Cloning and Molecular Characterization of the *Schistosoma mansoni* Genes RbAp48 and Histone H4

Patrícia P Souza, Débora N Santos, Sérgio D J Pena, Glória R Franco+

Laboratório de Genética Bioquímica, Departamento de Bioquímica e Imunologia, ICB, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, M G, Brasil

The human nuclear protein RbAp48 is a member of the tryptophan/aspartate (WD) repeat family, which binds to the retinoblastoma (Rb) protein. It also corresponds to the smallest subunit of the chromatin assembly factor and is able to bind to the helix 1 of histone H4, taking it to the DNA in replication. A cDNA homologous to the human gene RbAp48 was isolated from a Schistosoma mansoni *adult worm library and named SmRbAp48. The full length sequence of SmRbAp48 cDNA is 1036 bp long, encoding a protein of 308 amino acids. The transcript of SmRbAp48 was detected in egg, cercariae and schistosomulum stages. The protein shows 84% similarity with the human RbAp48, possessing four WD repeats on its C-terminus. A hypothetical tridimensional structure for the SmRbAp48 C-terminal domain was constructed by computational molecular modeling using the β-subunit of the G protein as a model. To further verify a possible interaction between SmRbAp48 and* S. mansoni *histone H4, the histone H4 gene was amplified from adult worm genomic DNA using degenerated primers. The gene fragment of SmH4 is 294 bp long, encoding a protein of 98 amino acids which is 100% identical to histone H4 from* Drosophila melanogaster.

Key words: RbAp48 - histone H4 - *Schistosoma mansoni* - gene cloning

The human protein RbAp48 was first identified as one of the major polypeptides from HeLa cell lysates that binds specifically to a putative functional domain of the carboxy terminus of the Rb protein, a known cellular tumor suppressor (Lee et al. 1991). Human RbAp48 protein shares sequence homology with MSI1, a negative regulator of the Ras-cAMP pathway in the yeast *Saccharomyces cerevisiae*. Overexpression of MSI1 gene suppresses the heat-shock sensitivity of *iraI* and $Ras2^{val19}$ mutant strains and reduces the cAMP levels in these mutants (Ruggieri et al. 1989). Furthermore, similarly to MSI1, the human RbAp48 supresses the heat-shock sensitivity of the same mutants (Qian et al. 1993). This finding demonstrates that there is a functional homology between both proteins (Qian et al. 1993). The yeast null mutant of MSI1 has been obtained and presents sensitivity to UV irradiation associated to a decrease in the silencing of telomere adjacent genes (Kaufman et al. 1997). RbAp48 is a nuclear protein and a member of the tryptophan/aspartate (WD) repeat family (Qian et al. 1995). Proteins constituted by at least four WD repeats can be clustered into this structural family of proteins, the members of which appear to perform regulatory functions in several cellular processes, such as cell division, cell fate determination, gene transcription, transmembrane signaling, mRNA modification, and vesicle fusion (Neer et al. 1994).

This work was supported by grants from Brazilian Research Council (CNPq) and Pró Reitoria de Pesquisa, UFMG.
+Corresponding author. Fax: +55-31-3499.2984. E-mail: gfranco@icb.ufmg.br.
Received
Accepted

Verreault and colleagues, in 1996, described a human chromatin assembly complex containing a chromatin assembly factor (CAF-1) and modified histones H4 and H3, acetylated in specific lysine residues. It was further verified that RbAp48 corresponds to the smallest subunit of CAF-1 and is able to bind to the helix 1 of histone H4, taking the later to the DNA in replication (Krude 1999). A related p48 protein in *Saccharomyces*, named Hat2p, is a constituent of a subunit of histone H4 acetyltransferase B type (Parthurn et al. 1996). These findings suggest that a family of p48 proteins may be involved in diverse aspects of histone functions in a variety of different organisms (Verreault et al. 1996).

Another function attributed to RbAp48 is its participation on the assembly of a basal repression complex, formed by histone deacetylases (HDAC) 1 and 2 and also RbAp46, recruited by a variety of co-repressors and repression associated factors to strength the transcriptional repression during the cell cycle (Knoepler & Eisenman 1999). Nicholas and colleagues (2000) found that RbAp48 belongs to the HDAC complex that associates with the Rb protein to repress the E2F transcription factor during the cell cycle.

Studies on the regulation of *S. mansoni* gene expression are still in the beginning. Molecular characterization of regulatory proteins involved in the control of transcription and DNA metabolism will contribute to a better understanding of the biology and development of this parasite. This communication reports on the cloning and molecular characterization of SmRbAp48 and SmH4 genes of *S. mansoni*. Evidences indicate that SmRbAp48 is expressed in different developmental stages of the parasite. Additionally, a hypothetical three dimensional model for the SmRbAp48 C-terminus was constructed by computational molecular modeling using the b-subunit of the G protein as a model.

MATERIALS AND METHODS

cDNA libraries and DNA purification - *S. mansoni* egg, cercariae, 3h schistosomulum and adult worm cDNA libraries were constructed in λZAP as part of the *Schistosoma* genome project (Franco et al. 2000). Genomic DNA was purified from *S. mansoni* LE strain adult worms as described previously (Simpson et al. 1982). Plasmids and Polymerase Chain Reaction (PCR) fragments were purified with the Wizard DNA Purification Systems™ (Promega).

DNA cloning - Two identical clones (MAAD0269 and MAAD0270) carrying cDNA fragments homologous to the human gene RbAp48 were isolated from a *S. mansoni* adult worm library after random clone selection. The histone H4 gene was amplified by PCR from *S. mansoni* LE strain adult worm genomic DNA using degenerated primers. Several strategies were used to obtain the full-length sequence of both strands of SmRbAp48 and SmH4 genes. Both SmRbAp48 cDNA clones were digested with the *Rsa*I restriction enzyme in internal sites of the insert. Specific primers were also designed and used to amplify by PCR internal regions of the cDNA fragment. The initial portion of the cDNA containing part of the Open Reading Frame (ORF) and the 5' untranslated region (5'UTR), that was not present in the original cDNA clones, was obtained by amplification of other cDNA libraries using an hemi-nested PCR strategy. The digestion fragments and all the PCR products were cloned into the *Sma*I site of pUC18 (Amersham Pharmacia Biotech) using the Sureclone Ligation kit (Amersham Pharmacia Biotech).

PCR and sequencing - Specific primers targeting regions of SmRbAp48 were used to amplify parts of the SmRbAp48 cDNA cloned into pBlueScript KS^+ (Stratagene) and the four cDNA libraries (egg, cercariae, 3h schistosomulum and adult worm). Degenerated primers designed based on histone H4 genes from different organisms were used to amplify the SmH4 gene from adult worm genomic DNA.

For the amplification of SmRbAp48, 100 µl reaction mixture was used containing approximately 4 µl of the clones, 10 mM Tris HCl pH 8.5, 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM each deoxynucleotide triphosphate, 200 nM each primer and 2.5 U of Taq polymerase. The conditions used for amplification were 96°C for 2 min, followed by a step cycle program set to denature at 96°C for 1 min, anneal at 54°C for 1 min, and extend at 72°C for 2 min for a total of 25 cycles. Amplifications of the cDNA libraries were performed in a 30 µl volume containing 1 µl of the cDNA library, 10 mM Tris HCl pH 8.5, 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM each deoxynucleotide triphosphate, 200 nM each primer and 2 U of Taq DNA polymerase. The conditions used for amplification were the same described above. PCR of genomic DNA was performed in a 30 µl volume containing 20 ng of genomic DNA, 10 mM Tris HCl pH 8.8, 75 mM KCl_2, 3.5 mM $MgCl_2$, 200 µM each deoxynucleotide triphosphate, 400 nM each primer and 2U of Taq DNA polymerase. The conditions used for the amplifications were 95°C for 2 min, followed by a step cycle program set to denature at 95°C for 1 min, anneal at 48°C for 1 min, and extend at 72°C for 2 min for a total of 35 cycles. The amplicons were analyzed in 1% agarose gel stained by ethidum bromide or in 6% polyacrylamide gels silver stained (Santos et al. 1993).

Sequencing reactions were performed using the Thermo Sequenase fluorescent labeled primer cycle sequencing kit™ with 7-deaza-dGTP (Amersham Pharmacia Biotech). M13 fluorescent primers targeting the margins of the cloning sites were used for DNA sequencing of both strands, using the A.L.F. Automated DNA Sequencer (Amersham Pharmacia Biotech).

Sequence analysis - Search for homologous sequences was undertaken using the BLAST program (Altschul et al. 1997, http://www.ncbi.nlm.nih.gov). Open reading frame (OFR) search and DNA translation were performed using the DNAsis program. The PredictProtein server (http://cubic.bioc.columbia.edu/predictprotein/) was used for prediction of secondary structure and post-translational modification sites. A BMERC "The WD repeat Familyof Proteins" Server (http://bmerc-www.bu.edu/bioinformatics/wdrepeat.html) was used for prediction of the number and localization of the WD repeats in the protein, as well as to produce a three-dimensional (3D) model for the SmRbAp48 gene product, based on its homology to the β-subunit of the G protein (PDB ID 1SCG). The 3D model visualization was performed using the RASMOL Program.

RESULTS

Cloning and sequencing of the S. mansoni RbAp48 cDNA - On the *S. mansoni* Gene Discovery Program, based on the production and identification of Expressed Sequenced Tags (ESTs) from cDNA libraries (Franco et al. 2000), two identical clones (MAAD0269 and MAAD0270) carrying a cDNA insert homologous to the human gene RbAp48 were isolated from an adult worm library. The new gene was named SmRbAp48. In order to obtain the full-length sequence from both strands of the cDNA, three strategies were used. First, both cDNA clones were digested with the *Rsa*I restriction enzyme, once there are two site for this enzyme in the middle of the insert and another one into the vector in a position bordering the cloning site. The two fragments of interest (sizing 689 bp and 185 bp) were purified from agarose gel, cloned into pUC18 vector in the *Sma*I site and fully sequenced (Fig. 1). Second, PCR was used to amplify parts of the SmRbAp48 insert using two pair of primers (Rb240 and T3 and Rb435 and M13 universal primers) (Table I). Rb240 and Rb435 anneal at the insert, while T3 and M13 universal anneal at vector sites. The fragments obtained (339 bp and 568 bp in size) were subcloned into pUC18 vector and sequenced (Fig. 1). Once the isolated cDNAs did not contain the initial part of the ORF and the 5' UTR of the SmRbAp48 transcript, the third strategy was to obtain those regions from amplification of adult worm, cercariae and schistosomulum cDNA libraries, using an heminested PCR technique. Rb240 and M13 reverse primers were used for the first step of amplification and Rb95 and M13 reverse primers were used on the second step (Table). Fragments obtained from the second step of amplification were subcloned into pUC18 vector and sequenced as described above (Fig. 1). Using the DNASIS program, all

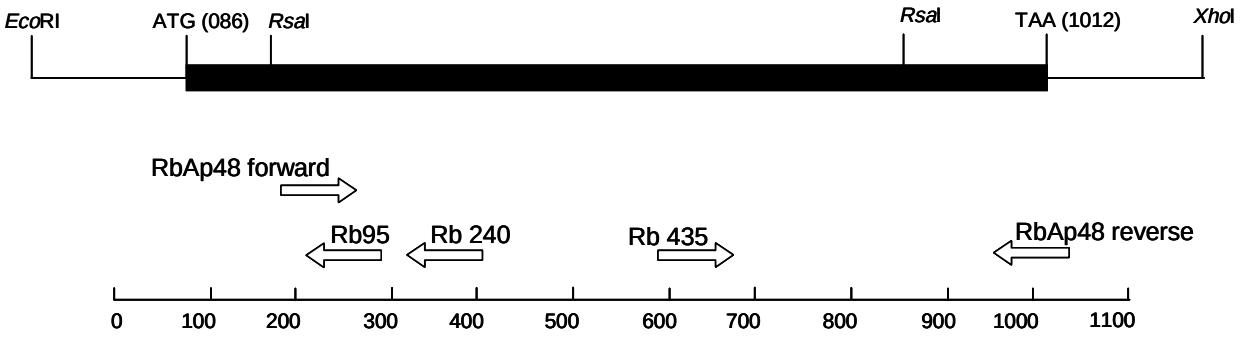

Fig. 1: schematic representation of the strategies used for cloning and sequencing the SmRbAp48 cDNA. Positions of start (ATG) and stop (TAA) codons, *Rsa*I restriction sites used for digestion and primers used for amplification of the cDNA are indicated.

TABLE

Set of primers used for cloning and sequencing of SmRbAp48 and SmH4 genes

Primers	Nucleotide sequence	Nucleotide position
Rb 435	5' CGTCTCAAGGGTCATCA 3'	596 - 612
Rb 240	5' TCACCTCGTTCACTATCA 3'	376 - 393
Rb 95	5' TAATCTCGCCCAGTCCT 3'	248 - 264
RbAp48 forward	5' GACGGATCCATGAGGAATACTCCGTTCTTG 3'	161 - 178
RbAp48 reverse	5' GAGCTGCAGGTTATTTGGTTATTTATGTG 3'	1006 - 1025
Hist4 forward	5' GGWMGWGGWAARGGWGGWAA 3'	1 - 20
Hist4 reverse	5' CCCRTAVAGVGTNCKKCCYTG 3'	274 - 294

Restriction sites inserted at the 5'end of some primers are underlined.

sequences were aligned and the collated sequence of both strands was obtained (Fig. 2).

Theoretical analysis and structural predictions - The full-length sequence of the SmRbAp48 cDNA is 1036 bp long, with an ORF of 927 bp which encodes a 308 aa protein with a calculated molecular mass of 34,965 Da and estimated pI of 5.34 (Figs 1, 2). The sequence revealed a 5' UTR of 85 bp and a poly-A tail of 36 bp. The putative ATG initiation codon was chosen due to the absence of another ATG codon on previous positions in the same frame. However, the context where this codon is present is not in agreement with the Kosak consensus sequence, that signals the translation initiation in eukaryotes (Kosak 1987). The amino acid sequence of the SmRbAp48 translation product is rich in leucine (8.4%), aspartate (8.1%) and serine (8.1%), but very poor in tyrosine and glutamine (2.9%), tryptophan (2.3%) and cysteine (1.6%).

Protein database searches revealed that the SmRbAp48 protein shows 84% identity with human RbAp48. Fifteen putative phosphorylation sites were identified: one for cAMP- and cGMP-dependent protein kinases (residues 25-28), six for protein kinase C (residues 56-58, 99-101, 111-113, 117-119, 162-164 and 259-261), eight for casein kinase II (residues 6-9, 56-59, 71-74, 95-98, 160-163, 194-197, 227-230 and 244-247), suggesting that the protein can be regulated by phosphorylation. No nuclear localization signal, which is responsible for recognition of the protein by nuclear transport factors, is found on SmRbAp48. However, this protein is not very big (< 50 kDa) and its small size could make possible for it to cross through the nuclear pore by a passive diffusion process.

SmRbAp48 also has at its carboxy terminus four WD repeat motifs (residues 121-154, 174-205, 224-255 and 270-301), two of them starting with GH and ending in WD. This characteristic is seen in classical WD repeat motifs found in members of this protein family. The motifs fold as β-sheets composed each one by four antiparallel β-strands, forming the blades of a β-propeller (Neer et al. 1994; Garcia-Higuera et al. 1996). An hypothetical 3D structure for the carboxy end of the SmRbAp48 protein was obtained by computational molecular modeling based on the homology to the G protein β-subunit, the archetypal of this structural family (Fig. 3). The model clearly demonstrated the presence of four blades forming a β-propeller. Each blade is a β-sheet formed by four antiparallel β-strands, three of them derived from one WD repeat and the fourth b-strand from the following WD repeat. It was not possible to obtain an hypothetical 3D model for the first 121 amino acids of SmRbAp48, once there is not any protein structure with enough similarity to this region of the parasite protein. Nevertheless, according to the secondary structure prediction, using the PredictProtein server, this region could form two α-helices (results not shown). The helices might stabilize the β-propeller structure of SmRbAp48, as has been shown for the G protein β-subunit, which needs to be associated with the two a-helices of the g-subunit to be crystallized (Sondek et al. 1996).

Expression of SmRbAp48 in S. mansoni life cycle stages - cDNA libraries of different stages of the parasite life cycle (egg, cercariae, 3h schistosomulum and adult worm) were amplified by PCR using the primers Rb435

and RbAp48 reverse (Table). Amplicons of 429 bp were detected in all the stages of the life cycle evaluated, indicating that this gene is expressed throughout the parasite development (Fig. 4).

Cloning and sequencing of Histone H4 gene of S. mansoni - In order to verify, in the future, if the protein SmRbAp48 plays a similar role as its human counterpart in histone metabolism, being able to interact with the Hi-

```
5' AAACTGACCGTTTACACTTAACAAAGCATCTGTGTAATTATTCAAAGGAGGTTACGCCGTTTCTGAAGCGTAT   73

   TCTATAATAAAA ATG ATG TTG CAT CCT TCG GAT TCT GAA GAC ATT GTC GAA GAG AGA  130
                Met Met Leu His Pro Ser Asp Ser Glu Asp Ile Val Glu Glu Arg   15

   GTA ATA AAC GAA GAA TAC AAG ATA TGG AAG AGG AAT ACT CCG TTC TTG TAC GAT  184
   Val Ile Asn Glu Glu Tyr Lys Ile Trp Lys Arg Asn Thr Pro Phe Leu Tyr Asp   33

   ATG CTG ATG TCA CAC TGC TTG GAA TGG CCA AGT TTA ACT GCC CAA TGG TTG CCA  238
   Met Leu Met Ser His Cys Leu Glu Trp Pro Ser Leu Thr Ala Gln Trp Leu Pro   51

   TCT GTG GAA AGG ACT GGG CGA GAT TAC TCC GTT CAT CGT TTA ATA CTC GGG ACT  292
   Ser Val Glu Arg Thr Gly Arg Asp Tyr Ser Val His Arg Leu Ile Leu Gly Thr   69

   CAC ACA TCT GAT GAG CAA AAT CAC TTG TTG ATA GTT ACG GTT CAT CTA CCA AAT  346
   His Thr Ser Asp Glu Gln Asn His Leu Leu Ile Val Thr Val His Leu Pro Asn   87

   GAC CAG GCG GAG TTT GAT GCA AGT GCT TAT GAT AGT GAA CGA GGT GAT TTC GGG  400
   Asp Gln Ala Glu Phe Asp Ala Ser Ala Tyr Asp Ser Glu Arg Gly Asp Phe Gly  105

   GGA TTT TAT TTT CCA TCT GGG AAG TTG GAA ATA TCA ATG AAA ATA AAT CAT GAA  454
   Gly Phe Tyr Phe Pro Ser Gly Lys Leu Glu Ile Ser Met Lys Ile Asn His Glu  123

   GGC GAA GTC AAT CGT GCT AGG TTT ATG CCA CAG AAC CCA GAC ATA ATA GCT ACC  508
   Gly Glu Val Asn Arg Ala Arg Phe Met Pro Gln Asn Pro Asp Ile Ile Ala Thr  141

   AAA ACA CCA AGT GGT GAT GTT TTA ATA TTC AAT TAT CCA AGA CAT CCA CCG AAA  562
   Lys Thr Pro Ser Gly Asp Val Leu Ile Phe Asn Tyr Pro Arg His Pro Pro Lys  159

   ACC CCA TCA GAC CGT GGT TGC CAA CCT GAT CTA CGT CTC AAG GGT CAT CAA AAA  616
   Thr Pro Ser Asp Arg Gly Cys Gln Pro Asp Leu Arg Leu Lys Gly His Gln Lys  177

   GAA GGT TAT GGT CTT TCA TGG AAT GTG TCT CTT AAT GGT CAT CTT CTT TCA GCG  670
   Glu Gly Tyr Gly Leu Ser Trp Asn Val Ser Leu Asn Gly His Leu Leu Ser Ala  195

   TCT GAT GAT CAG ACA ATT TGT TTA TGG GAT GTT AAT GCT GCT CCT TTA GAT GGC  724
   Ser Asp Asp Gln Thr Ile Cys Leu Trp Asp Val Asn Ala Ala Pro Leu Asp Gly  213

   TGT GAT CTA GAT GCG ATG GCT ATC TTT ACG GGT CAT CAT TCA GTA GTT GAG GAC  778
   Cys Asp Leu Asp Ala Met Ala Ile Phe Thr Gly His His Ser Val Val Glu Asp  231

   GTT TCC TGG CAC CTT TTC CAT GGA CAT ATT TTT GGT TCA GTA GCA GAT GAT AAT  832
   Val Ser Trp His Leu Phe His Gly His Ile Phe Gly Ser Val Ala Asp Asp Asn  249

   AAA CTT ATG GTT TGG GAT ACA CGG AGT TCA AAT CGT ACA AAA CCT CAG CAC CAA  886
   Lys Leu Met Val Trp Asp Thr Arg Ser Ser Asn Arg Thr Lys Pro Gln His Gln  267

   GTG GAT GCT CAT ACA GCC GAA GTC AAT TGT CTT GCT TTT AAT CCA TTT TCT GAG  940
   Val Asp Ala His Thr Ala Glu Val Asn Cys Leu Ala Phe Asn Pro Phe Ser Glu  285

   TTT ATT ATT GCT ACA GGA AGT GCG GAC AAA GTA ATT AAG TAT TTT ACC CTC GTA  994
   Phe Ile Ile Ala Thr Gly Ser Ala Asp Lys Val Ile Lys Tyr Phe Thr Leu Val  303

   TCT TTT TTT TAC ACA TAA ATA ACC AAA TAA CAT TCA TGC AGT 3'             1036
   Ser Phe Phe Tyr Thr ***                                                308
```

Fig. 2: nucleotide sequence of the SmRbAp48 cDNA with its deduced amino acids. The nucleotide sequence obtained by the hemi-nested PCR strategy is in bold italics. The four tryptophan/aspartate repeat motifs in the protein are double underlined and the fifteen putative sites for phosphorylation are in gray. Conserved residues essential for stabilization of the β-propeller fold are boxed. The SmRbAp48 cDNA sequence is available in GenBank™ with acession number AF297468.

Fig. 3: threotical three-dimensional structure of the SmRbAp48 carboxy terminus obtained by molecular modelling using the β subunit of G protein as a model (PDB ID 1SCG). A β-propeller of four blades, each one formed by a β-sheet composed by four antiparallel β-strands is shown.

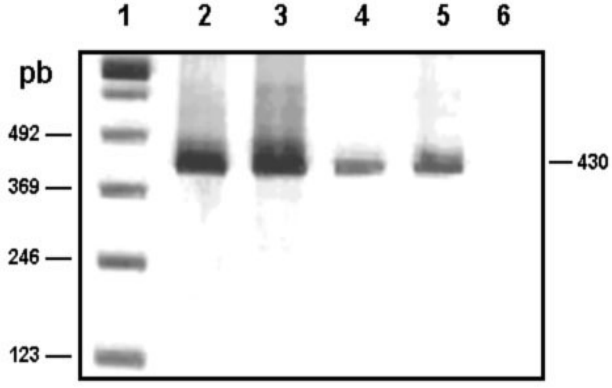

Fig. 4: PCR amplification of cDNA libraries of several stages of the *Schistosoma mansoni* life cycle using primers Rb 435 and RbAp48 reverse. Lane 1: 123 bp Ladder molecular weight marker (GIBCO, BRL); 2: egg; 3: cercariae; 4: 3h schistosomulum; 5: adult worm library; 6: negative control (without DNA). The arrow indicates the specific 429 bp amplicons.

stone H4 protein, degenerated primers were designed to amplify the Histone H4 gene from *S. mansoni*. The primers were designed based on a detailed inspection in the alignment of Histone H4 gene sequences from diverse organisms, using the Multalin Program (http://www.protein.toulose.inra.fr/multalin/multalin.html) (Fig. 5). The designed primers Hist4 forward and reverse (Table) target conserved regions of the gene, but the amplification product lacks the initial and final portions of the gene. The Histone H4 gene fragment was amplified from *S. mansoni* adult worm genomic DNA and from *D. melanogaster* and mice genomic DNA, which were used as controls in the experiment (Fig. 6). It is noticeable that all amplified bands have the same size, since Histone H4 genes do not contain introns and are highly conserved during evolution. The amplified fragment of the SmH4 gene was cloned into pUC18 vector and fully sequenced. The partial sequence of SmH4 is 294 bp, encoding a putative protein of 98 aa, with a calculated molecular mass of 10,887 Da and an estimated pI of 11.36. The translation product does not contain the first two and the last three amino acids present in Histone H4 from other organisms (Fig. 7). Database homology search results show 100% identity between SmH4 and the *D. melanogaster* protein. Theoretical analysis reveals the presence of four probable phosphorylation sites (residues 43-46, 70-73, 76-79 and 81-84), two nuclear localization signatures (residues 15-18 and 16-19) and four important lysine residues, conserved in all histones of this class (residues 4, 7, 11 and 15) (Fig. 7).

DISCUSSION

This work reports on the cloning and characterization of RbAp48 and histone H4 genes of *S. mansoni*. The SmRbAp48 gene encodes a putative nuclear protein, member of the WD repeat family, presenting four WD repeats on its carboxy end, which is probably regulated by phosphorylation. RbAp48 proteins are very conserved, and this high degree of conservation during the course of evolution is indicative of their functional importance in biological processes. The predicted amino acid sequence of SmRbAp48 shows 84% of similarity and 72% of identity to the human RbAp48.

The WD repeat was first described in the β subunit of heterotrimeric GTP-binding proteins, which transduce signals across the plasma membrane (Fong et al. 1986). WD repeat proteins are constituted of highly conserved repeating units, usually ending with Trp-Asp (WD) and have been found in all eukaryotes, but not in prokaryotes. The number of repeats vary from four to nine in different proteins (Garcia-Higuera et al. 1996). Detailed analysis of these repeats shows that there are four residues almost totally invariant, occupying strategic positions in loops and β-strand regions of the protein structure which in turn, by a network of hydrogen bonds, link the blades of the propeller and are essential for stabilization of the correct fold (Branden & Tooze 1999). They are boxed in the SmRbAp48 sequence (Fig. 2). The β-propeller structure is unstable as a monomer, as seen for the β subunit of G protein which becomes stable when associated with the γ subunit of this protein. The later one is unfolded in vitro and folds into two a-helices when forming a dimer with the β subunit. Their association is partially due to hydrophobic interactions between the long N-terminal α-helix of the β subunit and the N-terminal α-helix of the γ subunit of G protein (Branden & Tooze 1999). According to secondary structure predictions, the amino terminus of SmRbAp48 could form two a-helices. These helices might stabilize the β-propeller structure of SmRbAp48 after its association to another protein, as showed for the β-subunit of G protein, which needed to be associated to the γ-subunit to be crystallized (Sondek et al. 1996). The search for new SmRbAp48 partners and the understanding of the molecular mechanisms govern-

ing the interactions between them is of great interest to help elucidate the biological function of this protein in the parasite.

There are relatively few papers on the characterization of transcription factors and control of gene expression in *S. mansoni*, and none of them deals with the study of repression of gene transcription and histone metabolism. Another function attributed to RbAp48 is to be part of a basal repression complex, composed by HDAC-1 and 2 and RbAp46 (Knoepler & Eisenman 1999). This complex is recruited by a variety of co-repressors and repression associated factors to strength the transcriptional repression during the cell life cycle (Knoepler & Eisenman 1999). Thus, SmRbAp48 may consist of a factor participating in the repression of gene expression in the parasite. In a dbEST search, we found ESTs homologous to human HDAC3 (gi 5790729) and HDAC8 (gi 4224573, 4225199,

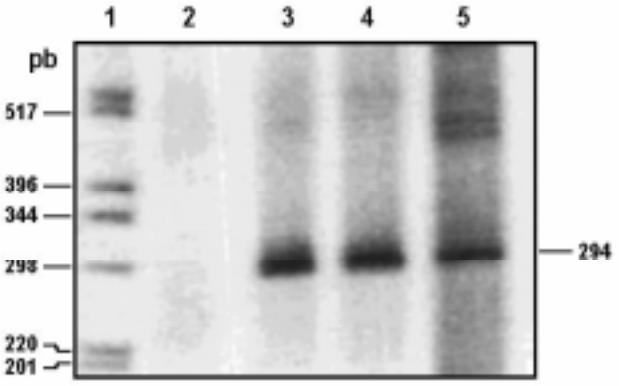

Fig. 6: PCR amplification of genomic DNA from diverse organisms using the degenerated primers Hist4 forward and reverse. 1: 1 Kb Ladder molecular weight marker (GIBCO, BRL); 2: negative control (without DNA); 3: *Drosophila melanogaster*; 4: mice; 5: *Schistosoma mansoni*. The arrow indicates the specific 294 bp amplicons.

HIST4 forward

```
           1                                30                                60                                90       100
           |--------------------------------+---------------------------------+---------------------------------+--------|
elegans    ATGTC GG CGTGG AAGGG GG AAAGG CT GG AAAGGAGG GCCAAGCGTCATCGCAA GT CTTCG GA AACATCCA GGTATCACCAAGCC G
Bos        ATGTC GG CGTGG AA GG GG AAAGG CT GG AAAGGAGG GC AAGCG CA CGCAAGGT CT CG GACAA ATCCAGGGTATCACCAAGCC G
chicken    ATGTC GG  G GG AAGGG GG AA GG CT GG AA GG GG GCCAAGCG CA CGCAAGGT CT CG GACAACATCCAGGG ATCACCAAGCC G
drosoph    ATG C GG CGTGG AAGGG GG AAAGGA TGGG AA GG GG GCCAAGCGTCATCG AAGGT CTTCG GA AACATCCAGGGTATCACCAAGCC G
arabidop   ATGTC GG CGTGG AAGGG GG AAAGG  TGGG AAAGGAGG GCCAAGCGTCA  G AAGGT CT  G GACAACATCCA GG ATCACCAAGCC G
apis       ATG C GG CGTGG AAGGG GG AAAGGA TGGG AA GGAGG GC AAGCGTCAT G AA GT CTT G GA AACATCCA GGTATCAC AA CC G
ascaris    ATGTC GG CGTGG AA GG GG AAAGGACTGGG AAAGG GG GC AAGCGTCATCGCAAGGT CTTCG GACAACATCCAGGG ATCAC AAGCC G
Consensus  ATGtC GG cGtGG AAgGG GG AAaGGacTgGG AAaGGaGG GCcAAGCGtCAtcGcAAgGT CTtcG GAcAAcATCCAgGGtATCACcAAgCC G

           101                              130                               160                               190      200
           |--------------------------------+---------------------------------+---------------------------------+--------|
elegans    C AT CG CG CT GC  G  GAGGTGG GT AAG G ATCTCTGG CTCATCTACGAGGA AC CG GG GT CT AAGGT TTCCT GAGAA GT AT
Bos        C ATCCG CG CTGGCTCGTCG GGTGGTGT AAGCG ATCTC GG CTCATCTACGAGGAGAC CG GG GT CT AAGGT TT CT GA AA GT AT
chicken    C ATCCG CG CTGGC CG CG GG GG GT AAGCG ATCTC GG CTCATCTACGAGGAGAC CGCGG GT CT AAGGT TTCCT GAGAACGT AT
drosoph    C AT CG CG  TGGCTCGTCG GG GGTGT AAGCGTATCTCTGG  T AT TACGAGGA ACTCGCGGTGT CT AAGGT TTCCT GAGAACGTTAT
arabidop   C ATCCG  G  TGGCTCGT GAGGTGG GT AAGCGTATC  TGG  T ATCTACGAGGAGACTCG GG GT CT AAG T TT CT GAGAACGT AT
apis       C AT CG CG  TGGCTCGTCG GG GGTGT AA CGTAT TCTGG  T AT TA GA GA ACTCGCGGTGT  T AA GT TTC T GA AA GTTAT
ascaris    C ATCCG CG CT GCTCGTCGAGGTGGTGT AAGCGTAT TCTGG CTCATCTACGAGGAGACTCGC GTGT CT AAGGT TT CT GA AACGTT T
Consensus  C ATcCG cG cTgGCtcGtcGaGGtGGtGT AAgcGtATctctGG cTcATcTAcGAgGAgACtCGcgGtGT cT AAggT TTccT GAgAAcGTtaT

           201                              230                               260                               290      300
           |--------------------------------+---------------------------------+---------------------------------+--------|
elegans    CCGTGATGC GTCAC TAC  CGAGCACGC AAG G AAGAC GT AC GCCATGGA GT GTCTA GC  T AAG G CA GGA G AC CT TACGGA
Bos        CCG GA GC GTCAC TACAC GAGCACGC AA CG AAGAC GT AC GC ATGGA GT GT TA GC CT AA  G CA GG  G AC CT TA GG
chicken    CCG GA GC GTCAC TACACCGAGCACGC AAG G AAGAC GT AC GCCATGGA GT GTCTA GC CT AAG G CA GGA G AC CT TACGG
drosoph    CCGTGA GC GTCAC TACACCGAGCACGC AAGCG AAGAC GT AC GCCATGGA GT GTCTA GC CT AAG G CA GGA G AC CT TACGGA
arabidop    CGTGATGC GTCAC TACAC GAGCACGC A G G AAGAC GT AC GCCATGGA GT GTCTA GC CT AAG G CA GGA G AC CT TACGGA
apis       CCG GATGC GT AC TACACCGA CAC C AA CG AA AC GT AC GC ATGGA GT GT TA GC CT AA    CA GGA G AC CT TA GGA
ascaris    C  T ATGC   C C  ACACC  G   GC  A CG    AC G   C GC  T  A
Consensus  ccgtgAtGC gtcaC tACaccgagcacgC aagcG aagAC Gt aC GCcaTggA gt gtcta gc ct aag g ca gga g ac ct tacgga

           301      312                                                                                  HIST4 reverse
           |----------|
elegans    TT GG GG TAA
Bos        TT GGCGG TAA
chicken    TT GGCGG TAA
drosoph    TT GGCGG TAA
arabidop   TT GGCGG TAA
apis       TT GGCGG TAA
ascaris
Consensus  tt ggcgg taa
```

Fig. 5: multiple alignment of histone gene sequences from diverse organisms using the Multalin program. Sites for annealing of the degenerated primers Hist4 forward and reverse are indicated by arrows. Elegans (*Caenorhabditis elegans* histone H4, gi 17541085); bos (*Bos taurus* histone H4, gi 2981287); chicken (chicken histone H4, gi 211905); drosoph (*Drosophila melanogaster* histone H4, gi 17975541); arabidop (*Arabidopsis thaliana* histone H4, gi 166741); apis (*Apis mellifera* histone H4, gi 1883000), ascaris (*Ascaris lumbricoides* histone H4, gi 1177237)

```
5' GGA AGA GGA AAG GGA GGA AAG GGC CTG GGG AAA GGG GGT GCC AAG CGC CAC CGC  54
    Gly Arg Gly Lys Gly Gly Lys Gly Leu Gly Lys Gly Gly Ala Lys Arg His Arg  18

    AAG GTC CTG CGC GAC AAC ATC CAG GGT ATC ACC AAG CCA GCC ATT CGG CGC CTT 108
    Lys Val Leu Arg Asp Asn Ile Gln Gly Ile Thr Lys Pro Ala Ile Arg Arg Leu  36

    GCT CGC CGC GGC GGC GTG AAG CGC ATT TCT GGC CTC ATC TAT GAG GAG ACC CGC 162
    Ala Arg Arg Gly Gly Val Lys Arg Ile Ser Gly Leu Ile Tyr Glu Glu Thr Arg  54

    GGA GTG CTG AAG GTG TTC CTG GAG AAC GTG ATT CGT GAT GCT GTG ACT TAC ACG 216
    Gly Val Leu Lys Val Phe Leu Glu Asn Val Ile Arg Asp Ala Val Thr Tyr Thr  72

    GAG CAC GCC AAA CGC AAG ACA GTG ACA GCG ATG GAT GTG GTC TAC GCG CTG AAG 270
    Glu His Ala Lys Arg Lys Thr Val Thr Ala Met Asp Val Val Tyr Ala Leu Lys  90

    AGA CAG GGC CGC ACT CTG TAT GGG 3'                                     294
    Arg Gln Gly Arg Thr Leu Tyr Gly                                         98
```

Fig. 7: nucleotide sequence of the SmH4 gene with its deduced amino acids. The four important lysines present in signatures for acetyltransferases and deacetylases recognition are in gray. The four probable sites for phosphorylation are double underlined. Genomic DNA sequence of SmH4 is available in GenBank^{TM} with accession number AF297469.

5788601, 5790341, 5788859), demonstrating the existence of the two different types of HDACs in schistosomes. The verification of a possible interaction between HDAC and SmRbAp48 is of interest, once Ahmad and colleagues (1999) described that chicken p48 subunit of CAF-1 interacts directely with HDAC1 and 2 in vivo through its WD repeat domain and Nicolas and colleagues (2001) found that HDAC3 interacts with human RbAp48, mediating its recruitment to the Rb protein.

Once human RbAp48 is homologous to the MSI1 gene of *S. cerevisiae*, a negative regulator of the Ras-cAMP pathway (Ruggieri et al. 1989) and SmRbAp48 is 84% similar to the human protein, it is also possible that SmRbAp48 may perform the same functions as human RbAp48 in the cell, and thus complement yeast mutant strains in which the gene MSI1 was deleted. It can be of interest to find other functions for RbAp48 in the parasite and to understand the polyvalent roles played by this protein in eukaryotic cells.

The second gene characterized was the histone H4 gene of *S. mansoni*, which shows 100% identity with histone H4 of *D. melanogaster*. It has four probable phosphorylation sites, two nuclear localization signatures and four important lysines, conserved in all H4 histones. These lysines are part of acetyltransferases and deacetylases recognition signatures – $GRGK_5GGK$, $KGGK_8GLG$ and $GLGK_{12}GGA$ – important on the process of recognition of recently modified histones for chromatin assembly (Ruiz-Carillo et al. 1975, Jackson et al. 1976). It is still unclear how human RbAp48 recognizes the helix 1 of histone H4, but the high degree of conservation of the WD repeats in RbAp48 proteins from different organisms is indicative of the functional importance of this motif on the interaction between the two proteins.

To our knowledge, this is the first report on a gene encoding a *S. mansoni* WD repeat protein and histone H4, and certainly the verification of a possible interaction between these two proteins demands further studies. This paper describes the structural characterization of two important genes for *S. mansoni*. Some conclusions were driven from theoretical predictions, and should be validated by future experimental research aiming to elucidate the functional role of RbAp48 protein in *S. mansoni*.

ACKNOWLEDGEMENTS

To Neuza Antunes Rodrigues and Kátia Barroso (UFMG, Minas Gerais, Brazil) for technical support and automated DNA sequencing and to Dr Mohammed Saber (TRBI, Cairo, Egypt) for supplying the egg, cercariae, 3h schistosomulum, and adult worm cDNA libraries.

REFERENCES

Ahmad A, Takami Y, Nakayama T 1999. WD repeats of the p48 subunit of chicken chromatin assembly factor-1 required for in vitro interaction with chicken histone deacetylase-2. *J Biol Chem 274*: 16646-16653.

Altschul SF, Thomas L, Madden AA, Schaffer JZ, Zheng Z, Webb M, David JL 1997. Gapped blast PSI-Blast; a new generation of protein database search programs. *Nucl Acids Res 25*: 3389-3402.

Branden C, Tooze J 1999. Signal transduction. In C Branden & J Tooze (eds), *Introduction to Protein Structure*, Garland Publishing, INC, New York, p. 251-281.

Franco GR, Valadão AF, Azevedo V, Rabelo EML 2000. The *Schistosoma* gene discovery program; state of the art. *Int J Parasitol 30*: 453-463.

Fong HK, Hurley JB, Hopkins RS, Miake-Lye R, Johnson MS, Doolittle RF, Simon MI 1986. Repetitive segmental structure of the transducin beta subunit: homology with the CDC4 gene and identification of related mRNAs. *Proc Natl Acad Sci USA 83*: 2162-2166.

Garcia-Higuera I, Fenoglio J, Lewis C, Panchenko MP, Reiner O, Smith TF, Neer EJ 1996. Folding of proteins with WD-repeats: comparison of six members of the WD-repeat superfamily to the G protein beta subunit. *Biochemistry 35*: 13985-94.

Jackson V, Shires A, Tanphaichitr N, Chalkley R 1976. Modifications to histones immediately after synthesis. *J Mol Biol 104*: 471-483.

Kaufman PD, Kobayashi R, Stillman B 1997. Ultraviolet radiation sensitivity and reduction of telomeric silencing in *Saccharomyces cerevisiae* cells lacking chromatin assembly factor-I. *Genes Dev 11*: 345- 357.

Knoepfler PS, Eisenman RN 1999. Sin meets NuRD and other tails of repression. *Cell 99*: 447-50.

Kosak M 1987. An analysis of 5'-coding region sequence from 699 vertebrate messenger RNAs. *Nucl Acids Res 15*: 8125-8148.

Krude T 1999. Chromatin assembly during DNA replication in somatic cells. *Eur J Biochem 263*: 1-5.

Lee WH, Hollingsworth RE Jr, Qian YW, Chen PL, Hong F, Lee EY 1991. RB protein as a cellular "corporal" for growth-promoting proteins. *Cold Spring Harbor Symp Quant Biol 56*: 211-217.

Neer EJ, Schmidt CJ, Nambudripad R, Smith TF 1994. The ancient regulatory-protein family of WD-repeat proteins. *Nature 371*: 297-300.

Nicolas E, Ait-Si-Ali S, Trouche D 2001. The histone deacetylase HDAC3 targets RbAp48 to the retinoblastoma protein. *Nucl Acids Res 29*: 3131-3136.

Nicolas E, Morales V, Jaulin LM, Harel-Bellan A, Richard-Foy H, Trouche D 2000. RbAp48 belongs to the histone deacetylase complex that associates with the retinoblastoma protein. *J Biol Chem 275*: 9797-9804.

Parthun MR, Widom J, Gottschling DE 1996. The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. *Cell 87*: 85-94.

Qian YW, Lee EYHP 1995. Dual retinoblastoma-binding proteins with properties related to a negative regulator of Ras in yeast. *J Biol Chem 270*: 25507-25513.

Qian YW, Wang YCJ, Hollingsworth RE, Jones D, Ling N, Lee EYHP 1993. A retinoblastoma-binding protein related to a negative regulator of Ras in yeast. *Nature 364*: 648-652.

Ruggieri R, Tanaka K, Nakafuku M, Kaziro Y, Toh-e A, Matsumoto K 1989. MSI1, a negative regulator os Ras-cAMP pathway in *Saccharomyces cerevisiae*. *Proc Nat Acad Sci USA 86*: 8778-8782.

Ruiz-Carrillo A, Wangh LJ, Allfrey VG 1975. Processing of newly synthesized histone molecules. *Science 190*: 117-128.

Santos FR, Pena SDJ, Epplen JT 1993. Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. *Hum Genet 90*: 655-656.

Simpson AJG, Sher A, McCutchan TF 1982. The genome of *Schistosoma mansoni*: isolation of DNA, its size, bases and repetitive sequences. *Mol Biochem Parasitol 22*: 169-176.

Sondek J, Bohm A, Lambright DG, Hamm HE, Sigler PB 1996. Crystal structure of a G-protein beta gamma dimer at 2.1A resolution. *Nature 379*: 369-375.

Verreault A, Kaufman PD, Kobayashi R, Stillman B 1996. Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. *Cell 87*: 95-104.