Journal of Molecular Graphics and Modelling 39 (2013) 29–38

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Modeling the zing finger protein SmZF1 from *Schistosoma mansoni*: Insights into DNA binding and gene regulation

Mainá Bitar[a,b,1], Marcela Gonçalves Drummond[b,c,1], Mauricio Garcia Souza Costa[a,1], Francisco Pereira Lobo[d], Carlos Eduardo Calzavara-Silva[e], Paulo Mascarello Bisch[a], Carlos Renato Machado[b], Andréa Mara Macedo[b], Raymond J. Pierce[c], Glória Regina Franco[b,*]

[a] *Laboratório de Física Biológica, Instituto de Biofísica Carlos Chagas Filho, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil*
[b] *Laboratório de Genética Bioquímica, Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil*
[c] *CIIL, Institut Pasteur de Lille, F-59019 Lille, France; Inserm, U 1019, F-59019 Lille, France; Univ. Lille Nord de France, F-59000 Lille, France; CNRS, UMR 8204, F-59021 Lille, France*
[d] *Laboratório de Bioinformática Aplicada, EMBRAPA Informática Agropecuária, Campinas, São Paulo, Brazil*
[e] *Laboratório de Imunologia Celular e Molecular, Centro de Pesquisas René Rachou, FIOCRUZ, Belo Horizonte, Minas Gerais, Brazil*

ARTICLE INFO

Article history:
Accepted 13 October 2012
Available online 23 October 2012

Keywords:
Schistosoma mansoni
SmZF1
Zinc finger
Transcriptional regulation
Comparative modeling
Molecular dynamics

ABSTRACT

Zinc finger proteins are widely found in eukaryotes, representing an important class of DNA-binding proteins frequently involved in transcriptional regulation. Zinc finger motifs are composed by two antiparallel β-strands and one α-helix, stabilized by a zinc ion coordinated by conserved histidine and cysteine residues. In *Schistosoma mansoni*, these regulatory proteins are known to modulate morphological and physiological changes, having crucial roles in parasite development. A previously described C_2H_2 zinc finger protein, SmZF1, was shown to be present in cell nuclei of different life stages of *S. mansoni* and to activate gene transcription in a heterologous system. A high-quality SmZF1 tridimensional structure was generated using comparative modeling. Molecular dynamics simulations of the obtained structure revealed stability of the zinc fingers motifs and high flexibility on the terminals, comparable to the profile observed on the template X-ray structure based on thermal b-factors. Based on the protein tridimensional features and amino acid composition, we were able to characterize four C_2H_2 zinc finger motifs, the first involved in protein–protein interactions while the three others involved in DNA binding. We defined a consensus DNA binding sequence using three distinct algorithms and further carried out docking calculations, which revealed the interaction of fingers 2–4 with the predicted DNA. A search for *S. mansoni* genes presenting putative SmZF1 binding sites revealed 415 genes hypothetically under SmZF1 control. Using an automatic annotation and GO assignment approach, we found that the majority of those genes code for proteins involved in developmental processes. Taken together, these results present a consistent base to the structural and functional characterization of SmZF1.

1. Introduction

Zinc fingers are structural motifs frequently observed in regulatory DNA-binding proteins often related to modulation of gene expression, among several other functions in the cell [1]. Zinc binding proteins correspond to up to 10% of the human proteome, most of them consisting of zinc finger containing proteins, highlighting the importance of such molecules for complex organisms [2]. The zinc finger structure is stabilized by a zinc ion coordinated by highly conserved cysteine and histidine residues [3].

In organisms that undergo different stages during a life cycle, regulatory proteins modulate morphological and physiological characteristics in response to environmental changes. This is the case with parasites such as *Schistosoma mansoni*, where these proteins are crucial throughout development [4]. *S. mansoni* is the etiologic agent of a neglected tropical disease, schistosomiasis, that affects 200 million people worldwide [5] and is the second most devastating disease concerning socioeconomic aspects [6]. In the course of its life cycle, *S. mansoni* dwells in both vertebrate and invertebrate hosts as well as in the external environment, going through six different stages with distinct morphological and physiological features. Therefore, this parasite represents a complex,

* Corresponding author at: Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais – UFMG, Avenida Antônio Carlos 6627, Belo Horizonte CEP 31270-010, MG, Brazil.
Tel.: +55 31 34092537; fax: +55 31 34092984.
E-mail address: gfrancoufmg@gmail.com (G.R. Franco).
[1] These authors contributed equally to this work.

1093-3263/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmgm.2012.10.004

Table 1
Primers used to amplify and sequence SmZF1 cDNA. Different combinations of forward and reverse primers were employed to amplify SmZF1 cDNA. Sequence reactions were carried out with appropriate primers according to the amplicon used.

Forward primers		Reverse primers	
Primer	Sequence	Primer	Sequence
SmZF1start	5′-ATGGAATTTTACTTCACA-3′	SmZF1stop	5′-CTGGCATACTTCACAT-3′
SmZF1exon3	5′-GTTTTCAAGTCAGAGTCTT-3′	SmZF1end	5′-ATTATACAATCTGGTTTCTCTT-3′
		SmZF1lower	5′-TGAAAGAATAATAAATGTA-3′

interesting, and challenging biological system for the study of gene regulation processes [4,7,8].

SmZF1 (GenBank ID: AAG38587) was initially described as a *S. mansoni* C_2H_2 zinc finger protein comprising 164 amino acids and encoded by a gene of 2182 base pairs (bp) and an opening reading frame (ORF) of 492 bp [9]. In a prior report we showed that the recombinant Maltose Binding Protein (MBP)-SmZF1 protein is capable of binding DNA and RNA oligonucleotides, with greater affinity for DNA [10]. In addition, we have previously shown the presence of this protein in the cells nuclei of *S. mansoni* during different life cycle stages and demonstrated its ability to activate gene transcription in a heterologous model [11]. Altogether, these findings support a role for SmZF1 as a transcription factor in *S. mansoni*.

Unexpectedly, we found an inconsistency on the originally reported SmZF1 cDNA sequence published by Eleutério de Souza et al. [9], which revealed a larger coding sequence than previously considered and also four instead of three zinc finger motifs. Here we characterize the new SmZF1 tridimensional structure with comparative modeling followed by molecular dynamics simulations. Further, we identified new putative DNA binding sites with sequence analysis and obtained a putative complex with docking. Finally, we investigated potential genes under SmZF1 control.

2. Materials and methods

2.1. cDNA sequencing

The SmZF1 cDNA was PCR amplified from either a *S. mansoni* adult worm cDNA library (AW2, from Minas Gerais Genome Network – http://rgmg.cebio.org/content/schistosoma-mansoni-transcriptome-project) or from cDNA originated from RNA of *S. mansoni* single-sex adult worms (the reverse transcription reaction was performed as previously described [11]). Reaction mixtures contained 200 μM of each dNTP, 0.2 μM of each primer (different combinations of those showed in Table 1), 1 U Taq DNA polymerase in reaction buffer 1B 1× (Phoneutria, Belo Horizonte, MG, Brazil), 0.05 U AccuPrime *Pfx* DNA Polymerase (Invitrogen, Carlsbad, CA, USA) and either 3 or 1 μL cDNA (arising from either newly synthesized cDNA or cDNA library, respectively) in a final volume of 10 μL. The amplification reactions were performed in a PT-100 thermal cycler (MJ Research, Waltham, MA, USA) using the cycling protocol: 95 °C for 6 min and 25 cycles of 95 °C for 1 min, 55 °C for 1 min, and 72 °C for 1 min, followed by a final step of 72 °C for 5 min. The amplicons were then cloned into the pGEM-T vector using the pGEM-T Vector System Kit (Promega, Madison, WI, USA) following manufacturer instructions. Ligation products were used to transform the *Escherichia coli* DH5α strain, and the rescued plasmids (300 ng) were sequenced using 10 pmol of appropriate primers (Table 1) and 4 μL of DYEnamic ET Dye Terminator Kit – MegaBACE (GE Healthcare, Waukesha, WI, USA). The sequencing products were analyzed in the MegaBACE 1000 DNA Sequencer (GE Healthcare). The resulting sequences were assembled by the software Multalin [12] and searched for similarity using blastn [13].

2.2. 5′ RLM RACE

The RLM RACE (RNA ligase-mediated rapid amplification of 5′ and 3′ cDNA ends) experiments were performed using the GeneRacer Kit (Invitrogen), according to manufacturer instructions. In the present work, only 5′ cDNA amplifications were performed using RNA (6.5 μg) from *S. mansoni* adult worms as starting material. RNA integrity was verified by agarose gel electrophoresis after each preparation step. The cDNA synthesis was performed using either random primers or an oligo dT primer.

To amplify the 5′ end of SmZF1 coding cDNA, a specific reverse primer was used (SmZF15′RACE1: 5′-ATTTGCATCTTCTTCACCATTACTTTTAC-3′), together with a forward primer capable of annealing to the RNA oligo added to the 5′ end of the cDNAs (GeneRacer 5′ Primer – 5′-CGACTGGAGCACGAGGACACTGA-3′). The PCR reaction mixtures were performed in a 50 μL final volume containing 0.2 μM of each primer, 400 μM of each dNTP, Advantage 2 Polymerase Mix 1× in Advantage 2 PCR Buffer (Clontech, Mountain View, CA, USA) and 1 μL cDNAs. The amplification was carried out using a cycling program based on step-down annealing temperatures, consisting of an initial denaturing step of 94 °C for 2 min followed by 5 cycles of 94 °C for 30 s, 60 °C for 30 s, and 68 °C for 1 min and 30 s; 5 cycles of 94 °C for 30 s, 55 °C for 30 s, and 68 °C for 1 min and 30 s; 25 cycles of 94 °C for 30 s, 50 °C for 30 s, and 68 °C for 1 min and 30 s, and a final extension step of 68 °C for 10 min. Control reactions were done to confirm the specificity of the generated product, using only one of each primer. Nested PCR was performed when oligo dT generated cDNA was used, following the same conditions described above, but using the following primers: SmZF15′RACE2 (5′-TCCCCATAGCGGCAAATCCACTTAGA-3′) and GeneRacer 5′ Nested Primer (5′-GGACACTGACATGGACTGAAGGAGTA-3′).

The specific bands originated from either PCR or nested PCR were excised from agarose gels using the Wizard SV Gel and PCR Clean-up System (Promega) according to manufacturer instructions. The DNA was then cloned into the PCR 2.1 TOPO vector, using the TOPO TA Cloning Kit (Invitrogen) according to the manufacturer protocol. Ligation products were used to transform *E. coli* One Shot TOP10 chemically competent cells, and the rescued plasmids were sequenced using the M13 Reverse Primer (Invitrogen). The sequencing reactions were performed by GATC Biotech (Constance, GE). The resulting sequences were analyzed using the software DNASTAR (http://www.dnastar.com/).

2.3. Protein sequence analysis

The SmZF1 cDNA sequence, redefined after the experiments described above, was used to predict the SmZF1 protein sequence using the Translate Tool in the Expasy Proteomics Server (http://ca.expasy.org/). Initially, the new sequence was used to visually define the putative zinc finger motifs and to characterize possible additional protein characteristics. To confirm the predicted motifs, in silico methods, such as the search tools from Pfam (Protein families database) [14] and an ab initio approach developed by Kaplan et al. [15], were used. Protein secondary structure

prediction based on position-specific scoring matrices was performed on the PSI-PRED server [16].

A more precise characterization of each exon and the protein regions coded by them was performed using BLAST [13] searches against the non-redundant (nr) database. The protein regions were submitted to the blastp program, automatically adjusting parameters to search for a short input sequence, and the results were manually inspected.

2.4. Comparative modeling

The tridimensional structure for the SmZF1 protein was built by comparative modeling. Initially, a BLAST search against the Protein Data Bank (PDB) [17] was performed to look for candidate structural templates. Two determinant factors were considered to define a suitable template structure for SmZF1: query coverage (since it was a limitation concerning the best hits set) and e-value score. The structure selected as template was a six-zinc finger protein artificially designed to recognize ANN triplets (PDB ID: 2I13) [18].

The sequences from both SmZF1 and the template structure were submitted to a pairwise alignment using PROMALS 3D [19] with default parameter values. The resulting alignment was inspected and minor corrections were performed. After generating a sequence alignment between SmZF1 and its template structure, molecular models were built using MODELLER [20]. As zinc ions are essential for the stability of zinc finger proteins, they were explicitly considered through all comparative modelingsteps.

As previously described [21], we applied a two-step protocol to obtain a SmZF1 model: firstly a set of one hundred preliminary candidate structures was generated using MODELLER [20]. We then applied the loop modeling protocol from MODELLER to refine protein regions not covered by template structure (residues 15–45, which includes an insertion fragment absent in the template structure), resulting in a refined set of candidate structures. These refined structures of SmZF1 were evaluated regarding stereochemical and native-like properties, according to Procheck Ramachandran plots [22] and ProSA *Z*-score [23], respectively. RMSD (root mean square deviation) calculations between the models and the template structure were performed after structural alignment using PyMOL align function [24]. Structures with the highest scores in all approaches (as described in Section 3) were manually analyzed and the best evaluated model was submitted to molecular dynamics simulations as described below. Visualization and manipulation of molecular images were performed with PyMOL [24].

2.5. Molecular dynamics

We performed two replicates of 10 ns molecular dynamics (MD) to assess the structural stability and convergence of the best-evaluated structure. To assure a correct representation of the zinc finger motifs we manually set the protonation states of the zinc coordinating residues. All cysteine residues responsible for zinc coordination were deprotonated: C46, C49, C106, C109, C135, C138, C164 and C167. In addition histidine residues in the proximity of the zinc ion were also deprotonated on the nitrogen closer to the metal [25,26]. Therefore, H62, H67, H122, H127, H151, H156 and H180 were deprotonated at Nε. Zinc parameters were taken from the Amber forcefield.

The GROMACS 4 package [27] was used to perform molecular dynamics simulations, energy minimization, and trajectory analyses under the AMBER99SB force field [28]. A 1.2 nm layer of explicit TIP3P water molecules [29] was added around the solute molecules, within a cubic water box, using periodic boundary conditions. Counter ions were inserted for system neutralization. LINCS [30] was applied to constraint solute bond, while Settle [31] was applied to constraint solvent bond. Temperature was maintained at 298 K by using a stochastic term to rescale velocities [32] and pressure was kept at 1 atm using the Berendsen approach [33]. Electrostatic interactions were treated with the PME method [34], using non-bonded cutoffs of 1.0 nm for Coulomb and 1.2 nm for Van der Waals. A 2 fs integration time was used throughout the MD.

A 3-step energy minimization protocol was used as previously described [35,36]: firstly, applying the steepest-descent algorithm: (i) 5000 steps with solute heavy atom positions restrained to their initial positions using a harmonic constant of 1 kJ/mol nm in each Cartesian direction, allowing free water and hydrogen movements and (ii) 5000 steps with all atoms free to move. Subsequently, the conjugated gradient algorithm was applied for further energy minimization until an energy gradient of 42 kJ/mol nm was achieved.

Subsequently, we performed a heating procedure from 20 to 298 K. For this proposal we performed a 500 ps MD keeping the protein heavy atoms restrained to their initial positions (using the same previous harmonic restraint potential). The velocities were assigned for an initial temperature of 20 K and we used the "annealing" option on the ".mdp" gromacs file to heat gradually the system until it reached the temperature of 298 K. To obtain distinct trajectories we have chosen distinct seeds for the random number generator when assigning velocities according to a Maxwell Boltzmann distribution. Therefore each simulation started with distinct initial velocities which resulted in two independent trajectories.

We then performed an equilibration procedure consisting of a preliminary MD (1.5 ns), gradually reducing the positional restraint potential in steps from 50 to 0 kJ/mol nm as follows: 100 ps MD steps for each potential – 50, 25, 10, 5, 2.5, 1, 0.5, 0.25, 0.1 and 0.05 kJ/mol nm – and 500 ps with no position restraint (Supplementary Fig. 1). This procedure allows the system to gently achieve solvent equilibration and avoid artifacts. Finally, we carried out a 10 ns production simulation without restraints.

2.6. Prediction of protein–DNA interaction sites

Aiming to identify potential DNA binding sites for SmZF1 interaction, we initially identified the amino acids most likely responsible for DNA interaction, i.e., the residues −1, 2, 3, and 6 (numbered in relation to the beginning of the α-helix) of each of the DNA-binding zinc finger motifs [37]. We started by performing a visual inspection of the previously generated 3D structure, and a comparison between SmZF1 sequence and other C_2H_2 proteins presenting well characterized DNA interacting residues. Here we have chosen the structure used as template for comparative modeling of SmZF1 (PDB entry: 2II3) [18] and Zif268, a mammalian transcription factor containing three C_2H_2 zinc fingers [37]. Additionally, we carried out a prediction of the protein residues responsible for DNA interaction using the DISPLAR software [38] which is based on a neural network architecture trained over a representative set of protein–DNA complexes experimentally solved.

After identifying the SmZF1 residues hypothetically responsible for DNA interaction, we selected them as inputs for three different algorithms designed to identify putative DNA binding sites for zinc finger proteins [15,39,40]. These algorithms are based on the previous knowledge of residues–nucleotide interactions from crystallographic studies of protein–DNA complexes [37,41–44]. Each of the programs retrieved some SmZF1 putative interacting sequences. The retrieved results were compiled, yielding a degenerate consensus sequence (5′-GCAGCGTAG-3′), which was further used in a search for genes under SmZF1 control.

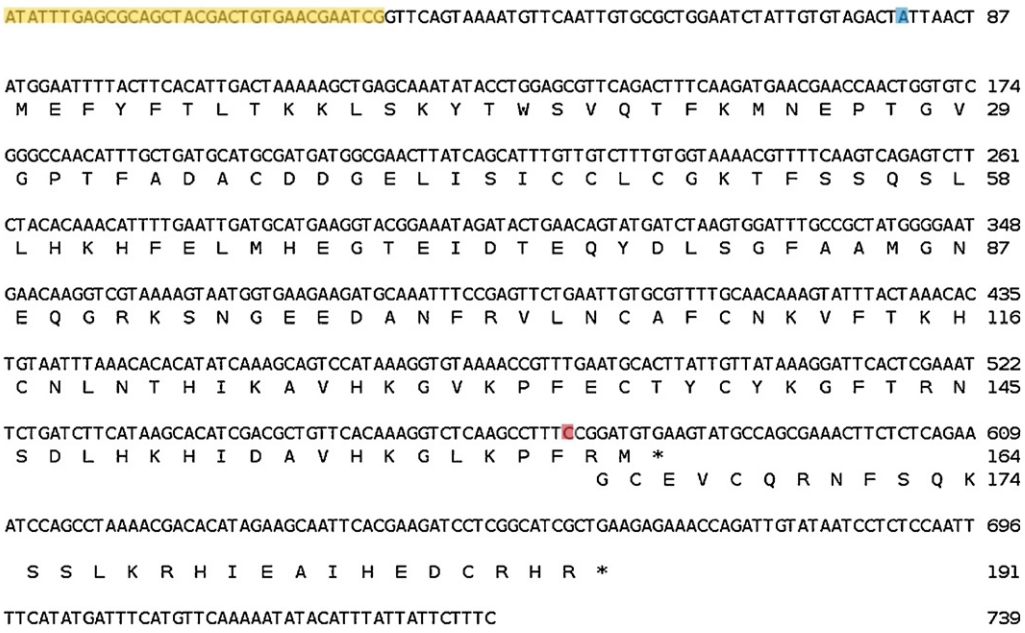

Fig. 1. New cDNA and protein sequences of SmZF1. The new SmZF1 cDNA sequence was defined from re-sequencing and 5′ RLM-RACE experiments. The protein sequence was obtained after cDNA translation using the Translate Tool in the ExPASy server (http://ca.expasy.org/). The 5′ UTR region added by 5′ RLM-RACE assays is highlighted in yellow. Blue indicates an adenine added to the sequence previously described [9] and red indicates a cytosine deleted from it. The new C-terminal region containing 29 additional residues originated from the frameshift caused by the cysteine insertion is shown in the amino acid sequence. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

2.7. DNA structure modeling

3D-DART (3DNA-Driven DNA Analysis and Rebuilding Tool) [45] was used to build the DNA structure for molecular docking calculations. This program generates custom 3D structural models of DNA using the rebuild functionality of the 3DNA software package [46] and additional internal tools to correct the global conformation of the DNA models. The input for 3D-DART was the 5′-GCAGCGTAG-3′ DNA sequence (identified as described above) and a base-pair step parameter file, which we allowed 3D-DART to automatically generate using canonical features. Molecule type was set to BDNA and all other parameters were kept as default.

2.8. Protein–DNA docking

All SmZF1-DNA docking calculations were performed with the HADDOCK web server using its guru interface [47]. We performed calculations on the SmZF1 model and also on the average structures of each 10 ns MD simulation. The SmZF1 active residues were selected as: S55, S57, L58, K61, K115, C117, D118, T121, R144, S146, D147, K150, Q173, S175, S176 and R179 corresponding to residues −1, 2 and 6 from each α-helix. The passive residues were defined as those within a 6.5 Å radius around the active residues. The protonation states of the cysteine and histidine residues involved in zinc coordination were the same as those applied on the MD simulations. The entire DNA sequence was defined as active residues in the docking process. All additional parameters were used in default values.

The docking protocol consisted of three steps. First a rigid body docking was carried out on a set of 1000 structures. We performed as previously described an extra sampling over 180° rotated solutions to avoid false positives [48]. Subsequently the best 200 solutions according HADDOCK score function were submitted to a semi-flexible refinement in torsion angle space. Finally, the top 100 structures were refined in a 8 Å shell of TIP3 water molecules [29]. A 7.5 Å RMSD cut-off value was set for clustering the final structures obtained which were then classified energetically.

2.9. Prediction of SmZF1 target genes

The predicted degenerate SmZF1-DNA binding sequence described in this work was used as a query to search the entire *S. mansoni* genome (retrieved from NCBI in GenBank format) in both strands using the fuzznuc program from the EMBOSS package [49]. As output we obtained the locations of each SmZF1 putative binding site in the genome. Subsequently, we developed an in-house perl script (version 5.10.0) using the Bio::Perl module (version 1.006001) to parse the *S. mansoni* genome in order to obtain the start positions of all known *S. mansoni* genes. This script also searched for SmZF1 putative binding sites located within 1000 bp upstream of the beginning of *S. mansoni* genes, therefore finding genes possibly under SmZF1 control. The coding regions of these genes were automatically annotated using the Blast2GO software [50] with default parameters. The most abundant Gene Ontology (GO) categories and protein family (Pfam) [14] domains found with the annotation were identified using the standard Blast2GO graphical outputs.

3. Results

3.1. SmZF1: a four zinc finger protein

Attempting to characterize SmZF1 cDNA expression in different life cycle stages of *S. mansoni*, we performed a search using BLAST [13] tools against a database containing expression sequence tags (ESTs) from the parasite (dbEST) [51]. Unexpectedly, we found a divergence between the SmZF1 coding sequences present at dbEST and that previously reported [9]. To validate our analysis, we extensively sequenced the SmZF1 cDNA obtained from adult worms and confirmed the insertion of an extra cytosine in the last SmZF1 gene exon in the published SmZF1 coding sequence (GenBank ID: AAG38587). The cytosine depletion revealed a different open reading frame (576 bp long) with a new stop codon further on in the sequence. This resulted in a distinct C-terminal sequence, the full length protein comprising 191 amino acids (Fig. 1).

Fig. 2. The new SmZF1 protein sequence contains four zinc finger motifs. The SmZF1 amino acid sequence was visually inspected to search for zinc finger motifs (first line). The conserved domains databank from Pfam [14] (second line) and the algorithm developed by Kaplan et al. [15] (third line) were used to confirm this visual inspection. The predicted zinc finger motifs, defined by the presence of $CX_{2-4}CX_{12}HX_{4-6}H$ elements, are highlighted. The protein sequence was also assessed for motifs coded by each exon of the gene. The region coded by the first gene exon is underlined in red. Blue, yellow, and green indicate the regions coded by the second, third, and fourth exons, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

To further confirm the transcription initiation site and the length of the 5′ UTR region we performed RLM-RACE experiments and found that this region is 34 bases longer than the original sequence [9]. This approach revealed another sequence inconsistency upstream of the translation start point, where an adenine is placed between two adjacent thymines (Fig. 1).

We next analyzed in detail the sequence and tridimensional structure of SmZF1. The sequence survey identified four, instead of the three previously reported putative zinc finger motifs [9]. All of them can be defined as members of the C_2H_2 zinc finger class, characterized by the presence of $CX_{2-4}CX_{12}HX_{4-6}H$ elements, where X represents any amino acid (Fig. 2) [1]. Interestingly, the first zinc finger motif is distant from zinc fingers 2 to 4 and is coded by the third exon, while the other three are coded by the fourth exon of the SmZF1 gene (Fig. 2). Subsequently, we generated molecular models of the SmZF1 protein to better characterize the three-dimensional structure of the protein. Initially, we searched for the best suitable template structure. In this step, we used the BLAST [13] algorithm with the blastp program to search against the Protein Data Bank (PDB) [17]. The subject structure with best coverage (73%), acceptable e-values (2e−10) and identity values (32% of identical residues and 47% of similar residues) was a six-zinc finger protein artificially designed to recognize ANN triplets (PDB ID: 2I13) [18]. This high-resolution structure (1.96 Å) was then selected as a template for comparative modeling of SmZF1. The initial structures generated by MODELLER [20] were recursively evaluated and refined, and the set of best candidate structures was submitted to energetical and stereochemical analysis. Finally, by comparing the obtained structures, we selected one that presented the highest score in both criteria (with 100% of residues in the allowed regions of Ramachandran plots [22], 91.8% of them being in the most favored region and a ProSA *Z*-score [23] of −1.93). Visual inspection of the SmZF1 model

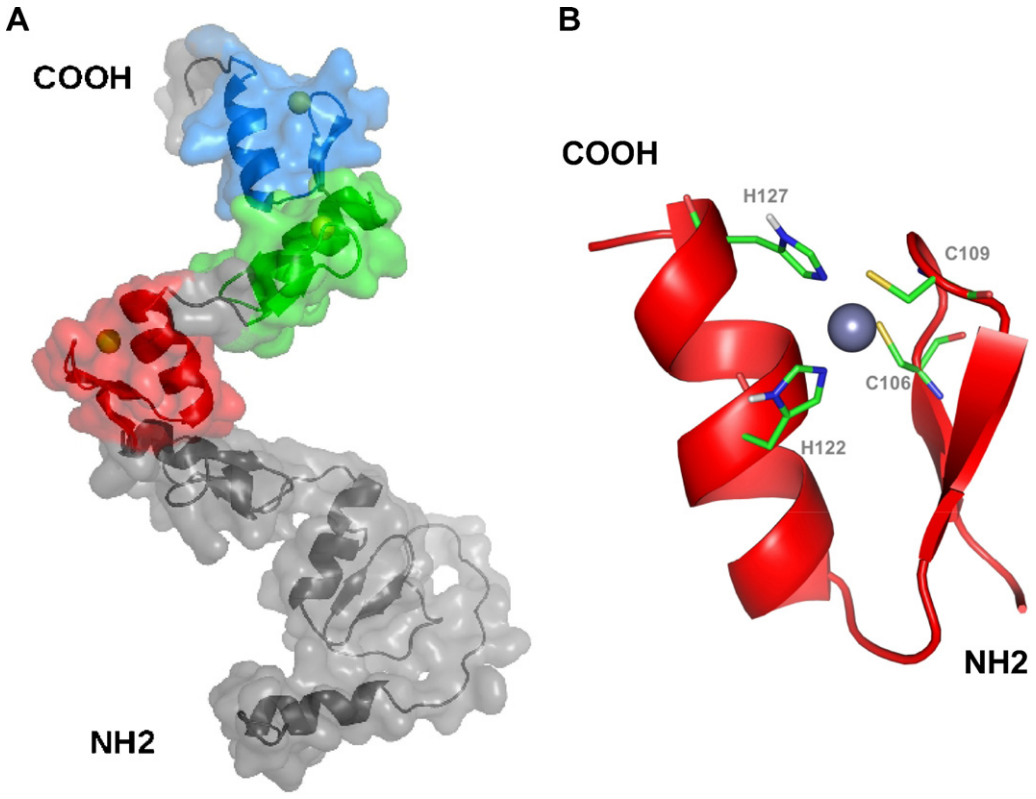

Fig. 3. SmZF1 molecular model confirms the presence of four zinc finger motifs. (A) The proposed tridimensional structure of SmZF1 was obtained by comparative modeling using MODELLER [20] with the Aart protein (PDB ID: 2I13) as a structural template and visualized with PyMOL [24]. The DNA binding zinc fingers are colored in red, green, and blue. (B) Detailed view of the zinc ion coordination by the histidine and cysteine residues in the third zinc finger from the SmZF1 model. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

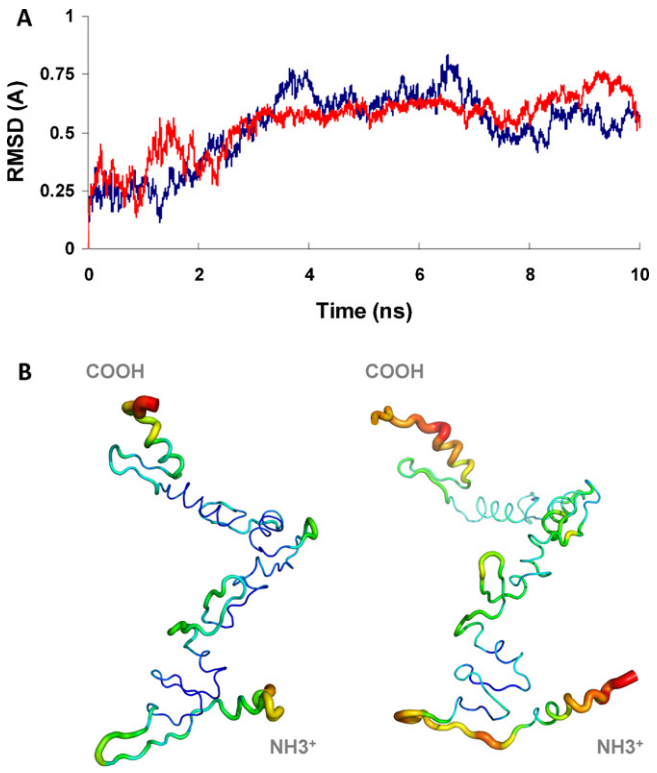

Fig. 4. Assessing the stability of the SmZF1 model by molecular dynamics simulations. Two MD simulations with different initial velocities were carried out using the GROMACS 4 package [27]. (A) RMSD (root mean square deviation) behavior throughout the MD simulation for the first 10 ns of both replicates. (B) Structures colored and thickened according to their root mean square fluctuations (RMSF). Residues displaying higher flexibility are thicker in this representation and show a more intense color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

structure confirmed the presence of the four zinc finger motifs predicted by sequence analysis (Fig. 3A). The cysteine and histidine residues were placed in conserved positions toward the zinc ions, suggesting that these motifs are well defined in the model, presenting a native-like ion coordination (Fig. 3B). The SmZF1 model was submitted to the Protein Model Data Base (PMDB) [52] (PDB ID: PM0076555).

The best evaluated model structure was further submitted to two molecular dynamics (MD) simulation replicates. Both MDs revealed a tendency for the stabilization of protein structure, assessed by the analysis of the time evolution of the root mean square deviation (RMSD) of SmZF1 backbone atoms throughout the trajectories (Fig. 4A).

Although we obtained relatively high deviations during the trajectories, it is worth noticing that the RMSD between the average structures and the model was small (1.8 Å for trajectory 1 and 2.1 Å for trajectory 2). A visual inspection of these structures revealed that the overall fold of the protein was maintained. If we compare the two trajectories, the values are even smaller (1.1 Å). We have also inspected the time-evolution of the solvent accessible surface (SAS) of SmZF1 during the trajectories. As displayed in Supplementary Fig. 2, both hydrophobic and hydrophilic exposed areas were stable throughout the entire simulations (Supplementary Fig. 2).

The root mean square fluctuation analysis by residue (RMSF) revealed a high flexibility at the N and C termini (Fig. 4B). Interestingly, this behavior is also observed on the template structure used in modeling (PDB ID: 2II3), since the analysis of the crystallographic b-factors also reveals high mobility on the terminals (Supplementary Fig. 3).

In addition, it was observed that the zinc finger motifs were stable as native-like structures during our simulations. Supplementary Table 1 displays the averaged Zn–S distance and S–Zn–S angle compared with values obtained by Zhang et al. [25]. We observed an average Zn–S distance of 2.14 Å while these authors have reported an average distance of 2.13 Å. Concerning the S–Zn–S angle, our results are closer to values obtained from the X-ray solved structure than to those reported on the aforementioned study (Supplementary Table 1). Lastly, it was observed and confirmed by secondary structure analysis that the folding of all four zinc fingers were maintained (data not shown).

3.2. New DNA binding sites on SmZF1

After the obtaining of a three-dimensional structure for SmZF1, the next step was to define potential DNA binding sites for the protein interaction, based on its newly revealed structural features. First, we predicted the interacting residues of each of the last three zinc finger motifs (since the first one was more likely to be involved in protein–protein interactions). This prediction was based on the well-established concept that residues at positions −1, 2, 3, and 6 (related to the beginning of the α-helix) of each zinc finger interact with a three-nucleotide site in the DNA molecule, in an antiparallel manner [37]. Thus, residues S55, S57, L58, K61, K115, C117, D118, T121, R144, S146, D147, K150, Q173, S175, S176 and R179 were identified according to these criteria. Additionally, we performed a prediction with the DISPLAR software, which is a neural network based method that allows identification of possible interacting residues on the protein surface [38]. Interestingly, the results obtained confirmed all of the above mentioned residues as possible DNA binding sites.

Next, we searched for the most likely DNA interacting sequence. To this purpose, we employed three different programs that specifically predict DNA binding sequences for C_2H_2 zinc finger proteins [15,39,40]. Each of the programs retrieved some SmZF1 putative interacting sequences, as follows: 5′-GNDSCGSAG-3′ for the algorithm developed by Kaplan et al. [15]; 5′-GVAVCGVAV-3′ for ZIFIBI, developed by Cho et al. [39]; and 5′-GCAGCGTGT-3′ for *Zifnet*, developed by Liu and Stormo [40] (where N = A, G, C or T; D = G, A or T; S = G or C; R = A or G and V = G, C or A, according to IUPAC definitions). These sequences were then used to define a degenerate one: 5′-GVARCGNAG-3′ likely to be involved in SmZF1 interaction.

In order to evaluate the stability of the interaction between SmZF1 zinc fingers 2–4 and one of the putative DNA binding sites defined in silico (5′-GCAGCGTAG-3′), we performed docking calculations using the HADDOCK software [47] as previously described elsewhere [53,54]. Calculations were performed using the generated protein model and also using the average structures obtained on each MD simulation.

Docking solutions were energetically classified and clustered using a 7.5 Å RMSD cut-off. The majority of the 1000 generated complexes presented the same topology, indicating this as the most likely protein–DNA binding conformation. Since the average structures obtained on the MD simulations were very similar to the model (see Section 2.5), the docking results obtained in all calculations were the same. The best-ranked complex (Fig. 5) presented an electrostatic energy value of −870.166 kcal/mol within a total energy value of −919.458 kcal/mol. The values are in accordance with the fact that electrostatic forces are the main component of protein–DNA interactions. Taken together, these results suggest that the predicted DNA binding site 5′-GCAGCGTAG-3′ is a plausible target for SmZF1.

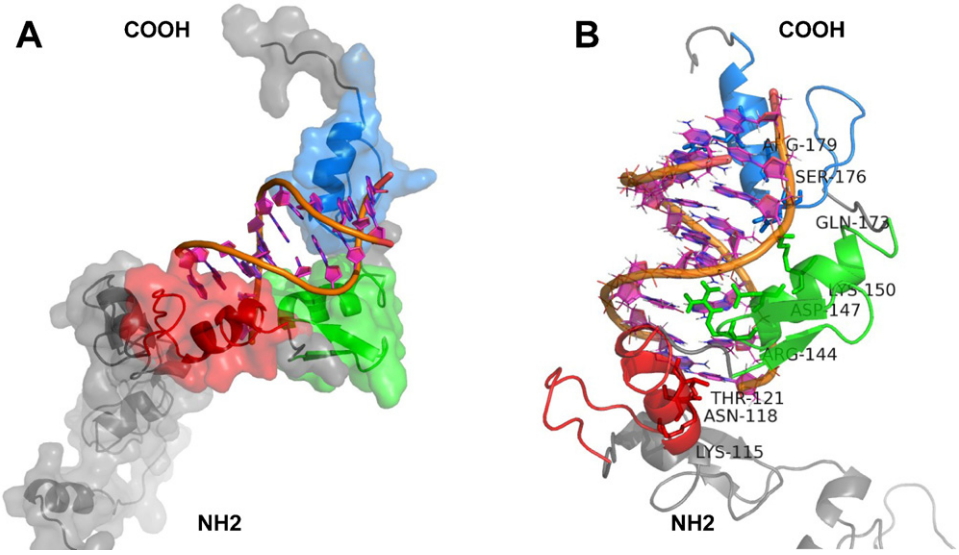

Fig. 5. Identification of a SmZF1 putative DNA binding site with docking calculations. Docking calculations were performed with the HADDOCK webserver [47] between the SmZF1 predicted structure and a putative DNA binding site defined in silico (5′-GCAGCGTAG-3′). Active residues were selected based to known interaction features of zinc finger proteins. The backbone of the DNA molecule is shown in orange and the three zinc fingers responsible for DNA interaction are colored in blue, green, and red. (A) The molecular surface of the protein is detailed. (B) The active residues are shown in licorice representation and the protein structure is shown in cartoon, evidencing secondary structure elements.

3.3. SmZF1 putative target genes are mainly involved in parasite development

We searched for the degenerate sequence defined as a consensus for SmZF1 binding (5′-GVARCGNAG-3′) in the *S. mansoni* genome to identify genes putatively under SmZF1 control. Those hits localized within gene promoters, hereby defined as the region comprising 1000 bp upstream of the initial ATG codon, were classified as putative *cis* elements under SmZF1 control. A total of 415 genes were, therefore, identified as potential targets for SmZF1 action. Taking advantage of an automatic annotation approach we were able to assign function to 124 of these genes, which were then categorized according to their coded protein molecular function (Fig. 6A), cellular compartment (Fig. 6B) and biological process (Fig. 6C). Interestingly, when the biological process was considered, approximately one quarter of these genes were directly involved with developmental mechanisms.

4. Discussion

In the present work we structurally characterized the zinc finger-containing protein SmZF1 that may act as a transcription factor in *S. mansoni* [11]. Using a combination of experimental and in silico methodologies we were able to describe four zinc finger motifs instead of the three previously reported for the protein [9].

The finding of an extra cytosine in the SmZF1 coding sequence drove the re-evaluation of the SmZF1 protein structure. Since this cytosine is placed in a dinucleotide region (5′-AAGCCTTTCCGG-3′) this mistake could be explained by the use of a gel-based DNA automated sequencer. The technology is not as precise as the more recently developed capillary sequencing equipments that generate outputs compatible with base-calling analysis. In addition, Eleutério de Souza et al. in 2001 [9] reported the sequence of a cDNA clone obtained from a phage library that could bear a sequence error. To fully characterize the SmZF1 coding sequence we also assessed its 5′ UTR and placed the transcription initiation site upstream in the sequence. These results indicate that the revised SmZF1 coding sequence may now be complete and precise.

After correcting the detected error in the protein, we determined a longer coding sequence and consequently a larger protein, with a different C-terminal. Analysis of the new protein sequence revealed four classical C_2H_2 zinc finger motifs that fit the general ($CX_{2-4}CX_{12}HX_{2-6}H$) [1] and even the more restricted consensus sequence for this kind of motif ($(F/Y)XCX_{2-5}CX_3(F/Y)X_{5\psi}X_2HX_{3-5}H$, with ψ being any hydrophobic residue) [43], except for the first F/Y restriction. It is well established that zinc ion coordination confers most of protein stability, even in the case of proteins bearing slight differences from the consensus, as is the case for SmZF1 [55,56]. A special attention was directed to the correct representation of the zinc ion motifs, both in the generation of three-dimensional models and also during the molecular dynamics simulations for SmZF1. When analyzing all obtained results, we were able to confirm that the four zinc-finger motifs were correctly coordinating zinc ions through their cysteine and histidine residues (Supplementary Table 1).

We questioned about the possibility of other zinc finger motifs on the SmZF1 structure, since the template structure contains 6 zinc ions. Following the same protocols employed here, we modeled a six-ion structure and carried out MD simulations to assess the stability of the putative new motifs. The results obtained revealed that the zinc ions were not properly coordinated and the putative additional zinc finger motifs were not stabilized (data not shown).

We found that the first SmZF1 zinc finger is distant from the other three motifs. Zinc finger clusters are frequently described in proteins capable of performing interactions with more than one type of molecule [57]. Also, several zinc finger proteins have been shown to perform protein–protein interactions in addition to their DNA binding abilities [58,59]. It is thus possible that motifs 2–4 are responsible for SmZF1-DNA binding, while the first one is involved in protein–protein interactions. According to Brayer and Segal, in multiple zinc finger proteins typically only 3–4 of them interact with DNA while the others are involved in other types of interaction [59]. Indeed, WT1, a protein containing four C_2H_2 zinc finger motifs, interacts with other proteins through its first motif and with DNA molecules through its last three zinc fingers [60]. Dimerization of DNA binding zinc finger proteins is known to facilitate communication between distant elements present in a DNA molecule as well as to enhance and diversify the DNA binding

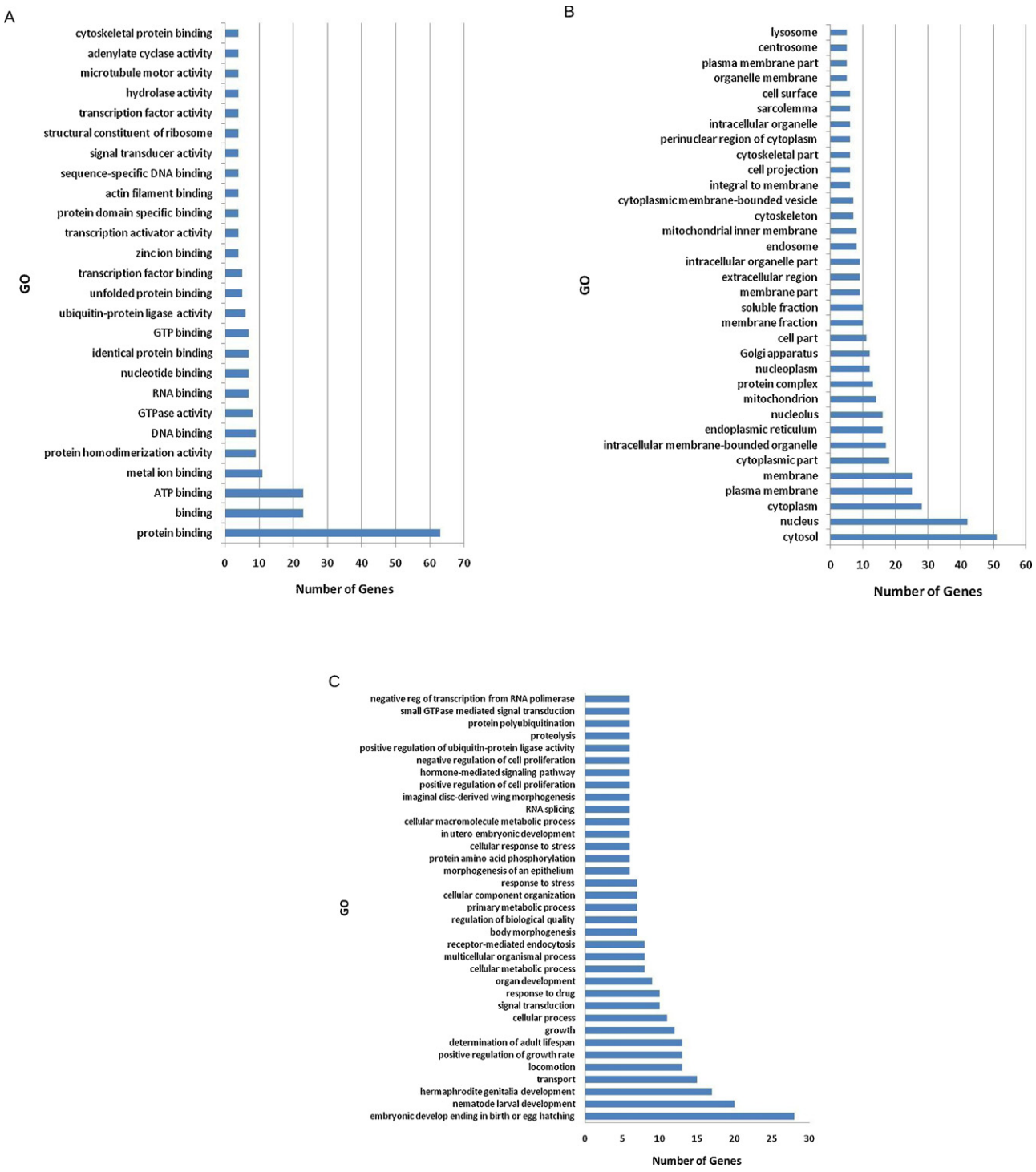

Fig. 6. Genes potentially controlled by SmZF1 were classified according to GO categories. Genes containing the predicted SmZF1 binding sites in their promoter region (defined here as the region comprising 1000 bp upstream of the initial ATG codon) were assumed to be under SmZF1 control. These genes were annotated based on similarity searches in the GO data bank using the Blast2GO tool and classified according to (A) the molecular functions; (B) the cellular compartments, and (C) the biological process of their coded protein.

ability of these proteins [61]. Interestingly, the last three SmZF1 zinc fingers are coded by the fourth exon of SmZF1 coding gene, while the first is coded by the third exon (Fig. 2). This suggests different origins for the putative protein–protein (zinc finger 1) and protein–DNA (zinc fingers 2–4) interaction motifs in the *SmZF1* gene structure.

Following the idea that the first zinc finger is responsible for protein–protein interactions, we defined in silico putative DNA binding sites for SmZF1 zinc fingers 2 through 4 and compiled them into a 9 bp degenerate sequence. It is worth noting that some positions in this degenerate sequence are quite well conserved and were retrieved by all programs used (5′-GVARCGNAG-3′), suggesting that they are crucial for DNA binding. Also, the obtained degenerate sequence shows a certain degree of conservation when compared with D1-3DNA (5′-GCGAGGGAGT-3′), an oligonucleotide used in previous SmZF1 functional studies [10,11].

Docking calculations performed between the generated protein model and the DNA sequence predicted to be one of the most likely for SmZF1 binding (5′-GCAGCGTAG-3′) has revealed major contributions of the electrostatic component in the protein–DNA interaction, as expected. Interestingly, similar results were obtained when using the average protein structure obtained from MD simulations in docking calculations (data not shown). Although we did not extensively investigated the energetical features of the protein–DNA binding, the topological assessment of this interaction is a very important result that will be useful in guiding future experiments (e.g. site-specific mutagenesis) in the absence of a crystal structure for the complex.

We can also speculate that the DNA molecule has an important role in stabilizing the SmZF1 protein structure. Furthermore, when we carried out docking calculations between SmZF1 and a 12 bp DNA molecule hypothetically capable of interacting with all zinc finger motifs, the DNA-binding was not stable (data not shown) and it was shifted to the last three zinc fingers. This observation supports the idea that the first zinc finger is more likely to be involved with protein rather than DNA binding.

When searching for putative *S. mansoni* genes under SmZF1 control, we found that most of them code for proteins that participate in developmental processes. It is assumed that each *S. mansoni* life cycle stage presents as many as 1000 differentially expressed genes [62]. Thus, proteins capable of regulating gene transcription, such as SmZF1, are extremely important for the maintenance of the complex parasite life cycle [4].

Several attempts have already been made to describe some regulatory proteins and their roles in parasite development [63–67], but much remains to be elucidated. Schistosomiasis is a public health problem, being endemic in 76 developing countries [68]. To fight this disease it is necessary to improve infrastructure and education, as well as to develop new drugs and vaccines [69]. Recently, Han et al. suggested that potential drugs against schistosomiasis should take into account proteins involved in processes such as DNA repair, transcription and replication [70], once again highlighting the importance of studying potential regulatory proteins such as SmZF1.

Funding source

This work received financial support from the Brazilian funding agencies CNPq (www.cnpq.br) and CAPES (www.capes.gov.br). The funders had no role in study design; collection, analysis, and interpretation of data; writing of the manuscript; nor in the decision to submit the manuscript for publication.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgements

The authors thank Dr. Priscila Grynberg for her contribution to data analysis and Dr. Alexandre M.J.J. Bonvin for valuable help on molecular docking protocols.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2012.10.004.

References

[1] S. Iuchi, Three classes of C_2H_2 zinc finger proteins, Cellular and Molecular Life Sciences 58 (2001) 625–635.
[2] C. Andreini, L. Banci, I. Bertini, A. Rosato, Counting the zinc-proteins encoded in the human genome, Journal of Proteome Research 5 (2006) 196–201.
[3] G. Parraga, S.J. Horvath, A. Eisen, W.E. Taylor, L. Hood, E.T. Young, et al., Zinc-dependent structure of a single-finger domain of yeast ADR1, Science 241 (1988) 1489–1492.
[4] M.R. Fantappie, F.M. de Oliveira, R.D. Santos, J.J. Mansure, D.R. Furtado, I.C. da Silva, et al., Control of transcription in *Schistosoma mansoni*: chromatin remodeling and other regulatory elements, Acta Tropica 108 (2007) 186–193.
[5] WHO, Parasitic Disease: Schistosomiasis, 2008.
[6] M. Reddy, S.S. Gill, S.R. Kalkar, W. Wu, P.J. Anderson, P.A. Rochon, Oral drug therapy for multiple neglected tropical diseases: a systematic review, JAMA: The Journal of the American Medical Association 298 (2007) 1911–1924.
[7] A. El-Ansary, S. Al-Daihan, Stage-specifically expressed schistosome proteins as potential chemotherapeutic targets, Medical Science Monitor 11 (2005) RA94–RA103.
[8] E.R. Jolly, C.S. Chin, S. Miller, M.M. Bahgat, K.C. Lim, J. DeRisi, et al., Gene expression patterns during adaptation of a helminth parasite to different environmental niches, Genome Biology 8 (2007) R65.
[9] P.R. Eleutério de Souza, A.F. Valadao, C.E. Calzavara-Silva, G.R. Franco, M.A. de Morais Jr., F.G. Abath, Cloning and characterization of SmZF1, a gene encoding a *Schistosoma mansoni* zinc finger protein, Memorias do Instituto Oswaldo Cruz 96 (Suppl.) (2001) 123–130.
[10] C.E. Calzavara-Silva, F. Prosdocimi, F.G. Abath, S.D. Pena, G.R. Franco, Nucleic acid binding properties of SmZF1, a zinc finger protein of *Schistosoma mansoni*, International Journal for Parasitology 34 (2004) 1211–1219.
[11] M.G. Drummond, C.E. Calzavara-Silva, D.S. D'Astolfo, F.C. Cardoso, M.A. Rajao, M.M. Mourao, et al., Molecular characterization of the *Schistosoma mansoni* zinc finger protein SmZF1 as a transcription factor, PLoS Negl Trop Dis 3 (2009) e547.
[12] F. Corpet, Multiple sequence alignment with hierarchical clustering, Nucleic Acids Research 16 (1988) 10881–10890.
[13] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, Journal of Molecular Biology 215 (1990) 403–410.
[14] R.D. Finn, J. Mistry, J. Tate, P. Coggill, A. Heger, J.E. Pollington, et al., The Pfam protein families database, Nucleic Acids Research 38 (2008) D211–D222.
[15] T. Kaplan, N. Friedman, H. Margalit, Ab initio prediction of transcription factor targets using structural knowledge, PLoS Computational Biology 1 (2005) e1.
[16] D.T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, Journal of Molecular Biology 292 (1999) 195–202.
[17] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, et al., The Protein Data Bank, Nucleic Acids Research 28 (2000) 235–242.
[18] D.J. Segal, J.W. Crotty, M.S. Bhakta, C.F. Barbas 3rd, N.C. Horton, Structure of Aart, a designed six-finger zinc finger peptide, bound to DNA, Journal of Molecular Biology 363 (2006) 405–421.
[19] J. Pei, N.V. Grishin, PROMALS: towards accurate multiple sequence alignments of distantly related proteins, Bioinformatics 23 (2007) 802–808.
[20] N. Eswar, B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M.Y. Shen, et al., Comparative protein structure modeling using MODELLER, Curr Protoc Protein Sci (2007), Chapter 2:Unit 2.9.
[21] M. Lery, M. Bitar, M. Costa, S. Rossle, P. Bisch, Unraveling the molecular mechanisms of nitrogenase conformational protection against oxygen in diazotrophic bacteria, BMC Genomics 11 (2010) S5–S7.
[22] A.L. Morris, M.W. MacArthur, E.G. Hutchinson, J.M. Thornton, Stereochemical quality of protein structure coordinates, Proteins 12 (1992) 345–364.
[23] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Research 35 (2007) W407–W410.
[24] R. Ordog, PyDeT, a PyMOL plug-in for visualizing geometric concepts around proteins, Bioinformation 2 (2008) 346–347.
[25] J. Zhang, W. Yang, J.-P. Piquemal, P. Ren, Modeling structural coordination and ligand binding in zinc proteins with a polarizable potential, Journal of Chemical Theory and Computation 8 (2012) 1314–1324.
[26] J. Wereszczynski, J.A. McCammon, Nucleotide-dependent mechanism of Get3 as elucidated from free energy calculations, Proceedings of the National Academy of Sciences of the United States of America 109 (2012) 7759–7764.
[27] B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl, GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation, Journal of Chemical Theory and Computation 4 (2008) 435–447.
[28] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, Proteins: Structure, Function, and Bioinformatics 65 (2006) 712–725.
[29] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, The Journal of Chemical Physics 79 (1983) 926–935.
[30] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations, Journal of Computational Chemistry 18 (1997) 1463–1472.
[31] S. Miyamoto, P.A. Kollman, Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models, Journal of Computational Chemistry 13 (1992) 952–962.
[32] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, Journal of Chemical Physics 126 (2007) 14101.

[33] H.J.C. Berendsen, J.P.M. Postma, W.F. Van Gusteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, Journal of Chemical Physics 81 (1984) 3684–3690.
[34] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, The Journal of Chemical Physics 103 (1995) 8577–8593.
[35] P.R. Batista, M.G.D.S. Costa, P.G. Pascutti, P.M. Bisch, W. de Souza, High temperatures enhance cooperative motions between CBM and catalytic domains of a thermostable cellulase: mechanism insights from essential dynamics, Physical Chemistry Chemical Physics: PCCP 13 (2011) 13709–13720.
[36] M.G.S. Costa, P.R. Batista, C.S. Shida, C.H. Robert, P.M. Bisch, P.G. Pascutti, How does heparin prevent the pH inactivation of cathepsin B? Allosteric mechanism elucidated by docking and molecular dynamics, BMC Genomics 11 (2010) S5.
[37] N.P. Pavletich, C.O. Pabo, Zinc finger-DNA recognition: crystal structure of a Zif268–DNA complex at 2.1 Å, Science 252 (1991) 809–817.
[38] H. Tjong, H.-X. Zhou, DISPLAR: an accurate method for predicting DNA-binding sites on protein surfaces, Nucleic Acids Research 35 (2007) 1465–1477.
[39] S.Y. Cho, M. Chung, M. Park, S. Park, Y.S. Lee, ZIFIBI: prediction of DNA binding sites for zinc finger proteins, Biochemical and Biophysical Research Communications 369 (2008) 845–848.
[40] J. Liu, G.D. Stormo, Context-dependent DNA recognition code for C_2H_2 zinc-finger transcription factors, Bioinformatics 24 (2008) 1850–1857.
[41] M. Elrod-Erickson, M.A. Rould, L. Nekludova, C.O. Pabo, Zif268 protein–DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions, Structure 4 (1996) 1171–1180.
[42] M. Elrod-Erickson, T.E. Benson, C.O. Pabo, High-resolution structures of variant Zif268–DNA complexes: implications for understanding zinc finger-DNA recognition, Structure 6 (1998) 451–464.
[43] S.A. Wolfe, R.A. Grant, M. Elrod-Erickson, C.O. Pabo, Beyond the "recognition code": structures of two Cys2His2 zinc finger/TATA box complexes, Structure 9 (2001) 717–723.
[44] Y. Choo, A. Klug, Physical basis of a protein–DNA recognition code, Current Opinion in Structural Biology 7 (1997) 117–125.
[45] M. van Dijk, A.M.J.J. Bonvin, 3D-DART: a DNA structure modelling server, Nucleic Acids Research 37 (2009) W235–W239.
[46] X.-J. Lu, W.K. Oslon, 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, Nucleic Acids Research 31 (2003) 5108–5121.
[47] C. Dominguez, R. Boelens, A.M. Bonvin, HADDOCK: a protein–protein docking approach based on biochemical or biophysical information, Journal of the American Chemical Society 125 (2003) 1731–1737.
[48] M. van Dijk, A.D.J. van Dijk, V. Hsu, R. Boelens, A.M.J.J. Bonvin, Information-driven protein–DNA docking using HADDOCK: it is a matter of flexibility, Nucleic Acids Research 34 (2006) 3317–3325.
[49] P. Rice, I. Longden, A. Bleasby, EMBOSS: the European molecular biology open software suite, Trends in Genetics 16 (2000) 276–277.
[50] A. Conesa, S. Gotz, J.M. Garcia-Gomez, J. Terol, M. Talon, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, Bioinformatics 21 (2005) 3674–3676.
[51] M. Boguski, T. Lowe, C. Tolstoshev, dbEST—database for "expressed sequence tags, Nature Genetics 4 (1993) 332–333.
[52] T. Castrignano, P. De Meo, D. Cozzetto, I. Talamo, A. Tramontano, The PMDB protein model database, Nucleic Acids Research 34 (2006) D306–D309.
[53] W. Liu, G. Vierke, A.-K. Wenke, M. Thomm, R. Ladenstein, Crystal structure of the archaeal heat shock regulator from *Pyrococcus furiosus*: a molecular chimera representing eukaryal and bacterial features, Journal of Molecular Biology 369 (2007) 474–488.
[54] D. Bessière, C. Lacroix, S. Campagne, V. Ecochard, V. Guillet, L. Mourey, et al., Structure–function analysis of the THAP zinc finger of THAP1, a large C2CH DNA-binding module linked to Rb/E2F pathways, The Journal of Biological Chemistry 283 (2008) 4352–4363.
[55] N.M. Luscombe, S.E. Austin, H.M. Berman, J.M. Thornton, An overview of the structures of protein–DNA complexes, Genome Biology 1 (2000) REVIEWS001.
[56] S.A. Wolfe, L. Nekludova, C.O. Pabo, DNA recognition by Cys2His2 zinc finger proteins, Annual Review of Biophysics and Biomolecular Structure 29 (2000) 183–212.
[57] J.M. Matthews, M. Sunde, Zinc fingers—folds for many occasions, IUBMB Life 54 (2002) 351–355.
[58] L. Sun, A. Liu, K. Georgopoulos, Zinc finger-mediated protein interactions modulate Ikaros activity, a molecular control of lymphocyte development, EMBO Journal 15 (1996) 5358–5369.
[59] K.J. Brayer, D.J. Segal, Keep your fingers off my DNA: protein–protein interactions mediated by C_2H_2 zinc finger domains, Cell Biochemistry and Biophysics 50 (2008) 111–131.
[60] R. Gamsjaeger, C.K. Liew, F.E. Loughlin, M. Crossley, J.P. Mackay, Sticky fingers: zinc-fingers as protein-recognition motifs, Trends in Biochemical Sciences 32 (2007) 63–70.
[61] J.P. Mackay, M. Crossley, Zinc fingers are sticking together, Trends in Biochemical Sciences 23 (1998) 1–4.
[62] S. Verjovski-Almeida, L.C. Leite, E. Dias-Neto, C.F. Menck, R.A. Wilson, Schistosome transcriptome: insights and perspectives for functional genomics, Trends in Parasitology 20 (2004) 304–308.
[63] R.L. de Mendonca, D. Bouton, B. Bertin, H. Escriva, C. Noel, J.M. Vanacker, et al., A functionally conserved member of the FTZ-F1 nuclear receptor family from *Schistosoma mansoni*, European Journal of Biochemistry 269 (2002) 5700–5711.
[64] B. Bertin, S. Sasorith, S. Caby, F. Oger, J. Cornette, J.M. Wurtz, et al., Unique functional properties of a member of the fushi tarazu-factor 1 family from *Schistosoma mansoni*, Biochemical Journal 382 (2004) 337–351.
[65] W. Wu, E.Y. Tak, P.T. LoVerde, *Schistosoma mansoni*: SmE78, a nuclear receptor orthologue of Drosophila ecdysone-induced protein 78, Experimental Parasitology 119 (2008) 313–318.
[66] C. Lu, E.G. Niles, P.T. LoVerde, Characterization of the DNA-binding properties and the transactivation activity of *Schistosoma mansoni* nuclear receptor fushi tarazu-factor 1alpha (SmFTZ-F1alpha), Molecular and Biochemical Parasitology 150 (2006) 72–82.
[67] W. Wu, P.T. Loverde, *Schistosoma mansoni*: identification of SmNR4A, a member of nuclear receptor subfamily 4, Experimental Parasitology 120 (2008) 208–213.
[68] D. Engels, L. Chitsulo, A. Montresor, L. Savioli, The global epidemiological situation of schistosomiasis and new approaches to control and research, Acta Tropica 82 (2002) 139–146.
[69] A. Loukas, J.M. Bethony, New drugs for an ancient parasite, Nature Medicine 14 (2008) 365–367.
[70] Z.G. Han, P.J. Brindley, S.Y. Wang, Z. Chen, Schistosome genomics: new perspectives on schistosome biology and host–parasite interaction, Annual Review of Genomics and Human Genetics 10 (2009) 211–240.