Molecular & Biochemical Parasitology 183 (2012) 105–114

Contents lists available at SciVerse ScienceDirect

Molecular & Biochemical Parasitology

Review

Schistosoma mansoni histones: From transcription to chromatin regulation; an *in silico* analysis

Letícia Anderson[a], Raymond J. Pierce[b], Sergio Verjovski-Almeida[a,*]

[a] *Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, 05508-000 São Paulo, SP, Brazil*
[b] *CIIL, Inserm U1019, CNRS UMR 8204, Institut Pasteur de Lille, Univ. Lille Nord de France, F59019 Lille, France*

ARTICLE INFO

Article history:
Received 4 January 2012
Received in revised form 29 February 2012
Accepted 1 March 2012
Available online 10 March 2012

Keywords:
Schistosoma mansoni
Histones
Histone mRNA processing
Histone modifying enzymes

ABSTRACT

Schistosoma mansoni is a human endoparasite with a complex life cycle that also infects an invertebrate mollusk intermediate host and exhibits many diverse phenotypes. Its complexity is reflected in a large genome and different transcriptome profiles specific to each life cycle stage. Epigenetic regulation of gene expression such as the post-translational modification of histones has a significant impact on phenotypes, and this information storage function resides primarily at histone tails, which results in a varied histone code. Evidence of transcription of the different histone families at all life stages of the parasite was detected by a survey of transcriptome databases; manual curation of each gene prediction at the genome sequence level showed errors in the coding sequences of three of them. The biogenesis of histones is coupled to DNA replication, and a detailed *in silico* analysis of the specialized machinery of histone mRNA processing in the *S. mansoni* genome reveals that it is as conserved as in other eukaryotes, consisting in transcription factors and stem-loop binding proteins which recognize the stem loop structure at the histone mRNA 3′UTR. Histone modifying enzymes (HMEs) such as histone acetyltransferases, methyltransferases and deacetylases (HDACs) have been described in *S. mansoni*, and their potential as new therapeutic targets was evidenced with the apoptotic phenotype that resulted from HDAC inhibition. However, the overall regulation of transcription coupled with gene expression profiles correlated to histone modifications has not yet been characterized. Besides the interaction of HMEs with histones, many factors involved in cellular processes are known to bind to histones, and were identified here by an *in silico* analysis of the *S. mansoni* genome. Knowledge of the histone families opens up perspectives for further studies that will lead to a better identification of their post-translational modifications, their gene regulation and to the possible characterization of HMEs as targets for the development of new drugs.

Contents

* Corresponding author. Tel.: +55 11 3091 2173; fax: +55 11 3091 2186.
E-mail address: verjo@iq.usp.br (S. Verjovski-Almeida).

0166-6851/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.molbiopara.2012.03.001

1. Introduction

Schistosomiasis is a parasitic disease infecting 200 million people and another 600 million live in endemic areas in developing countries. The disease is caused by blood-dwelling flukes of the genus *Schistosoma* (phylum Platyhelminthes) [1]. Schistosomiasis is a public health problem ranking second only to malaria among the parasitic diseases with regard to the number of people infected and those at risk, causing the annual loss of between 1.7 and 4.5 million disability adjusted life years (DALYs). The highest prevalence of schistosomiasis is in sub-Saharan Africa, usually affecting school-age children, adolescents and young adults [2].

The blood fluke *Schistosoma mansoni* is one of the three major human species, besides *S. haematobium* and *S. japonicum*, that exhibit dioecy and have a complex life cycle comprising several morphologically distinct phenotypes through two obligatory sequential hosts: fresh-water snails of the genus *Biomphalaria* as intermediate host, and man or rodents as definitive hosts. Adult parasites live in the mesenteric veins (for up to 10 years or more) laying 200–300 eggs per day, of which some are trapped in the microvasculature of the liver inducing a granulomatous inflammatory response and fibrosis [1].

2. Histones in *S. mansoni*

The complex life cycle and the diverse phenotypes of the parasite are reflected in its large genome (363 megabases) with eight chromosomes (seven autosomal and a ZW sex pair), with 11,807 genes encoding an estimated 13,197 transcripts [3]. Different profiles of gene expression are programmed into the various cell types and along the life-cycle stages of this complex eukaryote. Epigenetic mechanisms play a central role in programmed gene regulation and include a large number of histone post-translational modifications as well as genomic DNA methylation, small interfering RNAs, histone variants and post-translational modifications of proteins other than histones. In the present review we will concentrate on the histones as regulators of gene activity. Post-translational modifications at histones in eukaryotes involve several chromatin-based processes having a significant impact on gene expression profiles and phenotypes [4,5].

In eukaryotes, genomic DNA is packaged with histone proteins organized in nucleosomes, resulting in a macromolecular complex called chromatin. The nucleosome repeating unit is formed by two copies of each histone protein, H2A, H2B, H3 and H4, assembled in an octamer with 146 bp of DNA wrapped in roughly two superhelical turns and stabilized by the linker histone H1 [6]. The biogenesis of histones is tightly coupled to DNA replication, being encoded by two gene families: that of replication-dependent histone genes, which are coupled to DNA synthesis during S-phase in cells, and the replication-independent genes, which are constitutively expressed at a basal level throughout the whole cell cycle for putative chromatin lesion repair. Histone mRNAs must be expressed rapidly at the beginning of S-phase and persist at high levels to coincide with the replication of DNA. Moreover, the massive expression of histones is in a stoichiometric relation to DNA replication because an excess of histones during G1-, G2-, or M-phase is highly toxic to the eukaryotic cell [7–9].

In 2003 our group [10] published the transcriptome analysis of *S. mansoni* with 163,000 expressed-sequence tags (ESTs) from six developmental stages of the parasite (adult worms, eggs, miracidia, germ balls, cercaria and schistosomula). These ESTs were assembled into 31,000 different fragment sequences, representing an estimated 14,000 unique genes (an estimated 92% coverage of the transcriptome). Today there are 214,000 *S. mansoni* ESTs in the dbEST public database.

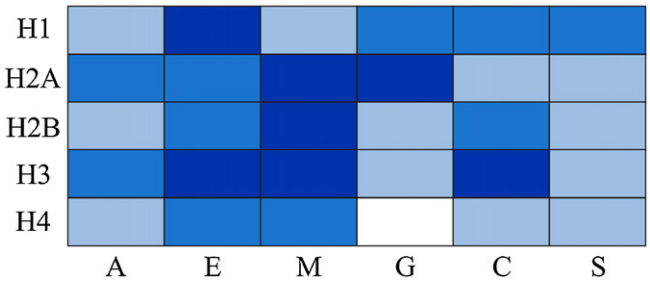

Fig. 1. Number of histones expressed-sequence tags through the six life stages of *S. mansoni*. Frequency of histone transcripts from NCBI database from each stage: A, adults; E, eggs; M, miracidia; G, germ balls; C, cercaria; S, schistosomula. Histone EST counts for each stage were normalized by the total number of ESTs for that given stage. Color scale indicates the number of normalized counts with white representing no count and different shades of blue representing normalized ranges: 0.1–4.0, 4.1–8.0 and >8.0.

Looking for evidence of expression of all *S. mansoni* histones among the entire set of public ESTs, we found that the five histone families (H1, H2A, H2B, H3 and H4) were detected at all six life cycle stages of the parasite (with the exception of H4 in germ balls, the second least sequenced stage) (Fig. 1). Given the high sequence identity among the different genomic copies of each histone (see below for details), one cannot uniquely map the ESTs to any of these copies; the read counts for each histone represent the non-redundant set of ESTs mapping to any member of that given histone family.

3. *In silico* analysis of *S. mansoni* histone families

In order to appraise the complement of *S. mansoni* histones, which comprise the multiple targets of the histone modifying enzymes (HMEs) reviewed below, we performed an *in silico* comparison among the *S. mansoni* histone sequences. The aim was to analyze in detail the available gene sequences and distinguish the different members within each family. Gene predictions (Smp sequences) derived from the genome sequencing project [3], together with a new assembly of the genome (version 5.2 available at GenBank as of December 13, 2011), along with transcriptome data from two large-scale EST sequencing projects [10,11] were used for the *in silico* analyses of the histone sequences, followed by manual inspection and curation.

The 29 histone Smp genes predicted in the genome are listed in Table 1. Alignment of the Smp genes against the public ESTs showed that the majority of them (21 genes) had their sequence predictions and expression in *S. mansoni* confirmed. Curiously, two different Smp numbers refer to the same genomic locus (see Table 1); therefore, 20 unique predicted Smp histones had evidence of transcription.

An additional five Smp predictions (marked in Table 1 with one asterisk) had no evidence of transcription; it is noteworthy that three of them have short incomplete conserved histone domains that cover only a small fraction of the human orthologs (Table 1). Smp_123850 has some similarity to an H2A.J human ortholog, however it covers only 33% of the human sequence; Smp_130880 similarity to human H2A.J extends only through 25% of the predicted Sm sequence and covers only 50% of the human sequence; Smp_026880 covers only 34% of the human H3.I sequence. They may represent wrong sequence predictions and/or wrong annotations of proteins that have partial histone domains and do not function as histones. Among the five predictions with no EST evidence, Smp_056420 is similar to H2A.J, and Smp_150540 is similar to H3.3B (Table 1).

The remaining 3 Smp histones had sequences predicted in the genome project that were discrepant from the sequences obtained