiardiaDB website, account for the possibility that side effects of encystation protocols might lead to over-estimation of the number of candidates. Analysis of encysting parasites isolated

from the gut of a host could serve as a benchmark. However, this would require synchronization of induction, which is not feasible in vivo.

Here, we implement two distinct but equally effective encystation protocols to test the hypothesis that induction of encystation in vitro activates the transcription of a very limited number of genes coding for products that are directly required for differentiation. We predict that transcriptional induction of many other factors is due to protocol-specific off-target effects, which are neither (directly) linked to, nor required for differentiation. To address this issue, we focused our analysis on the first 7 h of encystation. During this time, the bulk of CWPs is synthesized and exported from the endoplasmic reticulum to specialized organelles, the ESVs, for maturation. Comparative analysis of 7 h p.i. microarray datasets generated with two encystation protocols confirmed upregulation of a small core set of genes, and underscored the importance of the Myb binding sequence as a signature motif in promoters of encystation genes.

RESULTS AND DISCUSSION

Induction of encystation in vitro triggers a bipartite transcriptional response. Induction levels of CWP mRNA expression reported in the literature range from 100 to 300 fold with low background expression in trophozoites (Hehl *et al.*, 2000; Lujan *et al.*, 1995; Mowatt *et al.*, 1995; Sun *et al.*, 2003). Considering that there is a "spontaneous" rate of encystation of up to a few percent in standard trophozoite cultures (Su *et al.*, 2007), it is more likely that CWP expression is completely silenced at that developmental stage. Peak mRNA levels of the most abundant proteins exported to the cyst wall (CWP1-3) in induced parasites are detected at ~7 h p.i. (Hehl *et al.*, 2000; Lujan *et al.*, 1996b). To investigate the transcriptional induction of cargo exported to ESVs we therefore focused our analysis on this initial stage of the differentiation process, which is relatively synchronous. RNA was isolated from encysting trophozoites at 45 min, 3 h, and 7 h p.i and used for dual color hybridization to a full genome microarray with uninduced trophozoite RNA as the reference sample. For this time-dependent analysis we used the standard "two-step" method for induction (Boucher and Gillin, 1990; Hehl *et al.*, 2000), henceforth referred to as protocol A. The three datasets comprise 63 different genes with a signal increase of >2 (Figure 1A). Figure 1B shows that the "early" (45 min, 18 hits) and "late" (7 h, 29 hits) datasets do not overlap. The "intermediate" dataset (3 h, 16 hits) is the smallest and has four unique positions, two overlaps with the "early" dataset, and shares 10 positions with the "late" dataset. Together with the signal increase in all but two of these shared positions at 7 h p.i. this indicates that the transcriptional response is divided into an early phase, which disappears before 3 h p.i., and a late phase, which is sustained through 7 h p.i. Only four genes are unique to the 3 h time point. The mRNAs coding for the CWPs 1-3, an HCNCp, and the transcription factor Myb are strongly upregulated after 3 h p.i. In addition, several genes associated with synthesis of the CW glycan (Lopez *et al.*, 2003) are upregulated in the "late" fraction indicating that this dataset includes also genes coding for later steps in CW synthesis. In total, mRNAs from all previously identified regulated genes are included in the combined 3 h and 7 h datasets. Thus, looking at upregulated genes only, we find a qualitatively significant but

quantitatively limited, bipartite transcriptional response to the encystation stimulus in this microarray study. This fits with the idea that encystation requires relatively small changes in expression patterns except for the synthesis of the CW components (Marti *et al.*, 2003). While the late response is easily recognized as encystation-related, we currently have no explanation for the transient upregulation of 16 genes only at 45 min p.i. Two annotated but unconfirmed ORFs in this group (3595 and 16519) code for predicted iron-sulfur proteins. The ORF 3595 product could be involved in DNA damage repair, although the significance of this is not clear. Interestingly, no classical stress genes are upregulated before 3 h p.i., suggesting that induction of encystation is not associated with a stress response. However, five of the 16 genes, which are transiently upregulated at 45 minutes p.i. are hypothetical proteins with unknown function. Thus, we cannot determine whether any of those play a role in an early process of encystation or whether upregulation of all or some members of this group can be considered an off-target effect.

Two different protocols result in highly comparable induction of encystation

The 7 h dataset generated with cells encysted with protocol A contained all previously identified regulated genes. We therefore decided to use this time point for a comparative analysis of encystation protocols with the aim of filtering out off-target effects. Transferring cultured trophozoites in cholesterol (lipid)–free medium and elevated pH (protocol B) is an efficient (albeit more expensive) method to induce encystation. As a first step, we tested if the kinetics of induction elicited by protocols A and B were comparable. As criteria, we used expression of CWP1 protein, ESV formation, and appearance of cysts. FACS analysis of chemically fixed and detergent-permeabilized cells labeled with anti-CWP1 mAb showed that the kinetics of CWP1 production during the first 8 h p.i. was very similar with both protocols (Figure 2A). Immunofluorescence analysis at 4 h and 8 h p.i. confirmed equal formation of ESVs. Quantitation of encysting cells and cysts at 7 h and 24 h p.i. revealed a slightly higher cyst output with protocol B at 24 h p.i. This suggests that the second half of the encystation process might progress slightly faster in the latter. In these tests, we also confirmed the importance of an elevated pH (7.85) for the induction with protocol B (Figure 2B). Conversely, increase of the standard culture medium pH alone is not sufficient for induction.

It is likely that in vivo, elevated pH in combination with another signal involving lipids is required to induce encystation, indicating that differentiation in vivo is only triggered efficiently in the distal ileum. The combination of two signals suggests that temporal as well as spatial criteria are important. It is tempting to speculate that the latter may provide a way to segregate rapidly proliferating trophozoites from differentiating parasites. Since conditions change significantly between the duodenum and the distal ileum, an interesting question in this respect is how long inducing conditions have to persist for the cells to complete encystation. Preliminary in vitro data indicated that regardless of the protocol used, this time is >7 h but <20 h p.i. Parasites which are brought back to standard culture conditions before 7 h p.i. do not complete encystation and often become severely damaged (data not shown). This suggests that inducing conditions do not trigger a “fire and forget” response, but parasites need sustained stimulation to complete encystation.

Comparative analysis of induced genes generated with two protocols reveals a small core set of bona fide encystation genes

Microarray analysis of cells induced with protocol B yielded a set of 37 genes whose signal was increased >2 fold at 7 h p.i. (Figure 3A). A categorization of the genes identified in the 7 h datasets indicated clear protocol-specific differences (Figure 3B). Most notable was the high number of hypothetical proteins and the identification of two different HCMPs in the protocol B dataset.

A detailed comparative analysis of the upregulated genes in the two datasets revealed a greater number of divergent than intersecting positions (graphical depiction in Figure 3C). If we follow the definition that only genes significantly induced in both protocols are truly linked to encystation, the result suggests that each protocol produces major off-target effects. However, the intersecting dataset of 13 positions contained all confirmed, highly regulated encystation genes including CWPs and Myb. The HCNCp identified in a previous study (Davids *et al.*, 2006) did not appear in any of the datasets, and the three upregulated HCMP family members are not associated with encystation according to these criteria. Genes in the intersecting dataset show a comparable degree of signal increase in both protocols with the exception of ORF 92729 (annotated as fatty acid elongase 1) which shows an induction difference of >2 fold between datasets. Induction of CWP3 appears to be independent of the protocol used. Conversely, induction of CWPs 1 and 2 is lower with protocol A. This is consistent with the minor differences of CWP1 expression observed at early time-points in the comparative FACS analysis (Figure 2A), and could reflect differences in expression kinetics rather than the final amount of product.

The results of this comparative analysis underscore the importance of distinguishing between upregulation due to off-target effects and bona fide encystation genes. This is relevant as large datasets from genome-wide transcriptome analyses (SAGE and microarray) of differentiation processes (encystation/excystation) are available to the community via GiardiaDB, and information on stage-regulated expression is included in the datasheet of each giardial gene.

From a biological perspective, a direct comparative analysis of encystation with two distinct in vitro protocols is important. This allows narrowing down the region where encystation is triggered in the host's small intestine. Cultured parasites are exposed to complex changes in environmental conditions during in vitro encystation. Similarly, trophozoites, which reside in the duodenal region experience strong temporal and local fluctuations of bile and pancreatic secretions, in addition to changes induced by dietary factors. Importantly, there is a considerable pH gradient between the proximal and distal small intestine. The most consistent parameter of the two encystation protocols is the absolute requirement of a pH of 7.8 - 7.9. Because this is necessary but not sufficient for strong induction of encystation in vitro, it seems unlikely that encystation in vivo is triggered in the duodenal region where the pH is 6 – 6.5. The lumen of the jejunum and the ileum, however, is increasingly alkaline and reaches pH 7 - 8 in distal regions. Cholesterol and fatty acids are resorbed by enterocytes in the lower small

intestine thus, parasites in the distal regions of the ileum likely encounter conditions which trigger differentiation.

A Myb binding sequence is a signature motif for encystation genes

The cis-acting elements for control of *Giardia* gene expression are included in short ~100 bp sequences upstream of transcription start sites. The only element common to many giardial promoters is a short AT-rich sequence (initiator) close to the transcription start site (Elmendorf *et al.*, 2001). CWP promoters follow this rule, and in addition, contain binding sites for the *Giardia* Myb transcription factor. A sequence analysis of ~100 bp upstream of the 13 identified encystation genes revealed that all promoters contained at least one MBS. A MEME analysis using a motif-width between 4 and 12 nt showed that the MBS was the only motif which was over-represented in the promoter regions of encystation genes (Figure 4A). We also identified the position and orientation of 33 MBS in 13 promoter sequences (~100 bp of upstream sequence) (Figure 4B), and generated a more refined MBS 6-mer motif representation 5'-(C/T-T/A-A-C-A/T-G/A) -3' with this training set. The result is consistent with the motif identified in the scanning mutagenesis experiment to analyze binding properties of Myb performed by Sun and coworkers (Sun *et al.*, 2002). We tested if this motif was also over-represented in the 40 (16 + 24) non-overlapping genes in the two 7 h datasets. This analysis revealed one MBS each in the promoter sequences of five additional genes that appeared either in the protocol A or B dataset at 7 h p.i. These candidates were not in the intersecting group because they were eliminated for technical reasons (e.g. large variation between technical replicates, p value >0.01) or because they fell just below the cut-off of 2.0 fold signal increase (Supporting Material Figure S1). Nevertheless, taking into account additional criteria such as statistically significant signal increase of >1.9, detection of the MBS in the MAST analysis, upregulation in the 3 h time-point, or mention in the literature (Lopez *et al.*, 2003; Sun *et al.*, 2003), we decided to include these five candidates in a final set of 18 encystation genes (Figure 4C).

We expanded our analysis and found that none of the promoters of the 18 genes in the 45 min time-point (protocol A) contained a MBS. However, in 16 of these promoters, at least one AT-rich motif was identified by a MEME/MAST analysis, predominately in the distal half of the promoter (Supporting Material Figure S2A). This is additional support for the idea that the early response to protocol A is regulated separately and does not involve Myb. In the corresponding analysis of the 3 h time-point, a slightly divergent version of the MBS emerges in a MEME motif search (Figure S2B). Six out of nine genes in this group are included in the final set of encystation genes. Finally, a global MAST analysis of the promoter sequences of 4969 predicted genes in the *Giardia* genome found the 6 nt MBS sequence in 715 promoter regions (including the 18 bona fide encystation genes). This frequency of occurrence is bracketed by the statistically predicted frequencies of the 6-mer (sense orientation) of 121 if no variability is allowed, and 1941 if we allow one variation at positions 1, 2, 5, or 6 with equal probability for two nucleotides in the motif. Together with the highly significant over-representation in the

promoter regions of the encystation genes, this strongly suggests an otherwise random frequency of occurrence of the MBS motif in upstream regions of genes.

Myb is constitutively expressed but upregulated during encystation. Thus, since Myb is not the central regulator of encystation gene expression, the presence of the MBS in other promoters is not surprising. Positive autoregulation via MBS in the Myb promoter (Nicolaides *et al.*, 1991; Sun *et al.*, 2002) is a likely mechanism to ensure a rapid increase of Myb mRNA levels in differentiating cells. Sun et al. showed that three tandem MBS copies confer some upregulation to the constitutive Ran promoter in a reporter construct. Together with data demonstrating a cis-acting sequence in the promoter of CWP2 conferring tight transcriptional repression of the gene in trophozoites (Davis-Hayman *et al.*, 2003), this fits with the idea that MBS modulates but does not control stage-specific transcription. Indeed, the promoters of two of our final set of 18 encystation genes contain no MBS on the sense strand. The need for rapid upregulation of Myb during encystation could also be explained by competition for space on the short promoter sequence.

Taken together, the data strongly suggest that expression of all bona fide encystation genes identified here is modulated by recruitment of *Giardia* Myb to their promoter sequences. This alone could be sufficient to increase levels of gene products, which are also expressed in trophozoites. The positive autoregulation of Myb gene expression likely reinforces this effect. However, the CWPs at least, which are silenced in trophozoites, require additional mechanisms (e.g. derepression) for induction. Thus, we postulate that encystation genes can be divided into two groups: encystation-specific genes that are stage-specifically expressed (e.g. structural proteins such as CWPs), and genes whose expression is upregulated from basal levels in trophozoites after induction. A role of Myb as an amplifier is also consistent with our recently reported data indicating epigenetic control of gene expression upon differentiation (Sonda *et al.*, 2010).

Because the 42 upregulated genes reported in this work (Birkeland *et al.*, 2010) are distributed among the 301 hits with sense tags in the three categories “encystation” (35), “differentiation” (3), and “trophozoites & encyzoites” (262), we compared the microarray hits (Figure 3A, 4C) with the SAGE data only in a unidirectional manner. In addition, we limited our analysis to the SAGE data for the 4 h and 12 h time points. We found a better correlation with hits from protocol A (19/29) than with those from protocol B (18/37) (Figure S3A). The correlation was clearly improved when only the set of 18 encystation genes were considered (14/18). Conversely, the set of non-overlapping genes from protocols A and B showed the lowest correlation (13/35). This comparison should be taken with a grain of salt because neither the protocol used to induce differentiation nor the time points of the SAGE analysis are directly comparable. However, the relatively high number of encystation genes is consistent with the central prediction of our hypothesis.

The GS strain of *Giardia lamblia* has been reported to respond less efficiently to induction of encystation in vitro. Sequence differences in the promoter region of CWPs have been discussed as a possible explanation (Franzen *et al.*, 2009). To investigate this further, we analyzed the promoters of 14 GS homologs of encystation genes. We found mutations in one or more MBS motifs in 9 of 14 promoters, which appear to confirm this trend (data not shown). However, the

biological significance of this result and a link to less efficient encystation in vitro needs to be examined in more detail.

Pharmacologically induced histone hyperacetylation prevents upregulation of encystation genes

In a previous study (Sonda *et al.*, 2010) we showed that the general level of histone acetylation decreases in encysting trophozoites suggesting an increase in chromatin condensation in as yet undetermined loci as the cells differentiate to a dormant stage. We demonstrated that by preventing hypoacetylation in these parasites with a class I histone deacetylase (HDAC) inhibitor (FR235222), synthesis and export of cyst wall material was severely compromised and resulted in very low levels of cysts in these cultures. Having identified a set of bona fide encystation genes we used this existing microarray data to determine whether their induction was affected by FR235222, which would argue further for a common mechanism of regulation. Indeed, inhibition of the giardial HDAC in encysting cells lead to a decreased (>2 fold) induction of the nine most strongly expressed encystation genes at 7 h p.i. (Figure S3B). Induction of seven additional members of this set was detected as well, but expression of two of the most weakly induced genes (ORFs 21924 and 9115) did not change in the presence of the drug. The latter is most likely due to insufficient sensitivity of the microarray analysis. Overall, this is strong support for the idea that in the first 7 h of p.i., encystation genes are co-regulated by de-repression in the case of highly expressed genes coding for structural components, and/or positive modulation by binding of Myb to promoters. The combined data also suggests that histone deacetylation during encystation is (directly or indirectly) required for the induction of encystation gene expression.

Down regulated genes

Ten significantly down regulated genes were identified in the consolidated (i.e. after removal of deprecated genes) protocol A dataset at all three time points and only two for the 7 h time point of the protocol B dataset (Figure S4). Although this is much more difficult to interpret, taken together with the comparatively small changes in expression the data suggest that down-regulation may play a smaller but perhaps still significant role in encystation (see below). However, the fact that there is no overlap between the two 7 h datasets casts considerable doubt on the relevance of these genes with respect to differentiation. There is some agreement between the microarray and the SAGE data (Birkeland *et al.*, 2010): five of twelve identified genes are listed as down regulated. Interestingly, among those is ORF 112103 coding for arginine deiminase (ADI) (Knodler *et al.*, 1998; Ringqvist *et al.*, 2008), an important factor in host parasite interaction, which was reported to translocate to nuclei and inhibit CWP2 expression (Touz *et al.*, 2008). Its

relatively strong down regulation at 45 min and 3 h p.i. but not at the 7 h time point is consistent with this scenario and warrants further investigation.

Two hypothetical proteins are upregulated and localize to the endoplasmatic reticulum

In addition to several known and predicted genes, five newly identified encystation genes coding for hypothetical proteins are included in the final dataset. To determine if this group might contain structural proteins of the CW or the CW-plasma membrane interface we expressed the two predicted secreted proteins encoded by ORFs (3063, 32657) as HA epitope-tagged variants under the control of their respective endogenous promoters (Figures 5 and S5). Consistent with the microarray data and with previous experiments using CWP sequences, a short ~100 bp promoter sequence was sufficient for a strong stage-regulated expression of both tagged proteins. Immunofluorescence analysis showed both tagged products in the endoplasmatic reticulum (ER) of transgenic encysting parasites (note the distinct nuclear envelope staining)(Figure S5). The tagged ORF-32657 product also localized at the cell periphery, but not in the cyst wall, in addition to internal compartments (Figure 5A, B, S5B). The lack of distinct ESV staining makes it unlikely that these candidates are exported to the cyst wall. ORF 32657 encodes a multipass transmembrane protein, while the predicted ORF 3063 product is a type II transmembrane protein. Despite the completely different topologies, both have very similar expression kinetics and localizations in trophozoites. Taken together with the domain structure analysis of the final dataset, this strongly suggests that no major structural proteins in addition to the three CWP family members are involved in building the cyst wall.

MATERIALS AND METHODS

Biochemical reagents

Unless otherwise stated, all chemicals were purchased from Sigma and cell culture reagents were from GIBCO BRL.

Cell culture and transfection

Trophozoites of the *G. lamblia* strain WBC6 (ATCC catalog number 50803) were grown axenically in 11-ml culture tubes (Nunc, Roskilde, Denmark) containing Diamond's TYI-S-33 medium supplemented with 10% adult bovine serum and bovine bile at 37°C. Parasites were harvested by chilling the culture tubes on ice for 30 min to detach adherent cells and cells were collected by centrifugation at 1,000 x *g* for 10 min. Cells were then resuspended/washed in ice cold phosphate-buffered saline (PBS). Construction of expression plasmids and transfection was performed as described in (Sonda *et al.*, 2010).

Induction of encystation

Protocol A: Two-step encystation was induced as described previously (Boucher *et al* 1990, Gillin *et al* 1989). Briefly, cells were cultivated for 44h to confluency in bile free medium (preencysting medium) and subsequently in prewarmed encystation medium (preencysting

medium with an increase in pH to pH 7.85 and addition of porcine bile (0.25 mg/ml, Sigma B8631) as well as lactic acid (0.545 mg/ml, Sigma L-200)) to induce encystation.
Protocol B: Lipid starvation was induced as described in (Lujan *et al.*, 1996a). Briefly, bile-free medium containing delipidated FCS (PHM-L Lipsorb, Calbiochem) at a pH of 7.85 was used to induce differentiation of cells grown to confluency in standard trophozoite medium.

Protein analysis

Protein analysis was performed as described in (Stefanic *et al.*, 2009). *Giardia* parasites were harvested for gel electrophoresis by chilling culture tubes in ice and centrifugation at 1000 g. The cell pellet was washed once in ice-cold phosphatebuffered saline (PBS) and counted in a Neubauer chamber. The cell pellet was dissolved in SDS sample buffer to obtain $2 \cdot 10^5$ cells in 50 µl and boiled for 3 minutes. Dithiothreitol (DTT) was added to 7.75 µg/ml before boiling. SDS-PAGE on 12% polyacrylamide gels and transfer to nitrocellulose membranes was done according to standard techniques. Nitrocellulose membranes were blocked in 5% dry milk, 0.05% TWEEN-20 in PBS and incubated with primary antibody, a biotinylated - anti-HA (high affinity, 3Flo, Roche), 1: 500 in blocking solution. Bound antibodies were detected with horseradish peroxidase conjugated streptavidin antibody 1: 1000 (Thermo Scientific Pierce Protein Research Products) and developed using Western Lightning Chemiluminescence Reagent (PerkinElmer Life Sciences, Boston, MA). Data collection was done in a MultiImage Light Cabinet with AlphaEaseFC software (Alpha Innotech, San Leonardo, CA) using the appropriate settings.

Immunofluorescence analysis

Cells were harvested as described above. Fixation and preparation for fluorescence microscopy was done as described previously (Marti *et al.*, 2003). Briefly, cells were washed with cold PBS and fixed with 3% formaldehyde in PBS for 40 minutes at room temperature (20°C), followed by a 5 minute incubation with 0.1 M glycine in PBS. Cells were permeabilized with 0.2% Triton X-100 in PBS for 20 minutes at room temperature and blocked overnight in 2% BSA in PBS. Incubation of all antibodies was done in 2% BSA, 0.2% Triton X-100 in PBS. Mouse monoclonal Alexa Fluor 488-conjugated anti-HA (Roche Diagnostics, Manheim, Germany; dilution 1:30) or Cy3-conjugated anti-CWP1 (Waterborne, New Orleans, LA; dilution 1:80) were incubated for 50min on ice. Washes between incubations were done with 0.5% BSA, 0.05% Triton X-100 in PBS. Labeled cells were embedded with Vectashield (Vector Laboratories, Burlingame, CA) containing the DNA intercalating agent 4_-6-diamidino-2-phenylindole (DAPI) for detection of nuclear DNA. Immunofluorescence analysis was performed on a Leica DM IRBE fluorescence microscope with a 63x oil immersion objective.

Fluorescence activated cell sorting

For quantification of cells expressing CWP1 and cyst yield, parasites were induced to encyst with either protocol A or protocol B for the required amount of time. Cysts and detached trophozoites were collected as described above, pelleted by centrifugation at 1000 x *g* for 10 minutes and washed with PBS. Fixed and permeabilized cells were fluorescently labeled for 50

min on ice using monoclonal Cy3-conjugated anti-CWP1 antibody (Waterborne, New Orleans, LA; dilution 1:80). Before FACS analysis, cells were washed twice in PBS in resuspended in ice cold PBS. Unlabeled samples were used to determine background fluorescence, and subsequently, fluorescently labeled cysts were analyzed in triplicate on a FACS Calibur flow cytometer (Becton & Dickinson, Basel, Switzerland). All samples were analyzed in parallel by IFA to assess encystation efficiency and labeling and permeabilization quality.

Bioinformatic analysis

MEME/*MAST*: Protocol for the Motif Discovery and Search algorithms are available at *meme.sdsc.edu/.*

Microarray

Microarray experiments were performed as described in (Sonda *et al.*, 2010). Briefly, for RNA extraction trophozoites were encysted if required, and harvested as described above. Total RNA isolation was performed using the RNAeasy kit (Qiagen, Stanford, CA) following the "Animal Cells Spin" protocol. Residual genomic DNA was removed with DNAseI digestion according to the manufacturer's protocol. The integrity of the RNA was assessed using a Bioanalyser (Agilent Technologies Inc., Palo Alto, CA) with "Eukaryote Total RNA Nano Series II" settings.

The array slides used here were *Giardia lamblia* microarray version 2 (provided by JCVI free of charge). They are aminosilane surface coated glass slides with ss-oligo (70 mers) printed with a Genetix arraying machine. The array contains 19230 elements and covers the whole *Giardia lamblia* WB strain genome. For further information check *Giardia lamblia* Microarray Version2 (http://pfgrc.jcvi.org/index.php/microarray/array_description/ giardia_lamblia/version2.html).

For dual channel microarray analysis, extracted total RNA was processed using the Amino Allyl MessageAmp™II a RNA Amplification Kit (Ambion, Austin, TX) and labeled with N-hydroxysuccinimidyl ester-derivatized reactive dyes Cy™3 or Cy™5, according to the manufacturer's protocol. After purification, 2 µg each of Cy3 or Cy5 labeled aRNA were denatured, added to SlideHyb™ Buffer I (Ambion), and hybridized to *G. lamblia* microarrays version 2 (TIGR) in a Tecan HybStation at the Functional Genomics Centre Zurich, Switzerland. The arrays are aminosilane surface coated glass slides with 9115 oligonucleotides (70mers) designed to cover the whole *G. lamblia* WBC6 strain genome.

Prior to hybridization, slides were hydrated and blocked with 150 µl BSA-Buffer (0.1 mg/ml BSA, 0.1 % SDS in 3 x SSC-Buffer) for 1h at 50°C. After washing, samples were injected and hybridized for 16 h at 42°C. Slides were scanned in an Agilent Scanner G2565AA, using laser lines 543 nm and 633 nm for excitation of Cy3 and Cy5, respectively. Spatial scanning resolution was 10 µm, single pass. The scanner output files were quantified using the Genespotter Software (MicroDiscovery GmbH, Berlin, Germany) with default settings and 2.5 µm radius. The median spot intensities were evaluated with the Web application MAGMA (Rehrauer *et al.*, 2007) and normalized using the print-tip-wise loess correction of the LIMMA (Linear Models for Microarray Data) software package (Smyth *et al.*, 2005). Potential gene-specific dye-effects

were estimated from self-self hybridizations. Differential expression of genes is reported as the fold-change compared with control treated samples, as well as the p-value for differential expression as estimated by the empirical Bayes model implemented in *limma*. All reactions were performed in triplicate.

Expression vector construction and transfection

Stable chromosomal integration of the described constructs was performed using the pPacV-Integ expression vector (Stefanic *et al.*, 2009) using *XbaI* and *Pac*I restriction sites. Oligonucleotides (5'-3' orientation) used in this study were:
3063p_XbaI_s, CGTCTAGATGGGTCGTCGGCTCTACAG,
3063p_HA_EcoRI_as, CGGAATTCCGCGTAGTCTGGGACATCGTATGGGTATCCGTCTCCCGTAGACAG,
3063_EcoRI_s CGGAATTCAAGCCTCCAGTAGTGCCGTC
3063_PacI_as, CGTTAATTAATTACAGATTTTTACTAGTAGGGCAG,
32657_s, GCTAGCGGATTGTGACTCTTGTATTAGGTG,
32657_HA_as, TTAATTAACTACGCGTAGTCTGGGACATCGTATGGGTAATCCATTGTTGCGTACGAG.
For stable integration of plasmid vector DNA into the chromosome the vectors were linearized at the *Swa*I restriction enzyme recognition sites adjacent to the TPI homology region (Jimenez-Garcia *et al.*, 2008). Approx. 15 μg of digested plasmid DNA was electroporated (350V, 960μF, 800Ω) into trophozoites. Integration occurs by homologous recombination and stable transgenes are selected with the antibiotic puromycin.

Acknowledgements

The *Giardia lamblia* microarrays (version 2) are based on the *Giardia* (strain WB) genome sequence (Morrison *et al.*, 2007) and were kindly provided through NIAID's Pathogen Functional Genomics Resource Center, managed and funded by the Division of Microbiology and Infectious Diseases, NIAID, NIH, DHHS and operated by the J. Craig Venter Institute. The Functional Genomics Centre Zürich, Switzerland (www.fgcz.uzh.ch) is a joint facility of the ETHZH and the University of Zürich. We are grateful to Therese Michel for technical assistance and to Dr. Weihong Qi from the FGCZ for her help with data analysis. We acknowledge in particular the "Stiftung zur Förderung der Wissenschaftlichen Forschung an der Universität Zürich" for financial support for this project. C.S. was supported by the Roche and Novartis Foundation, and "Stiftung für Forschungsförderung" University of Zurich. Research in the Hehl laboratory is supported by the Swiss National Science Foundation (grant #31003A-125389).

REFERENCES

Bernander, R., Palm, J.E., and Svard, S.G. (2001) Genome ploidy in different stages of the Giardia lamblia life cycle. *Cell Microbiol* **3**: 55-62.

Birkeland, S.R., Preheim, S.P., Davids, B.J., Cipriano, M.J., Palm, D., Reiner, D.S., Svard, S.G., Gillin, F.D., and McArthur, A.G. (2010) Transcriptome analyses of the Giardia lamblia life cycle. *Mol Biochem Parasitol*.

Boucher, S.E., and Gillin, F.D. (1990) Excystation of in vitro-derived Giardia lamblia cysts. *Infect Immun* **58**: 3516-3522.

Bulik, D.A., van Ophem, P., Manning, J.M., Shen, Z., Newburg, D.S., and Jarroll, E.L. (2000) UDP-N-acetylglucosamine pyrophosphorylase, a key enzyme in encysting Giardia, is allosterically regulated. *J Biol Chem* **275**: 14722-14728.

Davids, B.J., Reiner, D.S., Birkeland, S.R., Preheim, S.P., Cipriano, M.J., McArthur, A.G., and Gillin, F.D. (2006) A new family of giardial cysteine-rich non-VSP protein genes and a novel cyst protein. *PLoS ONE* **1**: e44.

Davis-Hayman, S.R., Hayman, J.R., and Nash, T.E. (2003) Encystation-specific regulation of the cyst wall protein 2 gene in Giardia lamblia by multiple cis-acting elements. *Int J Parasitol* **33**: 1005-1012.

Elmendorf, H.G., Singer, S.M., Pierce, J., Cowan, J., and Nash, T.E. (2001) Initiator and upstream elements in the alpha2-tubulin promoter of Giardia lamblia. *Mol Biochem Parasitol* **113**: 157-169.

Franzen, O., Jerlstrom-Hultqvist, J., Castro, E., Sherwood, E., Ankarklev, J., Reiner, D.S., Palm, D., Andersson, J.O., Andersson, B., and Svard, S.G. (2009) Draft genome sequencing of giardia intestinalis assemblage B isolate GS: is human giardiasis caused by two different species? *PLoS Pathog* **5**: e1000560.

Gerwig, G.J., van Kuik, J.A., Leeflang, B.R., Kamerling, J.P., Vliegenthart, J.F., Karr, C.D., and Jarroll, E.L. (2002) The Giardia intestinalis filamentous cyst wall contains a novel beta(1-3)-N-acetyl-D-galactosamine polymer: a structural and conformational study. *Glycobiology* **12**: 499-505.

Gillin, F.D., Boucher, S.E., Rossi, S.S., and Reiner, D.S. (1989) Giardia lamblia: the roles of bile, lactic acid, and pH in the completion of the life cycle in vitro. *Exp Parasitol* **69**: 164-174.

Hehl, A.B., Marti, M., and Kohler, P. (2000) Stage-specific expression and targeting of cyst wall protein-green fluorescent protein chimeras in Giardia. *Mol Biol Cell* **11**: 1789-1800.

Huang, Y.C., Su, L.H., Lee, G.A., Chiu, P.W., Cho, C.C., Wu, J.Y., and Sun, C.H. (2008) Regulation of cyst wall protein promoters by Myb2 in Giardia lamblia. *J Biol Chem* **283**: 31021-31029.

Jarroll, E.L., Manning, P., Lindmark, D.G., Coggins, J.R., and Erlandsen, S.L. (1989) Giardia cyst wall-specific carbohydrate: evidence for the presence of galactosamine. *Mol Biochem Parasitol* **32**: 121-131.

Jimenez-Garcia, L.F., Zavala, G., Chavez-Munguia, B., Ramos-Godinez Mdel, P., Lopez-Velazquez, G., Segura-Valdez Mde, L., Montanez, C., Hehl, A.B., Arguello-Garcia, R., and Ortega-Pierres, G. (2008) Identification of nucleoli in the early branching protist Giardia duodenalis. *Int J Parasitol* **38**: 1297-1304.

Kane, A.V., Ward, H.D., Keusch, G.T., and Pereira, M.E. (1991) In vitro encystation of Giardia lamblia: large-scale production of in vitro cysts and strain and clone differences in encystation efficiency. *J Parasitol* **77**: 974-981.

Kim, J., Bae, S.S., Sung, M.H., Lee, K.H., and Park, S.J. (2009) Comparative proteomic analysis of trophozoites versus cysts of Giardia lamblia. *Parasitol Res* **104**: 475-479.

Knodler, L.A., Sekyere, E.O., Stewart, T.S., Schofield, P.J., and Edwards, M.R. (1998) Cloning and expression of a prokaryotic enzyme, arginine deiminase, from a primitive eukaryote Giardia intestinalis. *J Biol Chem* **273**: 4470-4477.

Konrad, C., Spycher, C., and Hehl, A.B. (2010) Selective condensation drives partitioning and sequential secretion of cyst wall proteins in differentiating Giardia lamblia. *PLoS Pathog* **6**: e1000835.

Lauwaet, T., Davids, B.J., Reiner, D.S., and Gillin, F.D. (2007) Encystation of Giardia lamblia: a model for other parasites. *Curr Opin Microbiol* **10**: 554-559.

Lopez, A.B., Sener, K., Jarroll, E.L., and van Keulen, H. (2003) Transcription regulation is demonstrated for five key enzymes in Giardia intestinalis cyst wall polysaccharide biosynthesis. *Mol Biochem Parasitol* **128**: 51-57.

Lujan, H.D., Mowatt, M.R., Conrad, J.T., Bowers, B., and Nash, T.E. (1995) Identification of a novel Giardia lamblia cyst wall protein with leucine-rich repeats. Implications for secretory granule formation and protein assembly into the cyst wall. *J Biol Chem* **270**: 29307-29313.

Lujan, H.D., Mowatt, M.R., Byrd, L.G., and Nash, T.E. (1996a) Cholesterol starvation induces differentiation of the intestinal parasite Giardia lamblia. *Proc Natl Acad Sci U S A* **93**: 7628-7633.

Lujan, H.D., Mowatt, M.R., Conrad, J.T., and Nash, T.E. (1996b) Increased expression of the molecular chaperone BiP/GRP78 during the differentiation of a primitive eukaryote. *Biol Cell* **86**: 11-18.

Marti, M., and Hehl, A.B. (2003) Encystation-specific vesicles in Giardia: a primordial Golgi or just another secretory compartment? *Trends Parasitol* **19**: 440-446.

Marti, M., Regos, A., Li, Y., Schraner, E.M., Wild, P., Muller, N., Knopf, L.G., and Hehl, A.B. (2003) An ancestral secretory apparatus in the protozoan parasite Giardia intestinalis. *J Biol Chem* **278**: 24837-24848.

Morrison, H.G., McArthur, A.G., Gillin, F.D., Aley, S.B., Adam, R.D., Olsen, G.J., Best, A.A., Cande, W.Z., Chen, F., Cipriano, M.J., Davids, B.J., Dawson, S.C., Elmendorf, H.G., Hehl, A.B., Holder, M.E., Huse, S.M., Kim, U.U., Lasek-Nesselquist, E., Manning, G., Nigam, A., Nixon, J.E., Palm, D., Passamaneck, N.E., Prabhu, A., Reich, C.I., Reiner, D.S., Samuelson, J., Svard, S.G., and Sogin, M.L. (2007) Genomic minimalism in the early diverging intestinal parasite Giardia lamblia. *Science* **317**: 1921-1926.

Mowatt, M.R., Lujan, H.D., Cotten, D.B., Bowers, B., Yee, J., Nash, T.E., and Stibbs, H.H. (1995) Developmentally regulated expression of a Giardia lamblia cyst wall protein gene. *Mol Microbiol* **15**: 955-963.

Nicolaides, N.C., Gualdi, R., Casadevall, C., Manzella, L., and Calabretta, B. (1991) Positive autoregulation of c-myb expression via Myb binding sites in the 5' flanking region of the human c-myb gene. *Mol Cell Biol* **11**: 6166-6176.

Palm, D., Weiland, M., McArthur, A.G., Winiecka-Krusnell, J., Cipriano, M.J., Birkeland, S.R., Pacocha, S.E., Davids, B., Gillin, F., Linder, E., and Svard, S. (2005) Developmental changes in the adhesive disk during Giardia differentiation. *Mol Biochem Parasitol* **141**: 199-207.

Pan, Y.J., Cho, C.C., Kao, Y.Y., and Sun, C.H. (2009) A Novel WRKY-like Protein Involved in Transcriptional Activation of Cyst Wall Protein Genes in Giardia lamblia. *J Biol Chem* **284**: 17975-17988.

Que, X., Svard, S.G., Meng, T.C., Hetsko, M.L., Aley, S.B., and Gillin, F.D. (1996) Developmentally regulated transcripts and evidence of differential mRNA processing in Giardia lamblia. *Mol Biochem Parasitol* **81**: 101-110.

Rehrauer, H., Zoller, S., and Schlapbach, R. (2007) MAGMA: analysis of two-channel microarrays made easy. *Nucleic Acids Res* **35**: W86-90.

Ringqvist, E., Palm, J.E., Skarin, H., Hehl, A.B., Weiland, M., Davids, B.J., Reiner, D.S., Griffiths, W.J., Eckmann, L., Gillin, F.D., and Svard, S.G. (2008) Release of metabolic enzymes by Giardia in response to interaction with intestinal epithelial cells. *Mol Biochem Parasitol* **159**: 85-91.

Smyth, G.K., Michaud, J., and Scott, H.S. (2005) Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**: 2067-2075.

Sonda, S., Morf, L., Bottova, I., Baetschmann, H., Rehrauer, H., Caflisch, A., Hakimi, M.A., and Hehl, A.B. (2010) Epigenetic mechanisms regulate stage differentiation in the minimized protozoan Giardia lamblia. *Mol Microbiol*.

Stefanic, S., Palm, D., Svard, S.G., and Hehl, A.B. (2006) Organelle proteomics reveals cargo maturation mechanisms associated with Golgi-like encystation vesicles in the early-diverged protozoan Giardia lamblia. *J Biol Chem* **281**: 7595-7604.

Stefanic, S., Morf, L., Kulangara, C., Regos, A., Sonda, S., Schraner, E., Spycher, C., Wild, P., and Hehl, A.B. (2009) Neogenesis and maturation of transient Golgi-like cisternae in a simple eukaryote. *J Cell Sci* **122**: 2846-2856.

Su, L.H., Lee, G.A., Huang, Y.C., Chen, Y.H., and Sun, C.H. (2007) Neomycin and puromycin affect gene expression in Giardia lamblia stable transfection. *Mol Biochem Parasitol* **156**: 124-135.

Sun, C.H., Palm, D., McArthur, A.G., Svard, S.G., and Gillin, F.D. (2002) A novel Myb-related protein involved in transcriptional activation of encystation genes in Giardia lamblia. *Mol Microbiol* **46**: 971-984.

Sun, C.H., McCaffery, J.M., Reiner, D.S., and Gillin, F.D. (2003) Mining the Giardia lamblia genome for new cyst wall proteins. *J Biol Chem* **278**: 21701-21708.

Touz, M.C., Ropolo, A.S., Rivero, M.R., Vranych, C.V., Conrad, J.T., Svard, S.G., and Nash, T.E. (2008) Arginine deiminase has multiple regulatory roles in the biology of Giardia lamblia. *J Cell Sci* **121**: 2930-2938.

Yang, H., Chung, H.J., Yong, T., Lee, B.H., and Park, S. (2003) Identification of an encystation-specific transcription factor, Myb protein in Giardia lamblia. *Mol Biochem Parasitol* **128**: 167-174.

FIGURE LEGENDS

Figure 1 Whole genome microarray analysis of encystation induced by Protocol A. A) List of statistically upregulated genes (cut-off > 2 fold signal increase) at 45 min., 3 h, and 7 h post induction. B) Graphical representation (Venn diagram) of overlaps in the three datasets. Note that no gene is present in all three datasets.

Figure 2 Comparison of the kinetics of the two encystation protocols. A) Flow cytometry analysis of 10^5 encysting cells each induced by protocol A (left column) or protocol B (right column). Chemically fixed and detergent permeabilized cells were labeled with an anti-CWP1 antibody coupled to Cy3. Gates for trophozoites (non-labeled, M1) and cells containing CWP1 in ESVs (M2) at four different time-points post induction are indicated together with calculated percentages of encysting cells. Representative IFA images of labeled cells (CWP1, red; DAPI, blue) are shown for the time points 4 h and 8 h p.i. Scale bars, 3µm. B) Quantitation of encystation at different conditions by microscopy (representative experiment). The percentage of encysting cells containing ESVs and cysts labeled with an anti-CWP1 antibody coupled to Cy3 at 7h (left column) and 24 h (right column) in each field of view was determined.

Figure 3 Comparative analysis of gene regulation induced by protocols A and B. A) The tabular depiction of ORFs with a statistically significant change of signal of >2.0 in the two datasets shows 13 common ORFs. B) Categorization of ORFs in the datasets according to function shows that the protocol B dataset contains significantly more hypothetical ORFs. C) Graphical representation (Venn diagram) of the two datasets. The intersecting positions are designated bona fide encystation ORFs.

Figure 4 Representation and distribution of the MBS in the promoter regions of regulated genes. A) MEME analysis and sequence logo of the MBS generated with the 13 encystation promoters as a training set. A single change to the alternative nucleotide at positions 1, 2, 5, and 6 is tolerated. B) The MEME analysis shows also the position of MBS in the sense (+) and antisense (-) orientation in promoters. Position 1 on the scale indicates the most proximal nucleotide next to the start codon of the ORF. C) Tabular depiction of the final set of encystation ORFs including five additional positions (darker shading).

Figure 5 Expression of predicted secreted proteins as tagged variants in encysting cells. A, B) Expression of HA-tagged ORFs 3063 and 32657 is controlled by the respective endogenous promoters. Western analysis of protein extracts separated on SDS-PAGE shows strong induction of the proteins during encystation (0 – 24 h p.i.). IFA analysis of chemically fixed and detergent-permeabilized cells. Representative cells at 7 h (encysting trophozoites, top rows) and 24 h p.i. (cyst, bottom rows) are shown. For quantitative analysis of colocalization see Figure S5. CWP1, Red; 3063-HA, green; DAPI, blue. Scale bars, 3 µm.

SUPPORTING MATERIAL

Figure S1 List of the five additional encystation genes and the criteria for their inclusion in the final set.

Figure S2 Graphical representations of MEME and MAST analysis results of promoters upregulated at early time points. A) An A/T-rich sequence is the most frequently identified motif in the 45 min. dataset. B) In the 3 h dataset the MBS motif emerges in 9 of 16 ORFs.

Figure S3 A) Correlation of microarray data with predicted stage-specific gene expression from SAGE data in GiardiaDB. Encystation genes detected with protocol A correlate better than those detected with protocol B. B) Microarray analysis of encysting cells at 7 h p.i treated with the HDAC inhibitor FR235222 (Sonda *et al.*, 2010). Expression changes in the set of 18 encystation genes are listed. Note that the effect is strongest for highly induced genes.

Figure S4 List of down-regulated genes in all datasets (deprecated genes have been deleted). The correlation with SAGE data on GiardiaDB is indicated: ±, partial correlation; -, no correlation; +, good correlation.

Figure S5 Confocal microscopy: analysis of the subcellular distribution of HA epitope-tagged products encoded by ORFs 3063 and 32657 (see also Figure 5). A) Signal overlap with the ER resident protein PDI1. Each row shows a single layer from a stack of optical sections. The scatter plot is a graphical representation of the degree of signal overlap in the three-dimensional voxel space of the entire stack. The degree of overlap is indicative of a localization in the endoplasmatic reticulum. B) The tagged proteins are not incorporated in the cyst wall. Although 32657-HA localizes in the periphery of cysts the lack of signal overlap in the three-dimensional analysis is not consistent with trafficking to the cyst wall. BF: bright field image. Scale bars: 5µm.

Figure1

A

Gene ID	GiardiaDB annotation	Fold change at 45 min	3 h	7 h
8245	Glucosamine-6-phosphate deaminase			5.3
10552	Hypothetical protein			3.7
14626	Oxidoreductase			3.3
92729	Fatty acid elongase 1			2.8
88581	Synaptic glycoprotein SC2			2.7
12082	Hypothetical protein			2.5
7134	Hypothetical protein			2.4
103785	Hypothetical protein			2.4
14759	6-phosphogluconate dehydrogenase			2.3
16069	Phosphoacetylglucosamine mutase			2.3
7388	Hypothetical protein			2.2
21924	Kinase, NEK			2.2
3063	Hypothetical protein			2.2
9115	Glucose-6-phosphate isomerase			2.1
4846	Protein 21.1			2.1
106496	Hypothetical protein			2.1
24412	Protein 21.1			2.1
102813	Protein 21.1			2.1
7139	Hypothetical protein			2.0
5435	Cyst wall protein 2		15.9	24.7
5638	Cyst wall protein 1		12.4	19.6
9046	Sugar transport family protein		6.1	9.1
8722	Myb 1-like protein		3.4	4.2
32657	Hypothetical protein		2.4	8.5
112432	High cysteine membrane protein Group 5		2.4	2.2
2421	Cyst wall protein 3		2.2	4.0
7260	Aldose reductase		2.2	2.1
137701	Kinase, NEK		2.1	2.8
88814	Protein kinase		2.0	2.2
3643	70 kDa peptidylprolyl isomerase		2.9	
98054	Heat shock protein HSP 90-alpha		2.7	
88765	Hsp70 (cytosolic)		2.4	
10429	Wos2 protein		2.0	
17012	Hypothetical protein	2.7	2.0	
10570	FKBP-type peptidyl-prolyl cis-trans isomerase	2.4	2.2	
9068	Hypothetical protein	2.8		
17012	Hypothetical protein	2.5		
3595	Endonuclease III	2.5		
16519	AstB/chuR-related protein	2.4		
17587	CTP synthase	2.4		
16602	Hypothetical protein	2.3		
10341	Hypothetical protein	2.3		
16498	Nucleolar GTPase	2.3		
13215	Kinase, NEK	2.2		
10568	Hypothetical protein	2.2		
103807	Protein 21.1	2.1		
16887	Probable ATP-dependent RNA helicase HAS1	2.1		
6242	Translationally controlled tumor protein homolog	2.1		
94653	Periodic tryptophan protein 2 homolog	2.0		
15048	ATP-dependent RNA helicase-like protein	2.0		
6436	Guanine phosphoribosyltransferase	2.0		

B

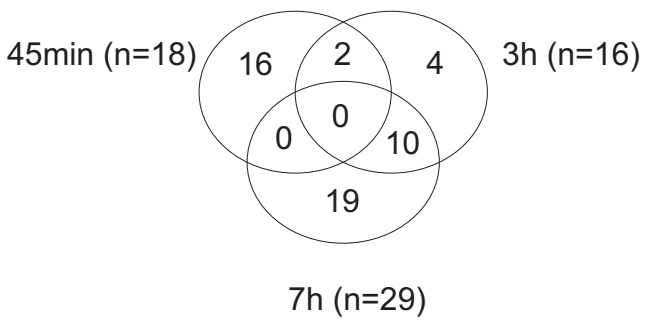


Figure 2

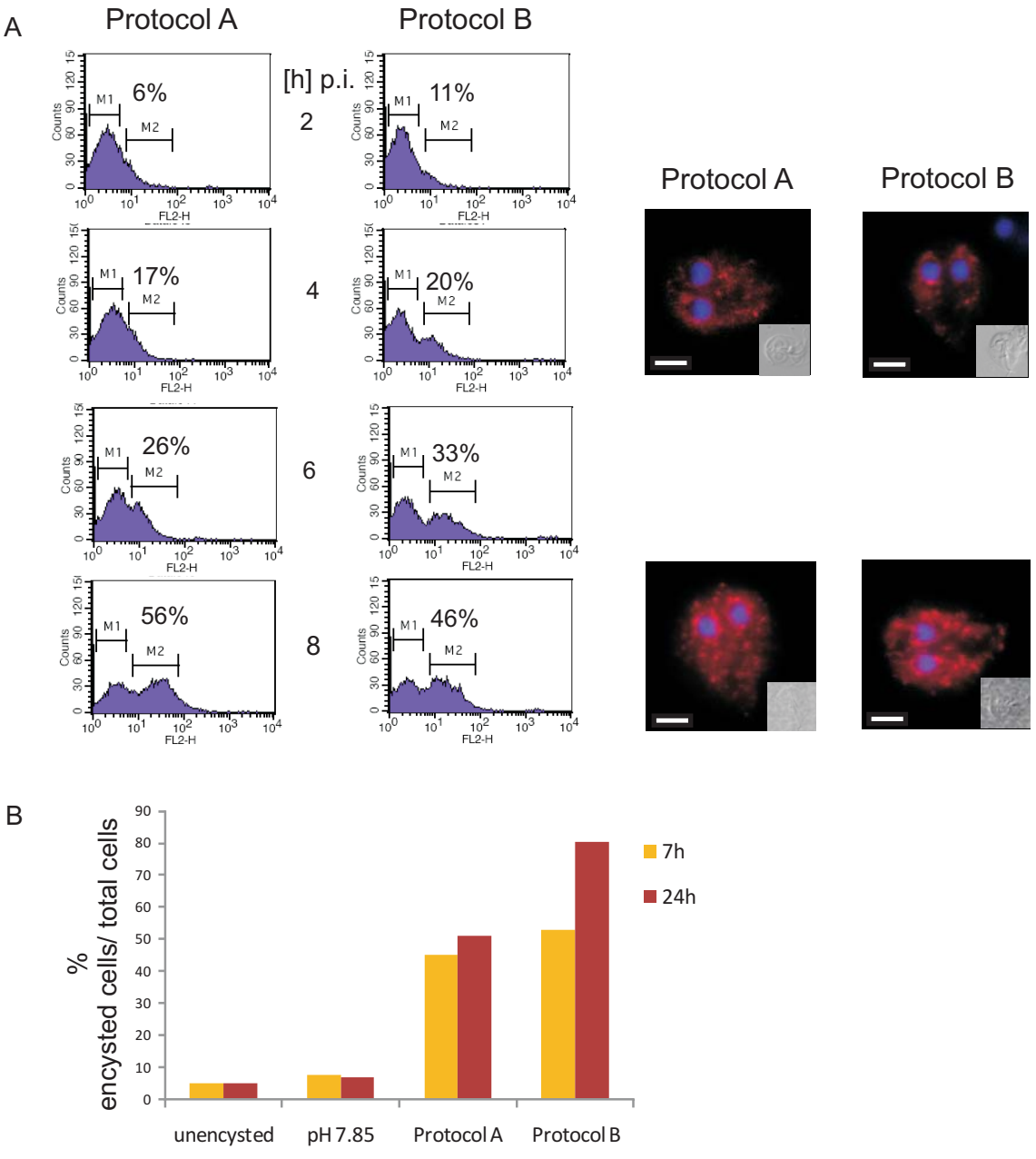
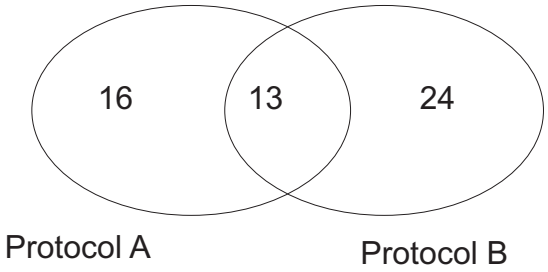

A
Protocol A
Protocol B
[h] p.i.
2
4
6
8
6%
11%
17%
20%
26%
33%
56%
46%
M1
M2
Counts
FL2-H
Protocol A
Protocol B
B
% encysted cells/ total cells
7h
24h
unencysted
pH 7.85
Protocol A
Protocol B

Figure 3

A

Gene ID	GiardiaDB annotation	Protocol A Fold change	Protocol B Fold change
5435	Cyst wall protein 2	24.7	15.1
5638	Cyst wall protein 1	19.6	15.6
9046	Sugar transport family protein	9.1	
32657	Hypothetical protein	8.5	11.9
8245	Glucosamine-6-phosphate deaminase	5.3	8.3
8722	Myb 1-like protein (Giardia Myb)	4.2	3.1
2421	Cyst wall protein 3	4.0	4.2
10552	Hypothetical protein	3.7	3.5
14626	Oxidoreductase	3.3	5.8
137701	Kinase, NEK	2.8	
92729	Fatty acid elongase 1	2.8	6.8
88581	Synaptic glycoprotein SC2	2.7	2.4
12082	Hypothetical protein	2.5	3.5
7134	Hypothetical protein	2.4	
103785	Hypothetical protein	2.4	
14759	6-phosphogluconate dehydrogenase	2.3	
16069	phosphoacetylglucosamine mutase	2.3	
7388	Hypothetical protein	2.2	
112432	High cysteine membrane protein Group 5 (HCMp)	2.2	
88814	Protein kinase	2.2	
21924	Kinase, NEK	2.2	
3063	Hypothetical protein	2.2	3.9
9115	Glucose-6-phosphate isomerase	2.1	
4846	Protein 21.1	2.1	
106496	Hypothetical protein	2.1	
7260	Aldose reductase	2.1	
24412	Protein 21.1	2.1	
102813	Protein 21.1	2.1	3.1
7139	Hypothetical protein	2.0	
15250	High cysteine membrane protein Group 6 (HCMp)		8.8
9620	High cysteine membrane protein Group 2 (HCMp)		7.4
14259	Glucose 6-phosphate N-acetyltransferase		5.3
93488	Hypothetical protein		4.3
5800	Hypothetical protein		4.2
27028	Hypothetical protein		3.8
8987	Hypothetical protein		3.6
89849	Hypothetical protein		3.5
5810	Hypothetical protein		3.0
9355	Hypothetical protein		3.0
35999	Hypothetical protein		2.9
32419	Hypothetical protein		2.8
37010	Hypothetical protein		2.8
7982	UDP-glucose 4-epimerase		2.7
14247	TM efflux protein		2.7
17043	Glyceraldehyde 3-phosphate dehydrogenase		2.6
11149	Hypothetical protein		2.5
8652	Hypothetical protein		2.4
23015	Serine palmitoyltransferase 1		2.3
101699	Protein 21.1		2.2
10425	Hypothetical protein		2.2
137680	Cathepsin L-like protease		2.1
28112	Hypothetical protein		2.1
6492	Hypothetical protein		2.1

B



C

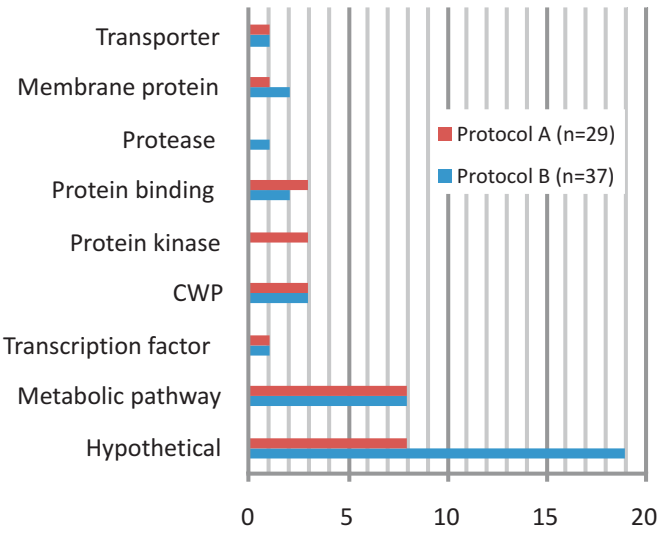
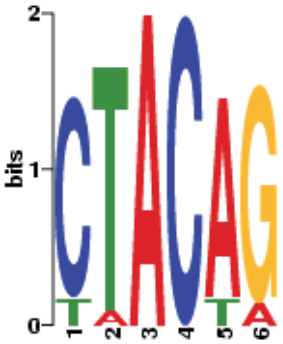


Figure 4

A



B

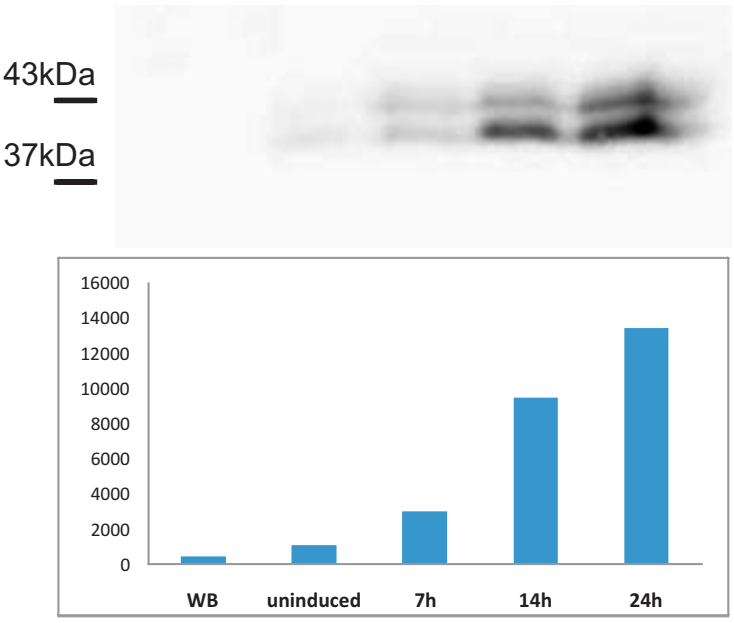
GiardiaDB ID	Lowest p-value	Motif position and orientation
GL50803_2421	2.12e-04	-1 +1 +1
GL50803_8722	2.12e-04	+1
GL50803_5638	2.12e-04	+1 +1 +1 +1
GL50803_5435	2.12e-04	-1 -1 +1 +1
GL50803_14626	2.12e-04	-1 -1
GL50803_8245	2.12e-04	-1 -1 +1 -1
GL50803_32657	2.12e-04	-1 +1
GL50803_102813	2.12e-04	-1 +1 +1
GL50803_3063	2.12e-04	+1 -1 -1
GL50803_12082	2.12e-04	+1 +1
GL50803_88581	2.12e-04	-1 +1 -1
GL50803_92729	2.12e-04	+1
GL50803_10552	2.12e-04	-1
SCALE		1 25 50 75 100

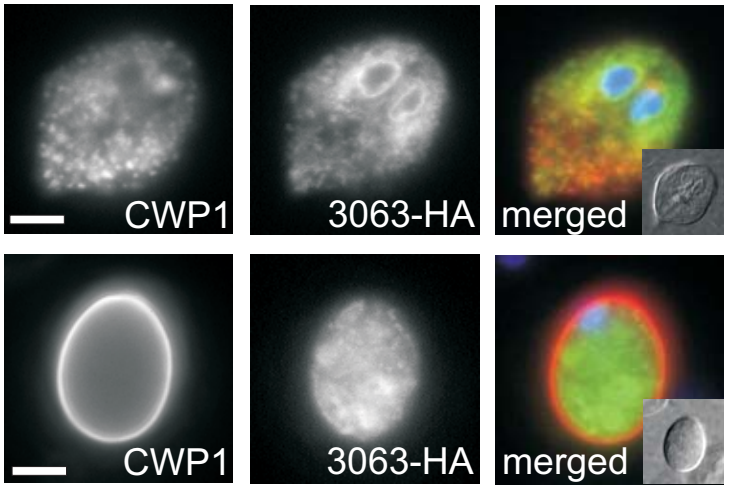
C

GiardiaDB ID	GiardiaDB annotation
GL50803_5638	Cyst wall protein 1
GL50803_5435	Cyst wall protein 2
GL50803_2421	Cyst wall protein 3
GL50803_8722	Myb 1-like protein
GL50803_92729	Fatty acid elongase 1
GL50803_88581	Synaptic glycoprotein SC2
GL50803_8245	Glucosamine-6-phosphate deaminase
GL50803_10552	Hypothetical protein
GL50803_3063	Hypothetical protein
GL50803_32657	Hypothetical protein
GL50803_14626	Oxidoreductase
GL50803_12082	Hypothetical protein
GL50803_102813	Protein 21.1
GL50803_9046	Sugar transport family protein
GL50803_9115	Glucosamine-6-phosphate isomerase
GL50803_16069	phosphoacetylglucosamine mutase
GL50803_27028	Hypothetical protein
GL50803_21924	Hypothetical protein

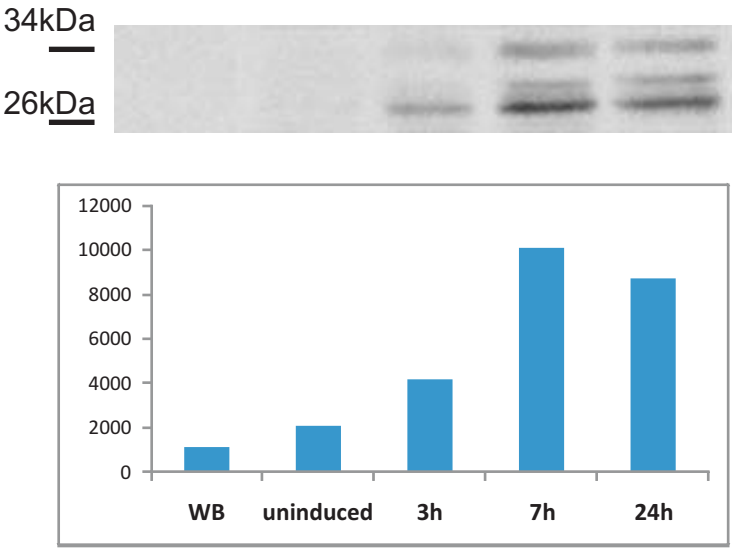
Figure 5

A
Protein 3063





B
Protein 32657



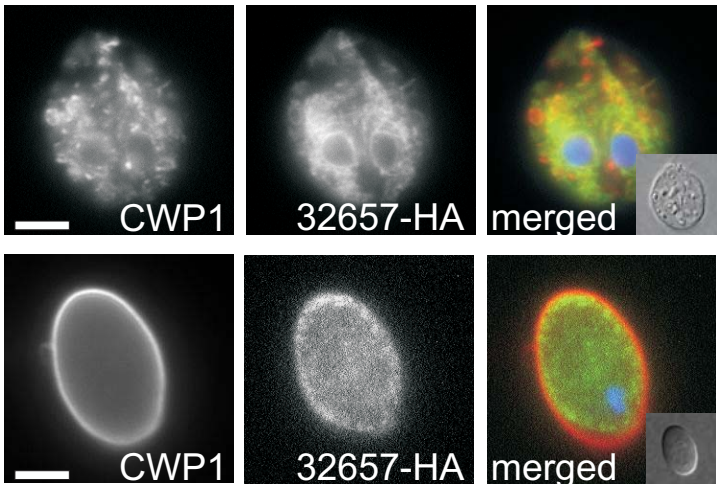


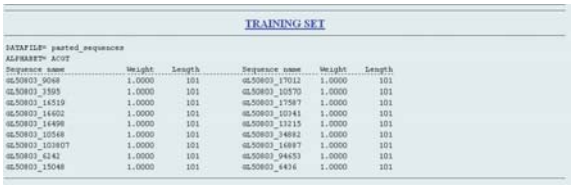
Figure S1

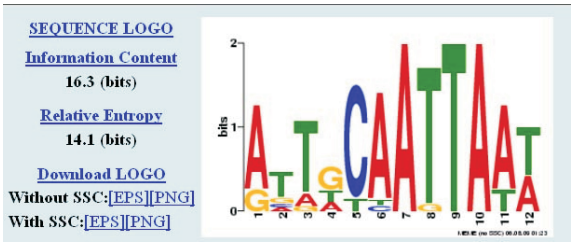
GiardiaDB ID	GiardiaDB annotation	Fold change in protocol A	Fold change in protocol B	Additional criteria	MBS
GL50803_9046	**Sugar transport family protein**	**9.14**	**nd: technical issue**	**up at 3 h**	**MAST**
GL50803_9115	**Glucosamine-6-phosphate isomerase**	**2.14**	**nd: technical issue**	**Sun *et al.* 2002**	**MAST**
GL50803_16069	**Phosphoacetylglucosamine mutase**	**2.26**	**nd: technical issue**	**Lopez *et al* . 2003**	**MAST**
GL50803_27082	**Hypothetical protein**	**1.99**	**3.79**		**MAST**
GL50803_21924	**Hypothetical protein**	**2.16**	**1.92**		**MAST**

Figure S2

A

45min
Protocol A

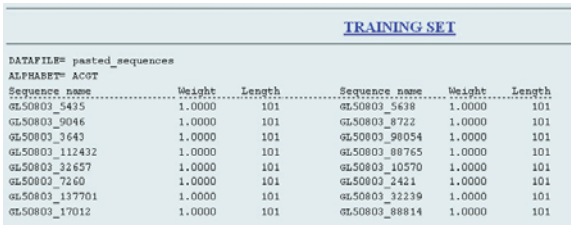




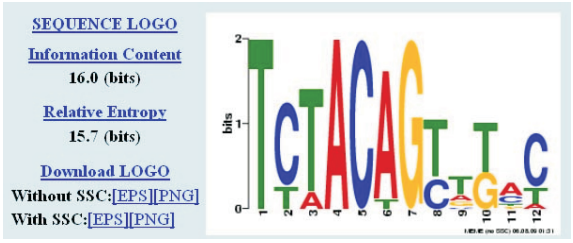
Name	Lowest p-value	Motifs
GL50803_10568	5.42e-07	+1
GL50803_6242	8.85e-07	-1
GL50803_94653	1.66e-06	-1
GL50803_16519	1.66e-06	+1
GL50803_17012	1.66e-06	-1
GL50803_16887	2.95e-06	+1 -1 -1
GL50803_6436	1.29e-05	-1
GL50803_15048	1.29e-05	+1
GL50803_16602	1.45e-05	-1
GL50803_17587	1.68e-05	-1
GL50803_3595	1.68e-05	+1
GL50803_13215	3.47e-05	+1
GL50803_10570	3.47e-05	-1 -1
GL50803_9068	3.78e-05	+1
GL50803_16498	5.53e-05	+1
GL50803_10341	5.87e-05	-1
SCALE		1 25 50 75 100

B

3h
Protocol A

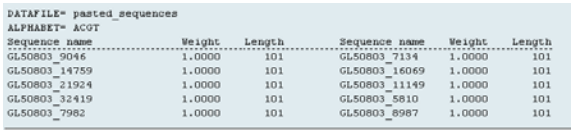


Sequence name	Weight	Length	Sequence name	Weight	Length
GL50803_5435	1.0000	101	GL50803_5638	1.0000	101
GL50803_9046	1.0000	101	GL50803_8722	1.0000	101
GL50803_3643	1.0000	101	GL50803_98054	1.0000	101
GL50803_112432	1.0000	101	GL50803_88765	1.0000	101
GL50803_32657	1.0000	101	GL50803_10570	1.0000	101
GL50803_7260	1.0000	101	GL50803_2421	1.0000	101
GL50803_137701	1.0000	101	GL50803_32239	1.0000	101
GL50803_17012	1.0000	101	GL50803_88814	1.0000	101

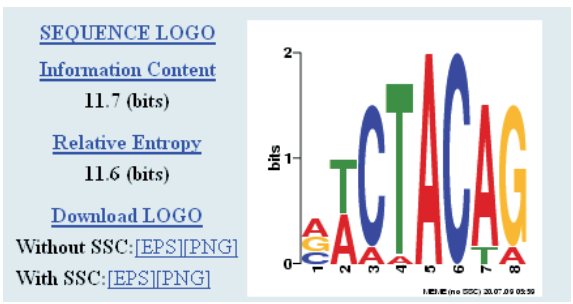


Name	Lowest p-value	Motifs
GL50803_9046	1.71e-07	+1 +1
GL50803_5638	1.71e-07	+1 +1 +1
GL50803_8722	4.11e-07	+1
GL50803_2421	1.16e-06	+1
GL50803_137701	1.63e-06	-1
GL50803_32657	6.14e-06	-1
GL50803_112432	1.14e-05	-1
GL50803_7260	1.36e-05	-1
GL50803_5435	1.99e-05	-1 -1 +1
SCALE		1 25 50 75 100

With promotor element but not in intersection



Sequence name	Weight	Length	Sequence name	Weight	Length
GL50803_9046	1.0000	101	GL50803_7134	1.0000	101
GL50803_14759	1.0000	101	GL50803_16069	1.0000	101
GL50803_21924	1.0000	101	GL50803_11149	1.0000	101
GL50803_32419	1.0000	101	GL50803_5810	1.0000	101
GL50803_7982	1.0000	101	GL50803_8987	1.0000	101



Name	Lowest p-value	Motifs
GL50803_21924	1.21e-05	+1 -1 -1
GL50803_14759	1.21e-05	+1 -1
GL50803_7982	2.43e-05	+1 +1
GL50803_7134	2.43e-05	+1
GL50803_9046	5.60e-05	+1 +1
GL50803_5810	8.78e-05	-1 -1 +1
GL50803_32419	8.78e-05	+1
GL50803_16069	8.78e-05	+1 +1 +1
GL50803_8987	1.44e-04	+1 -1
SCALE		1 25 50 75 100

Figure S3

A

Protocols A/B (7 h) vs. SAGE data (timepoints 4 h, 12 h)

Correlation with SAGE data:

Protocol	no correlation	partial	good correlation
A (n=29)	10	2	17
B (n=37)	19	1	17

Revised microarray datasets:

Dataset	no correlation	partial	good correlation
Overlapping (n=18)	4	-	14
Non-overlapping (n=35)	22	3	10

B

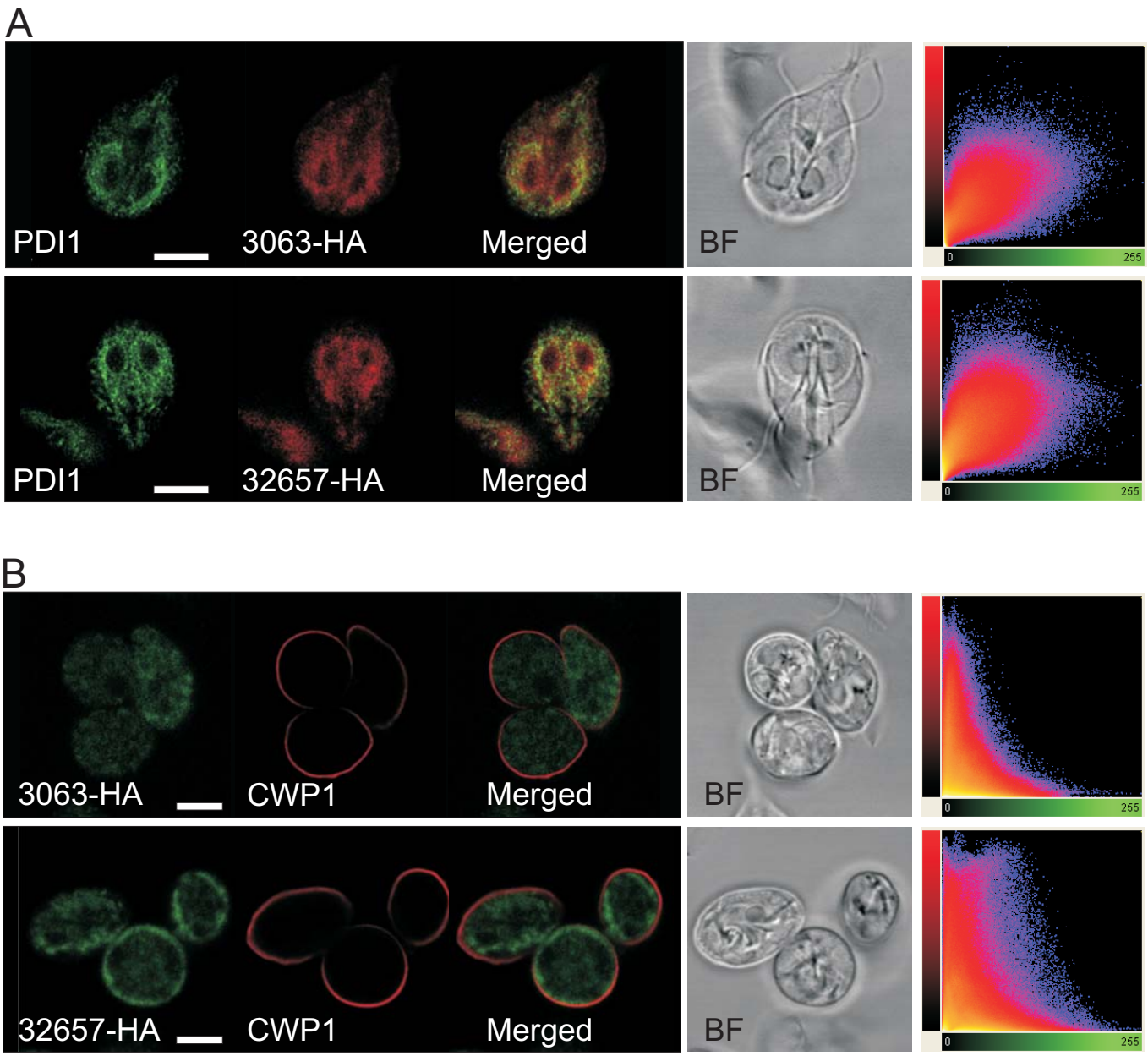
GiardiaDB ID	GiardiaDB annotation	Fold change
GL50803_5435	Cyst wall protein 2	-4
GL50803_2421	Cyst wall protein 3	-3.2
GL50803_9046	Sugar transport family protein	-3.2
GL50803_5638	Cyst wall protein 1	-3.0
GL50803_32657	Hypothetical Protein	-2.8
GL50803_8722	Myb 1-like protein	-2.6
GL50803_8245	Glucosamine-6-phosphate deaminase	-2.6
GL50803_10552	Hypothetical Protein	-2.3
GL50803_14626	Oxidoreductase	-2.1
GL50803_12082	Hypothetical protein	-1.9
GL50803_92729	Fatty acid elongase 1	-1.7
GL50803_88581	Synaptic glycoprotein SC2	-1.7
GL50803_102813	Protein 21.1	-1.6
GL50803_16069	Phosphoacetylglucosamine mutase	-1.6
GL50803_3063	Hypothetical Protein	-1.5
GL50803_27028	Hypothetical Protein	-1.5
GL50803_21924	Hypothetical Protein	1.1
GL50803_9115	Glucosamine-6-phosphate isomerase	1.1

Figure S4

Downregulated genes

		Protocol A			Protocol B	
		Fold change at				Correlation with SAGE
Gene ID	**GiardiaDB annotation**	**45 min**	**3 h**	**7 h**	**7h**	
100635	Hypothetical protein	-2.0	-2.1			±
9548	Cathepsin L precursor	-2.0	-2.7	-2.4		
7865	L-asparaginase	-2.1				
8044	Seven TM protein 1	-2.5				+
11448	Phosphomannomutase-2	-2.6				+
5908	Hypothetical protein	-2.8	-2.4			+
112103	Arginine deiminase	-2.8	-2.0			+
7103	Kinase, NEK	-3.1	-2.2			-
114777	Hypothetical protein		-2.8	-2.2		-
15125	Hypothetical protein			-2.2		-
7718	DNA-damage inducible protein				-2.0	-
88888	Hypothetical protein				-2.0	-

Figure S5


A
PDI1
3063-HA
Merged
BF
0
255
PDI1
32657-HA
Merged
BF
0
255
B
3063-HA
CWP1
Merged
BF
0
255
32657-HA
CWP1
Merged
BF
0
255