Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

Identification of a new *Schistosoma mansoni* membrane-bound protein through bioinformatic analysis

F.C. Cardoso[1], J.M.R. Pinho[1], V. Azevedo[2] and S.C. Oliveira[1]

[1]Departamento de Bioquímica e Imunologia,
Instituto de Investigação em Imunologia, Instituto do Milênio,
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
[2]Departamento de Biologia Geral, Universidade Federal de Minas Gerais,
Belo Horizonte, MG, Brasil
Corresponding author: S.C. Oliveira
E-mail: scozeus@icb.ufmg.br

Genet. Mol. Res. 5 (4): 609-618 (2006)
Received June 26, 2006
Accepted August 21, 2006
Published October 6, 2006

ABSTRACT. Progress in schistosome genome research has enabled investigators to move rapidly from genome sequences to vaccine development. Proteins bound to the surface of parasites are potential vaccine candidates, or they can be used for diagnosis. We analyzed 4342 proteins deduced from the *Schistosoma mansoni* transcriptome with bioinformatic computer programs. Thirty-four proteins had membrane-bound motifs. Within this group, we selected the Sm29 protein to be further characterized by *in silico* analysis. Sm29 was found to have a signal peptide made up of 26 amino acids, with a cleavage site between Ser26 and Val27. The glycosylation site search revealed three threonines (39, 132 and 133) with high probability of O-glycosylation and two asparagines (58 and 115) with high probability of N-glycosylation. Only one transmembrane helix was found in the C-terminal region of the protein from Leu169 to Lis191. The search for similarities and conserved motifs show that Sm29 is a protein with high identity to proteins present in *S. japonicum* (53, 52, 49, and 37% of identity) and it possesses disulfide-rich conserved domains. Apparently, Sm29 is a membrane bound

protein, and it may be an important molecule in host-parasite interactions.

Key words: *Schistosoma mansoni*, Transcriptome, Bioinformatics, Membrane-bound protein, Sm29, Vaccine

INTRODUCTION

Schistosomiasis infects more than 200 million people in the tropics and subtropics (van der Werf et al., 2003). *Schistosoma mansoni* is the most widespread species responsible for this parasitic disease. Parasite eggs laid in the hepatic portal vasculature are the principal cause of morbidity and mortality in schistosomiasis (Boros, 1989). Chemotherapy remains an important strategy of intervention against this parasitic disease (Harder, 2002), but rapid reinfection demands frequent treatment (Stelma et al., 1995; Ismail et al., 1999). Schistosome research has revealed that this is a complex parasite with capacity for immune evasion (Pearce and Sher, 1987) and modulation (Maizels and Yazdanbakhsh, 2003), which makes it harder to develop an anti-schistosome vaccine.

Since the last decade, schistosome membrane-bound antigens have been frequent targets of vaccine studies. Sm25 (Ali et al., 1991), Sm23 (Lee et al., 1995), Sm16 (Rao and Ramaswamy, 2000), and Sm13 (Abath et al., 2000) are examples of antigens that contain a signal peptide, one of the essential features of membrane-bound antigens. The immunological properties of these surface antigens have also been investigated. Koster and Strand (1994) characterized the immunoreactivity of human sera against the surface antigen Sm23, showing that anti-Sm23 antibody titers varied widely in infected patients. Abath et al. (2000) found that sera from schistosomiasis patients specifically recognized Sm13. Furthermore, recombinant Sm25 was tested in vaccination trials (Suri et al., 1997). Despite high anti-Sm25 antibody titers, protection against subsequent cercaria challenge was not observed.

With the advent of whole-genome sequencing and advances in bioinformatics, the schistosome vaccine research field has radically changed. Potential vaccine candidates should include proteins that are preferentially surface-exposed or exported and are expressed in intramammalian stages (Verjovski-Almeida et al., 2003).

Our laboratory, a part of the Minas Gerais *S. mansoni* transcriptome project, has identified 34 antigens in the adult worm cDNA library with membrane-bound motifs through bioinformatic analysis. One of the selected antigens termed Sm29, had its amino acid sequence characterized using bioinformatic tools, revealing a glycoprotein bound to the schistosome surface. This study demonstrated how computational biology has revolutionized the ways to develop new vaccines against complex parasitic diseases.

MATERIAL AND METHODS

cDNA library construction and DNA sequencing

The *S. mansoni* adult worm cDNA library used in this study was previously described

by Franco et al. (1997). DNA sequences were determined as described by Cardoso et al. (2006).

Sm29 homology searches

The Sm29 sequence was first compared with those already deposited in the GenBank (nr) and dbEST databases using the Basic Local Alignment Search Tool (BLAST) algorithm available at the NCBI website (http://www.ncbi.nlm.nih.gov/BLAST/). The search for conserved domains was performed using InterPro Scan, which integrates searches in the PROSITE, Pfam and PRINTS (http://www.ebi.ac.uk/InterProScan/) Databases.

Post-translation modification and topology prediction

The signal peptide prediction was performed using SignalP 3.0 server (http://cbs.dtu.dk/services/SignalP) and SignalP Neural Networks (SignalP-NN). N-glycosylation and O-glycosylation sites were analyzed using the NetNGlyc 1.0 (www.cbs.dtu.dk/services/NetNGlyc/) and YinOYang (www.cbs.dtu.dk/services/YinOYang), respectively. The prediction of subcellular localization was performed using WolfPSORT (http://wolfpsort.seq.cbrc.jp/). The transmembrane helices were analyzed by SOSUI (http://sosui.proteome.bio.tuat.ac.jp/sosuiframe0.html) and TMHMM, version 2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0/).

Primary and secondary structure analysis

Molecular weight (MW) and isoelectric point (pI) were calculated with the Compute pI/Mw tool (http://www.expasy.ch/tools/pi_tool.html). The secondary structure analysis was performed by GOR IV secondary structure prediction method (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html).

RESULTS

Membrane-bound protein selection

Among approximately 68,684 sequenced ESTs, 4342 proteins were deduced and classified according to the BLAST, KOG (Eukaryotic Orthologous Groups) and UNIPROT (Universal Protein Resource)/GOA (Gene Ontology Annotation) databases. These analyses showed 34 proteins with membrane-bound protein motifs. The predicted membrane bound proteins were sequenced to obtain the complete cDNA sequence and the Sm29 cDNA was selected to perform further analysis (Cardoso et al., 2006).

Sm29 has homology to *S. japonicum* proteins

Sm29 is a protein found in *S. mansoni* with low similarity to any other protein deposited in the databases, except for unknown proteins found recently in the *S. japonicum* genome. These proteins show 53% (SJCHGC03008), 52% (SJCHGC05668), 49% (SJCHGC05578),

and 37% (SJCHGC02532) identity with the Sm29 amino acid sequence. The *S. japonicum* protein that showed the best score (223) and E-value (4e-57) in the homology analysis was SJCHGC05668 (Figure 1).

```
Sm29          MKSGWEYIGIFLYIMVNILDKQRCHSVRCYVCDYCPIVTSVSISEENNCTSCSTAGYNYS  60
Consensus     M*SG+++*GIFLYI++NILDK**+**S*RCYVCDYCP+VT*VSIS**++CTSC+*+G*+Y
SJCHGC05668   MNSGFKFTGIFLYILINILDKYQIQSFRCYVCDYCPMVTDVSISPGSSCTSCAISGVDYD  60

Sm29          IHRICVFKDGIP-INFPNENRTQCNTDLCNGLTVDNTGKIPSVPIANPFRCYTCLNCTKS  119
Consensus     +HRIC++*+**P-INFP**N+**C**DLCNG**V*N****P**PIA*P**CY+CLNCT*+
SJCHGC05668   VHRICIYNNSNPPINFPKVNQKVCIGDLCNGEAVRNASGTPQGPIARPISCYSCLNCTNN  120

Sm29          NQKVLSGCGACVTTRGSGIISKFCGTTCERLYIDDQISCCSTDLCNGMTKLSIHRHVIIV  179
Consensus     NQ+V*+*CG*C+T***SG+**KFCG*TC+*+**D*++SCCST**CNGM*KLS**+**II*
SJCHGC05668   NQEVRNTCGGCLTVTSSGLKHKFCGPTCDAVRSDVEVSCCSTSFCNGMIKLSTQQSSIIS  180

Sm29          LFVCIGISKYIL  191
Consensus     +*+***IS*YIL
SJCHGC05668   VLIITAISGYIL  192
```

Figure 1. Amino acid sequence comparison between Sm29 and *Schistosoma japonicum* SJCHGC05668. Identity (Amino acid). Strong similarity (+). No identity (*). Gaps (-).

The search for conserved motifs showed interesting conserved domains related to the snake-like toxins (disulfide-rich fold) and small protein (metal ligand, heme and disulfide bridges) classes. The Sm29 mature sequence possesses 17 cysteines, which probably form eight disulfide bridges. This is a common characteristic found on the membrane surface and in secreted proteins that confer high stability to its structure.

Sm29 is a membrane-bound protein

The first feature that we analyzed was the signal peptide, which is essential for protein-membrane addressing. The complete amino acid sequence of Sm29 was submitted to the SignalP 3.0 server, revealing a signal peptide composed of 26 amino acids, with a cleavage site between Ser26 and Val27 (Figure 2).

The pI and MW of Sm29 in its mature form were 8.08 and 18.07 kDa, respectively; these were calculated using the Compute pI/Mw tool (http://www.expasy.ch/tools/pi_tool.html). The fragment from Val26 to Lis168 of Sm29 showed a secondary structure composed of 67.83% random coil and 32.17% extended-strand regions (Figure 3). Sites with a high probability of O-glycosylation were found on Thre39, Thre132 and Thre133, according to the YinOYang computer program (www.cbs.dtu.dk/services/YinOYang). This bioinformatic tool produces neural network predictions for O-β-GlcNAc attachment sites in eukaryotic protein sequences. Sites with a high probability of N-glycosylation were found on Asn58 and Asn115 according to the NetNGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc/). This server predicts N-glycosylation sites in proteins using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr. The graphs showing the O-glycosylation and N-glycosylation sites in the Sm29 amino acid sequence are shown in Figure 4A and B, respectively. The potential of each glycosylation site is shown in Table 1. These glycosylation sites found in Sm29 justify the predicted molecular mass

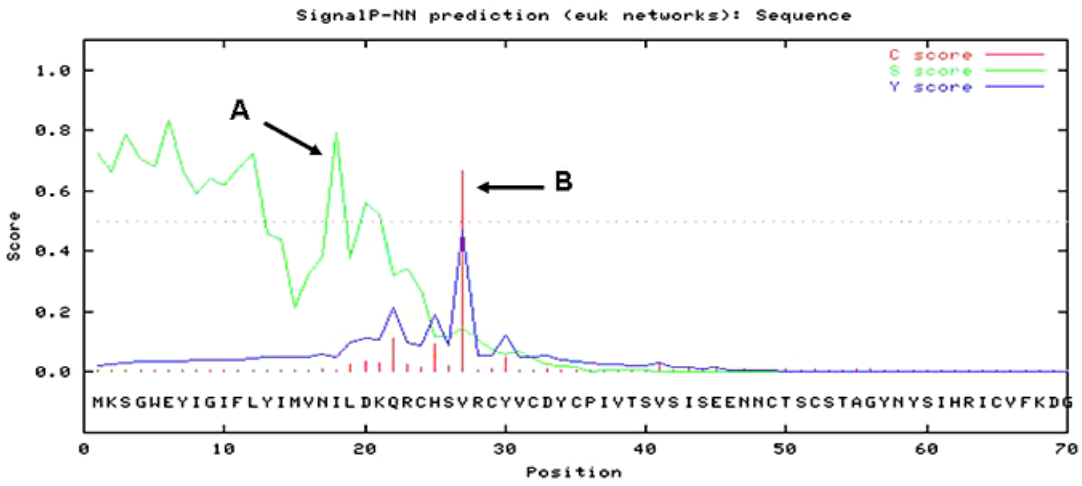


Figure 2. Sm29 signal peptide prediction using the SignalP 3.0 server according to SignalP Neural Networks (SignalP-NN). Line A corresponds to amino acids with high probability to compose the signal peptide. The cleavage site was found between Ser26 and Val27 and can be visualized in line B.

```
         10        20        30        40        50        60        70
VRCYVCDYCPIVTSVSISEENNCTSCSTAGYNYSIHRICVFKDGIPINFPNENRTQCNTDLCNGLTVDNT
ccccccccccceeceeccccccceeccccccccccceeeeeccccccccccceeeccccccccceeccc
GKIPSVPIANPFRCYTCLNCTKSNQKVLSGCGACVTTRGSGIISKFCGTTCERLYIDDQISCCSTDLCNG
ccccccccccccceeeeccccccceeecccceeeeccccceeeeeccccccceeeeccccceeeeeeeccc
MTK
eec
```

Figure 3. Sm29 secondary structure analysis according to the GOR IV secondary structure prediction method. **c** represents random coil regions and **e** extended strand regions.

of 29 kDa (Hankok K and Tsang VCW, unpublished results). Glycosylation trees confer hydrophilicity, a common characteristic of membrane-surface and secreted proteins.

The WolfPSORT subcellular localization analysis classified Sm29 into 32 different families of proteins, of which 21 belong to integral membrane protein families. The membrane helix prediction performed by SOSUI and TMHMM revealed one primary transmembrane helix in the C-terminal region from amino acids 169-191. This region is composed of highly hydrophobic amino acids (Figure 5).

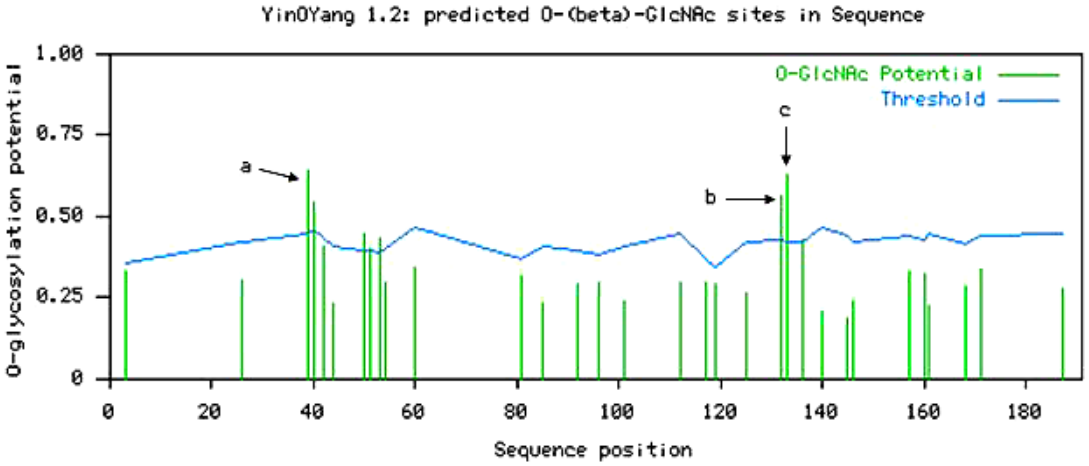
DISCUSSION

Research to develop an anti-schistosome vaccine has taken new directions with advances in molecular and computational biology. These tools provide a rapid and efficient means of gene identification and *in silico* characterization, facilitating finding vaccine candidates against this complex parasite. As described by Verjovsk-Almeida et al. (2003), potential vaccine candi-

Table 1. The putative sites for O-glycosylation and N-glycosylation in the amino acid sequence of Sm29.

Residue	Glycosylation type	Potential
Thre39	O-glycosylation	0.6431
Asn58	N-glycosylation	0.7272
Asn115	N-glycosylation	0.7534
Thre132	O-glycosylation	0.5634
Thre133	O-glycosylation	0.6298

A



B

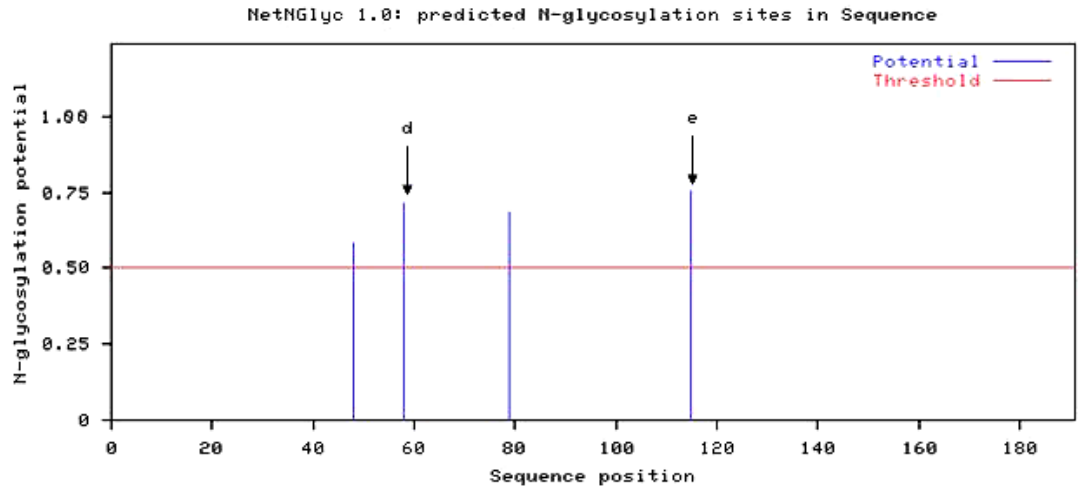
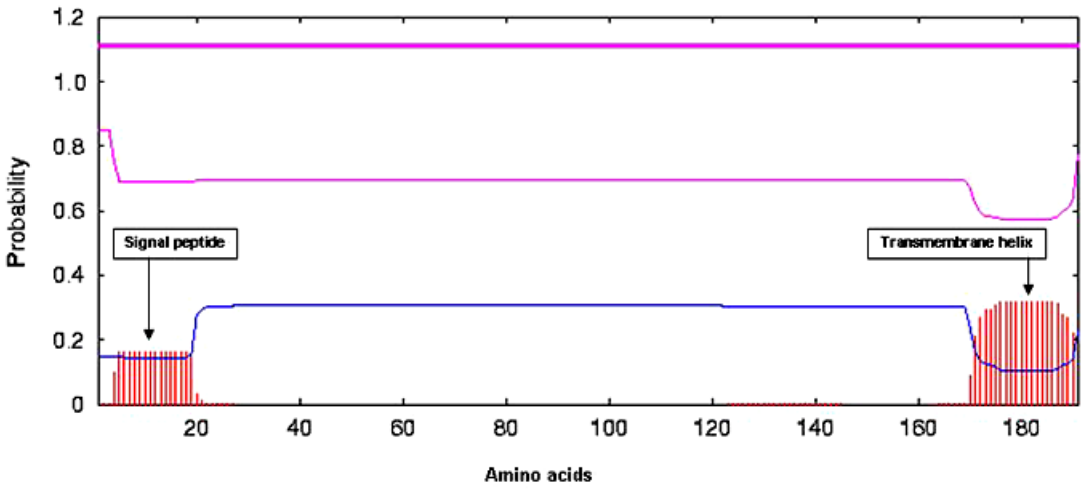


Figure 4. Graphic representation of O-glycosylation (A) and N-glycosylation (B) site analysis of the Sm29 amino acid sequence, using YinOYang and NetNGlyc 1.0 computer programs, respectively. Threshold = 0.5. The glycosylation sites are represented by the lines a, b, c, d, and e.

A



B

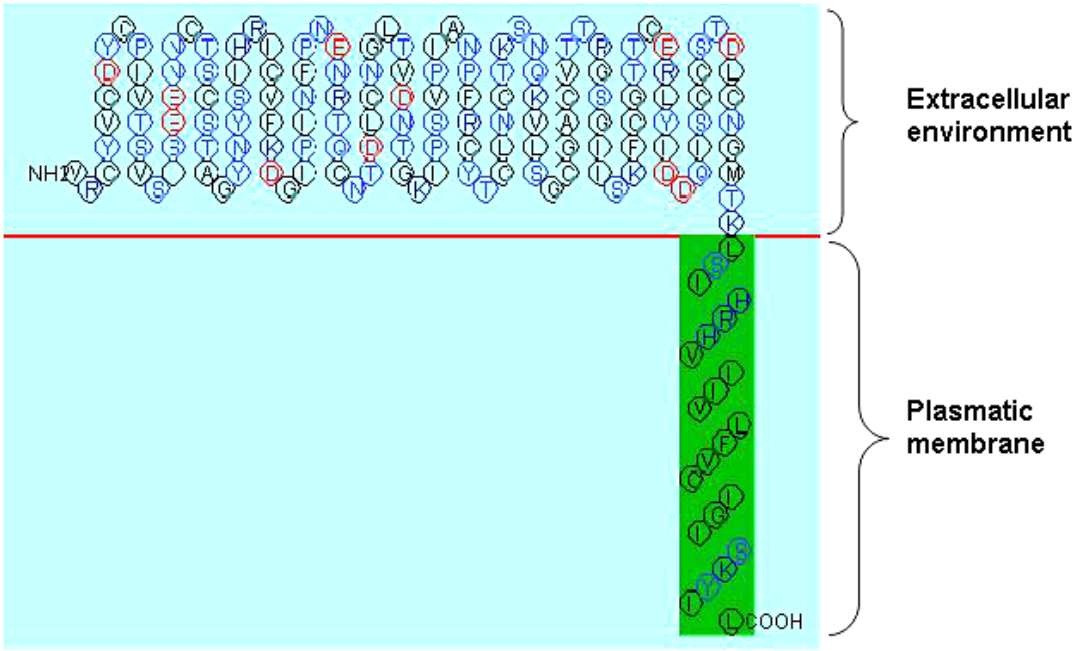


Figure 5. Topology prediction of Sm29. **A.** Hydropathy profile analysis (using the hidden Markov model TMHMM). The regions within the 1-26 amino acid and 169-191 amino acid sequences exhibited high hydrophobicity and included the signal peptide and transmembrane helix, respectively. **B.** The hypothetical transmembrane helix of Sm29 in the membrane using the SOSUI bioinformatic tool.

dates include proteins that are surface-exposed and expressed in intramammalian stages. These proteins are the first contact between the host immune system and the parasite, and they play a pivotal role in host-parasite interactions (Yoshino et al., 2001). Some recent studies using advanced molecular techniques have allowed the isolation of surface-exposed proteins (Smyth et al., 2003; Pearson et al., 2005), but their protective properties remain unknown.

We report the identification of a schistosome-membrane-bound potential vaccine candidate through bioinformatic analysis of a *S. mansoni* adult worm cDNA library. The main approaches used in this study to select membrane bound proteins were the finding of a signal peptide and membrane-bound motifs. The signal peptide is the most common characteristic analyzed by researchers when the aim is to find secreted or membrane-bound proteins in amino acid sequences (Gomez et al., 2000; Smyth et al., 2003; Pearson et al., 2005). The membrane-bound motifs in parasite surface proteins differ from those of anchored and secreted polypeptides, which is an important character to refine the search for vaccine candidates.

The Sm29 protein was selected from 34 putative membrane proteins obtained through bioinformatic analysis of the adult worm cDNA library. Bioinformatic analysis of the Sm29 amino acid sequence revealed a membrane-surface glycoprotein with an N-terminal signal peptide, three glycosylation sites and one C-terminal transmembrane helix. BLAST search analysis revealed that Sm29 has no identity with proteins with known functions in the database, except for unknown proteins found in *S. japonicum*. The search for conserved domains showed rich cysteine motifs, resulting in disulfide bridge formation, which confers stability to the protein.

Even though, bioinformatic analysis revealed that Sm29 is a membrane-bound protein, its function is completely unknown. Skelly et al. (2003) using RNAi experiments with *S. mansoni*, revealed that determining the function of unknown proteins in this organism is an important strategy. Analysis of Sm29 gene expression in different forms of the *S. mansoni* life cycle using RT-PCR and Western blot analysis may help to elucidate its biological function. Characterization of Sm29 three-dimensional structure is also important for understanding the biological importance of this molecule. This protein has no identity with proteins with defined tertiary structure. Therefore, it was necessary to perform crystallization and diffraction assays to elucidate its structure. Analysis of Sm29 secondary structure revealed the absence of alpha-helix and beta-sheets, demonstrating that Sm29 structure is probably dependent on disulfide bridge formation and sugar insertion in the glycosylation sites. Analysis of Sm29 gene orthologues and their expression in different *Schistosoma* species is also an important means to increase our knowledge about Sm29 and its importance in human schistosomiasis.

Recently, Cardoso et al., 2006 showed that Sm29 is a tegumental protein, based on immunolocalization assays made with male adult worms. The investigation of antibody isotype profiles specific to Sm29 in sera of patients living in endemic areas in Brazil showed that IgG1 and IgG3 antibodies were present at significant levels in the resistant to infection and re-infected individuals. Braschi et al. (2006a,b) studied the composition of *S. mansoni* tegument; they showed that Sm29 was one of the integral proteins consistently found in the tegument. These data demonstrate the great potential of bioinformatic analysis to select proteins of interest as vaccines. We also predicted immunodominant epitopes in vaccine-candidate antigens using algorithms, and these peptides have been shown to be immunoreactive to human T cells (Fonseca et al., 2004, 2005a,b).

With this approach, we were able to identify a new membrane-bound protein of *S. mansoni*. This protein showed some features that indicate that it is an important antigen of this

parasite (Verjovski-Almeida et al., 2003; Smyth et al., 2003; Pearson et al., 2005). Currently, we are testing recombinant Sm29 in vaccine trials in experimental animals.

ACKNOWLEDGMENTS

Research supported by CNPq, FAPEMIG, FAPESP, and PADCT/CNPq.

REFERENCES

Abath FG, Xavier EM, Allen R, Gomes YM, et al. (2000). Characterization of Sm13, a tegumental antigen of *Schistosoma mansoni*. *Parasitol. Res*. 86: 745-752.

Ali PO, Jeffs SA, Meadows HM, Hollyer T, et al. (1991). Structure of Sm25, an antigenic integral membrane glycoprotein of adult *Schistosoma mansoni*. *Mol. Biochem. Parasitol*. 45: 215-222.

Boros DL (1989). Immunopathology of *Schistosoma mansoni* infection. *Clin. Microbiol. Rev*. 2: 250-269.

Braschi S and Wilson RA (2006a). Proteins exposed at the adult schistosome surface revealed by biotinylation. *Mol. Cell Proteomics* 5: 347-356.

Braschi S, Curwen RS, Ashton PD, Verjovski-Almeida S, et al. (2006b). The tegument surface membranes of the human blood parasite *Schistosoma mansoni*: a proteomic analysis after differential extraction. *Proteomics* 6: 1471-1482.

Cardoso FC, Pacifico RNA, Mortara RA and Oliveira SC (2006). Human antibody responses of patients living in endemic areas for schistosomiasis to the tegumental protein Sm29 identified through genomic studies. *Clin. Exp. Immunol*. 144: 382-391.

Fonseca CT, Cunha-Neto E, Kalil J, Jesus AR, et al. (2004). Identification of immunodominant epitopes of *Schistosoma mansoni* vaccine candidate antigens using human T cells. *Mem. Inst. Oswaldo Cruz* 99: 63-66.

Fonseca CT, Cunha-Neto E, Goldberg AC, Kalil J, et al. (2005a). Identification of paramyosin T cell epitopes associated with human resistance to *Schistosoma mansoni* reinfection. *Clin. Exp. Immunol*. 142: 539-547.

Fonseca CT, Cunha-Neto E, Goldberg AC, Kalil J, et al. (2005b). Human T cell epitope mapping of the *Schistosoma mansoni* 14-kDa fatty acid-binding protein using cells from patients living in areas endemic for schistosomiasis. *Microbes Infect.* 7: 204-212.

Franco GR, Rabelo EM, Azevedo V, Pena HB, et al. (1997). Evaluation of cDNA libraries from different developmental stages of *Schistosoma mansoni* for production of expressed sequence tags (ESTs). *DNA Res*. 4: 231-240.

Gomez M, Johnson S and Gennaro ML (2000). Identification of secreted proteins of *Mycobacterium tuberculosis* by a bioinformatic approach. *Infect. Immun*. 68: 2323-2327.

Harder A (2002). Chemotherapeutic approaches to schistosomes: current knowledge and outlook. *Parasitol. Res*. 88: 395-397.

Ismail M, Botros S, Metwally A, William S, et al. (1999). Resistance to praziquantel: direct evidence from *Schistosoma mansoni* isolated from Egyptian villagers. *Am. J. Trop. Med. Hyg*. 60: 932-935.

Koster B and Strand M (1994). *Schistosoma mansoni*: Sm23 is a transmembrane protein that also contains a glycosylphosphatidylinositol anchor. *Arch. Biochem. Biophys*. 310: 108-117.

Lee KW, Shalaby KA, Medhat AM, Shi H, et al. (1995). *Schistosoma mansoni*: characterization of the gene encoding Sm23, an integral membrane protein. *Exp. Parasitol*. 80: 155-158.

Maizels RM and Yazdanbakhsh M (2003). Immune regulation by helminth parasites: cellular and molecular mechanisms. *Nat. Rev. Immunol*. 3: 733-744 (Review).

Pearce EJ and Sher A (1987). Mechanisms of immune evasion in schistosomiasis. *Contrib. Microbiol. Immunol*. 8: 219-232 (Review).

Pearson MS, McManus DP, Smyth DJ, Lewis FA, et al. (2005). *In vitro* and *in silico* analysis of signal peptides from the human blood fluke, *Schistosoma mansoni*. *FEMS Immunol. Med. Microbiol*. 45: 201-211.

Rao KV and Ramaswamy K (2000). Cloning and expression of a gene encoding Sm16, an anti-inflammatory protein from *Schistosoma mansoni*. *Mol. Biochem. Parasitol*. 108: 101-108.

Skelly PJ, Da'dara A and Harn DA (2003). Suppression of cathepsin B expression in *Schistosoma mansoni* by RNA interference. *Int. J. Parasitol*. 33: 363-369.

Smyth D, McManus DP, Smout MJ, Laha T, et al. (2003). Isolation of cDNAs encoding secreted and

transmembrane proteins from *Schistosoma mansoni* by a signal sequence trap method. *Infect. Immun.* 71: 2548-2554.

Stelma FF, Talla I, Sow S, Kongs A, et al. (1995). Efficacy and side effects of praziquantel in an epidemic focus of *Schistosoma mansoni*. *Am. J. Trop. Med. Hyg.* 53: 167-170.

Suri PK, Goldberg M, Madikizela M, Petzke MM, et al. (1997). Evaluation of recombinant protein r140, a polypeptide segment of tegumental glycoprotein Sm25, as a defined antigen vaccine against *Schistosoma mansoni*. *Parasite Immunol*. 11: 515-529.

van der Werf MJ, de Vlas SJ, Brooker S, Looman CW, et al. (2003). Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa. *Acta Trop.* 86: 125-139.

Verjovski-Almeida S, DeMarco R, Martins EA, Guimaraes PE, et al. (2003). Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. *Nat. Genet.* 35: 148-157.

Yoshino TP, Boyle JP and Humphries JE (2001). Receptor-ligand interactions and cellular signalling at the host-parasite interface. *Parasitology* 123: S143-S157 (Review).