JBC Papers in Press. Published on January 27, 2003 as Manuscript M211268200

Gene Structure and M20T Polymorphism of the *Schistosoma mansoni* Sm14 Fatty Acid-Binding Protein: Molecular, Functional and Immunoprotection Analysis.[1]

Celso Raul Romero Ramos §¶, Rita de Cassia Rossi Figueredo †, Thelma de Aguiar Pertinhez †, Mônica Magno Vilar ‡, Ana Lúcia Tabet Oller do Nascimento §¶, Míriam Tendler ‡, Isaías Raw §¶, Alberto Spisni †+, Paulo Lee Ho §¶#*

§ Instituto Butantan, Av. Vital Brasil 1500, CEP 05503-900, São Paulo, SP, Brasil; ¶ Instituto de Química, # Instituto Biociências, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, CEP 05508-900, São Paulo, SP, Brasil; † Centro de Biologia Molecular Estrutural, Labotarório Nacional de Luz Síncrotron, C.P. 6192, CEP 13084-971 Campinas, SP, Brasil; ‡Departamento de Helmintologia, Instituto Oswaldo Cruz-Fiocruz, Av. Brasil 4365, CEP 21045-900, Rio de Janeiro, RJ, Brasil; + Departament of Experimental Medicine, University of Parma, Italy.

* Correspondence to: P.L. Ho, Centro de Biotecnologia, Instituto Butantan, Av. Vital Brasil 1500, CEP 05503-900, São Paulo, SP, Brasil. Tel. (+55) 11 37267222 extension 2083;
Fax: (+ 55) 11 37261505; E-mail address: hoplee@butantan.gov.br.

Running title: Sm14, gene structure, polymorphism and functional properties.

Keywords: *Schistosoma mansoni*, fatty acid binding protein, gene structure, polymorphism, protein stability, vaccine.

Summary

The *Schistosoma mansoni* Sm14 antigen belongs to the fatty acid-binding protein (FABP) family and is considered a vaccine candidate against at least two parasite worms, *F. hepatica* and *S. mansoni*. Here, the genomic sequence and the polymorphism of Sm14 has been characterized for the first time. We found that the conserved methionine at position 20 is polymorphic, being exchangeable with threonine (M20T). To evaluate the function of the amino acid residue at this position, we have also constructed the mutant Sm14-A20 besides the two native isoforms (Sm14-M20 and Sm14-T20). The three purified recombinant 6xHis-tagged Sm14 proteins (rSm14-M20, rSm14-T20 and rSm14-A20) present a predominant β-barrel structure as by circular dichroism (CD) spectroscopy. Thermal and urea unfolding studies evidenced a higher structural stability of rSm14-M20 over the other forms (rSm14-M20>rSm14-T20>rSm14-A20). All Sm14 proteins were able to bind 11-(dansylamino) undecanoic acid (DAUDA), without substantial difference in the binding affinity. However, rSm14-M20 exhibited a higher affinity for natural fatty acids than the rSm14-T20 and rSm14-A20 proteins as judged by competitive experiments against DAUDA (rSm14-M20>rSm14-T20>rSm14-A20). The rSm14-M20 or rSm14-T20 isoforms, but not the rSm14-A20 mutant, were able to induce significant protection against *S. mansoni* cercariae challenge in immunized mice. The level of protection efficacy correlates with the extent of structure stability of the recombinant Sm14 isoforms and mutant. These results, besides suggesting that the M20T polymorphism may have potential functional consequences in the *S. mansoni* physiology, they also indicate that the Sm14-M20 isoform is a more suitable antigen for the development of a schistosomiasis vaccine.

Introduction

Schistosomiasis is the second major parasitic disease in the World after malaria and afflicts over 200 million people. This disease is caused by blood flukes belonging to the genus Schistosoma (*S. mansoni*, *S. japonicum* or *S. haematobium*). In America, only *S. mansoni* is found, introduced from Africa in the colonial time (1). These parasites lack the oxygen dependent pathways required for the synthesis of sterols and fatty acids, thus being entirely dependent on their hosts for these and other complex lipids supply. In fact, the intracellular fatty acid-binding protein Sm14 is particularly important for schistosomes in the uptake, transport and compartmentalization of host derived fatty acids (2). Since these proteins play a vital role in their physiology and survival, they can represent an ideal target for vaccine development. Indeed, this parasite antigen is considered a promising vaccine candidate for human schistosomiasis by the World Health Organization (3). The *S. mansoni* Sm14 showed already a protective activity against two parasite worms, *Fasciola hepatica* and *S. mansoni* (4). Recently, the T helper cells 1 (Th1) mediated immune response elicited by Sm14, has been associated with the resistance to schistosomiasis in individuals from endemic regions of Brazil (5). For this reason, the study of the *S. mansoni* gene structure and polymorphism of Sm14 is of pivotal importance in order to identify the isoform better suited to devise an efficient vaccine. The data presented in this work indicate that the sequence of the various Sm14 proteins is relatively conserved among the American strains of *S. mansoni* and provide the first experimental evidence for the existence of a reduced polymorphism, especially with respect to *S. japonicum* FABPs. In particular, the single mutation M20T is the principal example in terms of recurrence. Interestingly, this change induces important structural and functional modifications in the protein that can have direct consequence in the development of a schistosomiasis vaccine based on this FABP. In fact, the more structurally stable Sm14-M20 isoform appears to be the better vaccine candidate.

Experimental Procedures

Parasites - In the present work, we used both the male and female adult specimens of Brazilian endemic *S. mansoni* strains: LE, has been provided by the Laboratory of Helminthology - FIOCRUZ (Rio de Janeiro, Brazil), and BH by the Laboratory of Parasitology - Instituto Butantan (São Paulo, Brazil).

Sm14 encoding sequences - The genomic DNA was isolated from the *S. mansoni* worms following the standard protocols (6). The TRIzol reagent (Life Technologies, Rockville, MD, USA) was used for total RNA isolation from LE and BH schistosome strains. Two micrograms of total RNA were reverse transcribed by SuperScript II RT enzyme (Life Technologies) using an Oligo$(dT)_{18}$ primer. The first cDNA strands were used as template in the PCR reaction using the same sense and antisense primers designed for genomic Sm14 coding sequence: Forward Sm14 primer, 5′ AC CTCGAG GATATC CAT**ATG TCTAGTTTCTTGG** 3′, and Reverse Sm14 primer, 5′ TTTCCTTTT GCGGCCGC ACGCGT GAATTC GAGGCG**TTAGGATAGTCGTT** 3′. The restriction sites for XhoI, EcoRV and NdeI in the Forward primer and NotI, MluI and EcoRI in the Reverse primer, are underlined, respectively. In bold are the sequences derived from the Sm14 ORF. In addition, a sample from one cDNA library of *S. mansoni* BH strain, kindly provided by Dr. S. Verjovski-Almeida (Instituto de Química - Universidade de São Paulo, São Paulo, Brazil), was also used as template. A total of 3 independent PCR amplifications from reverse transcribed cDNAs were analyzed in this study. PCR products were purified from agarose gels, after electrophoretic separation, using the In Concert Rapid Gel Extraction System Kit

(Life Technologies, Rockville, MD, USA), cloned into pGEM-T vector (Promega, Madison, WI, USA), and sequenced using the T7 and SP6 promoter primers. The nucleotide sequence analysis of Sm14 cDNAs were performed on the ABI PRISM 377 sequencer using the ABI PRISM Big Dye Terminator Cycle Sequencing kit.

PCR in vitro mutagenesis - To create site-directed mutation in the cDNA encoding for Sm14, we used the PCR *in vitro* mutagenesis method (7). The amino acid residue at position 20, in this case methionine, was changed to alanine (M20→A20) using the forward F A20 5′ gctgtcGCgtcaaagctag 3′ and the reverse R 20 5′agctttgacGCgacagcatc 3′ primers (the nucleotide positions which were targets for mutagenesis are in upper case), in appropriate combination with Forward and Reverse Sm14 primers. The mutation was confirmed by DNA sequence analysis.

Expression and purification of the recombinant Sm14 proteins - The cDNA clones of the two native Sm14 isoforms (Sm14-M20 and Sm14-T20) and the mutant variant (Sm14-A20) cDNA clones were isolated from agarose gels after digestion with XhoI and MluI and ligated to the pAE expression vector (8) previously digested with the same enzymes. This vector allows the expression of recombinant proteins with a minimal N-terminal 6xHis-tag. The obtained constructs were used to transform *E. coli* BL21 (DE3) strain (Novagen, Madison, WI, USA). One Liter of 2xYT medium (1.5% casein hydrolisate, 1% yeast extract, 0.5% NaCl; all w/v) was inoculated with 40ml of the overnight recombinant bacteria culture from single colonies and grown until the OD at 600 nm reached 0.6. The expression of the recombinant protein was induced with 1mM isopropyl-β-D-thiogalactoside (IPTG) and cultivated for additional 3 hours at 37^{o}C. The inclusion bodies containing the recombinant 6xHis-tagged protein were isolated from

bacterial lysates and solubilized with 10 ml of buffer containing 8M urea, 0.05M Tris- HCl pH 8.0 and 0.005M 2-mercaptoethanol. The material was diluted 200x in the refolding buffer (0.005M imidazole, 0.5M NaCl, 0.05M Tris-HCl pH 8,0 and 0.005M 2-mercaptoethanol) and stirred for 16 hours at room temperature. After the refolding procedure, the recombinant protein was purified by metal-affinity chromatography in Chelating Sepharose Fast Flow resin (5 ml resin bed, 1 cm diameter column, Pharmacia, Uppsala, Sweden). Subsequently, the column was washed with refolding buffer and then the elution was carried out with a buffer containing 0.5M NaCl, 0.5M imidazole, 0.05M Tris-HCl, pH 6.8. The eluted protein was dialyzed against PBS. For the CD experiments, the protein was dialyzed against 10mM Na-phosphate buffer, pH 7.4. The protein purity was monitored by SDS-PAGE followed by Coomassie Brilliant Blue staining. The concentrations of the soluble recombinant proteins were estimated from absorbance at 280 nm, considering an extinction molar coefficient of ε^{280} =12325 $M^{-1}cm^{-1}$, based on the expected amino acid sequence of the recombinant protein.

Spectroscopic and protein stability studies - Equilibrium unfolding as a function of temperature was monitored by CD while the equilibrium unfolding as a function of denaturant concentration was studied by fluorescence spectroscopy. CD measurements were carried out on a Jasco J-810 Spectropolarimeter at 20^0C equipped with a Peltier unit for temperature control. Far-UV CD spectra were acquired using a 1mm path length cell at 0.5 nm intervals over the wavelength range from 190 to 260 nm. Five scans were averaged for each sample and subtracted from the blank average spectra. The protein concentration was kept at 10μM in 10mM Na-phosphate buffer as described above. The temperature range was from 15 to 75^0C. The loss of secondary structure was followed by measuring the molar ellipticity [θ] at 216 nm. Fluorescence changes were followed with a SLM-AMINCO-Bowman Series II Luminescence Spectrometer (Spectronic Instruments, Garforth, Leed, UK), with 1 ml samples in a quartz cuvette. The protein

concentration was 2μM and after each addition of urea the sample was equilibrated at 20^0C before measurements were made. The intrinsic fluorescence of the rSm14 proteins, mainly due to the two Trp residues, was recorded setting the excitation wavelength at 285 nm and monitoring the shift in the emission maximum in the range of 330-355 nm.

Fatty acid binding. The fatty acid dissociation constant of the three recombinant Sm14 proteins was determined by following the changes in fluorescence by increasing the concentration of the fluorescent fatty-acid analogue 11-(dansylamino) undecanoic acid (DAUDA) obtained from Molecular Probes (Molecular Probes, Inc., Eugene, OR, USA). The excitation wavelength used for DAUDA was 345 nm. A stock solution of 10mM of DAUDA in ethanol, kept in the dark at - 20°C, was freshly diluted in PBS to 1mM or 0.1mM before use in the fluorescence experiments. The protein concentration was 2μM in 1 mL PBS at 20°C. Fluorescence data were subtracted for the blank values (samples without proteins), and fitted by standard non-linear regression techniques (ORIGIN software version 6.1, Origin Lab Corporation, MA, USA) to a single non-competitive binding model to estimate the apparent dissociation constant (K_d) and maximal fluorescence intensity (F_{max}). Competitive experiments were designed to reveal a possible difference in affinity of the three rSm14 proteins for the various natural fatty acids. The miristic, palmitic, oleic and linoleic acids obtained from Sigma (Sigma, St. Louis, MO, USA) were stored and diluted as described above for DAUDA. In these experiments, the binding of 2μM of DAUDA to 2μM of protein was performed in the presence or absence of 2μM of each fatty acid.

Vaccination experiments - The three recombinant proteins, rSm14-M20, rSm14-T20 and rSm14-A20, were evaluated for vaccine efficacy in 4 - 6 week old female outbred Swiss mice as previously described (4, 9). Briefly, mice received three subcutaneous inoculations with 7 day

intervals of recombinant proteins in PBS emulsified in aluminum hydroxide as adjuvant. Control mice were injected with PBS plus adjuvant. The animals were further challenged percutaneously with 100 cercariae (LE strain)/mouse 60 days after the last immunization dose and perfused 45 days later. The overall protection efficiency was calculated by the equation [(*C* - *V*)/*C*] x 100, where *C* is the average number of worms in control animals and *V* is the average number of worms in vaccinated animals. Statistical analysis was done with Student's *t* test ($P<0.05$).

Results and Discussion

Sm14 genomic sequence and organization - The genomic sequence of Sm14 was amplified by PCR using DNA isolated from a pool of adult worms of the Brazilian endemic *S. mansoni* BH strain. The amplified product is larger than that amplified from RNA, indicating the presence of introns (data not shown). The amplified 1.7 kbp genomic DNA segment was cloned and sequenced from two independent PCR clones (GenBank accession AY055467). The sequence analysis revealed the sites of exon-intron junctions. All intron-exon boundaries obey the GT/AG rule (10) (Figure 1A). The exon nucleotide sequences were identical to that previously reported for Sm14 cDNA cloned from the Puerto Rican strain of *S. mansoni* (11). Like the other members of the FABP gene family, the gene for Sm14 contains four exons separated by three introns (12). Recently, the presence and position of the introns in the related FABP gene from *S. japonicum* was also determined by PCR and corresponded exactly to that observed here for *S. mansoni* (Figure 1B) (13). The comparison of helminth and mammalian FABPs genes showed that the sizes and positions of FABP exons are conserved among these organisms, whereas the size of the introns is quite variable (12) (Table I). The size of the FABP introns of *S. japonicum*, estimated by migration of PCR products in agarose gels (13), are larger than S. *mansoni* introns: aprox. 1200 vs 674, aprox. 850 vs 585 and aprox. 70 vs 42 nucleotides,

for introns 1, 2 and 3 of *S. japonicum* and *S. mansoni*, respectively. The intron 3 of *S. mansoni* is the smallest of all related introns found in the FABP gene family so far. Recently, a deletion variant of the *S. japonicum* FABP, called F25, was characterized by cDNA cloning (13), where the codons for the first 12 amino acids located in exon 2 were absent (Figure 1B). Taking into account the high identity in nucleotide sequence of the FABP genes of *S. mansoni* and *S. japonicum*, we speculated that in *S. japonicum*, the presence of a second 3′-acceptor site for the intron 1 at the beginning of the exon 2 would explain the generation of *S. japonicum* F25 transcript. This same alternative splicing could also be used to generate an orthologous *S. mansoni* F25 variant (shown by an arrow in the Figure 1A). After alternative splicing of intron 1 using this putative splice site, the first 12 amino acids of exon 2 could be excised while all the remaining codons would remain in frame, exactly as observed in *S. japonicum* F25. This second splice site for intron 1, however, lacks the characteristic pyrimidine rich sequence in the intron-exon junction (10), and therefore, as previously postulated for *S. japonicum* F25 (13), we do not expect this event to occur frequently.

When the obtained Sm14 genomic sequence was aligned with the nucleic acid sequences of other members of the FABP gene family, the positioning of each of the three introns was found to be conserved. As shown in Figure 2, the first intron always divides a codon after its first nucleotide, in contrast to the other exon/intron borders, which are located after the entire codon (14). It is worth to note, however, that, in spite of this highly conserved gene organization, FABPs do present few exceptions. In the desert locust *Schistocerca gregaria*, the orthologous gene does not contain intron 2 (15) and several putative lipid binding protein genes from the free-living nematode *Caernorhabditis elegans* do not have the intron 1 or possess additional introns (16).

Sm14 polymorphism - Several cDNA clones generated by independent RT-PCR reactions and cDNA libraries from two Brazilian *S. mansoni* strains were analyzed. A multiple alignment of the deduced amino acid sequences for the coding regions of all obtained Sm14 cDNA clones (a total of 30 cDNAs, 13 and 17 cDNAs from LE and BH strains, respectively) is shown in Figure 3. Eight clones showed a sequence identical to the Sm14 protein cloned from Puerto Rican *S. mansoni* strain (11), while other 14 clones presented a single mutation M20T change (ATG → ACG). Interestingly, methionine at this position is strongly conserved among the members of FABP protein family. In the rat intestinal and adipocyte FABPs, an important role for this amino acid residue in ligand binding function was demonstrated (17). Two clones, Sm14-M20-L114 and Sm14-T20-L114, displayed the synonymous mutation M114L and are considered to represent true polymorphic Sm14 proteins. Two clones, as deduced by comparison with the genomic Sm14 sequence (Figure 1A), showed the deletion of the amino acid sequence corresponding to the entire exon 3 (Sm14-ΔE3 clones), and they were probably generated by alternative splicing. Nevertheless, since the RT-PCR analysis showed a single amplification product corresponding to the Sm14 full length ORF, alternative splicing is expected to be a rare misprocessing of the mRNA precursor (results not shown). Five clones displayed silent mutations (CTA → CTG) at the Leu23 codon, which is located at 3´ end of the first exon of the Sm14 gene (data not shown). The remaining clones had Met20 plus single changes in positions where the amino acids were less conserved and are represented by only one clone. These clones may have been generated by PCR mutagenesis and for this reason, they were not further studied in this work. Figure 4 summarizes the possible splicing forms of Sm14 proteins.

Overall, almost all nucleotide sequences derived from the *S. mansoni* LE strain showed methionine at position 20 (twelve Met20 and one Thr20), while in the case of the clones obtained from one BH strain (Instituto Butantan, São Paulo), seven clones had the threonine at this position. In addition, the cDNA library derived from the another BH strain (obtained from

Instituto de Química, Universidade de São Paulo, São Paulo), yielded seven clones with threonine and three with methionine at position 20.

The polymorphism of the orthologous FABP protein was also characterized in Philippine strains of *S. japonicum* (13). In this case, the cDNA clones obtained had different sequences and none of them was identical to the previously reported FABP cDNA from the Chinese strain of *S. japonicum* Chinese (18). In conclusion, our results indicate that Sm14 proteins in both BH and LE strains are less polymorphic than the orthologous Sj-FABPs.

Significance of Sm14 M20T polymorphisms - The sequence analysis of the obtained clones indicates the existence of two main isoforms of *S. mansoni* FABP (Sm14-M20 and Sm14-T20). This polymorphism, besides the PCR clones from BH and LE strains, was also confirmed from isolated lambda cDNA library clones (11) from Puerto Rican strain (unpublished results). In order to understand the structural and functional features of these isoforms, we proceeded to verify if this amino acid position was conserved in the FABPs family using the primary sequence alignments previously reported (2, 12 and results not shown). The methionine at this position is changed for leucine in few FABP members or for valine only in testis lipid binding proteins of rat and mouse. These changes agree with the rule of "safe" residue substitutions (19), where Met can be substituted by Leu or Val without perturbing the protein structure and function. These mutations matched the changes observed for position 20 in the FABP protein family. The only exception is the Sm14-T20 isoform. It is worth to point out that the methionine residue at this position appears to play an important role in modulating the structural and functional properties of these proteins. In fact, the mutagenesis of this residue in two different FABPs (M20A in A-FABP and M18A in I-FABP) produced proteins characterized by a decrease of structural stability as well as of their affinity for fatty acids (17). Based on these evidences and since the nature of

the threonine is very different from the methionine in terms of hydrophobicity and chain size, we expect to observe significant differences between the two main isoforms of Sm14.

Structural analysis of Sm14 - The recombinant 6xHis-tagged Sm14 proteins (rSm14-M20 and rSm14-T20) were expressed and purified using a metal affinity chromatography. An additional mutant form (rSm14-A20) was also constructed in order to better evaluate the importance of this amino acid residue in the Sm14 protein. Single protein bands with molecular masses about 16 kDa were observed for rSm14-M20 and rSm14-T20 when subjected to SDS-PAGE analysis as well as for rSm14-A20 protein (data not shown). Figure 5A (black line) shows that the three rSm14 proteins present a CD spectrum typical of a protein possessing a secondary structure containing mainly β-structural elements in agreement with the prediction based on its primary sequence (4). Overall, the CD spectra evidence some differences in the chiroptical activity of the three proteins suggesting that indeed a single mutation at position 20 influences their structural features. The stability of the protein was assessed both by CD and fluorescence spectroscopy. Figure 5B shows the trace of thermal unfolding of the three rSm14 proteins. The sigmoidal shape indicates we are in the presence of a two state transition though, as also indicated by the CD spectra obtained after cooling down the protein, the transition is not completely reversible (Figure 5A, light grey line). The secondary structure recovery was higher for the rSm14-M20, followed by rSm14-T20 and rSm14-A20 (82.6%, 64.4% and 56.8%, respectively). The CD spectra at high temperature (Figure 5A, dark grey line) recorded when the ellipticity at 216 nm has reached a plateau, though they indicate a protein mainly unfolded, still present a negative shoulder at 222 nm, suggesting the presence of some residual helical determinants (Figure 5A). It is interesting to note that despite the fact that distinct fatty acid binding proteins have similar β-barrel structure, they may have different folding and unfolding intermediates (20). From the fitting of the CD traces recorded at increasing temperature (Figure

5B), the temperature of unfolding of each protein variant has been derived, showing that the rSm14-M20 isoform was more stable (Tm = 55.3 ^{0}C) than rSm14-T20 isoform and rSm14-A20 mutant (Tm = 44.6 and 45.8, respectively). The folding stability of the three rSm14 variants has also been investigated by following the change of their fluorescence emission maximum as a function of increasing concentration of urea. The data confirmed that indeed rSm14-M20 has a higher structural stability (results not shown).

Fatty acid binding analysis - The fatty acid binding capability of the three rSm14 proteins was examined by using the environment-sensitive fluorescent fatty acid analogue DAUDA that alters its fluorescence emission spectra and intensities on entry into binding proteins (21). The blue-shift of DAUDA fluorescence maximum from 543 to 538, 530 and 528 nm for rSm14-M20, rSm14-T20 and rSm14-A20, respectively (results not shown), and the concomitant increase of its fluorescence emission, both indicative of the entry of DAUDA into an apolar environment (22), confirmed the binding of the fluorescent fatty acid to the three rSm14 variants. These shifts are comparable with the results obtained from structurally related FABPs such as the heart FABP, brain FABP and adipocyte FABP where the fluorescence emission move to 536, 531 and 530 nm, respectively (23). The liver FABP and the intestinal FABP produce a shift to 500 and 496 nm, respectively (23), values closer to what is observed for *S. japonicum* F10 FABP (24). At large, the observed variability in the emission wavelength indicates that in each FABP there are local contacts with specific residues and this might justify their difference in affinity. Interestingly, while the stoichiometry of the binding of DAUDA to the three rSm14 variants shows a molar ratio of 1:1, similar to most of the FABPs (12), they exhibit quite similar Kd values (0.63, 0.82 and 0.66 for rSm14-M20, rSm14-T20 and rSm14-A20, respectively)(Figure 6). In order to highlight possible differences in the binding affinities of the three Sm14 proteins for other fatty acids, we performed competitive experiments using DAUDA as tracer and some natural fatty acids as competitors. The addition of myristic, palmitic, oleic and linoleic acid

indeed revealed a diverse affinity of the three mutants for these substrates represented by the different extent in the stoichiometric reversal of the fluorescence enhancement of DAUDA. Figure 7 shows a typical competitive curve where linoleic acid is used to displace DAUDA from rSm14-M20. The efficiency of each fatty acid to displace DAUDA from the rSm14 binding pocket is shown in Figure 8. A value of 100% indicates the complete displacement of DAUDA. The data show that the rSm14-M20 has the higher affinity for all the natural fatty acids. Overall, the results indicate that the efficiency in the displacement of DAUDA increases with the increase of both the fatty acid length and the number of insaturations (12) and, as observed for other FABPs, confirm the importance of the nature of the amino acid residue at position 20 in Sm14. It is interesting to observe that in the case of rSm14-A20, the myristic acid is not able to displace DAUDA evidencing, differently from the other rSm14 proteins, an absence of affinity of the mutant for this relatively short fatty acid. A possible key to interpret the similarities and differences in affinity, exhibited by the three Sm14 proteins with respect to DAUDA and to the various fatty acids, can be derived by the analysis of the crystallographic structure of human B-FABP that reveals that only long chain fatty acids are able to interact with Met20 (25). Based on this observation, we can hypothesize that due to its short chain, DAUDA (only 11 carbon atoms) will not be able to interact with any residue might be present in position 20, thus justifying the equivalent affinity exhibited by the three proteins for the fluorescent fatty acid probe (Figure 6). On the other hand, the competition experiments (Figure 8), where longer fatty acids have been used, revealed the existence of distinct affinities that very likely are due to selective contacts between the fatty acid and the distinct residue at position 20 of the rSm14 proteins.

Altogether, these results indicate that the nature of the amino acid at position 20 is very important for structure stability and for the binding features to fatty acids.

Vaccination experiments - The protection efficacy of the three rSm14 proteins against *S. mansoni* cercariae was evaluated in outbred Swiss mice as previously described (4). As shown in Figure 9, rSm14-T20 or rSm14-M20 proteins stimulated a significant protective response (44 and 67%, respectively). Animals vaccinated with these proteins showed a reduction in mean worm burden with values close to or higher than the 40% reduction value defined by WHO as a pre-requisite for an antigen to be considered a potential vaccine candidate against human schistosomiasis. Exception is the rSm14-A20 mutant that turned out to be a poor protective antigen, reaching protection values around 20% of worm burden reduction. Interestingly, however, in the Sm14 the amino acid at position 20, as shown by molecular modeling (2 and results not shown), is buried within the interior of the molecule like in the other FAPBs (2, 12, 23 and 25) and therefore, is not expected to be exposed as an epitope. Nonetheless, the data described here demonstrate the importance of this residue in modulating some structural and functional features of Sm14. Thus, the lack of significant protective ability for rSm14-A20 is probably the result of these properties rather than the change of the epitope exposure. The rSm14-A20 was the more unstable and less structured Sm14 protein (data not shown, Figure 5). The rSm14-M20 showed the highest secondary structure recovery (82.5%) after heating to 80^0C, followed by rSm14-T20 (64.3%) (Figure 5A) and was more resistant to denaturation by urea as followed by intrinsic fluorescence of tryptophane (results not shown). These properties correlate well with the level of protective ability observed (Figure 9). In fact, heating the rSm14-T20 just before the immunization of the animals resulted in a poor protective response, similar to that observed for rSm14-A20 mutant, thus suggesting that for immune protection, Sm14 has to possess the correct fold. This is an important observation for the development of a Sm14 based vaccine (3) and it implies that a quality control that take into account the structure integrity of the protein has to be included in the production of rSm14.

In conclusion, the data presented here show for the first time the genomic organization of the Sm14 protein and the existence of polymorphism. A preliminary evaluation of the structural and functional features of the some of these mutants has revealed that the Sm14-M20 isoform, characterized by a higher structural stability and by a more pronounced affinity for natural fatty acids, appears to be the more suitable antigen for the development of the schistosomiasis vaccine. This observation shall contribute for the Sm14 based vaccine scale up and development. In particular, we believe that the enhanced stability of the protein may be a key property since it can provide a better immune response due to the presence of the proper fold, an increase in the lifetime of the vaccine and, as a consequence, a decrease in the costs for vaccine storage, considered an important issue for the developing countries where the vaccine is targeted.

References

1. Despres, L., Imbert-Establet, D. and Monnerot, M. (1993) *Mol. Biochem. Parasitol.* **60**, 221-229.

2. Esteves, A., Joseph, L., Paulino, M. and Ehrlich R. (1997) *Int. J. Parasitol.* **27**,1013-1023.

3. Bergquist, N.R. and Colley, D.G. (1998) *Parasitology Today* **14**, 99–104.

4. Tendler, M., Brito, C.A., Vilar, M.M., Serra-Freire, N., Diogo, C.M., Almeida, M.S., Delbem, A.C., Da Silva, J.F., Savino, W., Garratt, R.C., Katz, N. and Simpson, A.S. (1996) *Proc Natl Acad Sci U S A*. **93**, 269-273.

5. Brito, C.F., Caldas, I.R., Coura Filho, P., Correa-Oliveira, R. and Oliveira, S.C. (2000) *Scand. J. Immunol.* **51**, 595-601.

6. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: a laboratory manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

7. Krajewski, J.L., Dickerson, I.M.and Potter, L.T. (2001) *Mol. Pharmacol.* **60**, 725-731.

8. Junqueira de Azevedo, I.L.M., Farsky, S.H.P., Oliveira, M.L.S. and Ho, P.L. (2001) *J. Biol. Chem.* **276**, 39836-39842.

9. Tendler, M., Pinto, R.M., Oliveira Lima, A., Gebara, G. and Katz, N. (1986) *Int. J. Parasitol.* **16**, 347-352.

10. Breathnach, R. and Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349-383.

11. Moser, D., Tendler, M., Griffiths, G. and Klinkert, M.Q. (1991) *J Biol. Chem.* **266**, 8447-8454.

12. Bernlohr, D.A., Simpson, M.A., Hertzel, A.V. and Banaszak, L.J. (1997) *Annu. Rev. Nutr.* **17**, 277-303.

13. Scott, J.C., Kennedy, M.W. and McManus, D.P. (2000) *Biochim. Biophys. Acta* **1517**, 53-62.

14. Matarese, V., Stone, R.L., Waggoner, D.W. and Bernlohr, D.A. (1989) *Prog. Lipid Res.* **28**, 245-272.

15. Wu, Q., Andolfatto, P. and Haunerland, N.H. (2001) *Insect Biochem. Mol. Biol.* **31**, 553-562.

16. Plenefisch, J., Xiao, H., Mei, B., Geng, J., Komuniecki, P.R. and Komuniecki, R. (2000) *Mol. Biochem. Parasitol.* **105**, 223-236.

17. Richieri, G.V., Low, P.J., Ogata, R.T. and Kleinfeld, A.M. (1998) *J. Biol. Chem.* **273,** 7397-7405.

18. Becker, M.M., Kalinna, B.H., Waine, G.J. and McManus, D.P. (1994) *Gene* **148**, 321-325.

19. Bordo, D. and Argos, P. (1991) *J Mol Biol.* **217**, 721-729.

20. Dalessio, P.M. and Ropson, I.J. (2000) *Biochemistry* **39**, 860-871

21. Kennedy, M.W., Brass, A., McCruden, A.B., Price, N.C., Kelly, S.M., and Cooper, A., (1995) *Biochemistry* **34**, 6700-6710.

22. Macgregor, R.B. and Weber, G. (1986*) Nature* **319**, 70-73.

23. Zimmerman, A.W., van Moerkerk, H.T. and Veerkamp, J.H. (2001) *Int. J. Biochem. Cell Biol.* **33**, 865-876.

24. Kennedy, M.W., Scott, J.C., Lo, S., Beauchamp and J., McManus, D.P. (2000) *Biochem. J.* **349**, 377-384.

25. Balendiran, G.K., Schnutgen, F., Scapin, G., Borchers, T., Xhong, N., Lim, K., Godbout, R., Spener, F. and Sacchettini, J.C. (2000) *J. Biol. Chem.* **275**, 27045-54.

Footnotes:

For Page 1:

1. This work was supported by grants from the Brazilian agencies Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação Butantan. C. R. Romero Ramos had a CAPES fellowship for this study. We would like to thank Dr. Sérgio Verjovsky-Almeida for providing us the *S. mansoni* cDNA library and Dr. Toshie Kawano for the adult worms of the *S. mansoni* BH strain.

Nucleotide sequence data reported in this paper are available in the GenBank™ database under accession numbers AY055467, AF492389 and AF492390.

2. Abbreviations used are: CD, circular dichroism; DAUDA, 11-(dansylamino) undecanoic acid; FABP, fatty acid binding protein; A-FABP, B-FABP, I-FABP, H-FABP, adipocyte-, brain-, intestinal- and heart- forms of FABPs, respectively; IPTG, isopropylthio-β-galactoside; ORF, open reading frame; PCR, polymerase chain reaction; RT-PCR, reverse transcribed-PCR; SDS-PAGE, sodium dodecyl sulfate containing polyacrylamide gel electrophoresis; Sj-FABPs, *S. japonicum* FABPs.

Figure Legends

Figure 1. Sequence of the Sm14 gene and alignment of the deduced protein with the *S. japonicum* F25 and F10 FABPs. (A) Sequence of the Sm14 gene. The complete sequence is in the GenBank database under accession number AY055467. The canonical GT/AG sequences for intron/exon boundaries are in bold. The internal EcoRI site is underlined and an arrow indicates the 3´alternative acceptor splice site for the intron 1. In this alternative splicing, a smaller Sm14 protein of the size of the *S. japonicum* F25, can be generated. (B) Amino acid sequence alignment of FABPs encoded by the *S. mansoni* Sm14 gene and by the *S. japonicum* FABP cDNA clones F10 and F25. The alignment shows the position of introns 1, 2 and 3. In the case of *S. japonicum* F25, the gap (-) represents the observed amino acid deletion.

Figure 2. Intron positioning in the FABP gene family. The nucleotide and amino acid sequences of a number of FABPs have been aligned to obtain the maximal positional codon identity. The position of the three introns falls at exactly the same points in each gene. * - The numbering of LBP-3 corresponds to the putative mature protein and the reported introns corresponding to introns 3, 4 and 5 of lbp-3 gene.

Figure 3. Multiple sequence alignments of the *S. mansoni* Sm14 isoforms. Numbering is according to the Sm14 sequence. **(-)** indicates identical amino acid, dots have been introduced to allow the best sequence alignment and to represent amino acid gap. M = Sm14-M20; T = Sm14-T20. The number of sequenced clones of each isoform is in parenthesis. Prediction of the secondary structure features are also shown below the sequences. The alignment includes the sequences of three *S. japonicum* FABPs (Sj-FABPs): Sj-FABP, Sj-F10 and Sj-F25 (13).

Figure 4. Schematic representation of the Sm14 genomic organization and of its splicing forms. The asterix (*) denotes the possible alternative splicing site that, in *S. mansoni,* would be responsible for the expression of a reduced form of Sm14 similar to Sj-F25.

Figure 5. CD evaluation of the thermal stability of the rSm14 proteins. (A) CD spectra of the three Sm14 proteins measured at 15 ^{0}C (black line), 80 ^{0}C (dark grey line) and after cooling down the protein to 15^{0}C (light grey line). Secondary structure recovery after heating the rSm14 proteins to 80 ^{0}C and cooling down to 15 ^{0}C was calculated for each protein considering the change of the wavelength value when [θ] = 0. The corresponding wavelength value considered for random structure was 191.5 nm. (B) Variation of the molar ellipticity [θ] measured at 216 nm throughout the temperature range 15-75°C for the three recobinant Sm14 variants. Tm values for each rSm14 proteins were calculated as the inflexion point of each curve according to Boltzman´s function using the Origin Program.

Figure 6. Titration curves of DAUDA binding to rSm14 proteins. Increasing amounts of DAUDA were added to a 2μM solution of the protein in PBS. The enhancement of DAUDA fluorescence upon binding to the rSm14 variants was obtained subtracting the fluorescence of the same concentration of DAUDA in the absence of the protein. All data were analyzed by a non-linear regression model according to Hill´s function using the Origin Program.

Figure 7. Displacement of DAUDA by linoleic acid in the binding to rSm14-M20. Competitive binding experiments were carried out using 2μM of DAUDA in the presence of 2μM of linoleic acid and 2μM of rSm14-M20.

Figure 8. Displacement of DAUDA binding to rSm14 proteins by natural fatty acids. The percentage of displacement is indicated above each bar. The same conditions as described for Figure 6 were used here, for each rSm14 protein (2μM) versus each fatty acid (2μM) in the presence of DAUDA (2μM).

Figure 9. Protection of vaccinated animals against *S. mansoni*. Protection was measured as the mean of worm burden in each group of animals and was expressed in percentage of protection as explained in Material and Methods. Animals were immunized with only adjuvante (Adj) or adjuvante plus rSm14-T20 (T), or plus rSm14-M20 (M), or plus rSm14-A20. One group of animals were immunized with adjuvant plus rSm14-T20 previously heated to 90^0C for 5 minutes just before the innoculation (^T). *P<0.05 when compared to the group of animals immunized with only adjuvant (Adj) and **P<0.05 when compared to the group of animals immunized with non-heated rSM14-T20.

Table I. Sizes of the introns and of the amino acids in the exons of several FABPs.

(*) The intron sizes of the *S. japonicum* FABP was estimated by PCR (13).

Table I

Gene	Amino acid sequence length/Exon				Intron length (bp)		
	1	2	3	4	1	2	3
Sm14	23	58	35	17	674	585	42
FABP-Sj (*)	23	58	35	16	1200	850	70
Murine KLBP	26	58	34	17	2300	300	500
Rat L-FABP	22	58	31	16	1450	1220	610
Human I-FABP	22	58	36	16	1200	1020	440
Mouse mP2	24	58	34	16	1200	200	1300
Mouse CRBPII	24	60	34	16	1200	1800	400

Figure 1

A

```
ATGTCTAGTTTCTTGGGAAAGTGGAAACTTAGCGAGTCACACAACTTCGAT   51
 M  S  S  F  L  G  K  W  K  L  S  E  S  H  N  F  D    17

GCTGTCATGTCAAAGCTAGgtaactacacactgttaaaatttatagttgta  102
 A  V  M  S  K  L                                     23

tacatcttcg    <---  600 bp of intron 1 --->   aggga  716

atttaatttatttatattttcctgtcagGTGTCTCATGGGCAACTCGACAG  767
                           G  V  S  W  A  T  R  Q     31

ATTGGGAACACAG↓TGACCCCAACTGTAACCTTCACAATGGATGGGGATAAA  818
 I  G  N  T  V  T  P  T  V  T  F  T  M  D  G  D  K    48

ATGACTATGTTAACAGAGTCAACTTTCAAAAATCTTTCTTGTACGTTCAAG  869
 M  T  M  L  T  E  S  T  F  K  N  L  S  C  T  F  K    65

TTCGGCGAGGAATTCGATGAAAAAACAAGTGACGGCAGAAATGTCAAGgta  920
 F  G  E  E  F  D  E  K  T  S  D  G  R  N  V  K       81

aataaact  <---  545 bp  of intron 2 --->  ctggctatt 1481

tgcttaattcccttgtagTCAGTTGTTGAAAAAAATTCCGAGTCGAAGTTA 1532
                  S  V  V  E  K  N  S  E  S  K  L     92

ACGCAAACTCAAGTAGATCCCAAAAACACAACTGTAATCGTTCGTGAAGTG 1583
 T  Q  T  Q  V  D  P  K  N  T  T  V  I  V  R  E  V   109

GATGGTGATACTATGAAAACGgtaacttatattttctcactcttgatatgt 1634
 D  G  D  T  M  K  T               intron 3          116

attccttttagACTGTGACTGTTGGTGACGTTACTGCCATTCGCAATTAT 1685
            T  V  T  V  G  D  V  T  A  I  R  N  Y   129

AAACGACTATCCTAA  1700
 K  R  L  S  *    133
```

B

```
Sm14 MSSFLGKWKLSESHNFDAVMSKL <-intron 1  -> GVSWATRQIGNTVTPTVTFTM
F25  MSSFLGKWKLSESHNFDAVMSKL <-intron 1? -> ------------VTPTVTFTM
F10  MSSFLGKWKLSETHNFDAVMSKL <-intron 1  -> GVSWPIRQMGNTVTPTVTFTM
◦
Sm14 DGDKMTMLTESTFKNLSCTFKFGEEFDEKTSDGRNVK  <-intron 2->   SVVEKNS
F25  DGDTMTMLTESTFKNLSVTFKFGEEFDEKTSDGRSVK  <-intron 2->   SVVTKDS
F10  DGDTMTMLTESTFKNLSVTFKFGEEFDEKTSDGRNVK  <-intron 2->   SVVTKDS
◦
Sm14 ESKLTQTQVDPKNTTVIVREVDGDTMKT <-intron 3->TVTVGDVTAIRNYKRLS 133aa
F25  ESKITQTQKDSKNTTVIVREIVGDTMKT <-intron 3->TVTVDDVTAIRNYKRL  120aa
F10  ESKITQTQKDAKNTTVIVREIVGDTMKT <-intron 3->TVTVDDVTAIRNYKRL  132aa
```

Figure 2

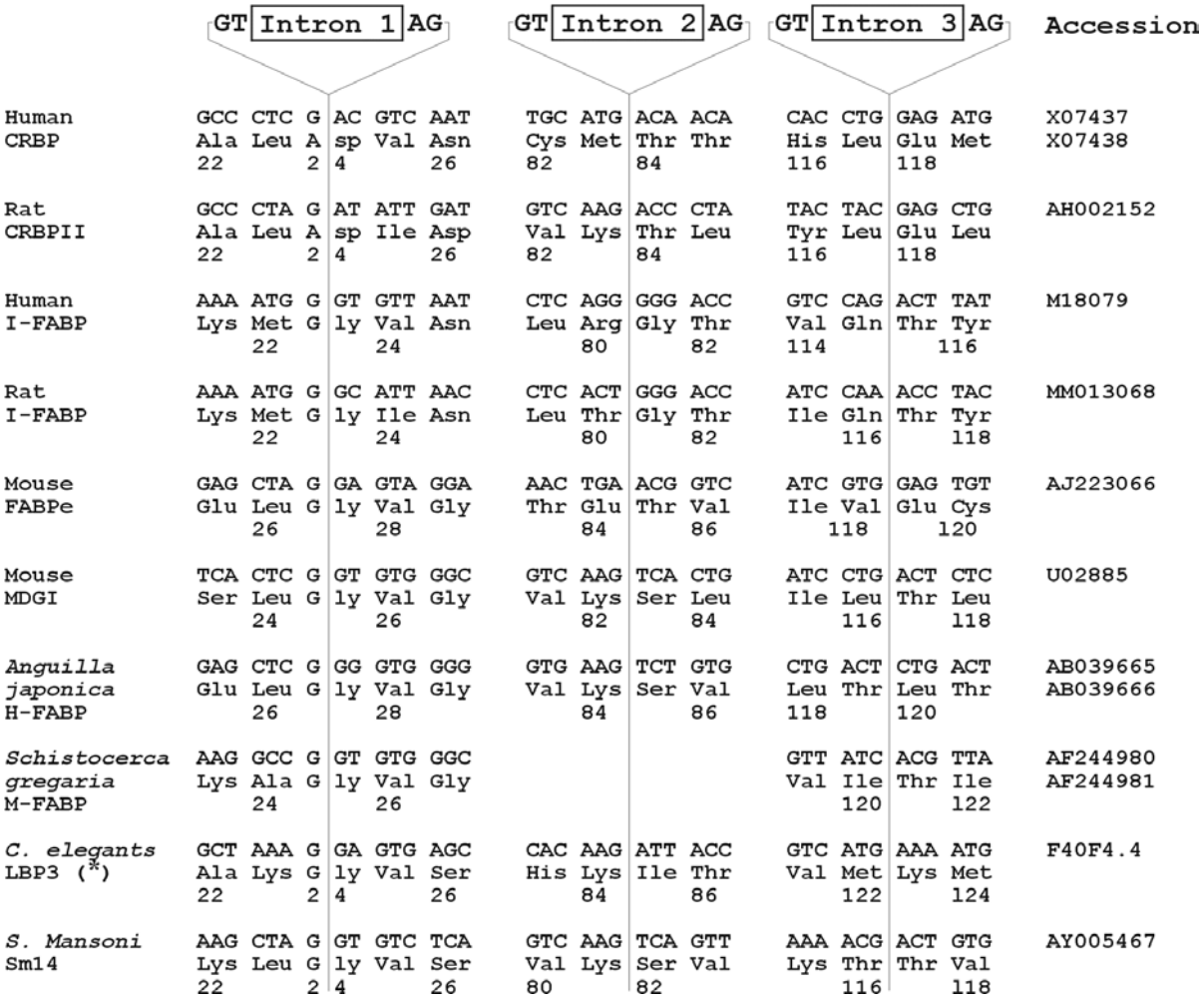


Figure 3

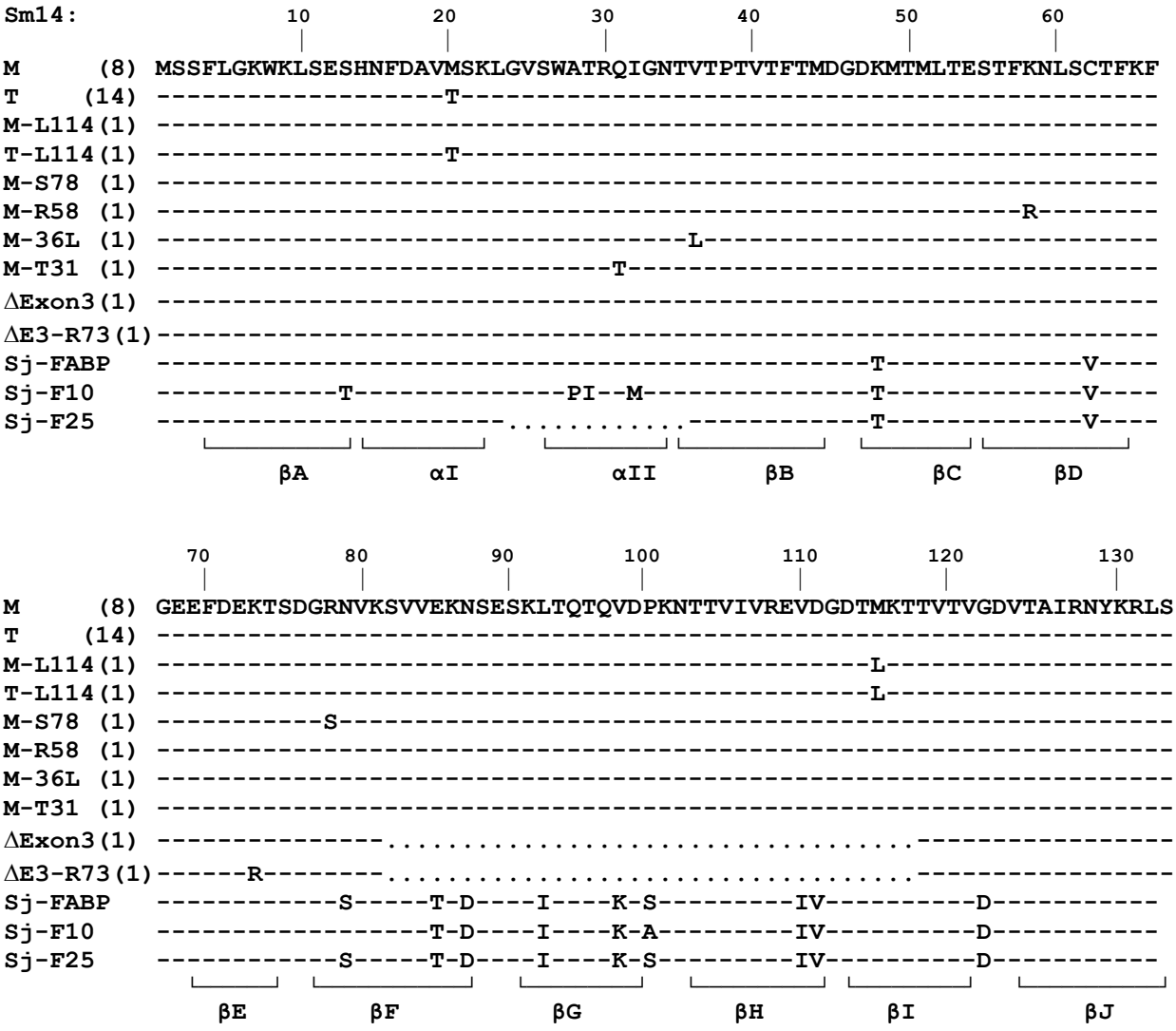


Figure 4

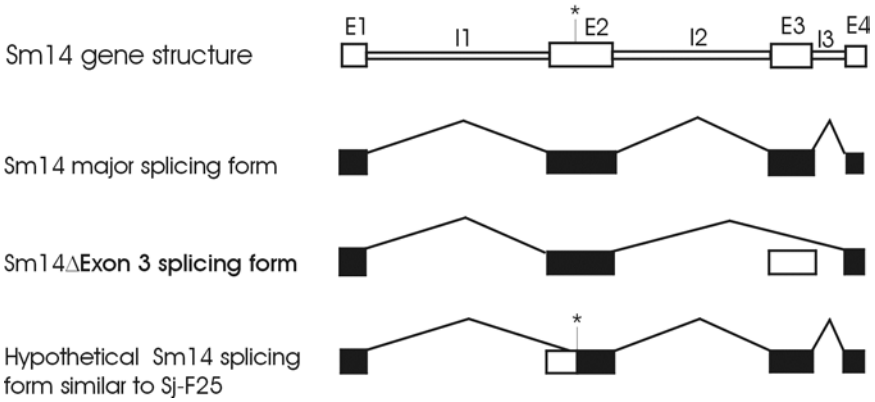


Figure 5

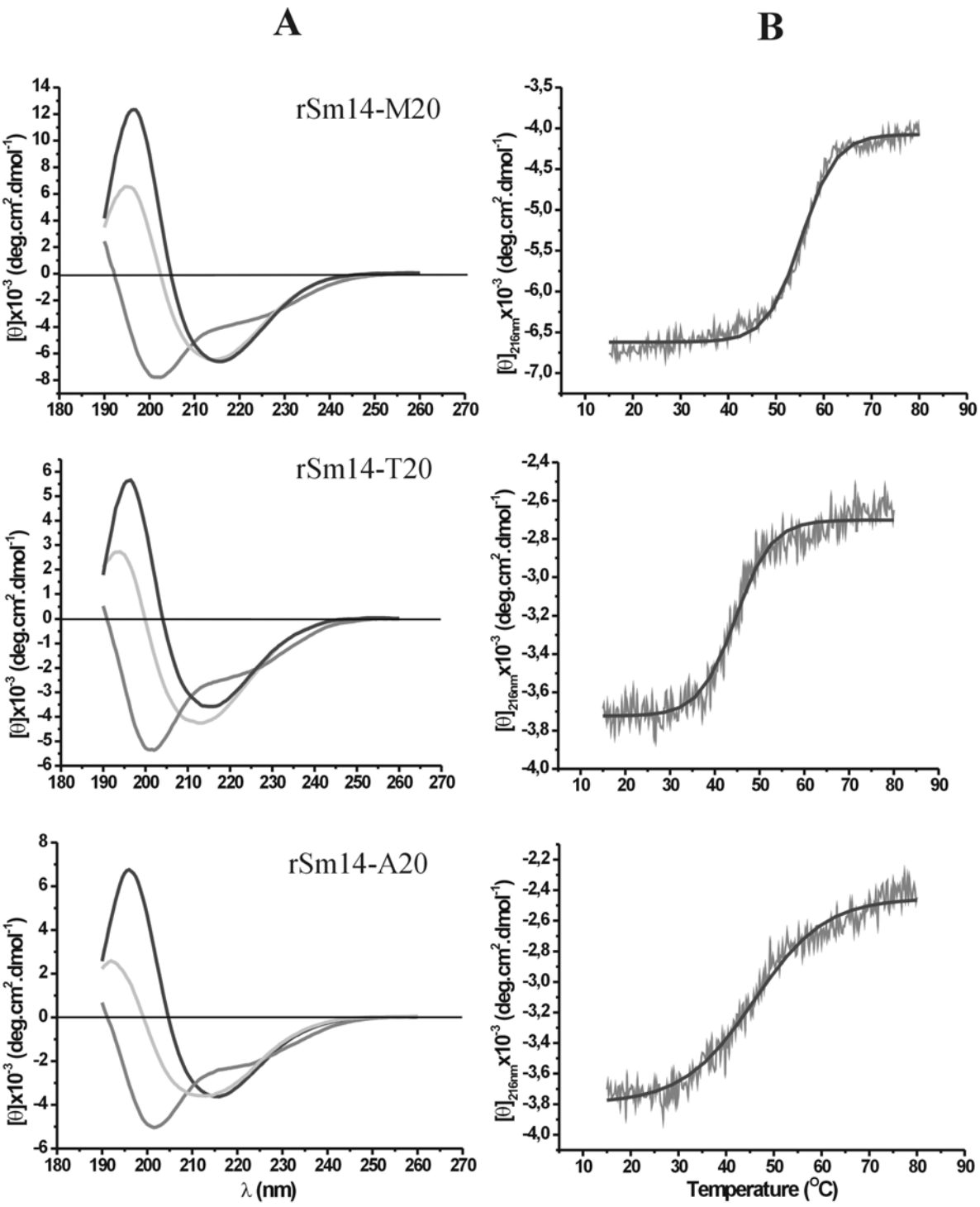


Figure 6

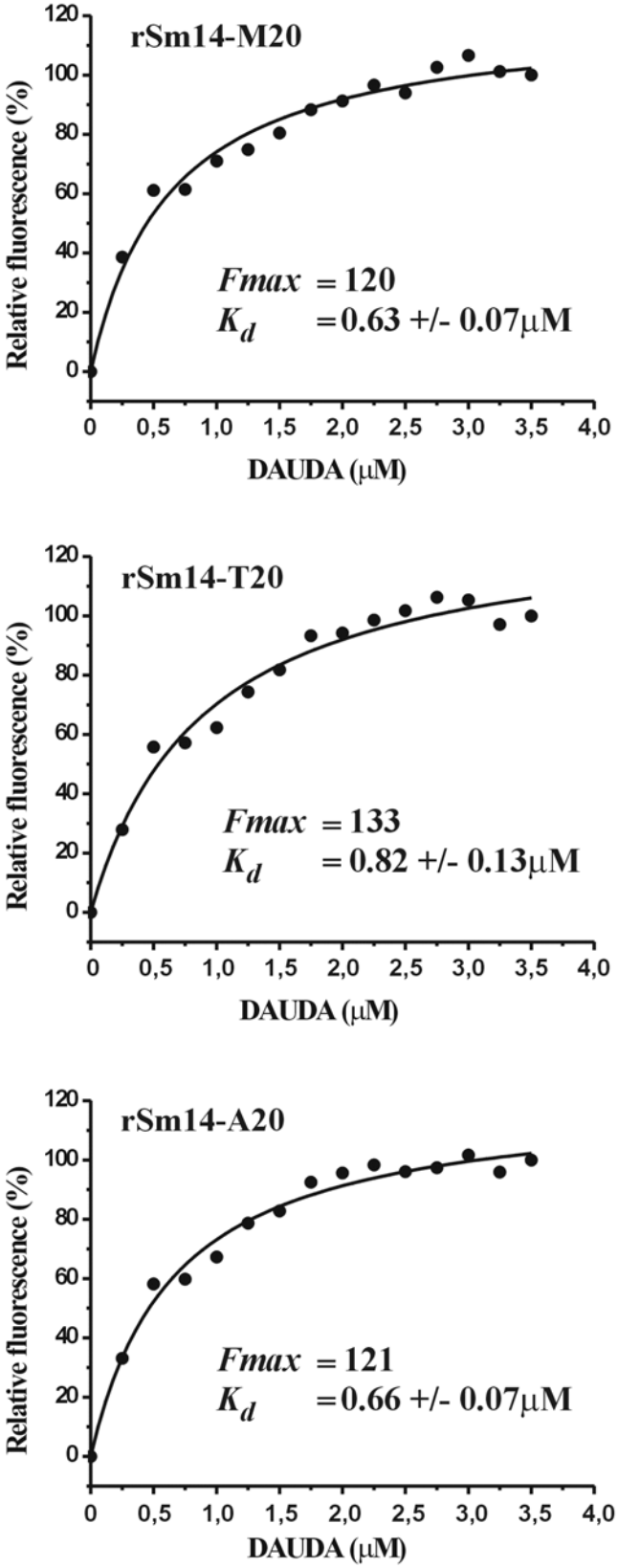


Figure 7

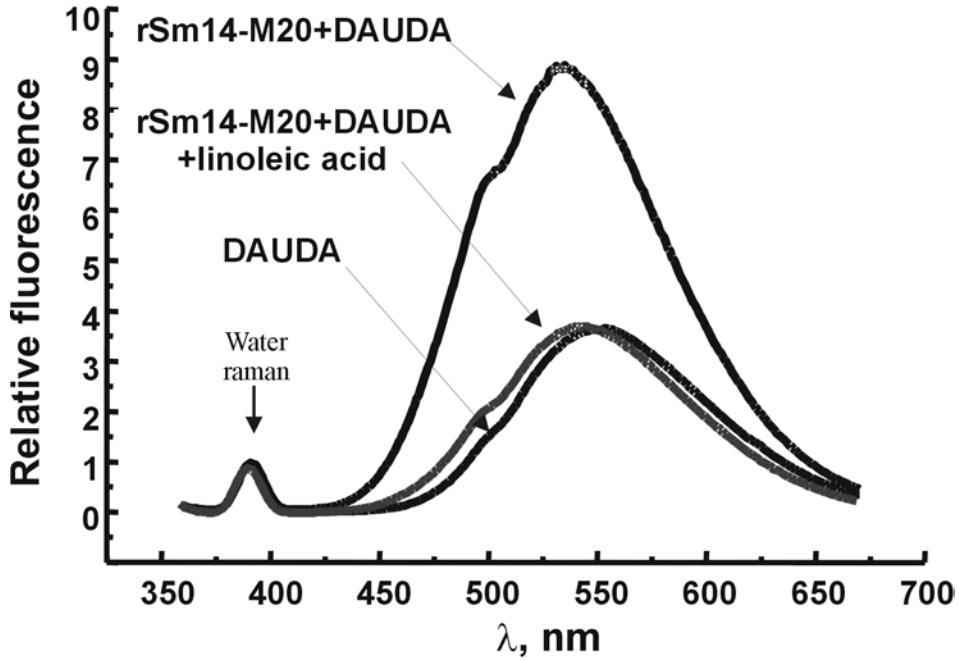


Figure 8

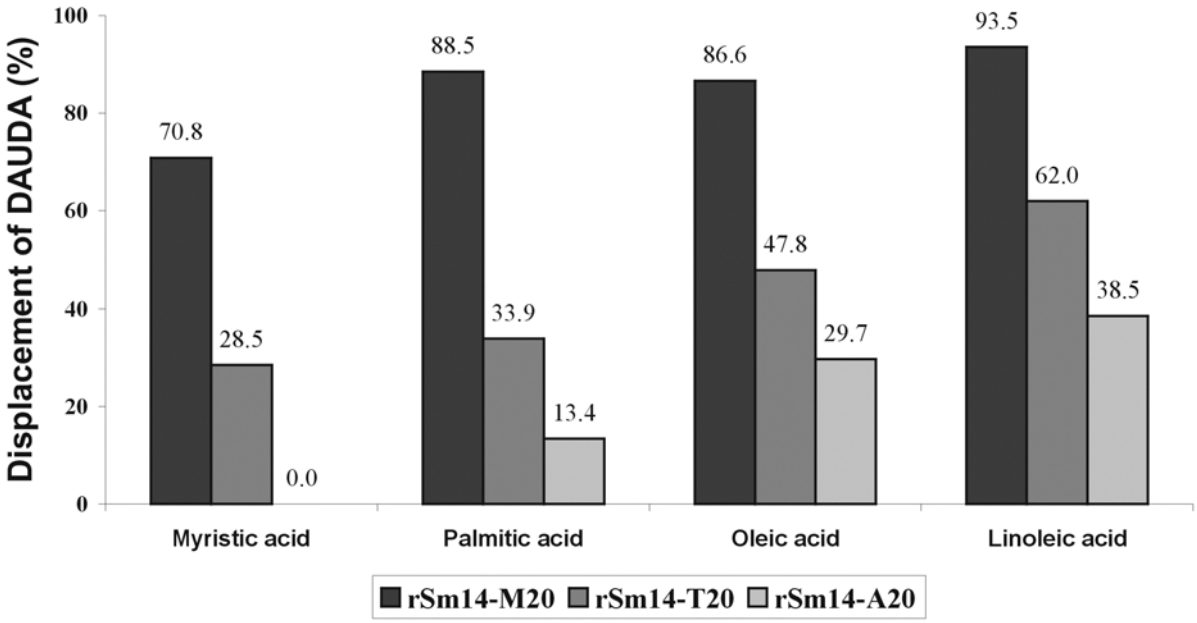
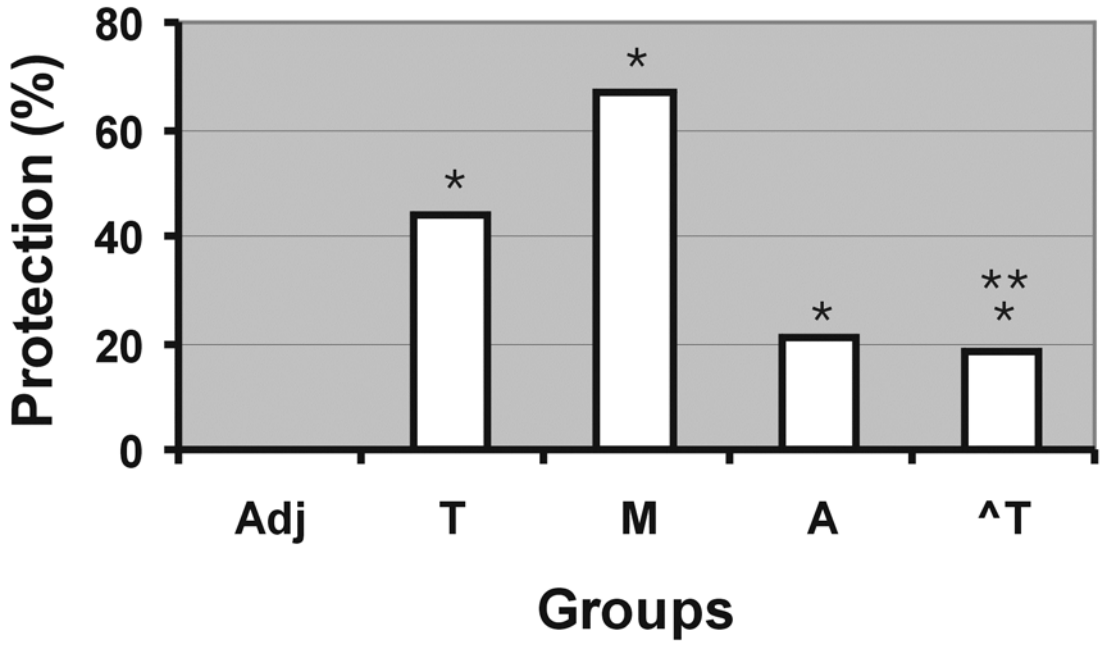


Figure 9



Gene structure and M20T polymorfism of the Schistosoma mansoni Sm14 fatty acid binding protein: Molecular, functional and immunoprotection analysis

Celso Raul Romero Ramos, Rita de Cássia Rossi Figueredo, Thelma de Aguiar Pertinhez, Monica Magno Vilar, Ana Lucia Tabet Oller do Nascimento, Miriam Tendler, Isaias Raw, Alberto Spisni and Paulo Lee Ho

J. Biol. Chem. published online January 27, 2003
