Cloning and Characterization of SmZF1, a Gene Encoding a *Schistosoma mansoni* Zinc Finger Protein

Paulo R Eleutério de Souza*, Analina F Valadão*, Carlos E Calzavara-Silva*, Glória R Franco*, Marcos A de Morais Júnior, Frederico GC Abath**/⁺

Laboratório de Imunopatologia Keizo Asami and Departamento de Genética, Universidade Federal de Pernambuco, Recife, PE, Brasil *Laboratório de Genética-Bioquímica, Departamento de Bioquímica e Imunologia, ICB, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil **Departamento de Imunologia, Centro de Pesquisas Aggeu Magalhaes-Fiocruz, Av. Moraes Rego s/nº, Cidade Universitária, 50670-420 Recife, PE, Brasil

The zinc finger motifs (Cys_2His_2) are found in several proteins playing a role in the regulation of transcripton. SmZF1, a Schistosoma mansoni *gene encoding a zinc finger protein was initially isolated from an adult worm cDNA library, as a partial cDNA. The full sequence of the gene was obtained by subcloning and sequencing cDNA and genomic fragments. The collated gene sequence is 2181 nt and the complete cDNA sequence is 705 bp containing the full open reading frame of the gene. Analysis of the genome sequence revealed the presence of three introns interrupting the coding region. The open reading frame theoretically encodes a protein of 164 amino acids, with a calculated molecular mass of 18,667 Da. The predicted protein contains three zinc finger motifs, usually present in transcription regulatory proteins. PCR amplification with specific primers for the gene allowed for the detection of the target in egg, cercariae, schistosomulum and adult worm cDNA libraries indicating the expression of the mRNA in these life cycle stages of* S. mansoni. *This pattern of expression suggests the gene plays a role in vital functions of different life cycle stages of the parasite. Future research will be directed to elucidate the functional role of SmZF1.*

Key words: DNA binding protein - gene cloning - *Schistosoma mansoni* - zinc finger protein

Schistosomiasis is a human disease caused by several trematodes of the genus *Schistosoma,* with approximately 200 milion people currently infected and a further 500-600 milion worldwide at risk of infection throughout tropical and subtropical areas of the world (Savioli et al. 1997). In South America, *S. mansoni* is the only causative agent of schistosomiasis. Schistosome worms have a complex life-cycle involving molluscus and vertebrate hosts as well as short periods of larvae swimming freely in water. The evolution from one developmental stage to another involves modifications in the morphology, physiology and biochemistry of the parasite, being associated with the activation/inactivation of stage-specific genes. Indeed, several stage-specific genes with different patterns of

expression during the schistosome life-cycle have been described (Simpson et al. 1984, Davis et al. 1985, Grossman et al. 1990, Chen et al. 1992, Abath et al. 1994, Mei & Lo Verde 1997).

The recognition of specific sequences of DNA by proteins is a central mechanism for a number of important biological processes, including gene expression, recombination, and DNA repair (Mitchell & Tjian 1989). Many of those proteins, including transcription factors, present zinc finger motifs in their structure, that are responsible for the binding to DNA sequences (Bernstein et al. 1994, Yokono et al. 1998, Clarke & Berg 1998). The classical C_2H_2 zinc finger motif consists of about 30 amino acids, four of which (two cysteines and two histidines - Cys_2His_2) coordinate tetrahedrically a single zinc atom, forming a loop of twelve amino acids between the second cysteine residue and the first histidine residue. The cysteine residues are separated by two or four amino acids and the two histidine residues are separated by three to five amino acids (Berg 1990, Bernstein et al. 1994). The sequence, number and organization of the zinc finger motifs, are important for the biological function of the protein.

The study of gene expression regulation in *S. mansoni* is still incipient. Thus, the molecular char-

acterization of these regulatory proteins in *Schistosoma* may contribute to a better understanding of the biology of the parasite as well as the evaluation of these proteins as targets for immunotherapy or drug therapy. The present communication report on the cloning and structural characterization of SmZF1, a gene encoding a putative transcription regulatory protein of *S. mansoni* containing three zinc finger motifs. In addition, evidence is provided indicating that SmZF1 is expressed in different life-cycle stages of the parasite.

MATERIALS AND METHODS

DNA purification and cDNA libraries - Genomic DNA was purified from *S. mansoni* LE strain adult worms as described previously (Simpson et al. 1982). The plasmids and polymerase chain reaction (PCR) fragments were purified with the Wizard DNA Purification Systems™ (Promega). *S. mansoni* adult worm cDNA libraries were constructed in λgt11 (Abath et al. 1993). In addition, egg, cercariae, 3 h schistosomulum, and adult worm cDNA libraries UUEE and SmZU were constructed in λZAP as part of the *Schistosoma* genome project (Franco et al. 2000).

PCR - A number of primers, targeting specific regions of the gene, were used to amplify DNA fragments for cloning into the vectors pBlueScript KS$^+$ (Stratagene) and pUC18 (Amersham Pharmacia Biotech), and further sequencing (Table, Fig. 1).

For the amplification of the SmZF1 gene from the λgt11 library, 100 µl reaction mixture was used containing approximately 4 µl of the cDNA library, 10 mM Tris HCl pH 8.4, 50 mM KCl, 1.5 mM MgCl$_2$, 200 µM each deoxynucleotide triphosphate, 250 nM each primer and 2.5 U of Taq polymerase. The conditions used for amplification were 94°C for 4 min, followed by a step cycle program set to denature at 94°C for 1 min, anneal at 55°C for 1 min, and extend at 72°C for 2 min for a total of 30 cycles. Amplifications of the other cDNA libraries were performed in a 30 µl volume containing 1 µl of the cDNA library, 10 mM Tris HCl pH 8.3, 75 mM KCl, 3.5 mM MgCl$_2$, 200 µM each deoxynucleotide triphosphate, 200 nM each primer and 2 U of Taq DNA polymerase. 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 94°C for 4 min, followed by a step cycle program set to denature at 94°C for 1 min, anneal at 52°C for 1 min, and extend at 72°C for 1 min for a total of 30 cycles. The amplicons were analyzed in 1% agarose gel stained by ethidum bromide or in 6% polyacrylamide gels silver stained (Santos et al. 1993).

DNA cloning and sequencing - An adult worm λgt11 cDNA library was screened by hybridization with the cDNA A157 (GenBank U67153) as a probe, following the instructions of the Digoxigenine-DNA labelling/Detection kit (Boehringer Mannheim). The PCR product of the cDNA clone isolated was cloned into the *EcoR*I digested pBlueScript KS$^+$. Other clones were inserted into the *Sma*I site of pUC18 using the Surclone ligation kit™ (Amersham Pharmacia Biotech).

The sequencing reactions were performed using the Thermo Sequenase fluorescent labelled primer cycle sequencing kit™ with 7-deaza-dGTP (Amersham Pharmacia Biotech). Fluorescent primers targeting the margins of the cloning sites were used for DNA sequencing of both strands, using the A.L.F. DNA Automated 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 posttranslational modification sites, whereas the "SDSC1" - SDSC Protein Structure Homology Modeling Server (http://cl.sdsc.edu/hm.html) was used for prediction of the three-dimensional structure of the SmZF1 gene product, using the Zif268 mouse protein (PDB accession 1mey_C) as a model, due to its high similarity with SmZF1. The hydrobicity plot was calculated according to Kyte and Doolittle (1982).

RESULTS

Cloning and sequencing of SmZF1, a gene encoding a S. mansoni zinc finger protein - The cDNA ZNF-17 was casually isolated from an adult worm λgt11 cDNA library during attempts to obtain the complete cDNA for Sm13, a *S. mansoni* tegumental antigen. This cDNA was amplified and cloned into pBlueScript KS$^+$, and DNA sequencing revealed that the 603 bp cDNA shared no significant homology to the A157 clone (GenBank U67153) of the Sm13 gene (Abath et al. 2000). Instead, it encoded a 88 amino acid protein which contained three zinc finger motifs, although lacking its N-terminal region.

On the basis of the sequence of the ZNF-17 cDNA, several primers were sinthesized to sequence the complete open reading frame of the gene and detect the existence of possible introns (Fig. 1). The 446 bp fragment SmZU-5R amplified from a λZAP adult worm cDNA library using primers ZnFUpper and ZFR1 allowed for the identification

of an additional sequence to the 5' region, while the 349 bp fragment UUEE-3F amplified from another λZAP adult worm cDNA library with primers ZFF1 and ZnFLower, identified the poly-A tail of the 3' region of the gene.

Characterization of the intron-exon pattern of SmZF1 - Primers were designed based on the collated SmZF1 cDNA sequence to characterize the genomic structure of the SmZF1 gene (Fig. 1, Table). The genomic fragment GZF17-I (obtained by amplification of genomic DNA with primers ZnFStart and ZnFMR) of approximately 1.7 kb was larger than the corresponding cDNA sequence, indicating the presence of introns in that region. The partial sequencing of the terminal regions of GZF17-I allowed the identification of restriction sites (*Hind*III and *Hinc*II) (Fig. 1) used in the subcloning strategies to sequence this DNA fragment. The

sequences of the genomic fragments GZF17-II (obtained using primers ZFF1 and ZnFLower) and GZF17-III (amplification with primers ZnFMF and ZnFStop) (Fig. 1) were identical to the cDNA sequence, indicating absence of introns. After complete sequencing of the clones and subclones, sequence alignment revealed the presence of three introns: the first intron is 1408 bp (nucleotides 109 to 1516), the second 32 bp (nucleotides 1578 to 1609) and the third 37 bp (nucleotides 1697 to 1733) (Figs 1, 2). The canonical donor (GT)/acceptor (AG) splicing sites are present at all exon/intron junctions (von Heijne 1987). The mean A + T content of the introns was around 68%, higher than that observed for the coding region. The 2181 bp collated sequence may represent the region that corresponds to the primary transcript of the SmZF1 gene (Figs 1, 2).
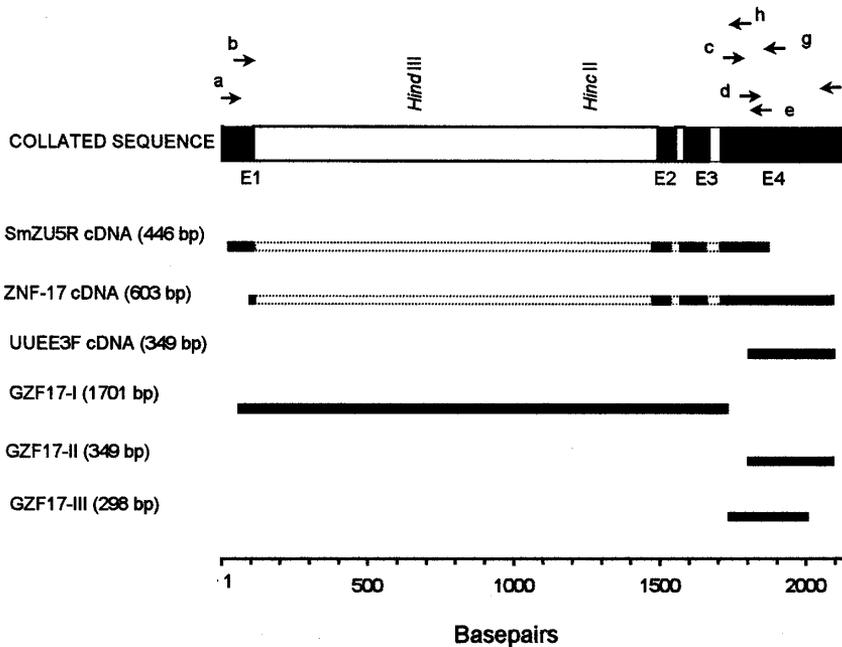
Fig. 1: schematic representation of the strategy for cloning and sequencing of the SmZF1 gene. Primers sites used for subcloning are shown. E1, E2, E3 and E4 represent exons. a: ZnFUpper; b: ZnFStart; c: ZnFMF; d: ZFF1; e: ZFR1; f: ZnFLower; g: ZnFStop; and h: ZnFMR

TABLE

Set of specific primers used for cloning and sequencing of the SmZF1 gene

Primer	Sequence	Nucleotide position
ZFR1	5'-GTTTTACACCTTTATGG-3'	Reverse (1906-1922)
ZFF1	5'-GAGTTCTGAATTGTGCG-3	Forward (1833-1849)
ZnFLower	5'-GAGGAATTCTGAAAGAATAATAAATGTA-3'	Reverse (2164-2181)
ZnFMF	5'-ACGGAAATAGATACTGAA-3'	Forward (1736-1753)
ZnFStart	5'-GAGGAATTCATGGAATTTTACTTCACA-3'	Forward (53-70)
ZnFStop	5'-GAGAAGCTTCTGGCATACTTCACAT-3'	Reverse (2018-2033)
ZnFMR	5'-TTCAGTATCTATTTCCGT-3	Reverse (1736-1753)
ZnFUpper	5'-GAGCTGCAGGTTCAGTAAAATGTTCAA-3'	Forward (1-18)

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

Theoretical analysis and structural predictions - The final SmZF1 cDNA sequence obtained revealed a 5' untranslated region (5' UTR) of 52 bp, an ORF of 492 bp which encodes a 164 aa protein with a calculated molecular mass of 18,667 Da (estimated pI of 7.62) and a 3' UTR of 161 bp without any characteristic polyadenylation site but presenting a poly-A tail. The putative ATG initiation codon is in agreement with the Kosak consensus sequence (Kosak 1987) (Fig. 2). The A+T content of the cDNA sequence, about 61% within the coding region and 65% in both 5' and 3' untranslated regions, and its codon usage are in accordance to the values described previously for *S. mansoni* genes (Milhon & Tracy 1995). The derived amino acid sequence is rich in lysine (9.8%), threonine (7.9%), phenylala-

```
GTTCAGTAAAATGTTCAATTGTGCGCTGGAATCTATTGTGTAGACTTTAACTATGGAATTTTACTTCACA   70
                                                     M  E  F  Y  F  T    6
TTGACTAAAAAGCTGAGCAAATATACCTGGAGCGTTCAGgtacgtttggggatacatcatatgttgcgtt  140
L  T  K  K  L  S  K  Y  T  W  S  V  Q                                  19
agtgtaacgcatattggtcgctacatttgatgacacttttgaatattccttaatcggattatctcgaagt  210
agaagcgtcatgggagaagaagcatttaggttttgtatttcaaatttggatataaagtaattagtatgcc  280
accatacatatttacggtgagctattttagtcgtcaaaaatatgagataatggaagcattttgaatgccg  350
attatgctatgttttttcgtcttcagcctcatgtttcactgaccggaaattcttgcttacgatagtcatcg  420
ggaatacagagttgtgttcactcgacacatctttctcaaatgccaggctaatgtataaaatcagtctact  490
aagtgaataaaattactttacattaggatgcttgtgctctttcaataacataatgctgtataacaaaaat  560
ctgtctatttcatgttcctaaagttatcagcaagtctatttgtcgttgtttttgtctcgagaaaagcttgt  630
tgtctaggtttcatcgtattctaggtattacaattaagacacaataccagccaacgtaataaaattggtc  700
ttccgtaataacctggttttttcgctgcgcatgaatatggtatatagttatataaaataactgatagtgtc  770
acgagctgctgtttggcctgataggtttgtcactgaatatctgctatctcccattcgaaattccctattt  840
gttaaaaaatccccttcaaacaaactagtaaaatatcgtgagtccaaatggtatgaaagtttggcgttga  910
tcattgtttctaaactgtagtttattgcctagatttaattagccagtaaatgacgatctctactccacat  980
ttattagtaacagacgttgagaattcggcatcatagatgacgtcaattactgaataagggttcagtcaat 1050
agaccaagatgtcatcagcccctactgccataagagcctatctagtagtagtgcgtccggacggtaccgg 1120
ttatcctacaaacatactagtaaagcagcactccaaatctttataaaaaacatccgcaacatacccatgt 1190
caccaaggttactcaactaacttatgtttccgcgtcttgatcctataagagtacatggccagcatatccc 1260
acgccagttaaagaatcagctattcaactctttaagcacgttaacagtaaacttgttgcaatatgatatt 1330
gtggttgcccccttcagattaaatactttaggtgaatgcttccgcaaacgcaaatataagttcgtagttcc 1400
catactacttctgtttactactgctactacgagttttttttttattcattttcctgtgtattcatccccgga 1470
ttcctagtatcccaatgctgttgatcttgtcgccttatatctgtagACTTTCAAGATGAACGAACCAACT 1540
                                             T  F  K  M  N  E  P  T    27
GGTGTCGGGCCAACATTTGCTGATGCATGCGATGATGgtaatctttaactataatcctatcgttttcagG 1610
G  V  G  P  T  F  A  D  A  C  D  D  G                                  40
CGAACTTATCAGCATTTGTTGTCTTTGTGGTAAAACGTTTTCAAGTCAGAGTCTTCTACACAAACATTTT 1680
 E  L  I  S  I  C  C  L  C  G  K  T  F  S  S  Q  S  L  L  H  K  H  F   63
GAATTGATGCATGAAGgtaactcgtcaagaggatttcaaacagtttaaattagGTACGGAAATAGATACT 1750
E  L  M  H  E  G                                        T  E  I  D  T  74
GAACAGTATGATCTAAGTGGATTTGCCGCTATGGGGAATGAACAAGGTCGTAAAAGTAATGGTGAAGAAG 1820
E  Q  Y  D  L  S  G  F  A  A  M  G  N  E  Q  G  R  K  S  N  G  E  E  D  98
ATGCAAATTTCCGAGTTCTGAATTGTGCGTTTTGCAACAAAGTATTTACTAAACACTGTAATTTAAACAC 1890
 A  N  F  R  V  L  N  C  A  F  C  N  K  V  F  T  K  H  C  N  L  N  T  121
ACATATCAAAGCAGTCCATAAAGGTGTAAAACCGTTTGAATGCACTTATTGTTATAAAGGATTCACTCGA 1960
 H  I  K  A  V  H  K  G  V  K  P  F  E  C  T  Y  C  Y  K  G  F  T  R  144
AATTCTGATCTTCATAAGCACATCGACGCTGTTCACAAAGGTCTCAAGCCTTTCCGGATGTGAAGTATGC 2030
N  S  D  L  H  K  H  I  D  A  V  H  K  G  L  K  P  F  R  M  *          164
CAGCGAAACTTCTCTCAGAAATCCAGCCTAAAACGACACATAGAAGCAATTCACGAAGATCCTCGGCATC 2100
GCTGAAGAGAAACCAGATTGTATAATCCTCTCCAATTTTCATATGATTTCATGTTCAAAAATATACATTT 2170
ATTATTCTTTC                                                           2181
```

Fig. 2: nucleotide sequence of the *Schistosoma mansoni* SmZF1 gene with its deduced amino acid sequence. Sequences corresponding to introns are in small letters. The three sequences corresponding to zinc finger motifs are double underlined. The two cysteine and histidine residues of each finger putatively involved in tetrahedral coordination of a zinc atom are represented in bold. A poly-A tail was identified following the last nucleotide. cDNA and genomic sequences reported in this paper are available in GenBank™ with accession numbers AF316827 and AF316828, respectively.

nine (7.9%) and glycine (7.3%), however very poor in arginine (2.4%) and tryptophan (0.6%).

Protein database searches revealed that the derived SmZF1 protein shows about 47% similarity with many eukaryotic zinc finger proteins that are involved in transcriptional regulation (data not shown). When EST database (dbEST) was searched, one *S. mansoni* cDNA sequence (GenBank accesion number AA559461) with similarity to the gene described herein was found, indicating the isolation of a homologous cDNA by the Schistosoma Genome Initiative. An attribute of the SmZF1 protein was the presence of four amino acid residues between the two histidines residues in all three zinc finger motifs (Fig. 2), whereas in the other proteins analyzed three residues predominated at that position (not shown). The three Cys_2His_2 zinc fingers motifs of 22 aa extension fit to the consensus sequence $Cys-X_{2-4}-Cys-X_3-Phe-X_5-Leu-X_2-His-X_{3-5}-His$ (Berg 1990, Bernstein et al. 1994). The putative three-dimensional structure of the protein was obtained by comparative molecular modelling using the Zif268 mouse protein (PDB accession 1mey_C). Computational modelling clearly demonstrated the presence of three zinc fingers motifs for the SmZF1 protein (Fig. 3), where each finger unit is composed by two antiparallel β-strands at the cysteine side and one α-helix at the histidine side. Together, these structures confer a semicircular C-shape arrangment for SmZF1 protein. A 27-amino acid deletion (residues 64 to 90) was generated by the modelling program at the end of the α-helix of finger 1, since there was no correspondence of this region in the Zif268 protein. Thus, this helix in the model presented here is formed by the junction of SmZF1 residues 56 to 63 and 91 to 95. According to the secondary structure prediction, using the PredictProtein server, another two putative α-helix can be present in the protein: a1 at the N-terminal region residues 11 to 14 and α3 between the first and the second fingers residues 89 to 95 (data not shown). Interestingly, fingers 1 and 2 are distant to each other, separated by the putative α3-helix, while fingers 2 and 3 are very close. A detailed analysis revealed the existence of two residues that can be involved in forming strong hydrogen bonds with nucleotides in the putative target DNA sequence: His116 at finger 2 and Arg144 at finger 3 (Fig. 3). Those residues are located just before the beginning of the α-helix structures of the zinc finger motifs and suggest the function of fingers 2 and 3 in binding DNA.

Six putative phosphorylation sites were identified: one for cAMP- and cGMP-dependent protein kinases (residues 9-12), two for protein kinase C (residues 8-10 and 20-22), two for casein kinase II (residues 32-35, 70-73 and 93-96) and one for tyrosine kinase (residues 131-137), suggesting the protein can be regulated by phosphorylation. No N-terminal signal peptide was recognized and the prediction of cellular location showed that SmZF1 seems to be preferentially cytoplasmatic. The hydrophobicity plot demonstrated this protein is very hydrophylic, without any apparent transmembrane domains (not shown). However, two regions spaning from Gly40 to Lys51 and Arg102 to Lys111 showed to be more hydrophobic. These regions included the β-sheet structures at the cysteine side of the two first fingers. The antigenic index analysis (Jameson & Wolf 1988) showed that this protein seemed to be very antigenic, excepting the two small hydrophobic regions described above (not shown).

Expression of SmZF1 in S. mansoni life cycle stages - cDNA libraries of different stages of the parasite development (egg, cercariae, 3h schistosomulum and adult worm) were amplified by PCR using the primers ZnFMF and ZnFStop (Table). Amplicons of 307 bp were detected in all the stages of the life cycle evaluated (Fig. 4), indicating that this gene is expressed during the whole life cycle of the parasite.

DISCUSSION

There are relatively few papers on the characterization of transcription factors in *Schistosoma,* although the promoter regions of several schistosome genes have been shown to contain putative recognition sites for a number of trans-acting factors (El-Sherbeini et al. 1991, Abath et al. 1994,
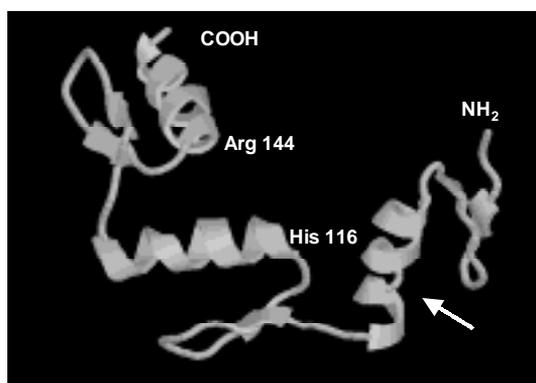
Fig. 3: theoretical three-dimensional structure of the SmZF1 protein obtained by molecular modelling using the Zif 268 mouse protein (PDB accession 1mey_C) as a model. Three classical zinc finger motifs are shown, each motif composed by two antiparallel β-sheets and one α-helix. His 116 at finger 2 and Arg 144 at finger 3 are shown. The region in the α-helix of the N-terminal finger, which is formed by the junction of SmZF1 residues 56 to 63 and 91 to 95, is indicated by an arrow.
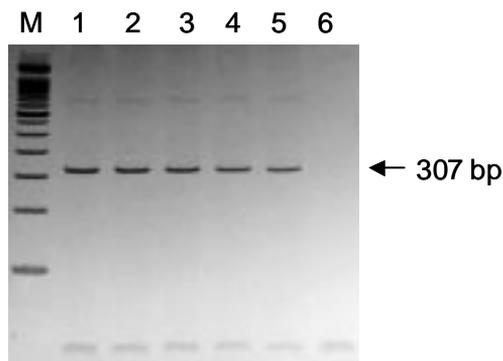
Fig. 4: PCR amplifications of cDNA libraries of several stages of the *Schistosoma mansoni* life cycle using the primers ZnFMF and ZnFStop. M: 100 bp ladder molecular weight marker (GIBCO BRL); 1: egg; 2: cercariae; 3: 3h schistosomulum; 4: adult worm library UUEE; 5: adult worm library SmZU; 6: negative control (without DNA). The arrow indicates the specific 307 bp amplicon.

Zemzoumi et al. 1995). Analysis of the 5´ flanking region of the gene encoding the 28-kDa glutathione S-transferase of *S. mansoni* (Sm28GST) have identified several regulatory sites able to bind nuclear factors such as AP-1, NF-Y and NF-AT-like proteins (Zemzoumi et al. 1995, Serra et al. 1999). The gene encoding the *S. mansoni* A subunit of the nuclear factor Y (SMNF-YA) was cloned and shown to be able to bind to the CCAAT boxes present in the promoter region of the Sm28GST gene (Serra et al. 1996, Zemzoumi et al. 1996). The gene encoding the *S. mansoni* hsp70 protein possesses two sequences similar to heat shock elements (HSE) and three inverted CCAAT boxes. Levy-Holtzman and Schechter (1994) demonstrated by electrophoretic mobility shift assays that *S. mansoni* extracts contain active transcription factors for hsp70 of *S. mansoni*. In fact, several cDNAs encoding the heat-shock transcription factor of schistosome have been cloned (Lantner et al. 1998). SMNF-YA is present in all the stages of the life cycle of the parasite, which is compatible with the idea that this transcription factor is involved with the regulation of several genes. On the other hand, HSE binding factor was detected only in schistosomula and adult worms, correlating with the pattern of hsp70 mRNA expression. In addition, a transcription factor homologous to the human Y-box binding protein was also cloned and structurally characterized in *S. mansoni* (Franco et al. 1997).

However, none of those yet characterized DNA binding proteins show zinc finger motifs in their structure. This work reports on the cloning and structural characterization of a gene that encodes a *S. mansoni* zinc finger protein. This class of proteins usually plays a role in the regulation of gene expression through interaction with nucleic acids, and have not been reported in *S. mansoni* previously. The derived amino acid sequence showed an ORF of 492 bp encoding a 164 amino acid protein interrupted by three Cys_2His_2 zinc finger motifs. The A+T content, high inside the ORF, and the codon usage are in agreement with data previously described for *S. mansoni* (Milhon & Tracy 1995).

The zinc fingers are among the most important structural motifs involved in the interaction between proteins and nucleic acids. Zinc-binding units such as the $Cys_2 His_2$ zinc finger domins are present in a large number of gene products, representing some of the largest protein families in the *Caenorhabditis elegans* genome (Clarke & Berg 1998). Although bacteria and archaea contain some proteins that bind zinc they appear to lack the large families of zinc-binding domains like those families in yeast, worms, and other eukaryotes. Proteins containing zinc finger motifs are involved in many aspects of the gene regulation in eukaryotes, for instance, the proteins that induce differentiation and growth, proto-oncogenes and transcription factors in general (Wingender 1993). Recently, a subgroup of zinc finger motifs associated to protein-protein interactions was described, the ring finger motifs, which are also able to interact with DNA (Tanimura et al. 1999). The three zinc finger motifs of SmZF1 protein fit to the consensus zinc finger sequence (Berg 1990, Bernstein et al. 1994). Computational modelling based on the crystal structure of the Zif268 mouse protein, that also contain three zinc fingers, demonstrated that the three SmZF1 zinc finger motifs are formed by finger structures with two β-strands to form one β sheet at the cysteine side and one α-helix at the histidine side, separated by a loop composed of 12 amino acid residues. The presence of four amino acids residues between the two histidines in all motifs might indicate that the SmZF1 fingers adopt a non-standard loop structure preceding the α-helix, as described for TFIIIA protein (Kochoyan et al. 1991). Elrod-Erickson et al. (1996) have demonstrated that α-helix structures of Zif268 protein fit to the major groove of B-DNA mainly by arginine-guanine and histidine-guanine interactions. Of particular importance are the arginine just preceding the α-helix structures. Our analysis demonstrated that SmZF1 residues His116 at finger 2 and Arg144 at finger 3, located just before the begining of the α-helix structures, might form strong hydrogen bonds to citosines in DNA sequences. This evidence suggests the function of fingers 2 and 3 in binding DNA, while the function of the first finger is not clear. Since zinc finger motifs can also mediate protein-protein interaction, one could speculate that the function of SmZF1 finger 1 is binding other co-associated factors.

Our results show that the SmZF1 mRNA is present in all stages of the parasite cycle analyzed, indicating a constitutive expression of SmZF1 throughout the developmental cycle of *S. mansoni*. However, it does not mean necessarily that the encoded protein is functionally active in all the parasite life cycle stages. In addition to transcriptional control, regulation of the protein activity by phosphorylation/dephosphorylation is the major intracellular control mechanism in eukaryotic cells. Many structural or regulatory proteins are targeted by protein kinases that introduce a phosphate group in specific residues of tyrosine, threonine and serine, resulting in inactivation, activation or intermediate forms of the protein (Cohen 1989). The presence of six potential phosphorilation sites in the SmZF1 protein suggests it might be submitted to such type of regulation, although this was not addressed in the present study.

To our knowledge, this is the first report on a gene encoding a *S. mansoni* zinc finger protein, and certainly the understanding of gene regulation in this human parasite will demand further studies on this category of proteins. In conclusion, this paper describes the structural characterization and expression of a *S. mansoni* gene encoding a putative zinc finger transcriptional regulatory protein. Some speculations were withdrawn from theoretical predictions, and should be validated by future experimental studies aiming at elucidating the functional role of this protein.

REFERENCES

Abath FGC, Hagan P, Jeffs A, Simpson AJG 1994. Partial characterization and kinetics of expression of Sm15, a *Schistosoma mansoni* tegumental antigen. *Parasitol Res 80*: 64-69.

Abath FGC, Hagan P, Jeffs SA, Schechter I, Meadows HM, Holder AA, Simpson AJG 1993. Structure of the gene encoding a putative *Schistosoma mansoni* tegumental antigen precursor. *Mol Biochem Parasitol 60*: 81-92.

Abath FGC, Xavier EM, Allen R, Gomes YM, Lucena-Silva N, Baliza M, Simpson AJG 2000. Characterization of Sm13, a tegumental antigen of *Schistosoma mansoni*. *Parasitol Res 86*: 745-752.

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.

Berg JM 1990. Zinc fingers and other metal-binding domains. Elements for interactions between macromolecules. *J Biol Chem 265*: 6513-6516.

Bernstein BE, Hoffman RC, Klevt RE 1994. Sequence-specific DNA recognition by Cys_2 His_2 zinc fingers. *Ann N Y Acad Sci 726*: 95-104.

Chen LL, Rekosh DM, Loverde PT 1992. *Schistosoma mansoni* p48 egg shell protein gene: characterization, developmental regulated expression and on parition to the p14 egg shell protein gene. *Mol Biochem Parasitol 52*: 39-52.

Clarke ND, Berg JM 1998. Zinc finger in *Caenorhabditis elegans*: finding families and probing pathways. *Science 282*: 2018-2022.

Cohen P 1989. The structure and regulation of protein phosphatases. *Rev Biochem 58*: 453-508.

Davis AH, Blanton R, Klich P 1985. Stage and sex specific differences in actin gene expression in *Schistosoma mansoni*. *Mol Biochem Parasitol 17*: 289-298.

Elrod-Erickson M, Rould MA, Nekludova L, Pabo CO 1996. Zif268 protein-DNA complex refined at 1,6 A°: A model system for understanding zinc finger-DNA interactions. *Structure 4*: 1171-1180.

El-Sherbeini M, Ramadan N, Bostian KA, Knopf PM 1991. Cloning and sequence analysis of the *Schistosoma mansoni* membrane glycoprotein antigen GP22. *Mol Biochem Parasitol 49*: 83-98.

Franco GR, Garratt RT, Tanaka M, Simpson AJ, Pena SDJ 1997. Characterization of *Schistosoma mansoni* gene encoding a homologue of the Y-box binding protein. *Gene 198*: 5-16.

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

Grossman Z, Ram D, Markovics A, Tarrab-Hazdai R, Lantner F, Ziv E, Schechter I 1990. *Schistosoma mansoni*: stage specific expression of muscle specific genes. *Exp Parasitol 70*: 62-71.

Jameson BA, Wolf H 1988. The antigenic index: a novel algorithm for predicting antigenic determinants. *Comp Appl Biosci 4*: 181-186.

Kochoyan M, Keutman HT, Weiss MA 1991. Alternating zinc finger in the human male-associated protein ZFY: HX_3H and HX_4H motifs encodes a local structural switch. *Biochemistry 30*: 9396-9402.

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

Kyte J, Doolittle RF 1982. A simple method for displaying the hydropatic character of a protein. *J Mol Biol 157*: 105-132.

Lantner F, Ziv E, Ram D, Schechter I 1998. Different forms of the mRNA encoding the heat-shock transcription factor are expressed during the life cycle of the parasitic helminth *Schistosoma mansoni*. *Eur J Biochem 253*: 390-398.

Levy-Holtzman R, Schechter I 1994. Schistosome extracts with heat shock factor activity revealed by the gel shift assay. *Parasitology 108*: 35-42.

Mei H, Lo Verde PT 1997. *Schistosoma mansoni*: the developamental regulation and immunolocalization of antioxidant enzymes. *Exp Parasitol 86*: 69-78.

Milhon JL, Tracy JW 1995. Update condon usage in *Schistosoma*. *Exp Parasitol 80*: 353-356.

Mitchell PL, Tjian R 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science 245*: 371-378.

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.

Savioli L, Renganathan E, Montresor A, Davis A, Behbehani K 1997. Control of schistosomiasis - A global picture. *Parasitol Today 13*: 444-448.

Serra E, Zemzoumi K, Trolet J, Capron A, Dissous C 1996. Functional analysis of the *Schistosoma mansoni* 28 kDa glutathione S-transferease gene promoter: involvement of SMNF-Y transcription factor in multimeric complex. *Mol Biochem Parasitol 83*: 69-80.

Serra EC, Lardans V, Dissous C 1999. Identification of NF-AT-like transcription factor in *Schistosoma mansoni*: its possible involvment in the antiparasite action of cyclosporin A. *Mol Biochem Parasitol 101*: 33-41.

Simpson AJ, Payares G, Walker T, Knight M, Smithers SR 1984. The modulation of expression of polypeptide surface antigens on developing schistosomula of *Schistosoma mansoni*. *J Immunol 133*: 2725-2730.

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.

Tanimura S, Ohtsuka S, Mitsui K, Shirouzu K, Yoshimura A, Ohtsubo M 1999. MDM2 interacts with MDMX through their ring finger domains. *FEBs Letters 447*: 5-9.

von Heijne G 1987. *Sequence Analysis in Molecular Biology: Treasure Trove or Trivial Pursuit*, Academic Press Inc, San Diego, 97 pp.

Wingender E 1993. *Gene Regulation in Eukaryotes*, VCH Verlagsgesellschaft mbH, Weinheim, 430 pp.

Yokono M, Saegusa N, Matsushita K, Sugiura Y 1998. Unique DNA binding mode of the N-terminal zinc finger of transcription factor Sp-1. *Biochemistry 37*: 6824-6832.

Zemzoumi K, Dissous C, Cochu A, Trolet J, Capron A, Mcnair A 1995. Interaction of nuclear extracts with the CCAAT-binding site revealed by gel shift assay. *Exp Parasitol 80*: 149-154.

Zemzoumi K, Serra E, Montovani R, Trolet J, Capron A, Dissous C 1996. Cloning of *Schistosoma mansoni* transcription factor NF-YA subunit: phylogenic conservation of thd HAP-2 homology domain. *Mol Biochem Parasitol 77*: 161-172.