Schistosoma mansoni encodes SMT3B and SMT3C molecules responsible for post-translational modification of cellular proteins

Fernanda J. Cabral, Olavo S. Pereira Jr., Camila S. Silva, Renata Guerra-Sá, Vanderlei Rodrigues

Abstract

The sumoylation pathway is a post-translational modification of nuclear proteins widespread among several organisms. SMT3C is the main protein involved in this process and it is covalently conjugated to a diverse assortment of nuclear protein targets. To date, 3 SUMO paralogues (SMT3C, A/B) have been characterized in mammals and plants. In this work we characterized two SUMO related genes, named SMT3B and SMT3C throughout Schistosoma mansoni life cycle. The SmSMTB/C encodes for proteins sharing significant amino acid homology with SMT3. Phylogenetical analyses revealed that both SmSMT3B/C are distinct proteins. Additionally, SmSMT3B and C are expressed in cercariae, adult worms, eggs and schistosomula however SmSMT3C gene showed an expression level 7 to 9 fold higher than SmSMT3B in eggs, schistosomula and adult worms. The comparison between the SmSMT3C genomic and cDNA sequences established that the encoding sequence is interrupted by 3 introns of 70, 37 and 36 bp. Western Blot has shown SMT3 conjugates are present in nuclear and total protein fractions of adults and cercariae. Therefore our results suggest a functional sumoylation pathway, and the presence of two paralogues also suggests the specificity of substrates for SMT3 in S. mansoni.

Keywords: Schistosoma mansoni; Post-translational modification; Sumoylation; SMT3

1. Introduction

Post-translational protein modifications (PTMs) play an important role on protein function by altering the activity, subcellular localization and/or ability to interact with other proteins. SUMO (Small ubiquitin modifier) is the best-characterized member of a growing family of ubiquitin-like proteins. Although it is structurally similar to ubiquitin, it does not target the modified proteins to proteolysis by 26 S proteasome [1]. In mammalian cells, modification by SUMO has been linked to the regulation of protein–protein interactions, transcriptional control, and increased of life-time of proteins [2].

SUMO 1, SUMO 2 and SUMO 3 also named SMT3C, SMT3B and SMT3A, are the three SUMO paralogues described in humans and mice. Whereas the mature forms of SMT3B and SMT3A are 95% identical to each other at the amino acid level, SMT3A/B subfamily proteins are only 47% identical to SMT3C [3]. Regardless of which molecule to be conjugated, the sumoylation pathway requires the same set of enzymes. Firstly, it relies on an enzymatic complex called AOS1/UBA2 heterodimer for SMT3 activation (E1) followed by a single E2 enzyme (UBC9) for conjugation to target substrates [4]. However,
in contrast to modification by SMT3C, SMT3B and A can be conjugated through the formation of poly-SMT3 chains [5]. At present, poor data is available regarding the abundance of SMT3B and A, since no tool is yet able to functionally distinguish these molecules within the cell [6].

**Schistosoma mansoni** is the agent of schistosomiasis, a parasitic disease affecting 300 million people in tropical and subtropical areas of the world [7]. **Schistosomes** are sexually dimorphic trematodes that reach maturity after alternating passages into two independent hosts. As a result, development through their life cycle is accompanied by several morphological and biochemical changes that guarantee adaptation. At the molecular level, ubiquitination has been implicated in the degradation of nuclear receptors in *S. mansoni*, suggesting the participation of this type of PTM in the regulation of gene expression [8]. A more recent study from our group implicated ubiquitination and proteasome degradation as key steps during the cercariae to schistosomula transformation [9]. Here we report the molecular characterization of transcripts encoding *S. mansoni* SMT3C (SmSMT3C) and *S. mansoni* SMT3B (SmSMT3B). The expression profile of both genes was investigated, highlighting the existence of a conserved sumoylation pathway in *S. mansoni*, and, consequently, adding to the repertoire of molecules the parasite utilizes to increase the functional diversity of its proteins.

2. Materials and methods

2.1. Parasites

*S. mansoni* LE strain was maintained by routine passage through * Biomphalaria glabrata* snails and BALB/c mice. Infected snails were induced to shed cercariae under light exposure for 2 h; the latter was recovered by sedimentation on ice.

Schistosomes were obtained by cercariae transformation from infected snails. Adult worms, cercariae, eggs and schistosomula were conducted as follows: initial denaturation for 5 min, following 20 cycles of denaturation at 95 °C for 1 min, annealing at 55 °C and extension at 72 °C for 1 min. Negative controls were performed using all PCR reagents without addition of the cDNA. Gene expression was normalized using alpha-tubulin (GenBank accession number M80214). The primers sequence for alpha-tubulin is: fw: 5′-CGTATTCGCAAGTTGGCTGACCA-3′ and rv:5′-CCATCGAAGCGCAAGTGACA-3′ and the transcript appears with 16 cycles of PCR reaction. Densitometry was performed using Gene Quantity software (Biorad). All PCR products were cloned in pGEMT-easy vector (FROMEGA) and were sequenced in ABI 3100 (Applied biosystem), using dye terminator kit. The results presented in this manuscript are representatives of three independent biological replicates.

2.2. Identification and analysis of SMT3B and SMT3C in *S. mansoni*

A cDNA encoding the open reading frame of SmSMT3B was obtained by RT-PCR, using the forward 5′-TCTAATTCGTTTCAAGAGAAGTCCCGC-3′ and reverse 5′-CGAAAAACGAGGTTTATTTGTAAGATTA-3′ primers used for amplification of the respective SMT3 gene in *Caenorhabditis elegans* suggested by [14]. The gene encoding SmSMT3C was identified by BLAST searches using the *S. mansoni* transcriptome database at [http://bioinfo.iq.usp.br/schisto](http://bioinfo.iq.usp.br/schisto). The corresponding predicted ORFs were aligned using the ClustalW algorithm available at [www2.ebi.ac.uk/clustalw](http://www2.ebi.ac.uk/clustalw). The pairs of primers were obtained from the consensus sequence: fw 5′-ATGACTGATTGCAATATAA-3′/rv 5′-CAATCCACAGTTGAGGCTGCT-3′. The semi-quantitative PCR reactions for SMT3C and SMT3B, for adult worms, cercariae, eggs and schistosomula were conducted as follows: initial denaturation for 5 min, following 20 cycles of denaturation at 95 °C for 1 min, annealing at 55 °C and extension at 72 °C for 1 min. Negative controls were performed using all PCR reagents without addition of the cDNA. Gene expression was normalized using alpha-tubulin (GenBank accession number M80214). The primers sequence for alpha-tubulin is: fw: 5′-CGTATTCGCAAGTTGGCTGACCA-3′ and rv:5′-CCATCGAAGCGCAAGTGACA-3′ and the transcript appears with 16 cycles of PCR reaction. Densitometry was performed using Gene Quantity software (Biorad). All PCR products were cloned in pGEMT-easy vector (FROMEGA) and were sequenced in ABI 3100 (Applied biosystem), using dye terminator kit. The results presented in this manuscript are representatives of three independent biological replicates.

2.3. Genomic DNA amplification

Genomic DNA was extracted from adult *S. mansoni* according to the CTAB method [16]. The oligonucleotides for SmSMT3C and SmSMT3B were used for amplification according to a program in the thermocycler which involved (1) initial denaturation at 95 °C for 5 min, (2) annealing at 55 °C and extension at 72 °C for 1 min during 35 cycles. The identity of the amplified products was confirmed by BLAST searches of their corresponding sequences.

2.4. Sequence analysis and construction of phylogenetic trees

Searches for DNA and protein identities were performed by BLAST. Analyses of the predicted protein sequences were conducted using the algorithms LAlign, PRODOM and Prosite at [www.expasy.org](http://www.expasy.org). Alignment of the predicted SmSMT3B and SmSMT3C was performed using ClustalW and formatted for presentation using BOXSHADE at [http://www.ch.embnet.org/software/BOX_form.html](http://www.ch.embnet.org/software/BOX_form.html). ClustalW alignments of the SmSMT3B and SmSMT3C predicted proteins sequences were used for the generation of the trees as a requirement for the utilization of the Mega software (version 4.0) described by [17], using maximum parsimony. The confidence of the branches was evaluated by bootstrap analysis using 1000 samplings. Alignments for phylogenetic reconstruction were done using BLOSUM matrix at ClustalW. The GenBank proteins sequences and accession numbers used in these analyses were as follows: SMT3B orthologues from: *S. mansoni* NY1817504; *Homo sapiens* NP008868; *Rattus norvegicus* P61959; *Danio rerio* Q6DHL4; *Xenopus laevis* Q28H04. SMT3C orthologs from: *S. mansoni* AAX30013; *R. norvegicus* SPQ510H3; *H. sapiens* gi|52783799; *Drosophila melanogaster* gi|6934292; *S. japonicum* gi|28341405; *C. elegans* NP490842; *Bombbyx mori* ABID3655; *D. rerio* AAH67553; *Pan troglodytes* XP_516035; *X. laevis* NP_001083717; *Saccharomyces cerevisiae* gi|123999147 was used as outgroup.
2.5. mRNA isolation and Northern blot analysis

Ten micrograms of total RNA from adult worms were electrophoresed in a 1% agarose/formaldehyde gel and transferred to nylon Hybond N+ (GE Healthcare). Specific SmSMT3B and SmSMT3C radiolabeled probes were constructed by using Random Primers DNA Labelling System (Invitrogen) and hybridization procedures executed as described [16].

2.6. Preparation of protein extracts

Crude cytosolic lysates were prepared from cercariae and adult worms. Briefly, the parasites were suspended in buffer A (5 mM Tris–HCl pH 8.0; 1% glycerol; 1 mM EDTA; 1 mM PMSF; 1 mM DTT and 1 mM MG132), homogenized in Polytron®, and centrifuged at 10,000 ×g for 30 min. The supernatant centrifuged at 40,000 ×g for 60 min. The protein content of the resulting soluble fraction was estimated by the BCA method using Pierce kit [18].

The nuclear protein extract from adult worms was prepared as described by Rabelo et al. 1992 [19]. Briefly, parasites was homogenized in buffer B (10 mM Tris–HCl pH 7.6; 10 mM NaCl; 3 mM MgCl2; 0.1% Triton X-100 and 1 mM PMSF), followed by centrifugation at 2000 ×g for 10 min at 4 °C. The supernatant containing the cytosolic fraction was stored at −70 °C. The pellet containing the nuclei was resuspended in 2 mL of buffer B and centrifuged as described above. Finally, the pellet containing the nuclei was resuspended for 45 min in buffer B plus 0.8 M NaCl and then centrifuged at 10,000 ×g for 45 min at 4 °C. The supernatant was dialyzed for 3 h in buffer C (20 mM HEPES pH 7.6; 5 mM MgCl2; 0.1 mM EDTA; 1 mM DTT and 10 mM PMSF). The protein content was determined as described above. Twenty micrograms of both cytosolic and nuclear fractions were loaded onto a 10% SDS-PAGE [20] for

Fig. 1. Identification and characterization of the SmSMT3B gene. (A) Local alignment between the SmSMT3B/C amino acids predicted sequence and its orthologues, using CLUSTALW and formatted using BOXSHADE. The arrows indicate the ubiquitin-like domain. (B) Phylogenetic tree for SmSMT3B/C and its orthologues in other organisms. The maximum parsimony tree was calculated using Mega 4.0. SUMO1 H. sapiens; SUMO1 R. norvegicus; SMT3 P. troglodytes; SMT3 S. cerevisiae; SUMO-1 X. laevis; SUMO1 D. rerio; SMT3 C. elegans; SMT3B S. mansoni; SUMO2 H. sapiens; SUMO2 D. rerio; SUMO2-2 X. laevis; SMT3 D. melanogaster; SMT3 B. mori; SMT3C S. mansoni; SMT3 S. japonicum. Linear distances between homologue sequences are inversely proportional to their amino acid sequence similarity.
electrophoretic separation. The protein content of the gel was evaluated by staining with Coomassie Brilliant blue.

2.7. Immunoblot analysis

Nuclear and total protein from adult worms and cercariae were transferred to a PVDF membrane (Hybond-Amersham Biosciences) according to [21]. Following an overnight blocking step in TBST buffer (10 mM Tris–HCl pH 7.5, 150 mM NaCl, 5% non-fat milk and 0.05% Tween 20), the blotted proteins were submitted to hybridization with an anti-SUMO 1 antibody (Santa Cruz Biotech) at 1:500 dilution for 4 h in the same buffer. Biotinylated anti-goat IgG antibody conjugated to alkaline phosphatase at 1:1000 dilution was used as the secondary antibody during 90 min of incubation. SMT3-conjugated proteins were visualized after developing the membrane in the presence of the enzymes substrates NBT/BCIP.

2.8. Statistical analysis

Relative gene expression in all stages was compared for one-way variance analysis (Tukey test). It was considered statistically significant \( P<0.001 \). The statistical analysis was performed using PRISM software.

3. Results

3.1. Sequencing and analysis of cDNAs coding for SmSMT3C and SmSMT3B proteins

The full-length cDNA sequence of SmSMT3B presents a 742-bp size, and its sequence was deposited on GenBank under the number AY817504. The position of the ATG start codon was inferred at position 327 encoding a protein of 94 amino acids with a predicted molecular mass of 10 kDa and a pI equal to 8.82.

Amino acids sequence comparison showed that SmSMT3B shares the Ub-fold motif, important for folding and conjugation of the protein to its targets. The conserved C-terminal Gly-Gly are also present (Fig. 1A). As a result, it is likely that SmSMT3B displays a 3D structure similar to the one described for human SUMO-1 despite substantial sequence divergence. In parallel the comparison between amino acid sequence of SmSMT3B and SmUbiquitin [22] revealed 16% of identity at the amino acid level. SmSMT3B also contains an extended N-terminal tail (19 residues) bearing two extra amino acids located after the C-terminal Gly-Gly motif. This feature is probably necessary for the action of proteolytic enzymes during the maturation of the protein [1].

In order to validate the \textit{in silico} predictions of the SmSMT3C gene, we used the molecular reconstitution approach described before (Material and Methods). A 270 bp cDNA fragment of the SmSMT3C gene was obtained from adult worms total RNA. This cDNA contained an open reading frame encoding a protein of 90 amino acid residues, and presents all the protein sequence features that we have already described for SmSMT3B sequence. Sequence analysis using LAlign showed that SmSMT3C and its paralogue in \textit{S. mansoni} SmSMT3B (Fig. 1A) present 48.8% of similarity of sequence, suggesting the existence of two paralogs of SUMO family in \textit{S. mansoni}, in agreement with data in other organisms [3].

Sequence comparison revealed similarities between SmSMT3B and its homologues from various organisms, particularly among metazoan species (Fig. 1A). The structural relationships between SMT3B proteins from \textit{S. mansoni}, human, rat, fish, frog, and yeast as an outgroup, were analysed (Fig. 1B). The phylogenetic tree revealed that SmSMT3B is grouped with SUMO-2 proteins, while SMT3C is grouped with SUMO-1 proteins.

In addition, the alignment of predicted protein sequence and its homologs from distinct taxa revealed is a conserved protein (Fig. 1B). Phylogenetic analyses showed that SmSMT3C is closely related to \textit{S. japonicum} orthologs, as expected due to the evolutionary proximity between members of the Lophotrochozoa superphylum.

3.2. Genomic structure of the SmSMT3C gene

Genomic amplification using SmSMT3C specific primers (the same used for RT-PCR) showed a product containing 413 bp. The alignment of predicted nucleotide sequences from cDNA and genomic DNA revealed that the coding region of SmSMT3C is interrupted by three introns with 70, 37 and 36 bp, respectively (Fig. 2).

3.3. Northern blot analysis

To confirm that both paralogues were transcribed, we performed Northern blot from adult worms. The SmSMT3B and
SmSM3TC probes recognized transcripts of 1.3 kb and 1.0 kb. In addition a second transcript of 1.8 kb was recognized by SmSMT3C probe (Fig. 3A).

3.4. Gene expression profile of SmSMT3B and SmSMT3C genes during the parasite life cycle

The endogenous mRNA expression of SmSMT3C/B genes was investigated in adult, eggs, schistosomula and cercariae stages and compared by semi-quantitative RT-PCR using α-tubulin gene as internal control (Fig. 3B). The results showed that both genes are expressed in all investigated stages. The levels of SmSMT3B/C transcripts in the adult worm, eggs and schistosomula were strikingly different (P<0.001). Whilst SmSMT3B showed similar levels of expression, up to a 9 fold difference was observed for SmSMT3C. In contrast, the expression of both genes in the cercariae stage did not differ significantly. The lowest expression level for SmSMT3C was found in the cercariae stage.

3.5. The SMT3C conjugation pathway is present in S. mansoni

Freshly perfused adult parasites were submitted to a differential protein extraction protocol in order to enrich for nuclear proteins. As shown in Fig. 4A, different electrophoretic migration patterns were obtained when the nuclear fraction was separated alongside the crude cytoplasmic fractions obtained from the adult and cercariae stages. To confirm that S. mansoni may potentially conjugate SMT3 to other proteins, we performed an immunoblot analysis with an antibody directed against the conserved N-terminal sequence of human SUMO1. As shown in Fig. 4B, cross-immunoreactive proteins could be observed in cercariae as well as in nuclear and total fractions of the adult parasite.

4. Discussion

Sumoylation is biochemically similar to ubiquitination, however, the targeted substrates are not directed for degradation by the Ubiquitin–Proteasome pathway [1]. Several studies have shown that SMT3 is a nuclear protein related to the control of important cellular processes such as genome integrity [23], transcriptional regulation by modification of transcription factors such as c-Jun and c-Fos [24], differential gene expression by modification of repression complexes [25], and trafficking of proteins from nucleus to cytoplasm through modification of the protein RanGAP-1 [26].

There are several unresolved issues regarding the distinctions between SMT3 paralogues. Firstly, it is not entirely clear whether or not conjugation of particular substrates is restricted to specific SMT3 variants or if different paralogues can be used interchangeably. For a few substrates, strong paralogue preference has been documented; for instance, RanGAP1 is modified almost exclusively by SMT3C in vivo [26]. Other substrates, such as the promyelocytic leukemia protein (PML), have been reported to be conjugated to both SMT3C [27] and SMT3B [28] in vivo. Analysis of S. mansoni transcriptome showed the existence of homologues for RanGAP1 and PML, suggesting that both modes of conjugation might occur in this parasite.
Our first attempt to characterize the existence of the sumoylation pathway in *S. mansoni*, at a time when the transcriptome and genome initiatives were still in their infancy, had to be based on finding the transcript(s) for SMT3 paralogues using a model organism (*C. elegans*) as a reference. RT-PCR for the related SMT3B gene in *S. mansoni*, revealed a 600-bp fragment corresponding to the SMT3B cDNA with a predicted ORF coding for a protein of 94 amino acids (Fig. 1A). Sequence analyses showed the C-terminal Gly-Gly conserved motif, responsible for attachment and conjugation of SMT3 on its substrates, and the presence of a flexible N-terminal extension which is absent on the ubiquitin sequence. Indeed, the comparison between the amino acid sequence of SmSMT3B and SmUbiquitin [22] revealed only 16% identity at the amino acid level.

Later, by mining the data generated by the *S. mansoni* genome database, we found another paralogue for SMT3 named as SMT3C. SmSMT3C also exhibits a C-terminal double glycine responsible for SMT3C conjugation to the targeted substrates (Fig. 1A) [5]. Using the strategy of molecular phylogeny, we grouped SmSMT3B and SmSMT3C with various orthologue sequences and two distinct proteins were revealed (Fig. 1B). The existence of two paralogues for SMT3 was not anticipated for lower eukaryotes such as Schistosomes. At present, only one gene coding for SMT3 has been demonstrated in *C. elegans* [14] and *D. melanogaster* [29]. The phylogenetic analysis of SmSMT3C has shown that the predicted protein has been conserved throughout evolution. In *S. mansoni*, the SMT3 paralogues share 48.8% identity, in agreement with published data from other organisms [4]. As expected, SmSMT3C sequence belongs to the same branch as its counterpart from *S. japonicum*.

Comparison between the genomic and transcriptome derived sequences for SmSMT3C showed the existence of three small introns of 70, 37 and 36 bp respectively, all flanked by the GT...AG splicing junction (Fig. 2). Likewise, SMT3B and SMT3A genes are interrupted by three introns in vertebrates. In the nematode *C. elegans* the sequence of the SMT3 gene is interrupted by 2 introns of 56 and 50 bp [14]. In vertebrates, up to four introns for SMT3C have been described [30]. Our results do not suggest the existence of introns for the SmSMT3B gene. This finding is in agreement with the predicted sequence for SmSMT3B, annotated under the accession number NP1435753 in the TIGR database.

The confirmation of these two transcripts expressed was accessed using Northern blotting (Fig. 3A). Even the experiments were performed using high stringency the specific probe for SmSMT3C recognized a 1.8 kb transcript, probably a result of cross-reaction with ubiquitin or other ubiquitin-like protein, such as Nedd8. In addition, the transcriptome data as well as previous data from our group have shown that poliubiquitin genes are highly expressed in adult worms [22].

In attempt to trace the expression profile of both genes, we decided to evaluate the levels of mRNA coding for SmSMT3C and SmSMT3B in larvae and adult stages of this parasite (Fig. 3B). The systematic higher levels of SmSMT3C in all the investigated stages points towards a wider or more general spectrum of activities performed by this paralogue. In contrast, SmSMT3B was found constantly in lower levels perhaps linked to specific and yet important cellular processes. In support to this hypothesis is the down regulation of SmSMT3C mRNA in the cercariae, whereas SmSMT3B levels remained equally comparable to those found in the adult, egg and larvae stage. In light of these results, further studies may potentially attribute functions for SMT3B such as maintenance of genome integrity that could help us to explain how the parasite succeeds under so many environmental stresses for the completion of its life cycle. Comparatively higher levels of SmSMT3C have also been described in mouse [30] and *Arabidopsis* [31].

At last, using an antibody targeted against the 19 N-terminal amino acids for human SMT3C, sumoylated conjugates in nuclear as well as in crude soluble preparations from adult worms and cercariae could be detected (Fig. 4B). Due to the degree of identity between SMT3 paralogues in *S. mansoni*, we believe that the observed reactivity might represent the total pool of sumoylated substrates conjugated via SMT3C and SMT3B. The parasite’s genome seem to be fully competent to encode the enzymes necessary for SMT3 processing and conjugation to its targets. Therefore, it was not surprising that the natural substrates of the SMT3 pathway could be detected by Western blotting. The identification of the latter will contribute to the complete characterization of this type of PTM in a helminth parasite.

Acknowledgements

This work was supported by the following Brazilian agencies of research: FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), CAPES (Coordenação de Apoio ao Pessoal de Nível Superior), and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico). We are glad to Dr. William de Castro Borges for helpful comments on this manuscript. We also thank Ms. Elenice A. Macedo and Ms. Olinda M. Brigato for technical assistance.

References


