



**HAL**  
open science

## **Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: Critical experimental parameters.**

Celine Cosseau, Abdelhalim Azzi, Kristina Smith, Michael Freitag, Guillaume Mitta, Christoph Grunau

### **► To cite this version:**

Celine Cosseau, Abdelhalim Azzi, Kristina Smith, Michael Freitag, Guillaume Mitta, et al.. Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: Critical experimental parameters.. *Molecular and Biochemical Parasitology*, 2009, 166 (1), pp.70-6. 10.1016/j.molbiopara.2009.02.015 . halsde-00459175

**HAL Id: halsde-00459175**

**<https://hal.science/halsde-00459175>**

Submitted on 1 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of *Schistosoma mansoni*: critical experimental parameters**

Running title:

**N-ChIP and ChIP-Seq of *Schistosoma mansoni***

Céline Cosseau<sup>a</sup>, Abdelhalim Azzi<sup>a</sup>, Kristina Smith<sup>b</sup>, Michael Freitag<sup>b</sup>, Guillaume Mitta<sup>a</sup>,  
Christoph Grunau<sup>a\*</sup>

<sup>a</sup> *Parasitologie Fonctionnelle et Evolutive,*

*UMR 5244, CNRS EPHE Université de Perpignan,*

*52 Ave Paul Alduy, 66860 Perpignan Cedex, France*

<sup>b</sup> *Center for Genome Research and Biocomputing and*

*Department of Biochemistry and Biophysics*

*Oregon State University*

*Corvallis, OR 97331-7305, U.S.A.*

\*Corresponding author: C. Grunau

*UMR 5244, CNRS EPHE Université de Perpignan,*

*52 Ave Paul Alduy, 66860 Perpignan Cedex, France*

Phone: +33 (0)4 68 66 21 80

Fax: +33 (0) 4 68 66 22 81

e-mail: christoph.grunau@univ-perp.fr

**Abstract:**

Histone modifications are important epigenetic marks that influence chromatin structure and consequently play a role in the control of eukaryotic transcription. Several histone modifying enzymes have been characterized in *Schistosoma mansoni* and it has been suggested that the regulation of gene transcription in schistosomes may require the action of these enzymes. However, the influence of chromatin structure on gene transcription in schistosomes has never been investigated. Chromatin immunoprecipitation (ChIP) is the technique of choice to study the relationship between histone modifications and gene expression. Although this technique has been widely used with cultured cells from model organisms and with many unicellular organisms, it remains challenging to apply this technique to non-conventional organisms that undergo complex life-cycles. In this work, we describe a native ChIP procedure that is applicable to all the stages of the *S. mansoni* life cycle and does not require expensive equipment. Immunoprecipitated DNA was analysed on a whole-genome scale using massively parallel sequencing (ChIP-Sequencing or ChIP-Seq). We show that ChIP-Seq and conventional quantitative PCR deliver comparable results for a life-cycle regulated locus, *smRHO*, that encodes a guanine-protein coupled receptor. This is the first time that the ChIP-Seq procedure has been applied to a parasite. This technique opens new ways for analyzing epigenetic mechanisms in *S. mansoni* at a whole-genome scale and on the level of individual loci.

**Introduction**

In eukaryotes, genotype and epigenotypes manifest themselves as a complex nuclear structure called chromatin. This nucleoprotein structure contains histone and non-histone proteins that interact with the genomic DNA. Chromatin exists either as a relaxed structure that is

permissive to gene expression and called euchromatin, or as a condensed structure that is typically silent and called heterochromatin [1]. The genotype refers to the DNA sequence of an organism. “Epigenotype” refers to the sum of chemical modifications of DNA (e.g., DNA methylation) and of DNA-associated proteins (e.g., histones) that can affect chromatin structure. Other mediators of epigenetic information are non-coding RNA and the location of genes in the nucleus [2, 3]. Posttranslational modifications of histones, for example methylation, acetylation and phosphorylation, are widely studied and have been shown to play a key role in chromatin compaction and control of gene transcription [4].

*Schistosoma mansoni* is a parasitic helminth whose life-cycle is characterized by passage through two obligatory sequential hosts: the fresh-water snail *Biomphalaria glabrata* (or dependent on the geographical location other *Biomphalaria* species) for the asexual stages, and human or rodents as hosts for the sexual stages. An estimated 200 million people in 74 countries suffer from schistosomiasis caused by *S. haematobium*, *S. japonicum*, and *S. mansoni*, and schistosomiasis is the most severe tropical disease in terms of morbidity after malaria [5]. Eggs of the parasite accumulate in the host’s liver and cause disease symptoms. Eggs are also released with the faeces and when they come into contact with water, free-swimming miracidia hatch and actively seek *B. glabrata* snails as intermediate host. After penetration into this host, the parasite develops via a primary “mother” sporocyst and a secondary “daughter” sporocyst generation into cercariae that can infect the vertebrate host.

It has been proposed that the regulation of gene transcription in schistosomes may require the action of factors that can modify chromatin [6]. Genomic DNA of *S. mansoni* is not methylated [7]. However, histone-modifying proteins are present. The histone arginine methyltransferase, *SmPRMT1*, for instance plays a role in nuclear receptor-mediated chromatin modification [8]. A predicted histone acetyltransferase, *SmGCN5*, displays high acetyltransferase activity with histone H3 as a substrate [9], *SmCBP1* acetylates histones with

a marked activity toward H4 [10] and several histone deacetylases (HDACs) have been characterized in schistosomes [11]. Transformation of miracidia into sporocysts can be blocked by HDAC inhibitors [12]. Posttranscriptional gene silencing (PTGS) can occur, both in miracidia and in adult stages of this parasite [13] [14]. PTGS shares pathway components with RNA-mediated heterochromatin formation identified in fission yeast and plants [15]. Taken together, these results suggest that the chromatin of *S. mansoni* may be compartmentalized as that of other eukaryotes and that the different levels of compaction may be developmentally regulated. It is therefore obvious that chromatin structure changes must be analyzed in *S. mansoni* and other metazoan parasites.

One of the methods to map chromatin along the genome is chromatin immunoprecipitation (ChIP). When we attempted to apply existing protocols to *S. mansoni* we realized that they were not adapted to the model, and we have therefore developed a native chromatin immunoprecipitation procedure (N-ChIP) for *S. mansoni* based on existing protocols [16]. *S. mansoni* undergoes a complex life cycle with larval stages for which biological samples are not readily available in sufficient amounts for many laboratory uses. Exhaustive optimization was thus required to apply the ChIP procedure to this organism. We describe here an optimized N-ChIP procedure for *S. mansoni*. This includes cell lysis, extraction of nuclei, chromatin fragmentation, choice of antibodies, and conditions for antibody/chromatin interactions. We optimized the sensitivity of the assay to reduce the amount of starting material and render the procedure feasible to all the life cycle stages of this parasite. In addition, immunoprecipitated DNA was analyzed at the whole-genome scale by high-throughput sequencing on an Illumina/Solexa 1G sequencing system. Results were validated through quantitative PCR analyses at the *SmRHO* locus, a previously characterized gene that encodes a guanine-protein coupled receptor (GPCRs) [17].

## **Methods**

### **Parasite strains and cell culture:**

A Guadeloupean strain of *S. mansoni* (GH2) was used in this study. The strain was maintained in its sympatric *B. glabrata* strain, and in hamsters (*Mesocricetus auratus*) as described previously ([18] and [19]). Eggs were axenically recovered from 50-day infected hamster livers and miracidia were hatched from eggs as previously described [19]. Miracidia were concentrated by sedimentation on ice for 15 min. Subsequently, miracidia were induced to transform to sporocysts *in-vitro* by incubation in sterile filtered *B. glabrata* embryonic (Bge) cell medium (Schneider's Drosophila Medium (Gibco BRL, #21720-016) with 4.5 % (w/v) lactalbumin hydrolysate (Gibco BRL, #18080-028), 10% (v/v) fetal calf serum (Sigma, #F-2442), 1.3% (w/v) galactose (Sigma, #G-5388), 1x antibacterial/antimicotic solution (Sigma, #A5955), pH 7.4) during 24 hours. Cercariae were recovered from infected snails (4 weeks post-infection) and harvested on ice by pipetting. They were pelleted by centrifugation (4000 g for 6 min). Eight-week adult worms were recovered by portal perfusion of hamsters with 0.9 % (w/v) NaCl and 0.8 % (w/v) trisodium citrate. For the ChIP procedure, miracidia, sporocysts, cercariae and adults were kept at -80°C. For RNA extractions, miracidia, sporocysts, cercariae and adults were pelleted and suspended immediately into 100 µl lysis buffer (Dynabeads mRNA DIRECT™ Micro kit, Dynal® Biotech, Invitrogen, #610.21) in RNase-free tubes and stored at -80°C.

Human lymphoblastic cells (CCRF-CEM) were grown as stationary suspension cultures in RPMI1640 medium with 2 mM l-glutamine, 10 % (v/v) heat-inactivated fetal calf serum, and 1x antibiotics (penicillin and streptomycin), at 37 °C in a humidified chamber (5% CO<sub>2</sub>).

### **mRNA Isolation:**

Messenger RNA was extracted from parasites (1000 miracidia, 1000 sporocysts, 1000 cercariae, 10 adults) using a Dynabeads® mRNA DIRECT Micro kit (Invitrogen, #610.21)

according to the manufacturer's instructions. At the end of the purification procedure, RNA was separated from the beads by heating to 80°C for 2 min in 20 µl 10mM Tris-HCl, pH 7.5. Resuspended mRNA was transferred to fresh RNase-free tubes.

#### **cDNA Synthesis:**

First strand cDNA was synthesized using 10 µl of the mRNA preparation in a final volume of 20 µl (0.5 mM dNTPs, 0.01 mM DTT, 1x first-strand buffer, 5 µM oligo-dT, 40 Units RNase out) with 200 Units of SuperScript™ II RT (Invitrogen, # 18064-014) as recommended per manufacturer instruction. After reverse transcription, the cDNA was purified with the Wizard® SV gel and PCR clean-up system (Promega, #A9281). cDNA was eluted from the column with the elution buffer of the kit in a final volume of 100 µl.

#### **Real time PCR analysis:**

Real time PCR analyses were performed using the LightCycler® 2.0 system (Roche Applied Science) and LightCycler® Faststart DNA Master SYBR Green I kit (Roche Applied Science, 12239264001). qPCR amplification was done with 2.5 µl of cDNA in a final volume of 10 µl (3 mM MgCl<sub>2</sub>, 0.5 µM of each primer, 1 µl of master mix). The primers were designed with the LightCycler® probe design software or the web-based Primer3 plus interface (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) (Supplementary Table 1). The following Light-Cycler run protocol was used: denaturation, 95°C, 10 min; amplification and quantification (repeated 40 times), 95°C for 10 s, 62°C for 5 s, 72°C for 16 s; melting curve, 60–95°C with a heating rate of 0.1°C/s and continuous fluorescence measurement, and a cooling step to 40°C. For each reaction, the crossing point (Ct) was determined using the “Fit Point Method” of the LightCycler Software 3.3. PCR reactions were done in duplicates and the mean value of the Ct was calculated. The  $\alpha$ -tubulin gene was used as an internal control [20]. The amplification of a unique band was verified by agarose gel electrophoresis for each qPCR product.

**Sequence analysis:**

Alignment of DNA sequences was performed using Sequencher<sup>TM</sup> software (Gene Codes Corporation).

**Western blot:**

Western blots were performed as described previously [12].

**Native chromatin immunoprecipitation (N-ChIP):**

Biological samples (miracidia, sporocysts, cercariae, adults and lymphoblasts) were centrifuged at 4000 rpm for 10 min at 4°C. The pellets were suspended in 1ml of buffer 1 (0.3 M sucrose, 30 mM KCl, 7.5 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 0.05 mM EDTA, 0.1 mM PMSF, 0.5 mM DTT, 7.5 mM Tris-HCl, pH 7.5) containing protease inhibitor cocktail tablets (two tablets for 50ml of buffer 1) (Roche Applied Science, #116974998001) and 5 mM sodium butyrate as histone deacetylase inhibitor (Sigma, #B5887). Samples were lysed by adding 1 ml buffer 1 with 0.8% NP40 and homogenized in a SZ22 tissue grinder tube (Kontes Glass Company, #885462-0022) using an SC tissue grinder pestle (Kontes Glass Company, #885451-0022) on ice for 3 min. The samples were kept on ice for 7 min after which lysates were overlaid on 8 ml of buffer 3 (1.2 M sucrose, 30 mM KCl, 7.5 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 0.05 mM EDTA, 5 mM sodium butyrate, 0.1 mM PMSF, 0.5 μM DTT, 7.5 mM Tris HCl pH 7.5) in Corex centrifuge tubes and centrifuged for 20 min (8500 rpm at 4°C) in a JJ-25 beckman coulter centrifuge using a TA-14-50 rotor. The buffer was removed and the pelleted nuclei were suspended in 1ml chromatin digestion buffer (0.12 M sucrose, 0.2 mM PMSF, 4 mM MgCl<sub>2</sub>, 5 mM sodium butyrate, 1 mM CaCl<sub>2</sub>, 0.05 M Tris HCl pH 7.5) and divided into aliquots of 500 μl in 1.5 ml Eppendorf tubes.

Chromatin digestion was performed for 4 min at 37°C with 1μl (15 U) of micrococcal nuclease (MNase) (USB, #70196Y). To stop the reaction, 20 μl of 0.5M EDTA was added and the tubes were immediately placed on ice. Samples were centrifuged for 10 min (13000



Xg at 4°C) and the supernatant (Fraction S1) was transferred into fresh tubes. The pellets (P1) were suspended in 100 µl dialysis buffer (200 µM EDTA, 200 µM PMSF, 5 mM sodium butyrate, 1 mM Tris HCl pH 7.5) and dialyzed overnight at 4°C in a Slide-A-Lyser MINI Dialysis Unit (cut-off at 3,500 Daltons) (Pierce, #69550). The fraction that remained in the dialysis tubes was centrifuged for 10 min (13000 Xg at 4°C). The supernatant (Fraction S2) was transferred into fresh tubes. Fractions S1 and S2 were then centrifuged three times for 10 min (13000 Xg at 4°C) and each time the supernatants were transferred into fresh tubes.

Chromatin from two combined fractions S1 and S2 were pooled and quantified by measuring the OD at 260nm. Thirty µg of chromatin were used and antibodies were added in excess ( $\geq$  saturating amount). An appropriate amount of stock solution was added to generate immunoprecipitation incubation buffer (150 mM NaCl, 20 mM sodium butyrate, 5 mM EDTA, 100 µM PMSF, 20 mM Tris HCl pH 7.5). Samples were incubated overnight at 4°C on a rotating wheel.

Fifty µl of protein A-sepharose (Sigma, #P3391) were added and incubated with the chromatin-antibody complexes for 4 hours at 4°C on a rotating wheel. The chromatin-antibody-protein A bead mixture was centrifuged for 10 min (11660 Xg at 4°C). The supernatant was transferred to a fresh tube, yielding the unbound fraction (“UB”).

Pellets (chromatin-antibody-protein A bead complex) were suspended in 10 ml washing buffer (50 mM Tris HCl pH 7.5, 10 mM EDTA, 5 mM sodium butyrate, 75 mM NaCl) and mixed gently for 10 min on a rotating wheel at 4°C. The mixture was centrifuged for 10 min at 4000 rpm at 4°C. The same procedure was repeated twice with increasing stringency conditions by using 125 mM and 175 mM NaCl as wash buffers, respectively. Finally, the pellets were suspended in 500 µl elution buffer (1% SDS, 20 mM Tris HCl pH 7.5, 50 mM NaCl, 5 mM EDTA, 20 mM sodium butyrate and 100 µM PMSF) and incubated for 15 min at

room temperature on a rotating wheel. The mixture was centrifuged for 10 min at 11600 g at 18°C, and the supernatants with the bound fraction (“B”) was transferred into fresh tubes.

DNA from the bound and unbound fractions was extracted with phenol/chloroform. DNA precipitation was done overnight at -80°C with 0.02 µg/µl glycogen, one volume isopropanol after adjusting the NaCl concentration to 250 mM. After centrifugation and a 70% ethanol wash, the pellets were suspended in 20 µl 10mM Tris HCl pH 7.5. The amounts of gene-specific sequences associated with each antibody were quantified by real-time PCR. A schematic representation and a detailed step-by-step procedure are described on our webpage: [http://methdb.univ-perp.fr/cgrunau/methods/native\\_chip\\_sm.html](http://methdb.univ-perp.fr/cgrunau/methods/native_chip_sm.html) .

#### **Generation of Illumina/Solexa libraries:**

DNA (30 ng) from native ChIP was polished to generate blunt ended fragments, add 3'-dA overhangs by treatment with Klenow polymerase and ligated to adapters as described (Pomraning et al., 2008). Fragments were size selected (150 bp to 500 bp) after migration through a 2% NuSieve agarose gel and sequencing libraries were generated by PCR amplification with 1 µl sample DNA, 1 X Phusion polymerase mix (Finnzymes, #F531), 2.25 µM of each Solexa PCR primer. We used a standard protocol: denaturation, 90°C, 30s; amplification (repeated 18 times), 98°C for 10 s, 65°C for 30 s, 72°C for 30 s; final elongation, 72°C, 5min. PCR products were purified on QIAquick PCR Purification columns (Qiagen, #28104) and eluted in 30 µl Qiagen elution buffer. We quantified the DNA library with a Nanodrop1000 spectrophotometer (Thermo Scientific) and clusters were generated with 4 pM of the 10 nM library dilution on the Illumina cluster station. A 36-cycle sequencing kit on the Illumina 1G analyzer was used according to the manufacturer's instructions [21].

#### **Processing 1G data:**

The resulting short reads were initially processed with the standard Illumina pipeline. The ELAND software trims all reads to 32 nt to decrease error rates on the Illumina 1G platform.

Only reads that matched a unique genomic locus with no (U0), one (U1) or two mismatches (U2) to the *S. mansoni* reference genome sequence ([http://www.sanger.ac.uk/Projects/S\\_mansoni/](http://www.sanger.ac.uk/Projects/S_mansoni/)) were considered for further analysis. We used “findpeaks” software to generate a lists of peaks and their genomic locations [22].

The rhodopsin gene (*SmRho*) was analyzed for enrichment of H3K9Ac by amplification of seven regions (amplicons R1 to R8) that were also used for validation by qPCR. We summed the U0 and U1 values for each bp in amplicons R1 to R8 (see Supplementary Table 1) and divided these sums by the length of each amplicon. For normalization, these values were divided by the value for R1 that is located 5.9 kb upstream of the transcriptional start site of *SmRho*.

#### **Quantitative real-time PCR:**

Real-time PCR was performed as previously described on a two-fold dilution of the N-ChIP samples with the primers described in Supplementary Table 1. Primer sets were designed to amplify amplicons shorter than 150 bp. Above this size, qPCR sensitivity is strongly affected, which results in background and amplification bias. The amount of target DNA recovered in the immunoprecipitated fraction was quantified by calculating the percent input recovery (% IR) normalized with the percent input recovery obtained with the  $\alpha$ -tubulin gene. The percent input recovery of the bound immunoprecipitated fraction for each amplicon was calculated as previously described [23] by the following formula: % input recovery =  $100 * E^{(Ct(input) - Ct(IP_{Bound}))}$ . The percent background was calculated by the following formula: % background =  $100 * E^{(Ct(input) - Ct(C_{-Bound}))}$ , where E is the primer efficiency designed to amplify the amplicon,  $Ct(IP_{Bound})$  is the Ct of the bound fraction obtained in the immunoprecipitated sample,  $Ct(C_{-Bound})$  is the Ct of the bound fraction obtained in the negative control (fraction without antibody), and  $Ct(input)$  is the Ct of the unbound fraction obtained in the negative control. It represents the quantity of chromatin that was used for the

study minus the fraction that bound non-specifically to the protein A Sepharose beads.

## **Results**

### ***Many commercial antibodies do not recognize histones of *S. mansoni* or bind unspecifically to proteins***

We used ChIP-grade antibodies when available. We tested antibodies raised against specific modifications of histone H3 or histone H4 from Abcam, Upstate and Active Motif (Table 1). These commercially available antibodies are raised against conserved peptides of histone proteins. Prior to any ChIP analysis, we performed western blots to confirm that these commercial antibodies were suitable for detection of modified histones of *S. mansoni*. The Upstate antibody raised against acetylated H3K9 (Cat# 07-352) resulted in a unique, specific band (supplementary figure 1). Other antibodies from Upstate or Active Motif (see Table 1) led to the detection of a band at the expected size but detected also several non-specific bands. These non-specific bands were not caused by secondary antibody interaction as this control alone did not reveal any signals (data not shown). Most Abcam antibodies did not recognize any *S. mansoni* proteins by western blot; an exception was the antibody against the C-terminus of H3 (Cat# ab1791) (supplementary figure 1). All optimization steps for the ChIP procedure were performed with the Upstate antibody raised against acetylated H3K9 (H3K9Ac).

### ***The antibody type and amount must be determined experimentally***

Since the western blot analysis with different antibodies did not always lead to unique bands, the amount of antibodies required in each N-ChIP reaction was determined by titration. N-ChIP reactions were performed starting with 1500 sporocysts for one immunoprecipitation assay and increasing concentrations of each antibody were used. Immunoprecipitated DNA was analyzed by qPCR on the  $\alpha$ -tubulin gene. The antibody from Upstate that specifically

recognized H3K9Ac in western blots was saturating at 4  $\mu$ l (supplementary figure 1A). In commercial antibody preparations, the antibody concentration (g/l) is not always indicated, however, the standardized production technology allows for reproducibly results based on volume values. The H3K9Me3 antibody from Abcam successfully immunoprecipitated *S. mansoni* chromatin but no signal was obtained with this antibody on western blots (supplementary figure 1B). Saturation for this antibody was reached at 4  $\mu$ g. The antibodies from Upstate that recognized H3K9Me3 and H3K27Me1 never reached saturation although up to 40  $\mu$ g of antibody were used for N-ChIP experiments (supplementary figure 1C and D). Apparently, these antibodies interact with additional proteins. This is in agreement with the western blots where additional high molecular weight bands were revealed. Their non-specific interaction may interfere with the immunoprecipitation and lead to background. Examination of *S. mansoni* sequence databases indicates that other proteins with conserved histone-like domains exist (data not shown), which may be recognized by these antibodies. Our findings show that the affinity of antibodies to *S. mansoni* histones can vary substantially, even if they have been tested on a wide range of model organisms. Thus, quantity and type of antibody for N-ChIP must always be determined experimentally.

***Optimization of the N-ChIP procedure for different S. mansoni life cycle stages:***

We established the *S. mansoni* N-ChIP protocol based on the previously published methods [16, 24]. Details of the procedure are described in the method section. We have optimized both lysis and extraction of nuclei for *S. mansoni* larvae (miracidia, sporocysts and cercariae) and adults. We found four minutes to be the optimum time for micrococcal digestion; this fragmented chromatin into mononucleosomes, thus providing the highest possible resolution for ChIP assays.

We also optimized washing steps and stringency conditions to reduce the background and increase signal to background ratio as much as possible. Triple washing, each time 10 min, of

fraction S1 and S2, and later of the chromatin-antibody-protein A complex (see Material and Methods) are essential.

To determine the minimum amount of biological material, the procedure was used with 3000, 1500, 750 and 375 miracidia for one immunoprecipitation with 2  $\mu$ l antibody against H3K9Ac. The immunoprecipitated DNA was analysed by qPCR of a segment within the  $\alpha$ -tubulin gene. This experiment indicates that with 3000, 1500 and 750 miracidia, the input recovery is 9%, with background levels below 1% (data not shown). Using 375 larvae, the background levels increase to 5%. We concluded that at least 750 larvae are necessary to successfully perform N-ChIP.

We quantified the immunoprecipitated DNA applying our N-ChIP procedure with an excess of antibody against H3K9Ac (8  $\mu$ l). Starting with 1500 miracidia, sporocysts and cercariae and a single adult, we were able to immunoprecipitate 100 ng, 170 ng, 110 ng and 120 ng of DNA, respectively, which allowed us to perform 16 qPCR analyses.

Alternatively, we applied a carrier N-ChIP procedure using lymphoblast cells as carrier cells. This procedure has been previously described to permit ChIP analysis of as few as 100 cells [25]. However, despite extensive washes and high stringency conditions, this procedure resulted in strong background and was therefore deemed not applicable to *S. mansoni* cells.

## **Profile of H3K9Ac enrichment along the rhodopsin gene**

### ***(I) ChIP-sequencing***

ChIP-sequencing (ChIP-Seq) combines chromatin immunoprecipitation with high-throughput parallel sequencing. This method is increasingly being used to map protein–DNA interactions *in-vivo* on a genome scale [26] [27] [28]. Our N-ChIP procedure was applied to miracidia, cercariae and adults using 8  $\mu$ l of the Upstate antibody against H3K9Ac. We then performed genome-wide sequencing of the immunoprecipitated DNA on an Illumina/Solexa 1G genome analyzer. We generated 23,758, 27,958 and 34,925 clusters for adults, cercariae and miracidia

respectively, corresponding to roughly 660 Mb. Although the length of the input DNA is approximately 150 bp, only 36 nt are sequenced for each read. Trimmed, 32-nt reads were mapped to the *S. mansoni* reference genome with the Solexa ELAND software and only reads that map to unique genomic locations in the reference genome (U0, U1 and U2) were considered for further analysis. Regions of high sequencing read density are referred to as peaks. We used the findpeaks software [22] to generate a list of peaks that comprises the genomic locations of sites inferred to be occupied by H3K9Ac. This data is not yet fully analyzed but will be made available in the future.

To validate our N-ChIP-seq analysis, we examined the profile of H3K9Ac along the *S. mansoni* rhodopsin gene (*SmRHO*). *SmRHO* (Smp\_104210) encodes a guanine-protein coupled receptor (GPCR), a member of an important family of eukaryotic receptor molecules that connects intracellular second messengers to extra-cellular inputs [29]. The rhodopsin subfamily of GPCRs encodes light-absorbing proteins that mediate dim light vision. *SmRHO* has been previously identified in *S. mansoni*. This gene shows large life-cycle-specific changes in gene expression [17]. We measured transcript levels of the rhodopsin gene at different developmental stages by qPCR with the R7 primer set (Figure 1). Rhodopsin is strongly expressed in cercariae. Transcript levels in miracidia and sporocysts are four- and tenfold lower, respectively. Rhodopsin expression was barely detectable in adults. This confirms previous analyses, where rhodopsin expression had been investigated by semi-quantitative RT-PCR [17].

The predicted coding region of *SmRHO* (GenBank accession number AF155134) was used to identify Smp\_scaff001984 as the scaffold of the current *S. mansoni* genome assembly (assembly 4) that carries the entire *SmRHO* gene. Alignment of the coding sequence with Smp\_scaff001984 was used to annotate the gene. *SmRHO* spans a 30 kb region from position 20699 to 50125 of Smp\_scaff001984 (Figure 2). The gene has seven exons. Exon size varies

from 85 bp to 414 bp, and intron size varies from 791 bp to 12,138 bp. We noted some discrepancies with the GeneDB *S. mansoni* annotation (version 4). Exon 1 of the gene is not annotated on Smp\_scaff001984. Furthermore, a 146 bp sequence corresponding to the 5' end of exon 5 is missing in the assembly of Smp\_scaff001984 although a genomic DNA sequence that corresponds to this missing part is present in the trace sequence “shisto6493g08.p1k”, which suggests incomplete assembly of this scaffold. The transcriptional start site is at position 20,699 and the predicted start codon is at position 31,407 of Smp\_scaff001984.

We added the U0 and U1 findpeaks output values that correspond to R1 to R8 locations along *SmRHO* (see Supplementary Table 1). In miracidia, two regions are enriched for H3K9Ac: region R2, located 2.2 kb upstream of the transcriptional start site, and R4, located 100 bp downstream of the transcriptional start site (Figure 2). In adults, H3K9Ac is underrepresented, which is consistent with the lack of expression of *SmRHO* in this stage (Figure 1). H3K9Ac is a histone mark mainly found in the promoters of active genes or within the coding sequence of active genes [30].

## **(II) ChIP and qPCR**

ChIP-qPCR was performed on miracidia, sporocysts, cercariae and adults to validate the ChIP-Seq analysis at the rhodopsin locus (Figure 2 and Supplementary Table 1). The percent input recovery was calculated for regions R1 to R8 as described in the Materials and Methods. The results were normalized with the percent input recovery of the  $\alpha$ -tubulin gene. All data were normalized to the R1 results. The ChIP-qPCR analysis showed enrichment in H3K9Ac at the R2 in cercariae, R3 in sporocysts and R4 in miracidia but no enrichment was observed on the ChIP product obtained from adults. These results are consistent with the expression analysis and confirm the results obtained on adults, cercariae and miracidia with the ChIP-Seq analysis.



## **Discussion:**

Chromatin immunoprecipitation is the technique of choice for studying protein-DNA interactions and to investigate the relationships between histone modifications and gene expression [4]. Particularly, ChIP in combination with high throughput sequencing analysis is of considerable interest to investigate the specific roles of individual histone modifications. Generally speaking, these analyses have been restricted to cultures of cells from model organisms [31][26][30]. ChIP has also been widely applied to unicellular parasites [32][33][34][35][36] and combined with whole genome analysis using the “ChIP-on-chip” technology [37][38]. Metazoan parasites are more difficult to study by these methods, as they go through complex life cycles. *Caenorhabditis elegans* is the only metazoan of a size comparable to *S. mansoni* for which a ChIP protocol was published [39].

In the present work, we describe for the first time a protocol to perform native chromatin immunoprecipitation of *S. mansoni*. We have optimized all individual steps of this procedure. We show the feasibility of the technique for the investigation of all life cycle stages of this parasite, at the same time minimizing the amount of biological samples required. A drawback of most current ChIP protocols is the requirement for high numbers of cells, which limits the feasibility of ChIP for samples with limited amount of starting material, as is the case with *S. mansoni* larvae. Therefore, we established conditions that are as sensitive as possible to minimize the amount of biological samples (750 larvae or one adult) required for each experiment. We have investigated the quality of histone-specific antibodies from different companies and show that each antibody must be tested carefully before it can be used for the ChIP procedure.

We have analyzed the immunoprecipitated products by a whole-genome approach, and we have validated the method by comparing the H3K9Ac profile at the *SmRHO* locus by both ChIP-qPCR and ChIP-Seq. It is evident from our results that H3K9 acetylation is high in the

5' region of the *SmRho* gene in those developmental stages in which *SmRho* is expressed. The level of acetylation corresponds to the level of transcription. In contrast, in the body of the gene, H3K9 acetylation is relatively low. These findings are fully compatible to what is known from the literature. Our results indicate also that, while in the majority of sites qPCR and ChIP-Seq confirm each other, and despite the fact that both methods deliver the above outlined 5'-3' profile, in some sites qPCR and ChIP-Seq enrichment factors can be different. We believe that this is due to the fact that bioinformatics analysis of ChIP-Seq data is still in its infancy and that artifacts are probably introduced by the data treatment. Fortunately, many laboratories have initiated the development of new or improved software tools [40] [41] . We intend to re-analyze our data when mature solutions become available. Nevertheless, we underline that by the very nature of the ChIP-Seq procedure (2 consecutive PCR, data treatment) for the moment the method should be considered explorative, and its results should always be confirmed by qPCR. Nevertheless, to our knowledge this is the first time that the ChIP-Seq technology has been applied to study protein-DNA interactions in a parasite.

Work by others has clearly demonstrated the influence of histone modifying enzymes on the control of gene expression in *S. mansoni* [6] [10]. Our work shows that *S. mansoni* chromatin states can be analyzed by ChIP using antibodies that recognize known euchromatic and heterochromatic marks. It will now be possible to investigate if different chromatin states are responsible for gene expression regulation during adaptation of the parasite to different environments, maturation from larval to adult stages, and sexual differentiation.

### **Acknowledgements:**

This work was funded by the "Programme Schistophepigen" from the French National Agency for Research. The Freitag Lab is supported by a grant from the American Cancer Society (RSG-08-030-01-CCG) and startup funds from the CGBI at OSU. Illumina/Solexa

high-throughput sequencing was carried out at the OSU CGRB core facility with the help of Mark Dasenko, Steve Drake and Christopher Sullivan. C.C. was a recipient of a fellowship from the Centre National de la Recherche Scientifique (CNRS). The authors are indebted to Yves Moné, Anne Rognon and Bernard Dejean for technical support, Jerome Guard (IGH, CNRS Montpellier) for helpful discussion, and Anne-Cécile Ribou (Université de Perpignan) for generously providing cultured lymphoblast cells.

## Legends:

### Supplementary Figure 1: Antibody titration for N-ChIP analysis on *S. mansoni*

N-ChIP analysis was performed with 1500 sporocysts and with increasing amounts of antibody that recognize H3K9Ac (Upstate) (A), H3K9Me3 (Abcam) (B), H3K27Me1 (Upstate) (C) and H3K9Me3 (Upstate) (D). Immunoprecipitated DNA was quantified by qPCR of a region in body of the  $\alpha$ -tubulin gene. Input Recovery as % of input (%IR, y axis) is shown for each tested antibody (in  $\mu$ l or  $\mu$ g, x axis). The value % IR background indicates % IR without antibodies. On the right side of each graph, the corresponding western blot is shown. For anti H3 (Abcam) no ChIP experiments were performed (E).

### Figure 1: Stage specific expression of rhodopsin

Ratios of transcript levels corresponding to *SmRHO* at different developmental stages of *S. mansoni*. Ratios were determined using real-time quantitative PCR and are expressed relative to mRNA  $\alpha$ -tubulin expression levels. Each histogram represents the average value of duplicates  $\pm$  S.D.

### Figure 2: Schematic representation of the rhodopsin coding sequence on *S.mansoni* scaffold 001984 and enrichment of acetylated H3K9 along the gene

Above: Seven exons were identified and are represented as gray boxes. The dashed box represents a 146 bp sequence that is present in the previously described rhodopsin sequence (Hoffman et al. 2001; AF155134) but that is absent in the Smp\_scaff001984. Intron sizes are indicated and the position of each exon on the Smp\_scaff001984. Primer sets R1 to R8 are shown (see supplementary table 1 for starting positions). Primers set R7 was used for

determination of the transcription level and spans exon 6 to 7, all other primers sets are for qPCR of immunoprecipitated DNA.

Below: Chromatin immunoprecipitation was performed with antibodies against H3K9Ac on chromatin extracted from adults, miracidia, and cercariae. Immunoprecipitated DNA was analyzed by high throughput sequencing analysis on an Illumina/Solexa 1G analyzer (full line) or by qPCR (dotted line). For high throughput sequencing analysis, U0, U1 and U2 peak values for the regions corresponding to each amplicon (R1 to R8) were summed and normalized to the amplicon size. qPCR was performed with primer sets R1 to R8, the percent input recovery of each target sequence was normalized with the recovery of  $\alpha$ -tubulin. qPCR results are the average of three independent experiments. All data were normalized to the R1 region. Connecting lines do not represent experimental data but are there to guide the eye.

## References :

- [1] Luijsterburg M S, White M F, van Driel R, Dame R T. The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit Rev Biochem Mol Biol* 2008;43:393-418.
- [2] Umlauf D, Fraser P, Nagano T. The role of long non-coding RNAs in chromatin structure and gene regulation: variations on a theme. *Biol Chem* 2008;389:323-31.
- [3] Dillon N. The impact of gene location in the nucleus on transcriptional regulation. *Dev Cell* 2008;15:182-6.
- [4] Berger S L. The complex language of chromatin regulation during transcription. *Nature* 2007;447:407-12.
- [5] Chitsulo L, Loverde P, Engels D. Schistosomiasis. *Nat Rev Microbiol* 2004;2:12-3.
- [6] Fantappie M R, de Oliveira F M, Santos R D, Mansure J J, Furtado D R, da Silva I C, Rumjanek F D. Control of transcription in *Schistosoma mansoni*: Chromatin remodeling and other regulatory elements. *Acta Trop* 2007;
- [7] Fantappie M R, Gimba E R, Rumjanek F D. Lack of DNA methylation in *Schistosoma mansoni*. *Exp Parasitol* 2001;98:162-6.
- [8] Mansure J J, Furtado D R, de Oliveira F M, Rumjanek F D, Franco G R, Fantappie M R. Cloning of a protein arginine methyltransferase PRMT1 homologue from *Schistosoma mansoni*: evidence for roles in nuclear receptor signaling and RNA metabolism. *Biochem Biophys Res Commun* 2005;335:1163-72.
- [9] de Moraes Maciel R, de Silva Dutra D L, Rumjanek F D, Juliano L, Juliano M A, Fantappie M R. *Schistosoma mansoni* histone acetyltransferase GCN5: linking histone acetylation to gene activation. *Mol Biochem Parasitol* 2004;133:131-5.
- [10] Bertin B, Oger F, Cornette J, Caby S, Noel C, Capron M, Fantappie M R, Rumjanek F D, Pierce R J. *Schistosoma mansoni* CBP/p300 has a conserved domain structure and interacts functionally with the nuclear receptor SmFtz-F1. *Mol Biochem Parasitol* 2006;146:180-91.
- [11] Oger F, Dubois F, Caby S, Noel C, Cornette J, Bertin B, Capron M, Pierce R J. The class I histone deacetylases of the platyhelminth parasite *Schistosoma mansoni*. *Biochem Biophys Res Commun* 2008;377:1079-84.
- [12] Azzi A, Cosseau C, Grunau C. *Schistosoma mansoni*: Developmental arrest of miracidia treated with histone deacetylase inhibitors. *Exp Parasitol* 2008;121:288-91

- [13] Boyle J P, Wu X J, Shoemaker C B, Yoshino T P. Using RNA interference to manipulate endogenous gene expression in *Schistosoma mansoni* sporocysts. *Mol Biochem Parasitol* 2003;128:205-15.
- [14] Ndegwa D, Krautz-Peterson G, Skelly P J. Protocols for gene silencing in schistosomes. *Exp Parasitol* 2007;117:284-91.
- [15] Lippman Z, Martienssen R. The role of RNA interference in heterochromatic silencing. *Nature* 2004;431:364-70.
- [16] O'Neill L P, Turner B M. Immunoprecipitation of native chromatin: NChIP. *Methods* 2003;31:76-82.
- [17] Hoffmann K F, Davis E M, Fischer E R, Wynn T A. The guanine protein coupled receptor rhodopsin is developmentally regulated in the free-living stages of *Schistosoma mansoni*. *Mol Biochem Parasitol* 2001;112:113-23.
- [18] Dissous C, Dissous C, Capron A. Isolation and characterization of surface antigens from *Schistosoma mansoni* schistosomula. *Mol Biochem Parasitol* 1981;3:215-25.
- [19] Theron A, Pages J R, Rognon A. *Schistosoma mansoni*: distribution patterns of miracidia among *Biomphalaria glabrata* snail as related to host susceptibility and sporocyst regulatory processes. *Exp Parasitol* 1997;85:1-9.
- [20] Bahia D, Avelar L, Mortara R A, Khayath N, Yan Y, Noel C, Capron M, Dissous C, Pierce R J, Oliveira G. SmPKC1, a new protein kinase C identified in the platyhelminth parasite *Schistosoma mansoni*. *Biochem Biophys Res Commun* 2006;345:1138-48.
- [21]
- [22] Fejes A P, Robertson G, Bilenky M, Varhol R, Bainbridge M, Jones S J. FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. *Bioinformatics* 2008;24:1729-30.
- [23] Edmunds J W, Mahadevan L C, Clayton A L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *Embo J* 2008;27:406-20.
- [24] Umlauf D, Goto Y, Feil R. Site-specific analysis of histone methylation and acetylation. *Methods Mol Biol* 2004;287:99-120.
- [25] O'Neill L P, VerMilyea M D, Turner B M. Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. *Nat Genet* 2006;38:835-41.
- [26] Barski A, Cuddapah S, Cui K, Roh T Y, Schones D E, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell* 2007;129:823-37.

- [27] Valouev A, Johnson D S, Sundquist A, Medina C, Anton E, Batzoglou S, Myers R M, Sidow A. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat Methods* 2008;
- [28] Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith O L, He A, Marra M, Snyder M, Jones S. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* 2007;4:651-7.
- [29] Yeagle P L, Albert A D. G-protein coupled receptor structure. *Biochim Biophys Acta* 2007;1768:808-24.
- [30] Wang Z, Zang C, Rosenfeld J A, Schones D E, Barski A, Cuddapah S, Cui K, Roh T Y, Peng W, Zhang M Q, Zhao K. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 2008;40:897-903.
- [31] Schones D E, Zhao K. Genome-wide approaches to studying chromatin modifications. *Nat Rev Genet* 2008;9:179-91.
- [32] Freitas-Junior L H, Hernandez-Rivas R, Ralph S A, Montiel-Condado D, Ruvalcaba-Salazar O K, Rojas-Meza A P, Mancio-Silva L, Leal-Silvestre R J, Gontijo A M, Shorte S, Scherf A. Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell* 2005;121:25-36.
- [33] Lavi T, Isakov E, Harony H, Fisher O, Siman-Tov R, Ankri S. Sensing DNA methylation in the protozoan parasite *Entamoeba histolytica*. *Mol Microbiol* 2006;62:1373-86.
- [34] Kulakova L, Singer S M, Conrad J, Nash T E. Epigenetic mechanisms are involved in the control of *Giardia lamblia* antigenic variation. *Mol Microbiol* 2006;61:1533-42.
- [35] Lowell J E, Cross G A. A variant histone H3 is enriched at telomeres in *Trypanosoma brucei*. *J Cell Sci* 2004;117:5937-47.
- [36] Lira C B, de Siqueira Neto J L, Khater L, Cagliari T C, Peroni L A, dos Reis J R, Ramos C H, Cano M I. LaTBP1: a *Leishmania amazonensis* DNA-binding protein that associates in vivo with telomeres and GT-rich DNA using a Myb-like domain. *Arch Biochem Biophys* 2007;465:399-409.
- [37] Gissot M, Kelly K A, Ajioka J W, Grealley J M, Kim K. Epigenomic modifications predict active promoters and gene structure in *Toxoplasma gondii*. *PLoS Pathog* 2007;3:e77.
- [38] Cui L, Miao J, Furuya T, Li X, Su X Z, Cui L. PfGCN5-mediated histone H3 acetylation plays a key role in gene expression in *Plasmodium falciparum*. *Eukaryot Cell* 2007;6:1219-27.



- [39] Whittle C M, McClinic K N, Ercan S, Zhang X, Green R D, Kelly W G, Lieb J D. The genomic distribution and function of histone variant HTZ-1 during *C. elegans* embryogenesis. *PLoS Genet* 2008;4:e1000187.
- [40] Li R, Li Y, Kristiansen K, Wang J. SOAP: short oligonucleotide alignment program. *Bioinformatics* 2008;24:713-4.
- [41] Campagna D, Albiero A, Bilardi A, Caniato E, Forcato C, Manavski S, Vitulo N, Valle G. PASS: a Program to Align Short Sequences. *Bioinformatics* 2009;