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 irdl and Ras2val19 mutant strains and reduces the cAMP levels in these 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).

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.

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
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 SystemsTM (Promega).

DNA cloning - Two identical clones (MAADD0269 and MAADD0270) 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 Rsal 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 Smal 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 MgCl2, 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 35 cycles. The amplicons were analyzed in 1% agarose gel stained by ethidium 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 kitTM 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 (ORF) 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 (MAADD0269 and MAADD0270) 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 Rsal 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 using the DNASIS program, all the PCR products were cloned into the Smal site of pUC18 vector and sequenced as described above (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). SmRbAp48 insert was amplified into the vector, 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 hemi-nested 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
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 structures, 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 β-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 by forming the blade of a β-subunit, which needs to be associated with the two α-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...
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 GenBankTM with accession number AF297468.

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-
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.toulouse.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). According to secondary structure predictions, the amino terminus of SmRbAp48 could form two α-helices after its association to another protein, as shown 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 strengthen 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, 4224528).

![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.](image)

![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.](image)
### 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 with accession number AF297469.

<table>
<thead>
<tr>
<th>GenBank Accession</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AF297469</td>
<td>Genomic DNA sequence of SmH4 gene</td>
</tr>
</tbody>
</table>

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 directly 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$_2$GGK, KGGK$_2$GLG and GLGK$_2$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


