

# **TESE DE DOUTORADO**

**UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS BIOLÓGICAS  
DEPARTAMENTO DE BIOQUÍMICA E IMUNOLOGIA**

## ***Schistosoma mansoni*: DESCOBERTA DE NOVOS GENES E ESTUDOS DE GENÔMICA FUNCIONAL DE UMA RHO GTPase**

**Orientado: Túlio Marcos Santos**

**Orientador: Prof. Dr. Sérgio Danilo Junho Pena**

**Co-orientadora: Profa. Dra Glória Regina Franco**

**Colaborador Principal: Prof. Dr. Carlos Renato Machado**

**Belo Horizonte – MG**

**Setembro, 2000**

**TÚLIO MARCOS SANTOS**

***Schistosoma mansoni*: DESCOBERTA DE NOVOS GENES E  
ESTUDOS DE GENÔMICA FUNCIONAL DE UMA Rho GTPase**

Tese apresentada ao Departamento de Bioquímica e  
Imunologia do Instituto de Ciências Biológicas da  
Universidade Federal de Minas Gerais como  
requisito parcial para a obtenção do grau de Doutor  
em Ciências.

**Orientador: Prof. Dr. Sérgio Danilo Junho Pena**

**Co-orientadora: Profa. Dra Glória Regina Franco**

**Colaborador Principal: Prof. Dr. Carlos Renato Machado**

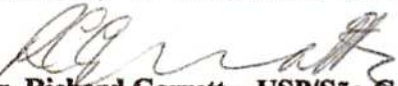
**Belo Horizonte – MG**

**Setembro, 2000**



Instituto de Ciências Biológicas-UFMG  
Curso de Pós-Graduação em Bioquímica e Imunologia

**ATA DA DEFESA DA TESE DE DOUTORADO DE TÚLIO MARCOS SANTOS.** Aos vinte e sete dias do mês de setembro de 2000 às 13:00 horas, reuniu-se no Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais a Comissão Examinadora da tese de doutorado, indicada pelo Colegiado do Curso em 22 de fevereiro de 2000, para julgar, em exame final, o trabalho intitulado "*Schistosoma mansoni*: Descoberta de Novos Genes e Estudos de Genômica Funcional de uma Rho GTPase", requisito final para a obtenção do grau de Doutor em Ciências: Bioquímica. Abrindo a sessão o Presidente da Comissão Prof. Dr. Sérgio Danilo Junho Pena da Universidade Federal de Minas Gerais, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores, com a respectiva defesa do candidato. Logo após a Comissão se reuniu, sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foram atribuídas as seguintes indicações: - Dr. Richard Garrett da Universidade de São Carlos-USP, aprovado; - Dr. Guilherme Corrêa Oliveira do Centro de Pesquisas René Rachou-Fiocruz, aprovado; - Dr. Paulo Marcos Zech Coelho da Universidade Federal de Minas Gerais, aprovado; - Dr. Alfredo Miranda de Góes da Universidade Federal de Minas Gerais, aprovado; - Dr. Sérgio Danilo Junho Pena da Universidade Federal de Minas Gerais, orientador, aprovado. Pelas indicações o candidato foi considerado **APROVADO**. O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar o Presidente da Comissão encerrou a reunião e lavrou a presente Ata que será assinada por todos os membros participantes da Comissão Examinadora. Belo Horizonte, 27 de setembro de 2000.


  
Dr. Richard Garrett - USP/São Carlos

  
Dr. Guilherme Corrêa Oliveira - CpRR-Fiocruz

  
Dr. Paulo Marcos Zech Coelho - UFMG

  
Dr. Alfredo Miranda de Góes - UFMG

  
Dr. Sérgio Danilo Junho Pena - Orientador-UFMG

  
Prof. Carlos Alberto Petalita Cavacas  
Coordenador do Curso de Pós-Graduação  
em Bioquímica e Imunologia ICB-UFMG

## AGRADECIMENTOS

*Fique advertido que se você desejar, como eu o desejo, construir uma sociedade na qual os indivíduos cooperem generosa e desinteressadamente para um bem comum, você poderá esperar pouca ajuda da natureza biológica. Tentemos ensinar generosidade e altruísmo, porque nascemos egoístas. Compreendamos o que nossos próprios genes egoístas tramam, porque assim, pelo menos, poderemos ter a chance de frustrar seus intentos, uma coisa que nenhuma outra espécie jamais aspirou fazer.*

Richard Dawkins em "O gene Egoísta"

Ao Prof. Sérgio Danilo Junho Pena, por quem tenho profundo respeito e admiração como o cientista que eu gostaria de ser, pela oportunidade ímpar de trabalhar no seu conceituado Laboratório de Genética Bioquímica (LGB) e pela precisa orientação, apoio, estímulo, ricos ensinamentos e paciência incomum que teve comigo;

À Profa. Glória Regina Franco, cientista e docente entusiasmada, a quem admiro e respeito muito pela sua destacada capacidade e dedicação incansável devotada à ciência e ao ensino, pela paciência de "mãe de filho rebelde" que teve ao me orientar e pela amizade que temos;

Ao Prof. Carlos Renato Machado, pela inestimável contribuição dada à parte de genômica funcional deste trabalho, pela eficiente orientação, paciência e ricas discussões que sempre tivemos a respeito de nossas leveduras e igualmente pela amizade e boa companhia na degustação às sextas-feiras, no buteco da biologia, daquilo que *Saccharomyces cerevisiae* produz de melhor: a cervejinha gelada!

Aos nossos colaboradores:

Prof. Vasco Azevedo, do Depto. de Biologia Geral do ICB-UFMG, pelos primeiros e fundamentais ensinamentos em Biologia Molecular, dedicada orientação, apoio e estímulo constantes além da amizade, de bons e divertidos papos científicos ou não;

Dr. Mohammed Saber e sua equipe, do Theodore Bilharz Research Institute, Cairo, Egito e ao Dr. Guilherme Corrêa Oliveira, do Instituto de Pesquisas René Rachou da FIOCRUZ, Belo Horizonte - MG, pela gentileza de nos cederem as bibliotecas de cDNA de vários estágios do desenvolvimento de *S. mansoni* usadas por nós no programa de descoberta de novos genes;

Dr. David Johnston e sua equipe do The Natural History Museum, Londres, Inglaterra, pelas ESTs de cercária da biblioteca 2 e também por suas correções e discussões que ajudaram a enriquecer muitíssimo o nosso trabalho de análise das ESTs de cercária;

Profa. Élide Mara Leite Rabelo, do Depto. de Parasitologia do ICB-UFMG, pela inestimável ajuda na produção e análise das ESTs, colaboração além da divertida amizade e sempre bem humorada companhia;

Prof. José Miguel Ortega, do Depto. de Bioquímica e Imunologia do ICB, UFMG, pela grande ajuda que nos deu na implementação e análises das ESTs pelo programa ICATOOLS, pela generosidade em nos fornecer meios de cultura, vetores e de ceder incondicionalmente sua sala de cultura para que pudéssemos trabalhar com nossas leveduras, pelas opiniões e idéias que ajudaram nosso trabalho, além da amizade alto astral;

Prof. Richard Garratt, do Instituto de Física da USP de São Carlos - SP e ao seu estudante de doutorado Humberto Muniz Pereira, pela importante contribuição que deram na parte de modelagem molecular computacional comparativa da proteína Rho GTPase. Ao Humberto quero agradecer em especial pela amizade e gentileza de me receber em sua casa durante o período que estive na cidade de São Carlos;

Prof. Marcelo Valle de Souza, do Centro Brasileiro de Sequenciamento de Proteínas da Universidade de Brasília - DF, pela sua preciosa colaboração e gentileza com que sempre nos atendeu, além da vaga de pós-doc aberta para mim no seu laboratório dentro desse mesmo espírito colaborativo;

Aos Professores Andréa Macedo e Fabrício Santos pelo apoio na administração do LGB, pela indispensável contribuição na discussão dos trabalhos apresentados nas reuniões semanais do laboratório e amizade;

Às funcionárias do LGB, Neuza Antunes (Neuzinha), Kátia Barroso (Katita) e Míriam Costa (Miroca), pelo trabalho zeloso e dedicado que sempre desempenharam para manter o laboratório organizado e funcionando da melhor maneira possível, para conforto do nosso trabalho (N. C. A.), e igualmente pela generosa dose de paciência, bom humor e a amizade que temos;

Aos membros da banca, pela disponibilidade e valiosas contribuições;

Aos Professores Maria Helena Perez Garcia e Ênio Cardillo Vieira, chefes do Departamento de Bioquímica e Imunologia durante o período de realização dessa tese, pela disponibilidade, atenção e amizade;

Ao Professor Carlos Alberto Pereira Tavares, coordenador do Curso de Pós-Graduação em Bioquímica e Imunologia pelo apoio, estímulo e amizade;

À Celise, secretária do Curso de Pós-Graduação em Bioquímica e Imunologia, pela disponibilidade, atenção e amizade;

Aos professores do Departamento de Bioquímica e Imunologia pela cordialidade e amizade além do elevado espírito crítico e colaborativo que muito contribuiu para minha formação e amadurecimento, em especial aos professores Vânia Prado, Marcelo Santoro e Alfredo Miranda pela atenção, disponibilidade e amizade com que sempre me brindaram;

Aos funcionários do Departamento de Bioquímica e Imunologia pelo eficiente trabalho, atenção e amizade, em especial à Maria Aparecida (Peu), Estér Soares, Rita de Cássia Soares França, Kátia Moraes e ao Jamil Silvano pelo sempre dedicado e eficiente trabalho e, sobretudo pela generosa e divertida amizade;

Aos professores, alunos e funcionários do Instituto de Ciências Biológicas por toda ajuda que puderam me dar em especial ao Professor Carlos Rosa, do Depto. de Microbiologia, pelas valiosas dicas de cultivo e biologia celular de leveduras e ao Prof. Paulo Marcos Zech Coelho e Sr. Alberto do GIDE pelas infecções dos nossos camundongos para obtenção de vermes adultos;

Aos estudantes e amigos do Departamento de Bioquímica e Imunologia em especial à Rosângela, Shauma e Suzana pela amizade e pelo que puderam me ajudar;

Aos meus bons amigos Márcio Arndt (Soiό) e Pedro Édson (Pedroco), à minha querida amiga Janaína (Jana) e à Grácia Rosinha pela amizade sempre divertida e bem humorada, companheirismo e apoio em todas as ocasiões;

Aos bons amigos do LGB, Denise (Dedê), Heloísa Pena, Rosane, Analina, Claudinha, Magdinha, Riva, Luciana Andrade (Lu A), Luciana Haddad (Lu H), Andréia Vercesi, Débora, Aléxia, Juliana Pimenta (Ju Pi), Renatinho, Patrícia Souza (Pacu), Flávia Parra, Luís Augusto, Renata Freitas, Cláudia, Mariana, Fabricinho, Jorge, Simone, Débora (Aline), Patrícia Pessoa, aos calouros Charles, André, Carlos Gustavo e Carlos Eduardo e, de modo especial, à turma do gel da meia-noite: Rinaldo, Juliana (Jumps), Wendell, Eduardo e Victor Pardini pela amizade, espírito solidário, sempre bem-humorada companhia, divertidos bate-papos permeadas de discussões científicas, políticas e filosóficas;

*Dedico este trabalho aos meus pais,  
Yvone e José,  
que muitas vezes deixaram de viver suas vidas  
para que eu pudesse bem viver a minha.*



# SUMÁRIO

	Página
<b>Lista de Figuras.....</b>	VIII
<b>Lista de Abreviaturas.....</b>	IX
<b>Resumo.....</b>	XI
<b>Abstract.....</b>	XIV
<b>Considerações Iniciais.....</b>	XVII
<b>1. Introdução Geral.....</b>	1
1.1. A esquistossomose .....	2
1.2. O <i>Schistosoma mansoni</i> e seu ciclo biológico.....	2
1.3. Biologia reprodutiva de <i>S. mansoni</i> e o papel das GTPases de baixo peso molecular no desenvolvimento sexual da fêmea do parasita.....	8
1.4. Diagnóstico, tratamento, estratégias de controle da esquistossomose e o desenvolvimento de vacinas.....	11
1.5. Projetos Genoma.....	14
1.6. Estudos pós-genômicos.....	19
1.6.1. Transcriptoma.....	20
1.6.2. Proteoma.....	21
1.6.3. Organismos modelos como sistemas experimentais para estudos da função gênica.....	23
1.7. O Projeto Genoma de <i>Schistosoma mansoni</i> .....	26
1.8. Projetos Genoma de outros parasitas.....	35
1.9. Projeto Genoma de <i>Schistosoma mansoni</i> : Estudos pós-genômicos	36
<b>Parte I: Descoberta de novos genes de <i>S. mansoni</i>.....</b>	38
I.1. Justificativa.....	39

I.2. Objetivos.....	39
I.3. Artigo 1: Update of the gene discovery program in <i>Schistosoma mansoni</i> with the expressed sequence tag approach .....	41
I.4. Artigo 2: Evaluation of cDNA libraries from different developmental stages of <i>Schistosoma mansoni</i> for production of expressed sequence tags (ESTs).....	47
I.5. Artigo 3: Analysis of the gene expression profile of <i>Schistosoma mansoni</i> cercariae using the Expressed Sequence Tag approach.....	58
<b>Parte II: Estudos de genômica funcional de uma Rho GTPase de <i>Schistosoma mansoni</i>: Caracterização Molecular e Funcional de um Gene de <i>S. mansoni</i> Homólogo a uma Ras-Like GTP Binding Protein (Rho).....</b>	<b>78</b>
II.1. Justificativa.....	79
II.2. Objetivos.....	79
II.3. Artigo 4: Characterization and Complementation in Yeast of a Rho GTPase Gene from <i>Schistosoma mansoni</i> (em preparação).....	80
<b>2. Considerações Finais.....</b>	<b>108</b>
2.1. Experimentos finais para complementação do artigo 4.....	108
2.2. Expressão da proteína Rho GTPase de <i>S. mansoni</i> em <i>E. coli</i> .....	110
<b>3. Perspectivas.....</b>	<b>112</b>
<b>4. Conclusões.....</b>	<b>114</b>
<b>5. Referências bibliográficas.....</b>	<b>117</b>
<b>Apêndices.....</b>	<b>A1</b>
A.1. Aminoácidos e suas propriedades.....	A1
A.2. Codon usage de <i>Schistosoma mansoni</i> .....	A2
A.3. Mapa do vetor pYEDP 60-2.....	A3
A.4. Mapa do vetor pMAL.....	A4

A.5. Mapa do vetor pYEplac112.....	A5
------------------------------------	----

## LISTA DE FIGURAS

	Página
Figura 1. Casal de vermes adultos de <i>Schistosoma mansoni</i> .....	4
Figura 2. O ciclo de vida de <i>Schistosoma mansoni</i> .....	5
Figura 3. Esquema para geração de mutações sítio-dirigidas pela PCR usando megainiciadores.....	109
Figura 4. Eletroforese em gel de poliacrilamida (SDS-PAGE) de extratos protéicos de <i>E. coli</i> expressando a proteína SMRHO clonada no vetor pMALc2.....	111

## LISTA DE ABREVIATURAS

**μm** – micrômetro

**μl** – micro litro

**2D** – eletroforese bidimensional

**AceDB** – A *Caenorhabditis elegans* Data Base (banco de dados sobre *C. elegans*)

**ATP** – adenosina trifosfato

**BAC** – Bacterial Artificial Chromosome (cromossomo artificial de bactéria)

**cDNA** – sequência de DNA complementar

**CISS** - Chromosomal *in situ* Suppression Hybridization (supressão da hibridização cromossômica *in situ*)

**dbEST** – EST Data Base (banco de dados de ESTs)

**DNA** – ácido desoxiribonucleico

**ESTs** – Expressed Sequence Tags (Etiquetas de Sequências Transcritas)

**FISH** - Fluorescent *In Situ* Hybridization (hibridização *in situ* fluorescente)

**GAPDH** – gliceraldeído 3-fosfato desidrogenase

**HGP** – Human Genome Project (Projeto Genoma Humano)

**IPTG** – isopropil-β-D(-)-tiogalactopiranosídeo

**kb** – quilo base

**kDa** – quilo Dalton

**M** – Molar

**Mb** – megabase

**MBP** – Maltose Binding Protein (proteína ligadora de maltose)

**mRNA** – RNA mensageiro

**OMS** - Organização Mundial da Saúde

**ORF** – Open Reading Frame (janela aberta de leitura)

**pb** – par de base

**PCR** – Polymerase Chain Reaction (Reação em Cadeia da Polimerase)

**pmole** – picomole

**RNA** – ácido ribonucleico

**RT-PCR** – Reverse Transcript PCR (PCR de transcritos reversos)

**SchistoDB** – *Schistosoma* Data Base (Banco de Dados de *Schistosoma*)

**SDS-PAGE** – Sodium Dodecil Sulfate Poliacrylamide Gel Electrophoresis (eletroforese em gel de poliacrilamida contendo o detergente dodecil sulfato de sódio)

**TDR** - Special Program for Research and Training in Tropical Diseases (Programa Especial para Treinamento e Pesquisa em Doenças Tropicais da OMS)

**WHO** – World Health Organization (Organização Mundial da Saúde)

**YAC** – Yeast Artificial Chromosome (cromossomo artificial de levedura)

## RESUMO

SANTOS, T. M. ***Schistosoma mansoni*: Descoberta de novos genes e estudos de genômica funcional de uma Rho GTPase.** 2000. 131 p. Tese (Doutorado) - Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, 2000.

*Schistosoma mansoni* é um trematódeo digenético que causa a esquistossomose, doença parasitária que constitui um importante problema de saúde pública no Brasil e em muitos outros países do mundo. O Projeto Genoma de *S. mansoni* teve início em 1992 como uma iniciativa brasileira que visa à descoberta de novos genes do parasita para o desenvolvimento de novas drogas e vacinas. No início do programa, centenas de Etiquetas de Sequência Transcritas (ESTs) foram geradas a partir de uma biblioteca de cDNA de vermes adultos. Entretanto, *S. mansoni* tem um complexo ciclo de vida o que tornou necessário a expansão do programa de descoberta de novos genes para outros estágios de desenvolvimento do parasita. No presente trabalho, oito bibliotecas de diferentes estágios do desenvolvimento do parasita, todas construídas utilizando-se o sistema  $\lambda$ ZapII, foram excisadas em massa para a obtenção de moldes de cDNA de fita dupla para que fosse possível a geração de grandes números de ESTs a partir de cada uma delas. A qualidade dos transcritos das bibliotecas excisadas foi então avaliada. Ao menos 30 colônias de cada biblioteca foram selecionadas para avaliação do tamanho médio dos insertos por PCR. A maior parte deles tinha um tamanho médio maior do que 500 pb. Na maior parte das bibliotecas menos 20% dos clones eram de sequências não informativas e mais de 50% deles eram de novos genes. Quando se considera apenas os genes presentes em mais de uma biblioteca, um total de 466 genes únicos foi obtido, o que corresponde a 427 genes novos de *S. mansoni*. Dos genes únicos, 20,2% foram identificados com base na sua homologia com genes de outros organismos, 8,3% eram genes já conhecidos do parasita e 71,5% do total era de genes desconhecidos. Após

estas avaliações iniciais, nossos esforços foram direcionados para a cercária, a forma larvária do parasita que infecta o homem e outros vertebrados. Duas bibliotecas de cDNA de cercárias de *S. mansoni* foram examinadas e sequências parciais foram obtidas de 957 clones selecionados ao acaso. Baseado em buscas por homologia, 551 (57,6%) das ESTs geradas não tinham homólogos nas bases de dados públicas enquanto que 308 (32,2%) foram identificadas totalizando 859 ESTs. As 98 restantes (10,2%) eram ESTs não informativas (RNA ribossomal e sequências mitocondriais não codificantes). Através de análises de clusters, fomos capazes de identificar 453 genes diferentes. Cento e dezenove genes identificados foram agrupados em 11 categorias funcionais onde genes associados com o metabolismo de energia foram os mais abundantes (13%) e diversos. A diversidade e a abundância de genes associados com a maquinaria de transcrição e tradução e com funções regulatórias e de sinalização celular também foram marcantes. Um transcrito de paramiosina foi identificado indicando que este gene não é expresso exclusivamente em vermes adultos e esporocistos (como tem sido sugerido na literatura). Dentre os novos genes de *S. mansoni* descobertos, identificamos um gene que codifica uma Rho-type GTP-binding protein o qual foi sequenciado inteiramente neste trabalho. A sua janela aberta de leitura é de 579 pb e codifica um polipeptídeo de 193 aminoácidos cujo peso molecular estimado é de 21,8 kDa. A sequência de aminoácidos deduzida apresenta alta homologia com Rho GTPases de várias espécies e possui todos os domínios conservados característicos de Rho GTPases. A superexpressão do gene de Rho GTPase de *S. mansoni* (*SMRHO*) complementa uma cepa mutante da levedura *Saccharomyces cerevisiae* que teve seu gene *rho1* selvagem eliminado artificialmente por nocaute gênico. A complementação ocorre mesmo em condições de alta temperatura e de concentração de  $Ca^{2+}$  que são restritivas para o crescimento da levedura. Entretanto, algumas alterações fenotípicas, tais como parada do ciclo celular, aumento do tamanho da célula e brotamento ao acaso, foram observadas indicando que a Rho GTPase de *S. mansoni* não é capaz de complementar perfeitamente todas as funções

executadas pela *RHO1* GTPase selvagem de *S. cerevisiae*. Os resultados de complementação junto com alinhamentos de sequências tornaram evidente que uma maior conservação de resíduos de aminoácidos específicos na região da hélice  $\alpha 3$  e loop 7 da Rho GTPase de *S. mansoni* podem explicar a complementação observada mesmo nas condições de alta temperatura e de concentração de  $\text{Ca}^{2+}$ . Os resultados obtidos durante este trabalho demonstraram que ESTs constituem uma abordagem versátil e eficiente para a descoberta de novos genes de *S. mansoni* nos seus diferentes estágios do ciclo de vida e que a levedura *S. cerevisiae* pode ser uma ferramenta poderosa em investigações sobre as funções dos genes deste complexo parasita.

**Palavras chaves:** Projeto genoma, descoberta de genes, ESTs, biblioteca de cDNA, expressão gênica, *Schistosoma mansoni*, verme adulto, cercária, Rho GTPases, levedura, *Saccharomyces cerevisiae*, nocaute gênico, complementação funcional.



## ABSTRACT

SANTOS, T. M. *Schistosoma mansoni*: Gene discovery and functional genomic studies of a Rho GTPase. 2000. 131 p. Thesis (PhD) - Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, 2000.

*Schistosoma mansoni* is a digenetic trematode that causes schistosomiasis, a parasitic disease that constitutes an important public health problem in Brazil and in many other countries throughout the world. The *S. mansoni* genome project was started in 1992 as a Brazilian initiative to discover new genes of this parasite for new drugs and vaccine development. At the debut of the program, hundreds of Expressed Sequence Tags (EST) were generated from an adult worm cDNA library. However, *S. mansoni* has a complex life cycle, which made necessary to expand the gene discovery program to other developmental stages of the parasite. In the present work, eight libraries from distinct developmental stages, all constructed using the  $\lambda$ ZapII system, were excised "en masse" to obtain double strand cDNA templates and generate large numbers of ESTs from each one of them. The quality of the transcripts from the excised libraries was then evaluated. At least 30 colonies from each library were selected to evaluate the average size of the inserts by PCR. Most of them had an average insert size greater than 500 bp. Most of them were shown to have less than 20% useless clones and more than 50% new genes. When considering only genes presents in more than one library, a total of 466 unique genes were obtained, corresponding to 427 new *S. mansoni* genes. From the total of unique genes, 20.2% were identified based on homology with genes from other organisms, 8.3% matched *S. mansoni* characterized genes and 71.5% represent unknown genes. After these initial evaluations, our efforts were directed to cercaria, the parasite larval form which infects humans and other vertebrates. Two *S. mansoni* cercarial cDNA libraries were examined and partial sequences obtained from 957 randomly

selected clones. Based on homology searches, 551 (57.6%) ESTs generated had no homologs in the public databases while 308 (32.2%) were putatively identified, totaling 859 informative ESTs. The remaining 98 (10.2%) were uninformative ESTs (ribosomal RNA and non-coding mitochondrial sequences). By clustering analysis we were able to identify 453 different genes. One hundred and nineteen identified genes were sorted into 11 functional categories where genes associated with energy metabolism were the most abundant (13%) and diverse. The diversity and abundance of genes associated with the transcription-translation machinery and with regulatory-signaling functions were also noticeable. A paramyosin transcript was identified, indicating that this gene is not exclusively expressed in adult worms and sporocysts (as had been suggested previously). Among the new genes discovered, we identified a *S. mansoni* Rho-type GTP-binding gene, which was entirely sequenced in this work. The gene open reading frame was 579 bp long and encoded a putative 193 amino acid polypeptide with an estimated molecular weight of 21.8 kDa. The deduced amino acid sequence was highly homologous to the Rho GTPases of several species and contained all Rho-type GTPase conserved domains. *S. mansoni* Rho GTPase gene (*SMRHO*) overexpression complements a *Sacharomyces cerevisiae rho1* null mutant strain. The complementation is seen even in high temperature and  $\text{Ca}^{2+}$  concentration conditions which are restrictive for the yeast growing. However, some phenotypical alterations, such as cell cycle arrest, cell enlargement, and random budding, were also observed indicating that *S. mansoni* Rho GTPase could not complement suitably all functions performed by the wild type *RHO1* GTPase of *S. cerevisiae*. The complementation results of a knockout *S. cerevisiae* strain with *SMRHO* gene along with sequence alignments made evident that a higher conservation of specific amino acid residues at  $\alpha$ 3-helix loop7 region of the *S. mansoni* Rho GTPase protein may explain the complementation seen even in high temperature and  $\text{Ca}^{2+}$  concentration conditions. Our results demonstrate that ESTs are a very versatile and efficient approach for gene discovery in the different developmental stages of the *S. mansoni* life cycle and

that *S. cerevisiae* may be a powerful alternative for gene function investigations in this complex parasite.

**Keywords:** Genome project; gene discovery, ESTs, cDNA library, gene expression, *Schistosoma mansoni*, adult worm, cercaria, Rho GTPases, yeast, *Saccharomyces cerevisiae*, gene knockout, functional complementation, gene function.

## CONSIDERAÇÕES INICIAIS

Este trabalho de tese de doutoramento está inserido dentro do programa de descoberta de novos genes do Projeto Genoma de *Schistosoma mansoni* que vem sendo desenvolvido no Laboratório de Genética Bioquímica do ICB-UFMG. Ele se justifica em razão do pouco conhecimento acumulado sobre o genoma e a biologia molecular deste importante parasita. Acreditamos que os novos genes descobertos e as observações experimentais obtidas a partir deste trabalho representam uma importante contribuição para uma maior compreensão sobre o genoma de *S. mansoni*, sua biologia celular e molecular e, particularmente, sobre a expressão gênica na cercária, fase larval do parasita que infecta o homem. Por razões didáticas, esta tese foi dividida em duas partes:

**PARTE I: Descoberta de Novos Genes de *Schistosoma mansoni* Usando a Estratégia das Etiquetas de Sequências Transcritas (ESTs).**

**PARTE II: Estudos de Genômica Funcional de *Schistosoma mansoni*: Caracterização Molecular e Funcional de um Gene de *Schistosoma mansoni* Homólogo a uma Ras-Like GTP Binding Protein (Rho).**

É importante ressaltar que, de acordo com as Normas Gerais que regem o Curso de Pós-Graduação em Bioquímica e Imunologia, do Departamento de Bioquímica e Imunologia do ICB-UFMG, tendo o aluno cumprido a exigência de publicar um mínimo de três trabalhos em revistas indexadas, sendo ele o primeiro autor em pelo menos um deles, estes trabalhos poderão ser usados para compor a sua tese num formato alternativo ao tradicional. Portanto, em cada parte estão os artigos publicados ou em preparação produzidos a partir dos resultados que foram obtidos. Neles estão expostos a introdução, materiais e métodos, resultados, discussões e conclusões relativos a cada parte. A ordem em que aparecem reflete mais a sequência lógica de desenvolvimento dos trabalhos do que propriamente a cronologia de submissão e de publicação dos mesmos.

## 1. INTRODUÇÃO GERAL

*Diante de tais conhecimentos, qual o perdão?  
Refleta. A história tem muitas passagens ardilosas  
e inventivos corredores. E questões...  
T. S. Eliot, "Gerontion"*

### **1.1. A Esquistossomose**

O gênero *Schistosoma* da família *Schistosomatidae* (LOOS, 1899) compreende os trematódeos digenéticos que vivem como parasitas no sangue do homem e de outros vertebrados. Deste gênero, as espécies *Schistosoma mansoni*, *S. haematobium*, *S. intercalatum*, *S. japonicum* e *S. mekongi* são as causadoras das principais formas da doença humana conhecida como esquistossomose ou bilharziose (NEVES, 2000).

Após a malária, a esquistossomose é a moléstia tropical mais prevalente no mundo estando essencialmente associada à pobreza e às más condições de vida e higiene das populações afetadas. A doença é endêmica em 76 países localizados nas regiões tropicais e subtropicais do globo onde mais de 600 milhões de pessoas estão sob risco de infecção. Mais de 200 milhões de pessoas residentes nas áreas rurais e peri-urbanas encontram-se infectadas. Destas, 120 milhões apresentam sintomas, 20 milhões sofrem severas consequências da doença e 20.000 morrem anualmente de causas associadas à esquistossomose (CHITSULO & COLS, 2000).

O *S. mansoni*, causador da esquistossomose intestinal, é a espécie com maior distribuição global ocorrendo em 54 países abrangendo a África, Península Arábica, Mediterrâneo Oriental, Brasil, algumas ilhas do Caribe, Suriname e Venezuela. Em 10 países da África central, outra forma de esquistossomose intestinal é causada pelo *S. intercalatum*. Na Ásia, a esquistossomose intestinal é causada pelo *S. japonicum*, endêmico na China, Indonésia e Filipinas, e pelo *S. mekongi* na bacia do rio Mekong. Já a forma urinária da esquistossomose, causada pelo *S. haematobium*, é endêmica em 53 países da África e Oriente Médio (WHO, 1996)

### **1.2. O *Schistosoma mansoni* e seu ciclo biológico**

O *Schistosoma mansoni* (SAMBOM, 1907) apresenta dimorfismo sexual assim como as demais espécies do seu gênero. Os machos têm as ventosas oral e ventral bem desenvolvidas, são achatados dorsoventralmente e mais robustos que as fêmeas que

são, por sua vez, mais cilíndricas e delgadas (Fig. 1). As extremidades laterais do corpo do macho se dobram ventralmente formando o canal ginecóforo. Durante seu ciclo de vida (Fig. 2) ocorre alternância de gerações onde a reprodução sexuada pelas formas adultas acontece no hospedeiro vertebrado e o seu estágio assexual multiplicativo ocorre em moluscos gastrópodes aquáticos (PESSOA & MARTINS, 1982; NEVES, 2000).

Os vermes adultos de *S. mansoni* instalam-se na luz das veias mesentéricas ou no sistema porta-hepático do hospedeiro onde ocorre o acasalamento. A fêmea, que vive a maior parte de sua vida alojada no canal ginecóforo do macho, é fecundada iniciando a postura de uma grande quantidade de ovos por dia. Parte destes ovos chega a atravessar a parede intestinal misturando-se às fezes.

O ovo maduro, eliminado junto com as fezes do hospedeiro, ao entrar em contato com a água de lagos ou cursos d'água, eclode liberando uma larva ciliada de vida livre, o miracídio. No ambiente aquático, o miracídio nada ativamente até encontrar um molusco gastrópode no qual penetra através da antena. No Brasil, *Biomphalaria glabrata*, *Biomphalaria tenagophila* e *Biomphalaria straminea* são as três espécies de caramujos que servem de hospedeiro intermediário ao *S. mansoni*.

Logo após a penetração no caramujo o miracídio torna-se sésil e começa a se diferenciar em esporocisto primário ou mãe. Células germinativas existentes no interior do esporocisto I proliferam por sucessivas divisões mitóticas e a partir delas desenvolvem-se os esporocistos filhos ou secundários.

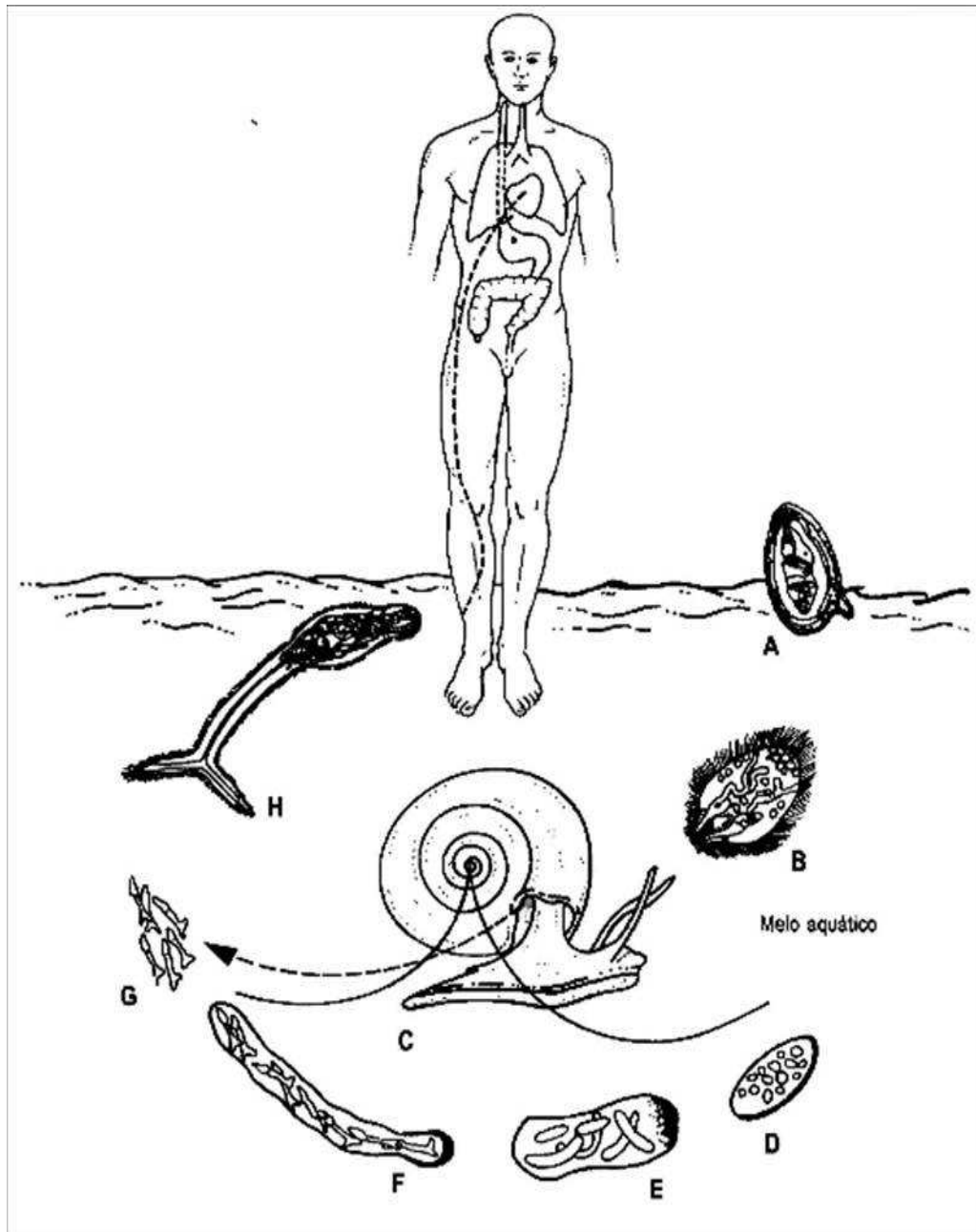
A maturação dos esporocistos filhos completa-se em torno do 15<sup>o</sup> dia após a infecção pelo miracídio. Neste período a maioria deles abandona a forma mãe e avança por vasos e tecidos do molusco com destino ao hepatopâncreas ou ovotestis. Preferencialmente, a larva se instala nos espaços linfáticos que cercam as glândulas digestivas do hepatopâncreas. Ao chegar ao seu destino final, os esporocistos filhos têm a forma de pequenos sacos alongados, contendo dezenas de células germinativas na sua luz. Estas células germinativas aumentam seu número por divisão mitótica podendo



**Figura 1.** Casal de vermes adultos de *Schistosoma mansoni*. A fêmea encontra-se alojada no canal ginecóforo do macho.

(Fonte: <http://www.biosci.ohio-state.edu/~parasite/home.html>)





**Figura 2. Ciclo do *S. mansoni*.** A) ovo com miracídio alcançando a água; B) miracídio nadando para um caramujo — *Biomphalaria*; C) penetração do miracídio nas partes moles do caramujo; D) esporocisto; E) esporocisto; F) esporocisto 2º com cercárias dentro; G) cercárias saindo do caramujo; H) cercárias nadando para novo hospedeiro. Extraído de NEVES (2000).

dar origem a cercárias, ou a massas germinativas que, por sua vez, produzem cercárias por poliembrionia. A infecção pelo *S. mansoni* provoca no caramujo uma diminuição na postura de ovos (BAI & COLS., 1997) além de uma erosão na superfície interna da concha (DAVIES & ERASMUS, 1984). Em 20 ou 30 dias cercárias maduras começam a deixar o esporocisto e migrar para o meio externo.

As cercárias são formas larvárias que medem ao redor de 350 µm. Apresentam um tegumento recoberto por espículas e têm o corpo ovalado ligado a uma cauda longa e bifurcada na sua extremidade posterior, que serve para natação. O verme adulto se desenvolverá a partir do corpo da larva, que contém células germinativas, as ventosas oral e ventral (acetábulo) e esboços embrionários de órgãos, além dos sistemas glandulares de escape, na porção anterior, e de penetração, ou acetabulares, na metade posterior.

Sob influência da luz e temperatura nas horas mais quentes do dia, centenas de cercárias saem ativamente do corpo do caramujo através do colar do manto. Após ter emergido do hospedeiro intermediário *Biomphalaria spp.*, a larva tem um curto período de vida na água para encontrar e penetrar na pele do hospedeiro vertebrado.

Estando em contato com a pele do hospedeiro definitivo, a cercária adere-se a ela por meio da ventosa oral e inicia a penetração com fortes movimentos de contração do corpo e agitação da cauda, secretando o material das suas glândulas acetabulares (STIREWALT & FREGEAU, 1966). Os principais estímulos para penetração e transformação da cercária em esquistossômulo parecem ser ácidos graxos insaturados livres existentes na superfície da pele do hospedeiro (SCHIFF & COLS., 1972). Em particular, o ácido linolênico é um potente estimulante da penetração bem como da liberação do conteúdo das glândulas das cercárias (SALAFSKY & COLS., 1984). Postula-se que papilas argentofílicas existentes na superfície apical da cercária funcionem como quimiorreceptores para a estimulação por lipídeos (KING & HIGASHI, 1992) ativando proteína kinase C não dependente de cálcio (MATSUMURA & COLS., 1991).

A cercária perde a cauda ao penetrar na pele do hospedeiro e começa a passar por profundas alterações na sua morfologia e fisiologia que determinam sua transformação em esquistossômulo. Este novo estágio larval caracteriza-se pela perda de grande parte do glicocálice, mudança na estrutura do tegumento, sensibilidade à água, esvaziamento das glândulas acetabulares e transição de um metabolismo energético aeróbico para anaeróbico, além de outras alterações (STIREWALT, 1974).

Após passarem pela camada córnea, as cercárias progridem através da epiderme, atravessam a membrana basal, atingindo a derme. Nas camadas mais profundas da pele, o esquistossômulo pode entrar num vaso periférico sangüíneo ou linfático. De 20 a 70 horas após a penetração, os esquistossômulos já se acumulam no pulmão para onde migraram através do sistema vascular (WILSON & COLS., 1978).

Do pulmão os esquistossômulos são levados para o lado esquerdo do coração pelas veias pulmonares, e daí para a circulação geral para, finalmente, chegar ao sistema porta no fígado por volta do 23<sup>o</sup> dia de infecção. No interior dos vasos portais os vermes jovens experimentam um rápido crescimento, e atingem a maturidade sexual. Os vermes adultos vivem acasalados no plexo venoso mesentérico e portal e para a oviposição migram para as veias mesentéricas inferiores. A postura dos ovos (de 200 a 2000 por dia) pode ter início 35-40 dias após a penetração da cercária.

O período pré-postural, que acompanha a penetração ativa da cercária pela pele do hospedeiro, também denominado de fase aguda da infecção, pode apresentar-se assintomático como quadro clínico discreto (ROCHA & COLS., 1993) ou como a forma toxêmica, exuberante em sinais e sintomas como urticária, edema, mialgia, febre, náusea, diarreia e outros (RASO, PEDROSO & NEVES, 1986)

Caracteristicamente, os ovos de *S. mansoni* apresentam uma projeção lateral bem afilada que lhes permite resistir ao fluxo sangüíneo, podendo auxiliar também na penetração da parede das vênulas. Enzimas histolíticas são liberadas pelo ovo à medida que ele passa através dos tecidos para a luz do intestino. Esta secreção histolítica é

produzida durante o período de amadurecimento do embrião, permitindo o deslocamento dos ovos em meio aos tecidos do hospedeiro.

A patologia da esquistossomose mansônica é caracterizada principalmente pela formação de reações granulomatosas eosinofílicas crônicas em torno dos ovos que ficam retidos no fígado e em outros tecidos. Tais reações inflamatórias aos ovos podem determinar na fase crônica da doença um quadro grave de fibrose hepática levando a hipertensão portal e hepato-esplenomegalia no hospedeiro, resultando não raramente em óbito. Durante a sua passagem para a luz intestinal os ovos provocam hemorragias e ulcerações além de granulomas no epitélio do órgão (NEVES, 1983; DUNNE, HAGAN & ABATH, 1995). Os ovos maduros que conseguem atravessar a parede intestinal são eliminados do corpo do hospedeiro em meio às fezes, completando-se assim o ciclo de vida do parasita.

### **1.3. Biologia reprodutiva de *S. mansoni* e o papel das GTPases de baixo peso molecular no desenvolvimento sexual da fêmea do parasita**

A patologia da esquistossomose é causada principalmente pelos ovos do parasita retidos nos tecidos do hospedeiro, como descrito anteriormente. O desenvolvimento sexual da fêmea de *S. mansoni* depende do acasalamento com o macho que induz atividades mitóticas essenciais para a diferenciação final e maturação dos órgãos reprodutores femininos (ERASMUS, 1973; POPIEL, 1986a,b). SHAW & COLS. (1977), com o objetivo de responder à questão de como o macho estimula tais processos de desenvolvimento na fêmea, demonstraram que o macho transfere um sinal indutor de natureza química, possivelmente um fator lipofílico, para a fêmea durante o acasalamento. Estudos subsequentes demonstraram a ocorrência de um aumento da síntese de DNA e da proliferação de células vitelínicas induzidos pelo acasalamento (DEN HOLLANDER & ERASMUS, 1985; ERASMUS, 1986).

Como o estímulo do macho que induz a proliferação de células vitelínicas na fêmea tem que ser propagado via transdução de sinal, SCHÜßLER, GREVELDING & KUNZ (1997) propuseram-se a identificar moléculas transdutoras de sinal do parasita que pudessem estar envolvidas na ativação de um programa mitogênico. Assim, através do uso de anticorpos monoclonais dirigidos contra epitopos conservados de Ras, Ras-GAP e MAP kinases, esses autores evidenciaram a participação destas moléculas no processo de desenvolvimento e maturação sexual da fêmea de *S. mansoni* induzido pelo acasalamento com o macho. SCHÜßLER, GREVELDING & KUNZ (1997) constataram que a expressão de Ras e Ras-GAP em fêmeas depende do acasalamento com o macho e que MAP kinases 1 e 2 são expressas preferencialmente em fêmeas.

Os primeiros estudos feitos acerca do mecanismo que regula a produção de ovos nas fêmeas do parasita indicaram que GTPases de baixo peso molecular da superfamília do protooncogene *RAS* podem desempenhar um papel importante na fisiologia reprodutiva de *S. mansoni*. Estes estudos conduziram à investigação da função da via que converte mevalonato em compostos isoprenóides na regulação da produção de ovos (VANDEWAA & BENNETT, 1987). Isto porque foi demonstrado que o efeito do inibidor mevinolina sobre a enzima hidroximetilglutaril CoA redutase (HMG-CoA redutase) resultava no decréscimo da produção de ovos pela fêmea do parasita tanto *in vitro* como *in vivo*, sem interferir com outros processos fisiológicos do parasita, e no bloqueio da patologia associada com a infecção em camundongos (VANDEWAA & COLS., 1989). O mesmo efeito era verificado pela ação do inibidor BZA-5B sobre a enzima prenilttransferase, responsável pela prenilação e geranilgeranilação de proteínas (BENNETT J.L. citado por LOEFFLER & BENNETT, 1996). Experimentos subsequentes demonstraram que os efeitos resultantes da inibição da enzima HMG-CoA redutase eram mediados pelos compostos finais isoprenóides da via do mevalonato (VANDEWAA & COLS., 1989; CHEN & BENNETT, 1993).

Posteriormente demonstrou-se que um grupo de proteínas de 43 kDa era farnesilada enquanto que um segundo grupo de proteínas de 25 kDa do parasita era geranilgeranilada (CHEN & BENNET, 1993). Estes mecanismos de prenilação (adição de uma calda de carbono na extremidade C-terminal) de ambos os grupos de proteínas foram bloqueados pelo inibidor BZA-5B. Por fim, demonstrou-se também que as proteínas de 25 kDa eram capazes de ligar GTP (CHEN & BENNET, 1993). Todos estes dados permitiram supor que proteínas de baixo peso molecular ligadoras de GTP poderiam desempenhar um papel importante na biologia sexual de *S. mansoni*. A caracterização destas proteínas tornou-se então crítica para entender melhor sua função na fisiologia reprodutiva de *S. mansoni*. LOEFFLER & BENNETT (1996) isolaram um cDNA de *S. mansoni* que codificava uma proteína de baixo peso molecular ligadora de GTP homóloga a RAB. Análises da sequência deste gene, ensaios de ligação a GTP e reações de geranilgeranilação permitiram identificar a proteína codificada por este gene como sendo um membro da família RAB.

Recentemente, um clone de cDNA codificando um gene com alto grau de homologia com o gene K-RAS de vários organismos foi isolado e caracterizado (OMER & COLS., 1999). O gene de *S. mansoni* homólogo de K-RAS, *smras*, é superexpresso em fêmeas e sua expressão foi detectada em todos os estágios de desenvolvimento do parasita. A proteína codificada por *smras* é farnesilada in vitro e foi identificada por imunocitocalização associada com o subtegumento do verme.

As GTPases de baixo peso molecular, pelas importantes funções que desempenham, são exemplo de uma classe de moléculas-alvo do parasita que requer o uso intensivo de abordagens genômicas e de genômica funcional para sua identificação e estudo. Até o presente momento poucos membros dessa família de proteínas em *S. mansoni* foram identificados. O conhecimento do papel dessa classe de moléculas na biologia reprodutiva do parasita pode abrir novas frentes de investigação para o

desenvolvimento de drogas a serem usadas no controle e tratamento da esquistossomose.

#### **1.4. Diagnóstico, tratamento, estratégias de controle da esquistossomose e o desenvolvimento de vacinas**

O uso de drogas para tratamento de pacientes com esquistossomose, preferivelmente em combinação com outras abordagens tais como controle do hospedeiro intermediário através do uso de moluscidas, educação sanitária, saneamento básico e fornecimento de água potável, constitui a estratégia básica recomendada pelo Comitê da OMS para Controle da Esquistossomose (WHO, 1996, 1998).

A demonstração de ovos do parasita na urina ou em amostras fecais (KATZ, COELHO & PELLEGRINO, 1970) fornece o diagnóstico definitivo para a doença. Para pessoas residentes em áreas não endêmicas, a presença de antígenos circulantes ou qualquer outro diagnóstico que detecte anticorpos pode ser suficiente para demonstrar a infecção. O diagnóstico precoce de infecções moderadas ou severas é importante para a prescrição de um tratamento médico adequado.

Três drogas efetivas e seguras para esquistossomose – praziquantel, oxaminiquine e metrifonato – encontram-se disponíveis na atualidade e fazem parte da lista modelo de drogas essenciais da OMS (WHO, 1996, 1998). Praziquantel é efetiva contra todas as formas da doença com poucos, e apenas passageiros, efeitos colaterais. Oxaminiquine é usada exclusivamente na África e América do Sul para tratamento da forma intestinal. Para o tratamento da forma urinária da esquistossomose o metrifonato provou ser um medicamento seguro e efetivo. Ainda que ocorra uma reinfecção após a quimioterapia, o risco de desenvolvimento de um quadro severo é diminuído ou mesmo revertido em crianças.

O uso de moluscidas nos programas de controle da esquistossomose depende muito das circunstâncias ecológicas e epidemiológicas locais, junto com recursos humanos e financeiros disponíveis. Entretanto, em comunidades de alta endemicidade, o uso focal de moluscidas é um recurso que pode ser de grande valia.

A distribuição global da esquistossomose tem mudado significativamente nos últimos 50 anos. A doença foi erradicada no Japão e nas Pequenas Antilhas, a transmissão foi interrompida na Tunísia e teve seu nível reduzido significativamente no Marrocos, Filipinas, Arábia Saudita e Venezuela (WHO, 2000).

No Brasil, dados da Fundação Nacional da Saúde mostram que de 1977 a 1993 a taxa de mortalidade devido à esquistossomose caiu de 0,65 para 0,30/100.000 habitantes, um decréscimo de mais de 50%. As internações por causa da doença no país também caíram de 3/10.000 em 1989 para 1,5/10.000 habitantes em 1995 (KATZ, 1998).

A despeito do sucesso obtido no controle da esquistossomose, recentes modificações ambientais ligadas ao manejo de recursos hídricos, o aumento da população, movimentos migratórios impulsionados pela fome, desemprego e o deslocamento de refugiados de guerras, têm ocasionado, sobretudo no continente africano, a dispersão da doença em áreas antes não endêmicas ou de baixa endemicidade. A estimativa atual é de que ao redor de 500 a 600 milhões de pessoas encontram-se sob risco de infecção nas regiões onde a doença ainda é prevalente (CHITSULO & COLS, 2000). O problema ganha dimensões preocupantes em vista de relatos que têm sido feitos nos últimos anos sobre o surgimento de cepas do parasita resistentes aos medicamentos usados no tratamento dos pacientes infectados (BRUCE & COLS., 1987; ISMAIL & COLS., 1996; COELHO, LIMA E SILVA & NOGUEIRA-MACHADO, 1997; WHO, 1998b).

Paralelamente, estudos têm sido feitos para desenvolver vacinas que resultem na diminuição da morbidade através da redução do número de vermes nas infecções e/ou da fecundidade dos parasitas em pessoas residentes nas áreas endêmicas. Os antígenos



candidatos à vacina atualmente conhecidos, cujos genes já foram clonados e extensivamente caracterizados, podem ser agrupados em três categorias: proteínas solúveis associadas à superfície do *Schistosoma*, glicoproteínas integrais de membrana e proteínas associadas ao tegumento mas não expostas na superfície do parasita.

No primeiro grupo estão incluídas enzimas como a glutationa S-transferase nas suas formas de 28 [Sm28GST] e 26 kDa [Sm26GST] (BALLOUL & COLS., 1987; TROTTEIN & COLS., 1992) e as enzimas da via glicolítica gliceraldeído 3-fosfato desidrogenase [GAPDH] (GOUDOT-CROUZEL & COLS., 1989) e triose fosfato isomerase [TPI] (SHOEMAKER & COLS., 1992).

O segundo grupo é constituído por proteínas ligadas por diferentes formas à membrana de *S. mansoni*. Fazem parte deste grupo as proteínas ligadas à membrana por âncoras de glicosilfosfatidilinositol (GPI), como é o caso de Sm38, Sm32 e Sm20 (PEARCE & SHER, 1989), proteínas ancoradas por uma ligação tioéster a um ácido palmítico como em Sm25 (PEARCE & COLS., 1991), proteínas ligadas à membrana por uma forma ainda não identificada de ancoragem, como visto em Sm15 (ABATH & COLS., 1993) e Sm14 (TENDLER, 1996) e também proteínas integrais de membrana, como é o caso de Sm23, que contém quatro domínios hidrofóbicos transmembrana (REYNOLDS, SHOEMAKER & HARN, 1992).

No terceiro grupo de antígenos candidatos à vacina estão incluídas as proteínas musculares paramiosina e miosina (cadeia pesada), também encontradas no tegumento do verme mas não expostas. A primeira foi isolada utilizando soro de camundongos imunizados com extrato total de cercária (PEARCE & COLS., 1988). A segunda proteína corresponde ao antígeno de 200 kDa rIrV-5 que foi isolado de uma biblioteca de cDNA utilizando anticorpos presentes no soro de camundongos vacinados com cercárias irradiadas (SOISSON & COLS., 1992). Este antígeno também está presente na superfície de esquistossômulos recém transformados.

Alguns dos antígenos acima citados (rIrv-5, GST-Sm28, MAP-4/TPI, paramiosina/Sm97, MAP-3/Sm23 e Sm14), considerados prioritários para finalizar testes independentes de valor protetivo em camundongos, foram incluídos nos programas de desenvolvimento de vacinas contra esquistossomose da OMS, financiados por este organismo através do Programa Especial para Pesquisa e Treinamento em Doenças Tropicais – TDR (WHO, 1997). Porém, a produção de uma vacina a partir de qualquer um dos antígenos candidatos requer ainda um longo processo de desenvolvimento.

A utilização das ferramentas da biologia molecular, voltadas principalmente para o estudo de genes codificadores de antígenos candidatos a uma vacina contra o *S. mansoni*, propiciou um ganho significativo de informações sobre seus processos fisiológicos, bioquímicos e a organização do seu genoma. Contudo, a quantidade de informações acumuladas ainda representa muito pouco em face da importância do parasita para a saúde pública mundial, justificando esforços ainda maiores no intuito de descobrir e caracterizar novos genes e seus produtos. Neste sentido, o conhecimento e tecnologias revolucionárias gerados pelo estudo sistemático de genomas, iniciado pelo Projeto Genoma Humano, abriu um campo de enormes possibilidades que, aplicadas ao estudo do genoma de *S. mansoni* e de uma variedade de outros patógenos, propiciará importantes avanços para o controle das doenças infecciosas e parasitárias que afligem sobretudo os países do terceiro mundo (PENA, 1996).

### **1.5. Projetos Genoma**

A obtenção de sequências completas de grandes porções do DNA genômico ganhou impulso com o advento dos métodos de sequenciamento por terminação da cadeia de DNA através da incorporação de didesoxinucleotídeos pela DNA polimerase (SANGER & COLS, 1977) e degradação química (MAXAN & GILBERT, 1977). Essas metodologias tornaram possível a obtenção da sequência completa do DNA mitocondrial humano (ANDERSON & COLS, 1981) e da sequência completa do genoma do

bacteriófago T7 (DUNN & STUDIER, 1983). *Haemophilus influenza* foi a primeira bactéria a ter seu genoma (~1,8 Mb) completamente sequenciado usando a estratégia de "shot-gun", um feito significativo com relação ao sequenciamento de grandes sêquências de DNA (FLEISHMANN & COLS., 1995). A estratégia de "shot-gun" trabalha com clones de DNA genômico em plasmídeos, bacteriófagos, cosmídeos, YAC (Yeast Artificial Chromossome) e BAC (Bacterial Artificial Chromossome), que são parcial ou totalmente sequenciados a partir de ambas as extremidades, e têm suas sequências sobrepostas em busca de regiões que permitam ordenar os clones e fazer montagens de sequências contínuas.

Ao longo da última década do século XX foram iniciados vários projetos cujos objetivos envolviam nada menos que o sequenciamento total do genoma de vários organismos eucariotos, identificar todos os seus genes e obter mapas físicos de alta resolução. A possibilidade de conclusão desses objetivos teve como bases principais a automação do processo de sequenciamento de DNA e o desenvolvimento da informática. Dentre todos, porém, o mais ambicioso projeto de pesquisa genômica é o Projeto Genoma Humano (HGP), que teve início oficialmente em 1990 nos Estados Unidos (COLLINS & GALAS, 1993) e é coordenado pelo Department of Energy (DOE) e o National Institutes of Health (NIH) desse país. A iniciativa foi logo seguida por outros países como França, Inglaterra e Japão.

O objetivo primário do HGP era o sequenciamento dos três bilhões de pares de base (~3.000 Mb) que compõem o genoma humano além de identificar e mapear todos os seus genes (~100.000) num espaço de tempo de 15 anos, contando com uma verba de U\$ 3 bilhões. Seus objetivos secundários incluíam a descoberta de novas formas de diagnóstico e tratamento para doenças genéticas assim como a transferência de tecnologia e conhecimento obtidos para outras áreas, onde poderão promover um forte estímulo em uma variedade de setores da pesquisa biológica, médica e biotecnológica (COLLINS & GALAS, 1993; GUYER & COLLINS, 1995).

O novo plano quinquenal (1998-2003) do HGP objetiva concluir e disponibilizar gratuitamente, de forma ordenada em bases de dados, a sequência de todo o genoma humano no final de 2003 (U.S. Human Genome Project 5-Year Research Goals 1998-2003, 1998). Mas, a Celera Genomics, empresa privada dirigida por Craig Venter assumiu um compromisso de conclusão em 2001. O grupo de Venter anunciou em abril de 2000 que, após ter sequenciado todas as bases, atualmente está trabalhando na ordenação do genoma humano. A sequência do cromossomo 22 foi publicada em dezembro de 1999 e mostrou ser um marco importante na descrição da primeira sequência completa do HGP. Esse cromossomo consiste de 12 sequências adjacentes de 33,4 Mb e contém 545 genes e 134 pseudogenes, fornecendo o primeiro mapa de um cromossomo humano e nos dando uma idéia do que pode ser encontrado no restante do genoma (DUNHAM & COLS., 1999).

Em 18 de maio de 2000 foi publicado pela revista *Nature* um artigo sobre a sequência completa do cromossomo 21, realizado principalmente por grupos japoneses e alemães. Esse cromossomo representa de 1 a 1,5% do genoma humano e é o menor dos autossomos. Uma cópia extra do cromossomo 21 provoca a síndrome de Down, a causa genética mais frequente e significativa de retardo mental, que afeta um em cada setecentos nascidos. Nesse trabalho relata-se com alta precisão a sequência do braço longo do cromossomo 21 com 33,5 Mb e 281 kb do braço curto o que cobre 99,7% do 21q. A análise do cromossomo revela 127 genes conhecidos, 98 possíveis genes e 59 pseudogenes (HATTORI & COLS., 2000).

Diante do imenso tamanho dos genomas dos organismos eucariotos, uma alternativa razoável seria a de priorizar o sequenciamento de regiões codificadoras que, apesar de geralmente representarem uma pequena fração do genoma, contém a parte mais significativa em termos de informação. Assim, o sequenciamento das extremidades de moléculas de cDNA para produção de etiquetas de sequências transcritas (ESTs), tem dado maior retorno e indubitavelmente tem proporcionado uma mudança revolucionária

nas estratégias usadas pelos geneticistas moleculares para identificação e clonagem de novos genes (STRAUSBERG, DAHL & KLAUSNER, 1997; MARRA & COLS., 1998).

ESTs são sequências de nucleotídeos geradas a partir das extremidades de clones de cDNA selecionados ao acaso de bibliotecas (ADAMS & COLS., 1991). As sequências parciais assim obtidas podem ser usadas para buscar homologias com sequências de DNA ou proteínas depositadas em bancos de dados na tentativa de identificar os genes dos quais elas derivam (ADAMS & COLS., 1991; ADAMS & COLS., 1995; HILLIER & COLS., 1996).

As ESTs produzidas, em geral, são depositadas em um banco de dados específico, criado dentro do GenBank em 1993 e denominado dbEST (banco de dados de sequências transcritas - BOGUSKI & COLS, 1993). Nesse banco de dados, além das sequências serem armazenadas, são procuradas e removidas regiões repetitivas ou porções de vetores que eventualmente foram depositadas como parte de ESTs. As sequências desse banco também são frequentemente comparadas com sequências já conhecidas, na tentativa de identificar ESTs depositadas (COOKE & COLS., 1996). A evolução do dbEST, em termos de número de sequências é tremenda. A versão liberada em 2 de janeiro de 1997 continha mais de um milhão de ESTs de oitenta organismos diferentes. Em 16 de junho de 2000, existiam mais de quatro milhões de ESTs disponíveis de mais de duzentas espécies (dbEST release 061600, 2000).

A notável expansão dos esforços de sequenciamento de cDNA levou nos últimos 8 anos à produção de mais de dois milhões de ESTs de diferentes tecidos e estágios do desenvolvimento de *Homo sapiens* (dbEST release 061600, 2000). O banco de dados UniGene contém mais de 81.951 grupos de sequências em junho de 2000, cada um representando um único transcrito (UniGene Human Release Statistics, 2000). Este número corresponde entre 70-80% do número total de genes que se acredita existir no genoma humano. Os grupos são gerados a partir da comparação entre genes conhecidos e ESTs para determinar as sequências que aparentemente derivam do mesmo gene.

Ainda, como parte dos objetivos do HGP, foram iniciados em paralelo programas de sequenciamento do genoma de organismos considerados modelos experimentais importantes para estudos evolutivos e determinação da função de cada um dos genes novos descobertos pelo HGP e por outros projetos genomas (BECK & STERK, 1998). De maneira geral o genoma destes organismos têm um tamanho bem menor que o humano e dentre os escolhidos estão o nematódeo *Caenorhabditis elegans* (~97 Mb), a levedura *Saccharomyces cerevisiae* (~12 Mb), a mosca da fruta *Drosophila melanogaster* (~180 Mb), a planta dicotiledônea *Arabidopsis thaliana* (~100 Mb) e o camundongo *Mus musculus* (~3.000 Mb). Alguns destes projetos já foram ou estão praticamente concluídos como os de *S. cerevisiae* (GOFFEAU & COLS., 1996), *C. elegans* (*C. elegans* SEQUENCING CONSORTIUM, 1998), que foi primeiro organismo pluricelular a ter o seu genoma completamente sequenciado, e da drosófila (ADAMS & COLS., 2000).

É impossível falarmos hoje de evolução de Projetos Genoma no Brasil sem citarmos o Projeto Genoma da *Xylella fastidiosa*, a bactéria causadora do amarelinho do clorose variegata dos cítricos (CVC). Esse projeto foi idealizado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e tem importância estratégica por seus frutos diretos e indiretos. Atualmente, O estado de São Paulo é uma das duas maiores regiões produtoras de cítricos do mundo, atividade que gera cerca de quatrocentos mil empregos. Uma das ameaças mais sérias a essa produção é a praga causada pela *X. fastidiosa*, que já se encontra disseminada em 30% das plantações de cítricos desse estado.

Em poucos meses o Projeto Genoma de *X. fastidiosa* foi estruturado em um comitê de altíssimo nível, equipando-se trinta e três laboratórios de diversas regiões do estado de São Paulo com sequenciadores automáticos de última geração. Foi criado então o primeiro grupo de sequenciamento em larga escala fora do primeiro mundo e estabelecido um centro virtual de sequenciamento de DNA, denominado ONSA (Organization for Nucleotide Sequencing and Analysis,

<http://watson.fapesp.br/Genoma3.htm>), que integra todos os participantes do projeto. A ONSA está entre os maiores grupos de estudos genômicos do mundo, sendo capaz de sequenciar mais de 1 milhão de nucleotídeos por dia (TEIXEIRA, 1999).

O anúncio da conclusão do projeto pioneiro do Programa Genoma da FAPESP faz do Brasil o único país do hemisfério sul a ter lugar no seleto grupo que detém a tecnologia que mais promete conquistas para o século XXI. O genoma de *X. fastidiosa* é um dos mais extensos genomas de microorganismos já completamente sequenciado e o primeiro de um organismo causador de doenças em plantas (SIMPSON & COLS., 2000). Outros projetos genoma estão sendo desenvolvidos pela Rede ONSA como o da cana-de-açúcar, do câncer humano e da bactéria *Xanthomonas citri*, um outro patógeno de cítricos.

## **1.6. Estudos pós-genômicos**

A proximidade do anúncio da conclusão do Projeto Genoma Humano e bem como a efetiva conclusão dos Projetos Genoma dos principais organismos modelos inaugura a era pós-genômica cujo objetivo é elucidar a função de todos os genes novos revelados pelos programas de sequenciamento. A nova fase representa um grande desafio que exigirá um intenso esforço colaborativo entre os membros da comunidade científica mundial.

Existem muitas rotas para se construir o perfil de um gene e sua função. Num primeiro momento elas incluem buscas em bancos de dados para comparação de similaridade com sequências do mesmo ou de outros organismos (ASHBURNER & GOODMAN, 1997; RASTAN & BEELEY, 1997) e busca por motivos estruturais potencialmente homólogos (BORK & KOONIN, 1998). Entretanto, abordagens mais contextuais para a descoberta das funções dos novos genes revelados pelos programas de sequenciamento de genomas, ou genômica funcional, envolvem outros níveis de investigação como a expressão gênica e o conjunto de proteínas produzidas (OLIVER &

COLS., 1998). Sem dúvida a utilização de organismos modelos será de fundamental importância para a interpretação da função dos genes descobertos, estudos evolutivos e de integração das diversas funções biológicas.

### **1.6.1. Transcriptoma**

Uma abordagem eficiente para análise do transcriptoma, ou seja, do conjunto de mRNAs produzidos por uma célula, tecido ou órgão, é pelo uso de SAGE (Serial Analysis of Gene Expression). Essa metodologia tem como base o uso de pequenas etiquetas (10-14 pb) geradas a partir da extremidade 3' de moléculas de cDNA que servem para identificar cada um dos genes representados em uma amostra de transcritos. Estas etiquetas são concatenadas em séries que são posteriormente clonadas e sequenciadas. As sequências geradas a partir destas séries fornecem informação sobre múltiplos genes. A quantificação do número de vezes que uma etiqueta em particular é observada em uma população de etiquetas de SAGE fornece a informação direta sobre o nível de expressão gene correspondente (BERTELSEN & VELCULESCU, 1998).

Alternativamente, “DNA arrays” têm sido usados em estudos de larga escala do perfil da expressão de genes durante o desenvolvimento e diferenciação ou da maneira como respondem a um estímulo (CASTELLINO, 1997; DeRISI, IYER & BROWN, 1997). De maneira geral, a técnica de “DNA array”, em suas várias implementações, baseia-se no uso de sequências de DNA (cDNA ou oligonucleotídeos) dispostos em arranjos de alta densidade sobre uma superfície sólida que pode ser de vidro ou silício ou mesmo em membranas de nylon, para hibridização com uma sonda preparada de RNA extraído de uma célula, tecido ou organismo. A sonda é produzida por transcrição reversa do mRNA ou mesmo do RNA total e marcada radioativamente ou por fluorescência (GRANJEAUD, BERTUCCI & JORDAN, 1999). O experimento pode dessa forma fornecer uma indicação dos níveis de expressão de milhares de genes ao mesmo tempo.



### 1.6.2. Proteoma

A despeito da elegância e velocidade com as quais a expressão gênica pode ser monitorada pelas novas tecnologias de análise do conteúdo transcricional em uma célula tecido ou órgão, os níveis de mRNA não correlacionam necessariamente com aqueles das proteínas sintetizadas devido a diferenças nas taxas de degradação entre mRNA e proteínas e porque muitos polipeptídeos sofrem modificações após serem traduzidos (ANDERSON & SEILHALMER, 1997). Em consequência disso, grandes esforços estão sendo dirigidos no sentido de produzir uma completa descrição de uma célula, tecido, órgão ou mesmo de um organismo em termos do conjunto de proteínas produzidas. Tal abordagem é hoje conhecida como “proteoma” (De FRANCESCO, 1999). O desafio é enorme tendo em vista que o padrão de proteínas expressas varia em função do tempo e do ambiente, envolvendo respostas a sinais do desenvolvimento, doenças, disponibilidade de nutrientes, etc.

No centro desse esforço estão os géis bidimensionais (2D), baseados no sistema descrito originalmente por O’FARRELL (1975), para separação de proteínas inicialmente pelo ponto isoelétrico e posteriormente pela massa molecular. A vantagem do uso de géis 2-D para estudo de proteomas reside na possibilidade de se poder, através desta tecnologia, separar e visualizar milhares de proteínas ao mesmo tempo (KLOSE & KOBALZ, 1995). Assim é possível acompanhar mudanças na expressão de proteínas ou alterações que uma proteína pode sofrer na sua forma devido a vários tipos de modificações pós-traducionais.

Outras análises podem ser aplicadas às proteínas separadas em gel, como a espectrometria de massa. Espectrômetros de massa consistem de três partes essenciais. A primeira, uma fonte de ionização, converte moléculas em íons em fase gasosa. Uma vez que os íons tenham sido criados, índices individuais de carga-massa ( $m/z$ ) são separados por um segundo aparelho, um analisador de massa, e transferido para o terceiro, um detector de íons. Um analisador de massa usa uma propriedade física

(campos elétrico ou magnético, ou tempo de voo [TOF]) para separar íons de um valor particular  $m/z$  que então colidem com um detector de íons. A magnitude da corrente que é produzida no detector em função do tempo é usada para determinar o valor  $m/z$  de cada íon. Esta metodologia permite determinar o tamanho e a composição de aminoácidos dos fragmentos resultantes da proteólise controlada de uma proteína para que, em seguida, sejam comparados com as mesmas propriedades de sequências de proteínas conhecidas depositadas nos bancos de dados, permitindo então fazer sua identificação (KUSTER & MANN, 1998; YATES, 2000).

Entretanto, as metodologias clássicas descritas acima para estudos de proteomas descartam informações de grande valor para estudos funcionais relacionados às interações proteína-proteína. Isto porque as condições desnaturantes utilizadas no preparo das amostras para 2D destroem essas interações. Desse modo, uma abordagem paralela para análise de proteomas é examinar não a abundância relativa de proteínas, mas as interações potenciais de umas com as outras (OLIVER, 2000).

Sistemas de vetores para duplo-híbrido e suas variações usando leveduras, ou mesmo bactérias, como células hospedeiras constituem-se atualmente em um dos métodos genéticos mais poderosos para se estudar interações físicas entre duas ou mais proteínas e mesmo entre DNA-proteína (BRACHMANN & BOEKE, 1997).

O sistema básico de duplo-híbrido tem três componentes principais: primeiro, um vetor para expressão da proteína candidata (ou "isca") fundida ao motivo de ligação a DNA de um fator de transcrição. Segundo, um outro vetor para expressão de uma biblioteca de sequências de proteínas que podem interagir com a candidata (denominadas "presas") fundidas ao domínio de ativação do fator de transcrição e, por fim, um terceiro vetor contendo um ou mais genes repórteres colocados após a sequência do promotor que é reconhecido pelo motivo de ligação ao DNA fundido à "presa". Leveduras têm dois sexos, chamados  $\alpha$  e  $a$ . Assim, "iscas" e "presas" podem facilmente ser introduzidas na mesma levedura por cruzamento. O sistema entra em

atividade na levedura quando a "presa" interage com a "isca" fazendo com que os dois motivos do fator de transcrição se unam, liguem-se ao promotor e induzam a expressão do gene repórter ( $\beta$ -galactosidase ou um marcador de auxotrofia), cuja função é sinalizar a ocorrência de uma interação "presa-isca" efetiva.

Algumas interações detectadas por duplo-híbrido podem, no entanto, nunca ocorrer naturalmente, ou seja, são interações entre proteínas sem qualquer relação funcional ou espacial no ambiente celular. Reitera-se, mais uma vez, a necessidade de novas abordagens para estudos de proteomas, que façam uso de procedimentos de separação que não destruam interações proteína-proteína e que envolvam métodos de espectrometria de massa e bioinformática para uma rápida e inequívoca identificação (OLIVER, 2000).

### **1.6.3. Organismos Modelos Como Sistemas Experimentais Para Estudos da Função Gênica**

A posição de *Saccharomyces cerevisiae* como organismo modelo eucariótico deve-se muito às suas vantagens intrínsecas como sistema experimental (GOFFEAU & COLS., 1998). A levedura é um organismo unicelular que (diferentemente dos eucariotas mais complexos) pode crescer em meio definido, dando ao pesquisador um controle completo sobre seu ambiente físico e químico. *S. cerevisiae* tem um ciclo de vida que é ideal para análises genéticas clássicas o que tem permitido a construção de um mapa genético detalhado que define o conjunto haplóide de 16 cromossomos. Além do mais, técnicas simples e muito eficientes têm sido desenvolvidas para permitir que qualquer dos seus 6.000 genes sejam substituídos por um alelo mutante ou mesmo suprimidos do genoma com absoluta precisão, com a finalidade de explorar o fenótipo resultante, ou ainda de serem substituídos por genes exógenos para estudos funcionais por complementação heteróloga (WINZELER & DAVIS, 1997; BRACHMAN & COLS., 1998).

A facilidade de manipulação genética e manutenção em laboratório do nematódeo *C. elegans*, aliada ao seu pequeno genoma e anatomia simples, faz deste verme um organismo modelo igualmente atraente para o estudo de genes novos (KUWABARA, 1997). Se no passado a clonagem de genes de *C. elegans* era um problema para sua caracterização genética, este obstáculo foi superado agora que se dispõe da sequência completa do seu genoma e de um número abundante de ESTs. Estima-se que o genoma de *C. elegans* contenha mais de 19.000 genes. Este número é mais que o dobro de genes codificado pelo genoma de *S. cerevisiae* e provavelmente inclui muitos genes e famílias gênicas que podem ser encontrados apenas em metazoários e organismos pluricelulares.

Atualmente é possível começar um estudo com a sequência de um gene de *C. elegans* e trabalhar com técnicas de genética reversa (do gene para o fenótipo) para conseguir entender sua função e regulação por meio de mutações geradas em genes para os quais ainda não existem mutantes disponíveis (KUWABARA, 1997). *C. elegans* pode ser usado também para verificar se um gene de outro organismo pluricelular pode substituir o seu homólogo no nematódeo e assim determinar sua função observando se sua atividade ou regulação são conservados (KUWABARA, 1997). Porém, a injeção de RNA constitui uma metodologia inicial rápida e eficiente para se obter indícios sobre a função de um gene selvagem através da sua supressão (ERICKSON, 1999).

Já a drosófila é um dos organismos mais intensamente estudados em biologia e serve como sistema modelo para a investigação de muitos processos celulares e de desenvolvimento comuns nos eucariotos superiores, incluindo o homem. Seu genoma codifica aproximadamente 13.600 genes, número um pouco menor que o encontrado em *C. elegans*, mas com uma diversidade funcional bastante semelhante (ADAMS & COLS., 2000).

Dentre os organismos modelos, contudo, o camundongo é aquele que assumirá o papel mais importante nos estudos sobre a função dos genes de mamíferos na era pós-

genômica. Para tanto, a rota através da qual se espera maior contribuição é a da mutagênese. O desafio é desenvolver protocolos de mutagênese aplicáveis em toda a extensão do genoma para o estudo da função gênica no animal (BROWN & NOLAN, 1998). Neste sentido, existem duas abordagens: uma delas é dirigida ao genótipo (genética reversa) e a outra ao fenótipo (genética direta). Abordagens genótipo-dirigidas são voltadas para a sequência podendo envolver a deleção de genes alvos produzida por meio de estratégias de inserção ou substituição por construções clonadas em vetores integrativos introduzidos em células tronco embrionárias (DAVIES, GALLAGHER & SAVIOZ, 1994).

Uma abordagem mais sistemática pode ser feita usando a estratégia da armadilha gênica (gene trap). Nela uma construção apropriada selecionável é introduzida em células tronco com o objetivo de se obter um grande número de linhagens portadoras de unidades de transcrição rompidas e assim construir um banco de mutações (EVANS, CARLTON & RUSS, 1997). Mutações precisas podem igualmente ser introduzidas por meio de vetores integrativos permitindo dessa forma abordar questões específicas a respeito da regulação e das relações estrutura-função. Para as abordagens fenótipo-dirigidas o foco está no uso de sistemas de mutagênese ao acaso para obtenção de camundongos com alterações em um determinado fenótipo a partir dos quais os genes relevantes e as vias regulatórias envolvidas com aquela função podem ser identificados.

Os geneticistas que trabalham com camundongo têm voltado sua atenção para a mutagênese química usando o composto *N*-etil-*N*-nitrosouréia - ENU (deANGELIS & COLS., 2000; NOLAN & COLS., 2000). As abordagens fenótipo-dirigidas, porém, requerem a aplicação de procedimentos de varredura apropriados a cada caso com o objetivo de identificar os fenótipos de interesse.

### 1.7. O Projeto Genoma de *Schistosoma mansoni*

A esquistosomose, assim como uma grande variedade de doenças parasitárias e infecciosas que afligem sobretudo os países pobres, poderão ser mais bem controladas por meio de vacinas e procedimentos terapêuticos desenvolvidos a partir do conhecimento obtido do estudo sistemático do genoma desses patógenos. Foi com essa expectativa que em 1992 teve início no Brasil o Projeto Genoma de *Schistosoma mansoni*, por iniciativa do Professor Sérgio Pena, do Departamento de Bioquímica e Imunologia do ICB-UFMG, com a colaboração do Dr. Andrew Simpson, à época no Centro de Pesquisa Rene Rachou (FIOCRUZ), Belo Horizonte, e apoio logístico fornecido pelo Dr. J. Craig Venter, então no The Institut for Genomic Research (TIGR), Gaithersburg, EUA, objetivando a descoberta de novos genes e o mapeamento do seu genoma (PENA, 1996).

Até o início do Projeto Genoma de *S. mansoni* pouco se sabia sobre o seu genoma e apenas cerca de 100 genes do parasita eram conhecidos. O *S. mansoni* é um organismo diplóide, contendo 7 pares de cromossomos autossômicos e um par de cromossomos sexuais. O sexo homogamético é o macho (ZZ) e o heterogamético é a fêmea (ZW). O tamanho do genoma haplóide do parasita foi estimado em 270 Mb, cerca de um décimo do genoma humano. Deste total, 4% consiste de DNA altamente repetitivo (>1000 cópias) e 35 a 40% de DNA de média repetição (SIMPSON & COLS., 1982). Perfis de desnaturação térmica e cálculos de densidade de flutuação revelaram que o genoma de *S. mansoni* possui um conteúdo médio de A+T (66%) bastante elevado (HILLYER, 1974). O índice deste conteúdo varia dependendo se o DNA é de região codificadora (60%), ficando em torno de 70% na região não codificadora (MEADOWS & SIMPSON, 1989; MILHON & TRACY, 1995). Com base no tamanho do seu genoma e posição evolutiva, calcula-se que o *S. mansoni* contenha de 15.000 a 20.000 genes expressos (ALI & COLS., 1991). Análises do uso de códons em genes deste organismo demonstraram que purinas são mais freqüentemente encontradas na primeira posição,

61%, e pirimidinas na terceira posição dos códons, 73% (MILHON & TRACY, 1995). A existência de vícios na representatividade dos códons foi demonstrada em genes onde existem códons altamente representados e outros de baixa representatividade (MUSTO & COLS., 1998).

A estratégia básica adotada para o programa de descoberta de novos genes de *S. mansoni* foi a das etiquetas de sequências transcritas (ESTs). Isto porque a metodologia é simples, de custo moderado sendo portanto acessível a pequenos laboratórios como o nosso. Além de permitir trabalhar apenas com sequências expressas e identificar rapidamente uma grande quantidade de genes por meio de comparações feitas com sequências depositadas nos bancos de dados existentes na internet, as ESTs têm muitas outras aplicações como, p. ex., no monitoramento da expressão gênica, na construção de mapas físicos e na caracterização de grandes sequências genômicas (ZWEIGER & SCOTT, 1997; MARRA & COLS., 1998).

O programa de descoberta de novos genes de *S. mansoni* tornou-se realidade a partir do trabalho de tese de doutoramento da Profa. Glória Franco, do Departamento de Bioquímica e Imunologia da UFMG, iniciado em 1992 e concluído em 1996 (FRANCO, Tese de Doutorado, 1996). Nos 2 primeiros anos do programa foram obtidas 607 ESTs de uma biblioteca inicial de cDNA de verme adulto especialmente preparada para este fim. Desse total, 169 eram genes diferentes sendo que 154 deles eram genes ainda não detectados em *S. mansoni* (FRANCO & COLS., 1995).

Utilizando a metodologia de RAP-PCR (RNA Arbitrarily Primed PCR), que permite a identificação de genes diferencialmente expressos, para gerar mini-bibliotecas normalizadas, DIAS NETO & COLS. (1996) contribuíram com mais 185 ESTs para o programa de descoberta de novos genes do parasita. Desse total 70% não apresentaram homologia com qualquer sequência depositada nos bancos de dados e o nível total de redundância foi de apenas 2%.

O trabalho de DIAS NETO & COLS. (1996) serviu de base para o desenvolvimento de uma nova e revolucionária metodologia para produção de ESTs denominada ORESTES (Open REading frame ESTs), inteiramente desenvolvida no Brasil e anunciada recentemente (DIAS NETO & COLS., 2000). Essa metodologia baseia-se na amplificação de transcritos gênicos por RT-PCR usando iniciadores escolhidos arbitrariamente resultando na amplificação preferencial da porção central dos transcritos. Tal abordagem será de grande valor para a identificação de uma série de transcritos homólogos de genes conhecidos, de vez que amplia a chance de se encontrar a janela aberta de leitura do gene, e facilitará a montagem de sequências adjacentes dos transcritos já catalogados.

Vários genes novos descobertos pela estratégia das ESTs já foram ou estão sendo estudados no nosso laboratório com o objetivo de elucidar sua função no parasita. FRANCO & COLS. (1997b) caracterizaram um novo gene de *S. mansoni* pertencente à família de proteínas que se ligam ao elemento Y-box (YBP), que são fatores de transcrição que se ligam em motivos CCAAT presentes em vários promotores de genes eucarióticos. Este gene foi isolado de uma biblioteca de cDNA de verme adulto e identificado como sendo homólogo ao fator de transcrição Y-B1, uma das proteínas pertencentes à família YBP (FRANCO & COLS., 1995). O gene, denominado SMYB1, codifica para uma proteína de 217 aa contendo um domínio de ligação a DNA (o domínio de choque frio) que é 64% idêntico ao domínio correspondente em outras YBPs. A proteína recombinante é capaz de se ligar a oligonucleotídeos que consistem de partes de promotores de genes de *S. mansoni* e de outros organismos que contém CCAAT em diferentes contextos (VALADÃO, Tese de Mestrado, 1998).

FRANCO & COLS., 1995 isolaram também, da mesma biblioteca de cDNA de verme adulto, um gene de *S. mansoni* que foi identificado como sendo homólogo ao gene que codifica a proteína básica conservada de mama (BBC1) no homem. Este tipo de proteína apresenta alta homologia com a proteína ribossomal 60S L13 de rato, podendo



por isso fazer parte da constituição do ribossomo (OLIVEIRA & COLS., 1994). A proteína contém ainda sinais de localização nuclear e seu gene foi isolado pela primeira vez de uma biblioteca de cDNA de carcinoma de mama humano. Foi visto também que a expressão desta proteína é mais abundante em fibroadenomas do que em carcinomas, levantando a hipótese de que este gene pudesse atuar como supressor de tumor (ADAMS & COLS., 1992). BBC é expresso também em uma variedade de tecidos de mamíferos e é bastante conservado em outras espécies tais como levedura, *Drosophila* e plantas, com indícios de que o gene possa estar envolvido no controle do crescimento celular e desenvolvimento (ADAMS & COLS., 1992; HELPS & COLS., 1995). A caracterização do gene de BBC de *S. mansoni* feita por FRANCO & COLS. (1998) mostrou que ele codifica uma proteína de 184 aminoácidos, aproximadamente 55% idêntica a outras proteínas ribossomais L13 eucarióticas. Estudos de expressão feitos por reação em cadeia da polimerase de transcritos reversos (RT-PCR) mostraram que o transcrito é expresso em vários estágios do ciclo de vida do parasita. O gene foi mapeado nos cromossomos 3 e W por supressão da hibridização cromossômica *in situ* (CISS).

O gene que codifica a enzima purina nucleosídeo fosforilase (PNP), um importante componente da via de salvação para biosíntese de purinas em *S. mansoni*, também foi isolado e caracterizado a partir de uma da EST obtida de uma biblioteca de verme adulto (PEREIRA, Dissertação de Mestrado, 1999). A sequência completa do gene da PNP de *S. mansoni* tem 1057 pb e codifica uma proteína de 287 aminoácidos que apresenta 45% de identidade em relação à uma enzima homóloga existente em humanos. A estrutura tridimensional da enzima determinada por modelagem computacional comparativa é semelhante à estrutura das enzimas homólogas encontradas em humano e bovino. A partir dos estudos estruturais e funcionais poderão ser desenhados inibidores específicos para a PNP do parasita com o intuito de se produzir uma nova droga auxiliar no combate à esquistosomose.

A alta frequência com que alguns genes "desconhecidos" foram encontrados em bibliotecas de cDNA de vários estágios sugere que eles possam desempenhar importantes funções na biologia de *S. mansoni* sendo portanto candidatos para o desenvolvimento de novas drogas ou produção de vacinas.

Análises de agrupamento feitas com ESTs de diferentes bibliotecas de cDNA usando o programa ICATTOOLS FRANCO & COLS. (1997a) permitiram a identificação de um grupo composto de 16 ESTs de ambas extremidades do cDNA que correspondem a um gene desconhecido cuja expressão mostrou-se bastante frequente em uma biblioteca de cDNA de verme adulto. Este gene foi denominado AUT1 (transcrito abundante desconhecido 1). MEIRA & COLS. (1998) sequenciaram e fizeram a tradução de todas as 6 possíveis molduras abertas de leitura (ORFs) do cDNA de AUT1. A maior ORF encontrada foi de 1005 pb codificando para uma proteína de 335 aminoácidos. Análises da estrutura primária da provável proteína codificada por AUT1 mostrou um sítio potencial para N-glicosilação, 6 sítios de fosforilação por proteína kinase C, 5 sítios de fosforilação por caseína kinase II e 8 sítios para proteína kinase dependente de AMP cíclico, sugerindo que a proteína pode ser fosforilada no parasita. Porém, todas as buscas feitas em diversos bancos de dados na tentativa de identificar domínios estruturais resultaram infrutíferas. Verificou-se também no mesmo estudo que o gene AUT1 é expresso em todas as fases do ciclo do parasita.

Dentre os genes novos descobertos pela estratégia das ESTs uma classe nos interessa particularmente que é a dos genes envolvidos na regulação da embriogênese. Assim, RABELO & COLS. (1998) selecionaram uma sequência que apresenta alta homologia com um gene de *Drosophila* denominado mago nashi. Em díptera este gene está envolvido no processo de montagem dos componentes do plasma germinativo e sua mutação resulta na esterilidade da prole F1 e também na formação de eixos perpendiculares (BOSWELL, PROUT & STEICHEN, 1991). É concebível que um gene como este possa atuar nas muitas mudanças morfológicas observadas durante o ciclo

de vida do parasita. O cDNA do gene escolhido codifica uma proteína de 146 aminoácidos com 84% de homologia em relação à proteína codificada pelo gene de *Drosophila*. O mago nashi de *S. mansoni* é expresso no ovo, esquistossômulo e verme adulto.

Novos genes com potencial antigênico foram isolados pela estratégia das ASTs (etiquetas de sequências antigênicas). A estratégia consiste em selecionar clones de uma biblioteca de cDNA de verme adulto que codifiquem proteínas antigênicas por meio de uma varredura feita com soro de pacientes infectados com *S. mansoni*. De 30 clones isolados após as varreduras, 29 puderam ser identificados com base na homologia com sequências de *S. mansoni* depositadas nos bancos de dados (28 clones) ou com sequências de outros organismos (1 clone). Apenas 1 dos clones não pode ser identificado (ZOUAIN & COLS., 1998).

Os resultados iniciais obtidos pelo Projeto Genoma de *S. mansoni* encorajaram pesquisadores do Brasil e de outras partes do mundo a se unirem ao projeto. Colaborações importantes foram estabelecidas com a Dra Manami Tanaka (University of Tsukuba, Japão), o Dr. Mohammed Saber (Theodore Bilharz Research Institute, Cairo, Egito) e com o Dr. David Rollinson (The Natural History Museum, Londres, Inglaterra).

Desde 1994 o Projeto Genoma de *S. mansoni* recebe apoio adicional do consórcio OMS/Banco Mundial/Programa das Nações Unidas para o Desenvolvimento (WHO/World Bank/UNDP) através do programa TDR. A coordenação pela OMS da rede genoma de *S. mansoni* está a cargo do Dr. Philip LoVerde (State University of New York at Buffalo, EUA), secretariado pelo Dr. David Johnston (The Natural History Museum, Londres), e cinco centros estão envolvidos no programa de descoberta de genes usando a estratégia de ESTs: Universidade Federal de Minas Gerais e Centro de Pesquisas René Rachou-FIOCRUZ no Brasil, The Natural History Museum na Inglaterra, Theodore Bilharz Research Institute no Egito e State University of New York at Buffalo nos EUA.

Conseqüentemente o programa se expandiu contando agora com vários sequenciadores automáticos de DNA.

A partir de 1996, após uma análise extensiva de ESTs geradas de bibliotecas de vermes adultos, o programa de descoberta de novos genes de *S. mansoni* redirecionou a produção de ESTs para exploração de bibliotecas de cDNA de outros estágios do ciclo de vida do parasita. O objetivo principal era fazer comparações acerca dos padrões de expressão gênica nos vários estágios de desenvolvimento do organismo. Inicialmente, foi feito em nosso laboratório um trabalho de avaliação da qualidade de 7 diferentes bibliotecas de cDNA de 4 estágios do ciclo de vida e do perfil de expressão gênica gerado por cada uma delas. Foram produzidas 1401 ESTs cujos detalhes das análises estão no artigo 2 (FRANCO & COLS., 1997) desta tese. O próximo passo foi dar início à exploração das bibliotecas avaliadas como sendo de boa qualidade para o programa de produção de ESTs. Começamos por duas bibliotecas de cercária cujos detalhes das análises estão no artigo 3 (SANTOS & COLS., 1997) desta tese.

Em seguida foram exploradas duas bibliotecas de ovo (FARIA, Dissertação de Mestrado, 2000). Foram produzidas 1.335 ESTs que em sua grande maioria (72,6%) não tinha similaridade com qualquer sequência depositada nos bancos de dados e por isso podem representar genes específicos de *S. mansoni*. Análises de agrupamento do conjunto das ESTs úteis feitas por ICATOOLS revelou a presença de 1.104 genes distintos, o que ressalta um baixo nível de redundância nessas bibliotecas. Do total de genes, 1.060 eram genes que ainda não haviam sido identificados no parasita.

A iniciativa genoma de *S. mansoni* inclui também um programa de descoberta de genes para *S. japonicum* que vem sendo conduzido por grupos colaborativos na Austrália, China e Filipinas (JOHNSTON, 1997). Bibliotecas de verme adulto, ovo e miracídio de cepas chinesas e filipinas têm sido examinadas (FAN & COLS., 1998). Atualmente 12.675 ESTs de várias fases do ciclo de vida do *S. mansoni* e 1.672 ESTs de *S. japonicum* estão depositadas (dbEST release 061600, 2000). Análises de

agrupamento por homologia de 13.154 sequências de *S. mansoni* e 1.533 de *S. japonicum* revelaram que a existência de 7.269 e 1.017 transcritos diferentes de cada um dos parasitas, respectivamente ([http://www.nhm.ac.uk/hosted\\_sites/schisto/clusters/intro.html](http://www.nhm.ac.uk/hosted_sites/schisto/clusters/intro.html)).

Em conjunto, os números parciais indicam que ESTs para cerca de 15-20% do conteúdo de genes expressos pelo *S. mansoni* já se encontram disponíveis. Alguns genes foram obtidos apenas uma vez enquanto que outros foram isolados várias vezes da mesma e de outras bibliotecas. A taxa de redundância como um todo foi relativamente baixa para ambos organismos (45% para *S. mansoni* e 34% para *S. japonicum*) indicando que o potencial das bibliotecas utilizadas ainda não foi totalmente exaurido e que a geração de ESTs permanece sendo uma abordagem eficiente para a descoberta de novos genes.

A construção de um mapa físico do genoma de *S. mansoni* se inclui também entre os objetivos do projeto. Ele foi iniciado em 1992 pelo grupo da Dra. Manami Tanaka. Uma biblioteca de YAC, com cerca de 3.000 clones e insertos de 358 kb em média, cobrindo 2,6 equivalentes genômicos, foi construída com DNA de cercária (TANAKA & COLS., 1995). Marcadores tais como elementos repetitivos, genes individuais e ESTs foram usados na caracterização da biblioteca. Vários clones de YAC foram ordenados a partir da sua localização em cromossomos metafásicos usando a técnica de hibridização fluorescente *in situ* (FISH), gerando o primeiro mapa físico de baixa resolução de *S. mansoni*. A biblioteca está disponível para outros usuários na forma de filtros de alta densidade. O grupo da Dra. Tanaka também construiu uma biblioteca genômica em cosmídeo e fez a montagem de contíguos de YAC e cosmídeos cromossomo-específicos para os cromossomos 3, Z e W. A visualização destes contíguos por FISH deu origem a um mapa físico de segunda geração.

O progresso significativo do mapeamento genético de *S. mansoni* foi possível graças ao desenvolvimento e refinamento de novas técnicas. HIRAI & LoVERDE (1995)

publicaram protocolos para cromossomos metafásicos e FISH para *S. mansoni*. Outra técnica, a do iniciador *in situ* (PRINS), que permite a localização de sequências específicas através da reação em cadeia da polimerase (PCR) tem sido adaptada para cromossomos de *S. mansoni* (HIRAI, não publicado). Cromossomos podem ser isolados através de microdissecção e podem então ser utilizados na construção de bibliotecas cromossomo-específicas (HIRAI, não publicado).

Uma modificação da técnica FISH, denominada hibridização cromossomal *in situ* por supressão – CISS, permite usar plasmídeos, cosmídeos e YACs (cromossomo artificial de levedura) como sondas (TANAKA & COLS., 1997). A utilização de CISS tem contribuído para a construção de mapas físicos de alta resolução. Recentemente foi anunciada a construção de uma biblioteca genômica de *S. mansoni* em cromossomo artificial de bactéria (BAC) contendo insertos de 100 kb em média, representando uma cobertura de cerca de 8 vezes o genoma do parasita (Le PASLIER & COLS., 2000). Tais bibliotecas são valiosas não apenas como fonte de insertos para serem utilizados na montagem de sequências adjacentes e na preparação de mapas físicos de alta resolução, como também para a geração de sequências genômicas usando a estratégia de sequenciamento terminal de insertos de BAC.

Assim como em outros Projetos Genoma, a construção de um banco de dados de acordo com o modelo do banco de dados de *C. elegans* (AceDB) foi também implementada – o SchistoDB, cuja base está no The Natural History Museum em Londres, Inglaterra e tem como curador o Dr. David Jonhston. Este banco de dados serve como um repositório de toda a informação gerada pelo sequenciamento de cDNAs, análises de agrupamentos de ESTs, mapeamento cromossomal de clones de YAC e cosmídeos por FISH, tão bem quanto informações sobre grupos e laboratórios da “*Schistosoma* Genome Network” e metodologias experimentais ([http://www.nhm.ac.uk/hosted\\_sites/schisto/informatics/SchistoDB\\_info.html](http://www.nhm.ac.uk/hosted_sites/schisto/informatics/SchistoDB_info.html)). Recentemente foi criado pelo The Institute for Genomic Research (TIGR), EUA, o *Schistosoma mansoni* Gene Index

(<http://www.tigr.org/tdb/smgi/>) que integra os dados gerados pelo programa internacional de geração de ESTs do parasita. O objetivo desse banco de dados é apresentar uma visão não redundante de todos os genes de *S. mansoni* e dados sobre padrões de expressão, funções celulares e relações evolutivas.

### **1.8. Projetos Genoma de Outros Parasitas**

Em 1994 o programa TDR da OMS destinou cerca de U\$ 800.000 para que se desse início a cinco projetos internacionais para estudo do genoma dos parasitas *S. mansoni*, filária, *Trypanosoma cruzi*, *Leishmania* e tripanosomas africanos. Estes projetos deveriam ser conduzidos como colaborações entre laboratórios de países desenvolvidos e em desenvolvimento. Os principais objetivos desta iniciativa da OMS eram: o desenvolvimento de mapas de baixa resolução, a implementação de tecnologias novas e custo-efetivas, o desenvolvimento de bancos de dados públicos, assim como programas de gerenciamento de dados, o desenvolvimento de sistemas para análise genética de parasitas e por fim, o envolvimento e treinamento de cientistas de países em desenvolvimento que são endêmicos para estas parasitoses (WHO/TDR REPORT, 1998c).

Durante os primeiros anos de atividade do Comitê de Genoma de Parasita, cinco redes genoma foram implementadas e consolidadas. São elas:

Rede Genômica de *Schistosoma*:

[http://www.nhm.ac.uk/hosted\\_sites/schisto/](http://www.nhm.ac.uk/hosted_sites/schisto/)

Rede Genômica de *Leishmania*:

<http://www.dbbm.fiocruz.br/genome/LGN/pagesidx.html>

Rede Genômica de *Trypanosoma cruzi*:

<http://www.dbbm.fiocruz.br/genome/tcruzi/tcruzi.html>

Rede Genômica de Tripanosomas africanos:

<http://parsun1.path.cam.ac.uk/>

Rede Genômica de Filária:

<http://helios.bto.ed.ac.uk/mbx/fgn/filgen.html>

Encontros de planejamento de participantes destas redes têm sido feitos anualmente para selecionar cepas para pesquisa e traçar linhas mestras no sentido de cumprir os objetivos do Programa. Todas as redes têm cumprido com sucesso o objetivo de construir bibliotecas de cDNA estágio-específicas e bibliotecas de grandes fragmentos genômicos (YAC, BAC, cosmídeos) e o processo de mapeamento de cromossomos está começando (UNDP/WORLD BANK/WHO-TDR, 1999, <http://www.who.int/tdr/workplan/genome99.html>). Milhares de ESTs (dbEST release 061600, 2000) e marcadores de cromossomos também já foram obtidos. Uma expansão significativa destas atividades está sendo aguardada para os próximos dois anos, na esperança de que, com a introdução de sistemas robotizados de alta precisão, os genomas dos parasitas representativos sejam sequenciados nos próximos 5-10 anos. O conjunto destes dados constituirá uma valiosa base para as iniciativas de estudos genômicos funcionais.

O rápido progresso experimentado em muitos dos projetos genoma de parasitas levou o Comitê Genoma de Parasita a considerar maneiras pelas quais esta informação pode ser mais bem utilizada. Nos próximos anos a OMS dará uma ênfase maior a projetos de investigação em genômica funcional. Tais projetos estarão voltados para análise dos dados de sequências de DNA já disponíveis para identificação de moléculas importantes para o desenvolvimento dos parasitas e sobrevivência deles nos seus respectivos hospedeiros.

### **1.9. Projeto Genoma de *Schistosoma mansoni*: Estudos Pós-Genômicos**

O programa de descoberta de novos genes de *S. mansoni* encontra-se em um estágio bastante avançado fazendo-se necessário agora dar início aos estudos de



genômica funcional. Nesse sentido, as atenções estão voltadas para a produção de “DNA arrays”, nocaute e complementação em levedura e proteoma como diretrizes futuras do projeto (FRANCO & COLS., 2000). Uma análise mais aprofundada do perfil transcricional do parasita em diferentes condições poderá ser obtida através do uso dos clones de cDNAs das ESTs do programa em “DNA arrays” de tal forma que possam servir como alvos para sondas de mRNA ou cDNA de várias fontes. Desse modo será possível verificar o nível de transcrição de cada gene durante o desenvolvimento normal, que diferenças existem com relação à expressão gênica nas várias fases do ciclo de vida, entre vermes machos e fêmeas, entre cepas sensíveis ou resistentes a drogas, entre cepas crescidas em diferentes hospedeiros ou submetidas a diferentes tipos de estresses. Tais estudos poderão ser complementados e enriquecidos através do emprego das metodologias de proteoma para estudo do conjunto de proteínas produzidas pelo verme.

## **PARTE I:**

**Descoberta de novos genes de *Schistosoma mansoni* utilizando a estratégia das Etiquetas de Sequências Transcritas (ESTs)**

## I.1. JUSTIFICATIVA

Esta primeira etapa do trabalho se justifica em vista da importância que o parasita *Schistosoma mansoni* tem para a saúde pública no Brasil e no mundo como o agente etiológico da esquistossomose. A descoberta de novos genes de *S. mansoni* pela estratégia das ESTs, a partir de bibliotecas de cDNA de diversos estágios do ciclo de vida do parasita, pode contribuir para uma maior compreensão da interação parasita/hospedeiro assim como para revelar novos alvos para o desenvolvimento de medicamentos, vacinas e estratégias de controle epidemiológico.

## I.2. OBJETIVOS

### **Objetivo geral:**

Descobrir novos genes de *Schistosoma mansoni* e analisar o perfil de transcrição gênica do parasita a partir de clones isolados de bibliotecas de cDNA de diferentes estágios do seu ciclo de vida utilizando-se para isto a estratégia das Etiquetas de Sequências Transcritas (ESTs).

### **Objetivos específicos:**

1. Excisar em massa oito bibliotecas de transcritos de vários estágios do ciclo de vida de *S. mansoni*, que haviam sido construídas utilizando-se o sistema  $\lambda$ ZapII, para obtenção de moldes de cDNA dupla fita (dsDNA) no plasmídeo pBluescript SK capazes de serem sequenciados a partir de ambas as extremidades e identificados gerando ESTs;
2. Avaliar as bibliotecas de cDNA excisadas quanto à qualidade dos transcritos e o seu potencial de aproveitamento no programa de descobertas de novos genes do Projeto Genoma de *S. mansoni*;
3. Realizar um estudo comparativo do perfil de expressão gênica nos diferentes estágios do desenvolvimento de *S. mansoni* a partir das ESTs geradas de cada biblioteca de cDNA excisada;

4. Identificar, dentre as ESTs geradas, genes que possam servir como novos alvos para o estudo de aspectos ainda pouco conhecidos da biologia de *S. mansoni* e o desenvolvimento de novos medicamentos e vacinas;

### **I.3. ARTIGO 1**

**Update of the gene discovery program in *Schistosoma mansoni* with the expressed sequence tag approach.**

# Update of the Gene Discovery Program in *Schistosoma mansoni* with the Expressed Sequence Tag Approach

Élida ML Rabelo, Glória R Franco\*, Vasco AC Azevedo\*\*, Heloisa B Pena\*, Túlio M Santos\*, Wendell SF Meira\*, Neuza A Rodrigues\*, José Miguel Ortega\*, Sérgio DJ Pena\*/+

Departamento de Parasitologia \*Departamento de Bioquímica e Imunologia \*\*Departamento de Biologia Geral, ICB, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, MG, Brasil

*Continuing the Schistosoma mansoni Genome Project 363 new templates were sequenced generating 205 more ESTs corresponding to 91 genes. Seventy four of these genes (81%) had not previously been described in S. mansoni. Among the newly discovered genes there are several of significant biological interest such as synaptophysin, NIFs-like and rho-GDP dissociation inhibitor.*

Key words: *Schistosoma mansoni* - genome project - expressed sequence tags

The study of the *Schistosoma mansoni* genome has a high priority in the effort to understand the multiple facets of this complex parasite. Accordingly, a *S. mansoni* genome project was started in 1992 as a Brazilian initiative. From the original cDNA library used, 607 ESTs were generated (Franco et al. 1995a,b) leading to the identification of 154 new genes. This increased considerably the number of known genes in *S. mansoni*.

The *S. mansoni* genome project was adopted by the World Health Organization (WHO) and has been partially funded by this institution in a collaborative international project with the aim of undertaking a co-ordinated gene discovery program in *S. mansoni*. We wish to report the sequencing of 363 further templates by our laboratory, with identification of 91 genes, 74 of which have not been previously described in *S. mansoni*.

## MATERIALS AND METHODS

Plasmidial DNA preparation, sequencing and analysis of the sequences were done essentially as described by Franco et al. (1995a).

## RESULTS

Tables I and II summarise the data found in the 363 templates sequenced, which produced 205 useful ESTs derived from 185 clones. Some clones were sequenced in both directions using the forward and reverse primers and thus producing two ESTs for the same gene. That is the reason why there are 205

ESTs for 185 clones. All sequences were submitted to homology searches in DNA and protein databanks showing that the 205 ESTs corresponded to 91 different genes, 81% of which had not been previously described in the parasite. The sequences were grouped according to the following criteria: sequences presenting homology to previous identified *S. mansoni* genes (Tables II and III), sequences identified by homology with other organisms (Tables II and IV) and sequences with no homology with any gene deposited in database banks (Table II). Sequences showing low homology with other organisms were identified as partial matches (Tables II and V).

TABLE I

Information about sequencing of the *Schistosoma mansoni* cDNA library

Number of sequenced templates	363
Number of ESTs	205
Number of clones	185
Number of different genes	91
Average EST size	372nt
Average polyA tail length	23nt

TABLE II

EST categories of Sm cDNA library

	No. of clones	(%)
Putatively identified		
Sm match	47	22.7
Non Sm match	62	35.7
Not identified		
Non Sm partial match	4	2.2
Non database match	57	31.3
Mitochondrial	0	0
rRNA	12	6.5
Vector without insert	3	1.6
<b>Total</b>	<b>185</b>	<b>100.0</b>

This investigation received financial support from the following institutions: UNDP/WORLD BANK/WHO Special Programme for Research in Tropical Diseases (TDR) (ID no. 940325, 940751); FAPEMIG Process no. CBS 1190/95; CNPq and PADCT.

+Corresponding author. Fax: +55-31-441.5409

Received 16 April 1997

Accepted 30 June 1997

TABLE III  
 Characterization of ESTs: database match of ESTs to Sm genes

EST name	dbEST Acces	Homology with (accession)	Identity (%)	Length (nt)	Number of clones	Program	Score	Probability
SMPBE13F	W06737	alpha-tubulin (GB:S98950)	98.5	274	3	FASTA	1068	
SMPBD26F	W06720	Calreticulin (GB:SCMCALRET)	98.3	460	1	BLASTN	2230	5.7e-183
SMPBE11F	W06736	Fructose 1,6 biphosphate aldolase (GP:SCMALDO_1)	98.0	432	7	FASTA	1156	
SMPBG05F	W06784	GAPDH (GB:SCMGAPDH)	99.6	252	11	FASTA	989	
SMPBE62Fwazzu	W06751	GST (GB:SCMANT28K)	98.6	291	9	FASTA	1148	
SMPBF02F	W06765	HSP 86 (GB:SCMHSP86)	99.0	387	1	FASTA	1532	
SMPBG92F	W06807	Myosin heavy chain (GB:SCMMYH)	99.6	262	2	BLASTN	1301	1.2e-100
SMPBF01F	W06764	ER -luminal cysteine protease ER 60 gene (GB:SMF1PSPCX)	97.9	380	1	BLASTN	1830	9.9e-145
SMPBE24F	W06741	Sm750 gene (GB:SCMSM750)	96.2	240	2	FASTA	902	
SMPBE51F	W06747	Triose phosphate isomerase (GB:SCMSGTPI05)	100.0	256	2	BLASTN	1280	1.6e-101
SMPBE92F	W06761	Tropomyosin (GB:SCMTROPO)	99.7	369	2	BLASTN	1836	1.2e-145
SMPBE03F	W06734	Actin (GB: M80334)	96.6	304	14	BLASTN	1138	3.4e-87
SMPBF30F	W06772	Fibrillin 2 mRNA	89.2	337	1	BLASTN	673	1.7e-99
SMPBG29F	W06788	Elongation factor 1-alpha (EMB:SMF1ALPH)	97.5	405	3	BLASTN	1902	5.5e-151
SMPBD40R	W06733	Y-box binding protein (GB: U398831)	97.8	358	10	BLASTN	1536	4.9e-125
SMPBE59F	W06749	BBC-1 (GB:U57003)	98.9	373	2	BLASTN	1400	6.8e-147
SMPBG90F	W06805	Tropomyosin (GB:SCMTPM)	100.0	200	1	BLASTN	1000	9.0e-76

TABLE IV  
Characterization of ESTs: database match of EST to non-Sm genes

EST name	dbEST	Homology with (accession)	Similarity (%)	Identity (%)	Length (aa)	Number of clones	Program	Score	Probability
SMPBF59F	W06782	Aldose reductase(PDB:1DLA)	72.5	53.8	80	1	BLASTX	219	1.4e-23
SMPBF07F	W06768	Asp-tRNA synthetase (SP:SYD_CAEEL)	85.8	74.2	155	1	BLASTX	625	1.0e-80
SMPBE38F	W06746	<i>C. elegans clone</i> C16C10.10 (GP:CEC16C10_4)	75.0	57.4	68	1	BLASTX	210	2.3e-22
SMPBG74F	W06795	Dihydroliipoamide Acetyltransferase (GP:RATPDCE2_1)	83.3	72.9	48	2	BLASTX	192	5.1e-39
SMPBD30R	W06722	DNAJ homolog (PIR:S42031)	79.2	69.8	53	2	BLASTX	194	4.1e-38
SMPBE17F	W06739	Enolase (SP:ENO_SCHJA)	95.9	87.6	73	2	BLASTX	348	1.6e-41
SMPBE97R	W06821	Glutamine Synthetase (SP:GLNA_HUMAN)	78.6	66.0	103	1	BLASTX	413	5.0e-51
SMPBG67F	W06794	H+-transporting ATP synthase alpha-chain (PIR:S14516)	80.2	64.8	91	1	BLASTX	303	3.3e-37
SMPBD33R	W06725	Homo sapiens 9G8 splicing factor (GP:HUM9G8SF_1)	70.7	63.8	58	2	BLASTX	184	4.9e-37
SMPBE61F	W06750	Human Alu subfamily (SP:ALU7_HUMAN)	76.7	67.4	43	1	BLASTX	154	1.2e-14
SMPBF15F	W06771	Hypothetical protein 5 Xanthobacter sp (PIR:S47055)	60.0	44.0	125	1	BLASTX	271	2.6e-31
SMPBE28F	W06744	Lactate dehydrogenase (GP:MUSLDHB_1)	75.4	45.9	61	3	BLASTX	158	1.8e-15
SMPBE82F	W06757	NIFS-like 54.5KD protein (SP:NFS1_YEAST)	65.6	53.6	125	1	BLASTX	328	1.2e-38
SMPBE27F	W06743	Phosphoglycerate mutase (SP:PMG1_ECOLI)	1 77.6	67.2	67	1	BLASTX	239	1.3e-26
SMPBD34R	W06727	Polyadenylate binding protein (DBJ:HUMPOLYABP)	76.3	74.8	127	1	BLASTX	400	2.3e-23
SMPBD23F	W06714	Purine nucleoside phosphorylase (PDB:1IULA)	35.5	55.5	45	2	BLASTX	76	2.9e-15
SMPBF65F	W06783	Ribosomal protein L5 (SP:RL5A_XENLA)	73.4	52.1	94	3	BLASTX	252	5.0e-28
SMPBH15F	W06814	Ribosomal protein S4 (SP:RS4_HUMAN)	81.1	72.6	107	1	BLASTX	441	3.5e-56
SMPBD32R	W06723	rho-GDP dissociation inhibitor (GP:MUSGDPDI_1)	75.0	52.1	48	1	BLASTX	138	4.4e-19
SMPBE16R	W06818	Synaptophysin(PIR:A60548)	60.5	39.5	43	1	BLASTX	87	9.3e-08
SMPBE57F	W06748	Tubulin beta chain (PIR:S18457)	98.1	94.4	107	1	BLASTX	543	6.8e-70
SMPBF13F	W06770	Polyubiquitin (X60390)	99.0	99.0	105	2	BLASTX	517	6.1e-67
SMPBF52R	W06824	Vacuolar ATP synthase - subunit B(SP:VAT_DROME)	97.0	94.0	68	1	BLASTX	335	1.4e-39
SMPBE65R	W06819	Yeast hypothetical 103.7KD Protein (SP:YBM7_YEAST)	74.1	51.9	27	1	BLASTX	71	4.4e-11



TABLE V  
 Characterization of ESTs: database partial match of EST to non-Sm genes

EST No	dbEST	Homology with (accession)	Similarity (%)	Identity (%)	Length aa	Length nt	No. of clones	Program	Score	Probability
SMPBF32F	W06774	14 ORF YJR83.9 gene product [ <i>S. cerevisiae</i> ](GP:IX87611)	65.4	47.3	26		1	BLASTX	126	3.2e-12
SMPBG91F	W06806	RNA binding protein (PIR:S53050)	45.6	43.4	46		2	BLASTX	95	3.1e-08
SMPBH26F	W06816	<i>D. discoideum</i> plasmid Ddp2 trans-acting factor gene (GB:DDIDDP2)		68.1		62	1	BLASTN	194	4.7e-6

## DISCUSSION

Tables IV and V show that some of the genes with homology either to *S. mansoni* or with other organism were sequenced more than twice. This is the case of GAPDH (11 times selected from the library), GST (9 X), fructose 1,6 biphosphate aldolase (7 X), actin (14 X) and Y-box binding protein (10 X) showing a certain degree of redundance of the library as had been previously reported (Franco et al. 1995a). Adams et al. (1995) defined some parameters to define whether a library has sufficient quality for the purpose of generating ESTs. These parameters include the proportions of: vectors without inserts, contaminant of the library with others cDNAs (host or bacteria), presence of mitochondrial DNA or rRNA, number of new genes, number of genes matching other organism genes and number of genes with homology to the original organism (*S. mansoni* in our case). Although some redundancy was found, the library is still considered of very good quality, especially when we take in account the fact that 81% of the genes identified were new. Thus, we believe that it is worth pursuing further work with this library for the generation of new ESTs.

Among the new genes identified by homology with other organisms several stand out for having significant biological interest. Thus, an EST with homology to synaptophysin gene was found. In mammals, synaptophysin is one of the major integral membrane proteins of synaptic vesicles (McMahon et al. 1996) it is speculated that synaptophysin may function as a gap junction-like pore or channel (Calakos & Scheller 1994). Another homology that called our attention is the one with a NIFS-like protein from yeast. The NIFS-like protein is supposed to be involved in both tRNA-processing and mitochondrial metabolism (Kolman & Soll 1993), two interesting targets for design of new drugs.

The EST approach is also contributing for adding new members into gene families. Members of the family of GDP dissociation inhibitors (GDI) for the ras-related rho-subtype proteins appear to take part in the regulation of a number of biological processes, including cell growth and differentiation. We have identified an EST with a high similarity (75%) with a murine D4 cDNA, a new member of the GDP-GDI family (Adra et al. 1993).

From these initial identifications, one cannot ascertain to the *S. mansoni* genes the same functions of those founds to the homologue genes in other organisms. However, these identifications open new avenues to further characterise these genes and through functional studies obtain a correlation between gene function and homology with the most diverse organisms.

Through the sequencing of this adult cDNA library, a great number of new genes were identified in *S. mansoni*, showing the high efficiency of the EST approach. However, this parasite presents a complex life cycle with enormous changes in its morphology. Obviously, one would expect that such great changes are accompanied by changes at the gene expression level. Furthermore, if one considers the acquisition of information about the worm gene expression in the perspective of designing new drugs and/or vaccines, the young stages cannot be overlooked. Actually, the schistosomula stage is recognized as the main target to the host immune system attack (Smithers & Terry 1965). With the foregoing in mind it is our future aim to study the expression pattern of the different life stages of *S. mansoni* by sequencing cDNA libraries for the distinct stages.

#### ACKNOWLEDGEMENTS

To Kátia Barroso for carrying out automated DNA sequencing.

#### REFERENCES

- Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, Kirkness EF, Weinstock KG, Gocayne JD, White O, Sutton G, Blake JA, Brandon RC, Chiu MW, Clayton RA, Cline RT, Cotton MD, Earle-Hughes J, Fine LD, FitzGerald LM, FitzHugh WM, Fritchman JL, Geoghagen NSM, Glodek A, Gnehm CL, Hanna MC, Hedblom E, Hinkle Jr. PS, Kelley JC, Klimek KM, Kelley JC, Liu LI, Marmaros SM, Merrick JM, Moreno-Palanques RF, McDonald LA, Nguyen DT, Pellegrino SM, Phillips CA, Ryder SE, Scott JL, Saudek DM, Shirley R, Small KV, Spriggs TA, Utterback TR, Weidman JF, Li Y, Barthlow R, Bednarik DP, Cao L, Cepeda MA, Coleman TA, Collins EJ, Dimke D, Feng P, Ferrie E, Fischer C, Hastings GA, He WW, Hu JS, Huddleston KA, Greene JM, Gruber J, Hudson P, Kim A, Kozak DL, Kunsch C, Ji H, Li H, Meissner PS, Olsen H, Raymond L, Wei YF, Wing J, Xu C, Yu GL, Ruben SM, Dillon PJ, Fannon MR, Rosen CA, Haseltine WA, Fields C, Fraser CM, Venter CJ 1995. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* 377 (Suppl). 3-174.
- Adra CN, Leonard D, Wirth LJ, Cerione RA, Lim B 1993. Identification of a novel protein with GDP dissociation inhibitor activity for ras-like proteins CDC4Hs and rac I. *Genes Chromosom Cancer* 8: 253-261.
- Calakos N, Scheller RH 1994. Vesicle-associated membrane protein and synaptophysin are associated on the synaptic vesicle. *J Biol Chem* 269: 24534-24537.
- Franco GR, Adams MD, Soares MB, Simpson AJG, Venter JC, Pena SDJ 1995a. Identification of new *Schistosoma mansoni* genes by EST strategy using a directional cDNA library. *Gene* 152: 141-147.
- Franco GR, Simpson AJG, Pena SDJ 1995b. Sequencing and identification of expressed *Schistosoma mansoni* genes by random selection of cDNA clones from a direction library. *Mem Inst Oswaldo Cruz* 90: 215-216.
- Kolman C, Soll D 1993. SPL1-1, a *Saccharomyces cerevisiae* mutation affecting tRNA splicing. *J Bacteriol* 175: 1433-1442.
- McMahon HT, Bolshakov VY, Janz R, Hammer RE, Siegelbaum SA, Sudhof TC 1996. Synaptophysin, a major synaptic vesicle protein, is not essential for neurotransmitter release. *Proc Natl Acad Sci USA* 93: 4760-4764.
- Smithers S, Terry RJ 1965. The infection of laboratory hosts with cercarial of *S. mansoni* and the recovery of adult worms. *Parasitology* 55: 695-700.

## **I.4. ARTIGO 2**

**Evaluation of cDNA libraries from different developmental stages of *Schistosoma mansoni* for production of Expressed Sequence Tags (ESTs)**

## Evaluation of cDNA Libraries from Different Developmental Stages of *Schistosoma mansoni* for Production of Expressed Sequence Tags (ESTs)

Glória R. FRANCO,<sup>1</sup> Élide M. L. RABELO,<sup>2</sup> Vasco AZEVEDO,<sup>3</sup> Heloisa B. PENA,<sup>1</sup> J. Miguel ORTEGA,<sup>1</sup> Túlio M. SANTOS,<sup>1</sup> Wendell S. F. MEIRA,<sup>1</sup> Neuza A. RODRIGUES,<sup>1</sup> Carlos M. M. DIAS,<sup>2</sup> Richard HARROP,<sup>5</sup> Alan WILSON,<sup>5</sup> Mohamed SABER,<sup>6</sup> Hannan ABDEL-HAMID,<sup>6</sup> Michelyne S. C. FARIA,<sup>7</sup> Maria Elizabeth B. MARGUTTI,<sup>4</sup> Juçara C. PARRA,<sup>7</sup> and Sérgio D. J. PENA<sup>1,\*</sup>

*Departamento de Bioquímica e Imunologia,<sup>1</sup> Departamento de Parasitologia,<sup>2</sup> Departamento de Biologia Geral<sup>3</sup> and Departamento de Microbiologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil 31270-010,<sup>4</sup> Department of Biology, University of York, York, YO15DD, UK,<sup>5</sup> Theodore Bilharz Research Institute, Cairo, 12411, Egypt<sup>6</sup> and Centro de Pesquisas René Rachou, Belo Horizonte, Brazil 30190-002<sup>7</sup>*

(Received 7 April 1997)

### Abstract

A comparative study of the gene expression profile in different developmental stages of *Schistosoma mansoni* has been initiated based on the expressed sequence tag (EST) approach. A total of 1401 ESTs were generated from seven different cDNA libraries constructed from four distinct stages of the parasite life cycle. The libraries were first evaluated for their quality for a large-scale cDNA sequencing program. Most of them were shown to have less than 20% useless clones and more than 50% new genes. The redundancy of each library was also analyzed, showing that one adult worm cDNA library was composed of a small number of highly frequent genes. When comparing ESTs from distinct libraries, we could detect that most genes were present only in a single library, but others were expressed in more than one developmental stage and may represent housekeeping genes in the parasite. When considering only once the genes present in more than one library, a total of 466 unique genes were obtained, corresponding to 427 new *S. mansoni* genes. From the total of unique genes, 20.2% were identified based on homology with genes from other organisms, 8.3% matched *S. mansoni* characterized genes and 71.5% represent unknown genes.

**Key words:** Key Words: *Schistosoma mansoni*; developmental stages; cDNA sequencing analysis; expressed sequence tags

### 1. Introduction

*Schistosoma mansoni* (*Sm*) is a digenetic trematode worm responsible for schistosomiasis, a parasitic disease that is estimated to affect at least 300 million people in tropical and subtropical areas of the world (WHO, 1985).

Communicated by Kenichi Matsubara

\* To whom correspondence should be addressed. Departamento de Bioquímica e Imunologia, ICB/UFMG. Av. Antônio Carlos, 6627, Belo Horizonte, MG 31270-010, Brazil. Tel. +5531-227-3496, Fax. +5531-227-3792, E-mail: spena@dcc.ufmg.br

† EST sequences were deposited in dbEST and GenBank with the following accession numbers: Adult 1 library (T14340→T14651; T18616→T18626; T24126→T24150; W06712→W06824); Adult 2 library (AA185747→AA185837); Adult 3 library (AA218448→AA218524); Adult 4 library (AA125663→AA169943); Egg library (AA140558→AA140638); Cercariae library (AA143808→AA143896); Lung stage schistosomula library (AA125668→AA125734).

Despite intense efforts dedicated to eradicating schistosomiasis through sanitary measures, suppression of the intermediate host and drug treatment, the prevalence of the disease has not decreased. No vaccine is yet available and control of the disease is primarily by chemotherapy. However, reinfection of patients is common and we need new approaches to treatment and prevention, since *S. mansoni* is becoming increasingly resistant to drug therapy. It is hoped that detailed information about the genome of *S. mansoni* might uncover key gene products that may constitute new targets for drug and vaccine development.

Accordingly, in 1992 we started a systematic gene discovery program study in *S. mansoni* using the strategy of partial sequencing of cDNA ends to generate expressed sequence tags (EST).<sup>1</sup> Initially, we utilized an adult worm cDNA library, from which 607 ESTs were

obtained, corresponding to 169 different genes, 15 previously known in *S. mansoni* and 154 new genes.<sup>2</sup> This increased considerably the number of genes identified in the parasite. However, we felt that studying only adult worms was insufficient. *S. mansoni* has a complex life cycle with several morphologically very diverse stages (ova, miracidia, cercariae, schistosomula and adult worms), during which different sets of genes are expressed. Obviously, if one considers the acquisition of information about the worm gene expression in the perspective of designing new drugs and vaccines, the young stages can not be overlooked. Actually, the schistosomula stage is increasingly recognized as one of the main targets for the host immune system.<sup>3</sup>

With this in mind, we planned to extend our EST program to the other life stages of *S. mansoni*. For that, stage-specific cDNA libraries were needed, some of which, unfortunately, are very difficult to construct because of difficulties in obtaining the necessary amounts of pure mRNA. Thus, before embarking on large-scale studies, we decided to evaluate the libraries that were already in existence, comparing them with our original adult worm library. We here report our results with seven different cDNA libraries constructed from four distinct stages of the parasite life cycle, from which a total of 1401 ESTs were generated, totaling 466 different genes, 427 of which are newly describe in *S. mansoni*. From the total of identified genes, we can start to outline a pattern of gene expression, with some genes expressed in a stage-specific manner and others, housekeeping ones, in all developmental stages.

## 2. Methodology

### 2.1. Construction of cDNA libraries and sequencing

The following seven cDNA libraries were used in this study: four libraries (Adult 1-4) from adult worms and one library each from ova (Egg), cercariae and lung stage schistosomula (Lung stage). The construction of the Adult 1 cDNA library, plasmidial DNA preparation and sequencing of clones from this library have been previously described.<sup>2</sup> The other six libraries were constructed in  $\lambda$ ZapII (Stratagene), according to the manufacturer's instructions. Total RNA was isolated from distinct *S. mansoni* developmental stages by the guanidinium thiocyanate-phenol-chloroform method<sup>4</sup> and poly(A)+ RNA was obtained by chromatography on an oligo (dT) column.<sup>5</sup> Double-stranded cDNA was cloned into *EcoRI/XhoI* restriction sites of  $\lambda$ ZapII. pBluescript SK+ phagemids were obtained by "en masse" *in vivo* excision of  $\lambda$ Zap clones,<sup>6</sup> by co-infecting *Escherichia coli* XL-1 Blue cells with the ExAssist helper phage (Stratagene). The excised phagemids were used to infect *E. coli* SOLR<sup>TM</sup> cells (Stratagene) for production of

double-stranded DNA (dsDNA) templates. Transformants were plated onto LB agar containing ampicillin, X-gal and isopropyl- $\beta$ -D(-)-thiogalactopyranoside (IPTG). White colonies were selected and grown for 16 hr in 3 ml of Luria broth (LB) supplemented with ampicillin. Aliquots of the cultures (200  $\mu$ l) were mixed with the same volume of 30% glycerol in LB and frozen at  $-70^{\circ}\text{C}$  in 96-well plates. The rest of the cultures were used for plasmidial DNA preparation using the Wizard Plus Mini Prep DNA Purification System (Promega). dsDNA was sequenced by dideoxy chain-termination sequencing<sup>7</sup> using the Thermo-Sequenase Cycle Sequencing kit (Amersham) and M13 Reverse or M13-40 fluorescent-labeled primers (Pharmacia). Single-pass runs of the sequencing reactions were performed on an A.L.F. automated DNA sequencer (Pharmacia).

### 2.2. Data analysis

Sequences were manually edited to eliminate vector regions, poly(A) tails and lower quality data at the end of the sequence. ESTs containing less than 150 bp and more than 4% ambiguity were rejected. ESTs were compared to DNA and protein sequences deposited in non-redundant databases using the Basic Local Alignment Search Tool (BLAST) programs<sup>8</sup> at the National Center for Biotechnology Information (NCBI). Alignments scoring more than 200 for BLASTN and 100 for BLASTX were selected and after meticulous visual inspection on the biological significance of the alignment, ESTs were named as putative identification for the gene. ESTs with no significant database matches or showing only partial homology with database sequences were grouped as non-identified genes.

### 2.3. Clustering analysis:

Sequences sharing local similarities were clustered with the ICATOOLS set of programs<sup>9</sup> (freely available at ftp.ebi.ac.uk). Initially, each library was independently analyzed. The module ICAass was used to create an index of clustered sequences (threshold and ktup set to 25 and 8, respectively). One singular sequence was added to the cluster with ICAass and used to run the module ICAtool, under the same threshold and ktup settings. This was followed by the run of ICAtool with all sequences in the library. ICAprint was used to generate the output file, that was manually inspected since some clones had been sequenced in both orientations and/or led to the same identification when submitted to homology search. A second round of analysis was conducted with all libraries concomitantly in order to join the clusters that had been previously formed, but for this purpose only ICAass followed by ICAtool with a singular sequence was executed.

**Table 1.** Information about the sequencing of different *S. mansoni* cDNA libraries.

	<i>S. mansoni</i> cDNA libraries							
	Egg	Cercariae	Lung stage	Adult 1	Adult 2	Adult 3	Adult4	Total
Number of ESTs	106	110	107	812	94	101	71	1401
Number of sequenced clones	107	107	107	617	94	101	71	1204
Number of usable ESTs <sup>a)</sup>	80	98	67	657	91	78	52	1123
Number of usable clones <sup>a)</sup>	80	98	67	504	91	78	52	970

<sup>a)</sup> ESTs/clones analyzed by ICATOOLS. These numbers correspond to the total number of ESTs/clones after removing sequences of vector, mitochondrial DNA, rRNA and contaminating sequences from other organisms.

### 3. Results and Discussion

#### 3.1. Quality control of the cDNA libraries

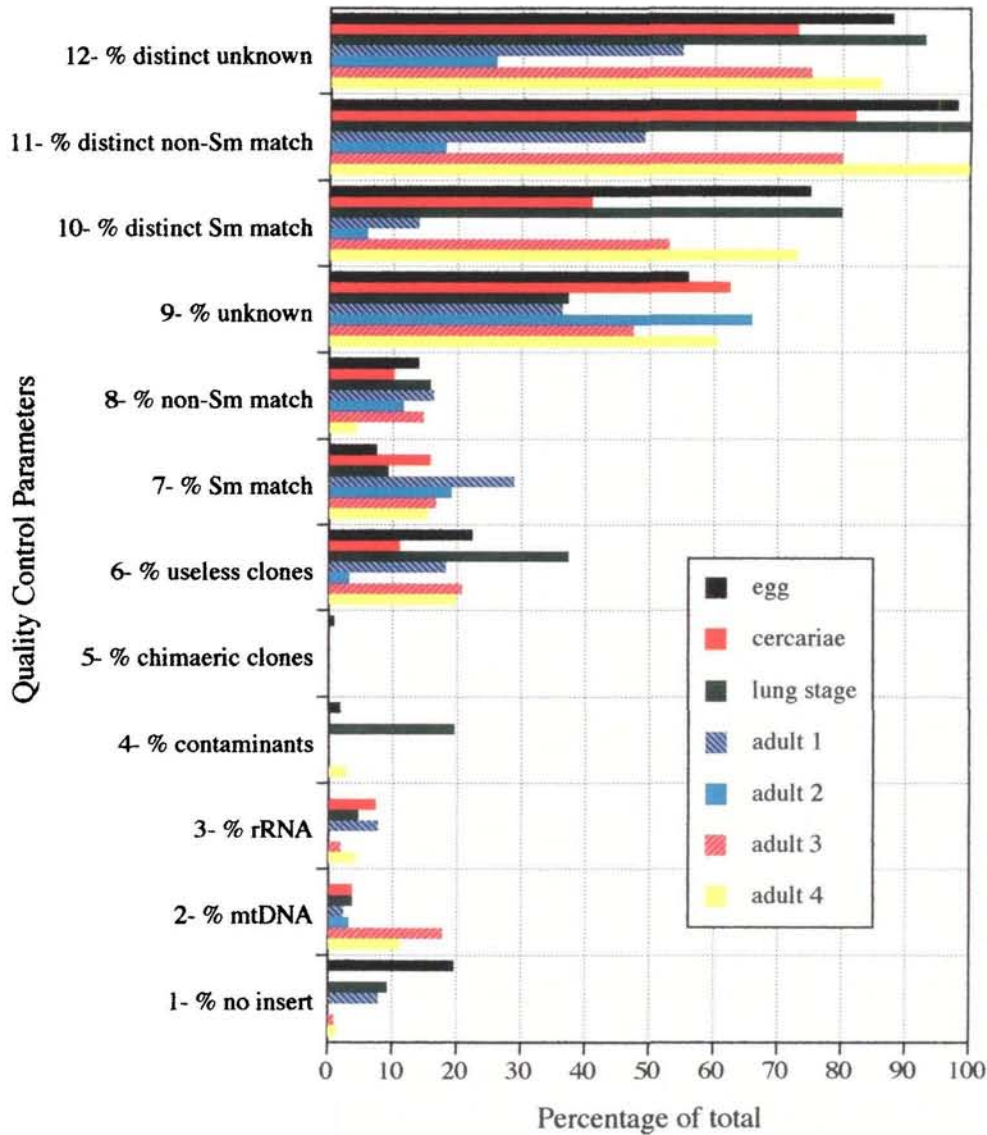
Since the start of *S. mansoni* genome project, one of our main focuses has been the large-scale sequencing of cDNA to produce ESTs, in an attempt to identify new genes of this organism. Initially, we used an adult worm cDNA library, from which we generated 607 ESTs corresponding to 154 new *S. mansoni* genes.<sup>2</sup> The good quality of this library was attested by the diversity of genes that were isolated, even after the discovery of a significant degree of redundancy (65% of the sequenced clones corresponded to 49 redundant genes).<sup>2</sup> The success of this approach prompted us to extend the sequencing program to include other libraries. We started with eight libraries from distinct developmental stages, all of them constructed using the λZap system (Stratagene): one egg, two cercariae (the human-infecting larvae), three adult worms, one 7-day schistosomula (the lung stage) and one from 25-day old worms. All libraries were excised "en masse" and at least 30 colonies from each library were selected to evaluate the average size of the inserts by polymerase chain reaction (PCR). Most of them had an average insert size greater than 500 bp, except for one cercariae and the 25-day worm libraries. Thus, we decided to use all three adult worm cDNA libraries and the Egg, one cercariae and the 7-day schistosomula libraries in this study.

Table 1 summarizes data obtained from the sequencing of the distinct libraries. A total of 1401 ESTs were produced from one or both ends of 1204 clones. The data from the Adult 1 library are cumulative since the beginning of the program and includes ESTs published by Franco et al., 1995.<sup>2</sup> In the Egg library, the number of clones exceeds the number of ESTs and this is due to the sequencing of a chimeric clone from which two ESTs were generated. Both ESTs were eliminated from subsequent analysis. After homology searches in non-redundant databases using BLAST programs<sup>8</sup> and elimination of ESTs corresponding to useless sequences (vector, mitochondrial DNA, rRNA and contaminating sequences from other organisms), 1123 ESTs derived from 970 clones were submitted to clustering analysis, using ICATOOLS program,<sup>9</sup> resulting in a list of distinct genes.

Adams et al.<sup>10</sup> proposed criteria to evaluate the quality of the libraries used in large-scale EST analysis. They state that the sequencing of 100–200 clones from a library is sufficient to assess the quality of this library and to detect problems that might have occurred during library construction. A useful library should contain no more than 20% useless sequences, at least 50% new genes and a broad variety of transcripts. We used their criteria to evaluate the seven cDNA libraries used in this study (Fig. 1). The first five parameters are a measure of the proportion of useless clones. In general, the libraries were of good quality with respect to these parameters, except for the Lung stage, Egg and Adult 3 libraries. The Egg library contains 20% clones without an insert, even though a previous blue/white selection of clones had been performed. The Adult 3 library is enriched in clones corresponding to mitochondrial DNA sequences. Most of them correspond to a polymorphic minisatellite sequence of 620 bp,<sup>11</sup> that contains part of an *S. mansoni* nuclear transcript denominated SM750.<sup>12</sup> This transcript is composed of a invariable region that is followed by five copies of a 62-bp polymorphic repeat element (PRE). Interestingly, five or more copies of the 62-bp PRE were seen solely or as part of the mitochondrial minisatellite in all libraries analyzed except the Egg library. This fact implies that PRE is a very frequent element in the genome of the parasite and that it could be part of a nuclear sequence that was incorporated into the mitochondrial genome.<sup>11</sup> None of the libraries contains excessive number of sequences derived from ribosomal RNA. The Lung stage library contains almost 20% contaminating sequences from other organisms. These contaminating sequences are derived either from *E. coli* or other bacteria, probably due to the contamination of the worm samples during the 7-day period of *in vitro* cultivation necessary to mature to lung stage schistosomula.

The quality of the construction of each library was also analyzed. All of them were shown to be unidirectional (most ESTs had matches to database sequences on the expected strand), composed of a high proportion of inserts longer than 500 bp, composed of inserts with short poly(A) tails and containing no chimeric clones. The only exception was the Egg library, where we found a single





**Figure 1.** Evaluation of the cDNA libraries according to the criteria of Adams et al.<sup>10</sup> Parameters 1 to 5 indicate the percentage of the total of clones in each library that produced useless ESTs and this set of data is totaled in parameter 6. The percentage of the total of clones that are identified either by homology with previously reported *S. mansoni* genes (*Sm* match), putatively identified by homology with genes from other organisms (non-*Sm* match), or with partial homology with genes from other organisms and non-database match sequences (unknown) is also shown (parameters 7 to 9). The percentage of useful clones that are distinct for each category of genes was determined by clustering analysis and is shown in parameters 10 to 12.

chimeric clone (parameter 5). The sixth parameter is the sum of the first five parameters and totals the frequency of useless clones in each library. Three out of the seven libraries exceed 20% non-useful clones: Lung stage (37%), Egg (22%) and Adult 3 (21%), and this is mainly due to the reasons discussed above. However, when analyzing the gene content in each of these three libraries, we verified that they have a high percentage of distinct genes and a low proportion of redundant genes (see below). This

fact justifies the continuation of using of these libraries in the EST sequencing program, but with the inclusion of a previous selection step to eliminate abundant useless clones.

Parameters 7 to 9 of Fig. 1 concern to the analysis of the composition of the libraries after EST homology searches in non-redundant databases. Most libraries showed a low proportion of cDNA clones with exact match to previously described *S. mansoni* genes

**Table 2.** Gene content of the cDNA libraries after random-sampling of clones.

	<i>S. mansoni</i> cDNA libraries							Total
	Egg	Cercariae	Lung stage	Adult 1	Adult 2	Adult 3	Adult 4	
Distinct genes	73	65	62	198	19	57	48	522
New genes	67	58	54	173	18	48	40	458
% of distinct genes per total of sequenced clones <sup>a)</sup>	68.2	60.7	57.9	32.1	20.2	56.4	67.6	43.4
% of new genes per total of sequenced clones <sup>a)</sup>	62.6	54.2	50.5	28.0	19.1	47.5	56.3	38.0

<sup>a)</sup> for the number of sequenced clones see Table 1.

(less than 20%), except for the Adult 1 library (parameter 7). This can be explained by the fact that this library is enriched in clones corresponding to the *S. mansoni* glycolytic enzyme glyceraldehyde 3-phosphate dehydrogenase (GAPDH),<sup>13</sup> the most redundant gene found in this library. Moreover, as the Adult 1 library was the most sequenced library in this program, it is possible that it better represents the profile of genes expressed in adult worms. Remarkably, all adult worm libraries had, in general, more cDNA matching *S. mansoni* known genes than the libraries constructed from other developmental stages. This is particularly interesting, since it reflects the sort of *S. mansoni* genes that have been deposited in public databases. Most of them are isolated from adult worms. However, the Cercariae library attained the same proportion of clones matching *S. mansoni* genes as the adult libraries. This can be explained by the presence of a very abundant transcript in this category, the calcium-binding protein (CaBP),<sup>14</sup> that corresponds to 10% of the total of useful clones. Most probably this protein is very important for the cercariae metabolism and may be involved in movement. Few clones in all libraries could be putatively identified by significant homology with genes from other organisms (parameter 8) and the great majority of clones in each library (>35%) could not be identified (parameter 9). These last ones correspond to cDNA that had only partial matches to sequences from other organisms or non-database match cDNA.

Parameters 10 to 12 consist of the number of distinct genes divided by the number of useful clones in each category and measure the diversity of transcripts. To obtain the number of distinct genes, each library was submitted to clustering analysis, using the program ICATOOLS. The program grouped together as a single cluster clones with a high degree of identity; each cluster was treated as an independent gene. The veracity of such clusters was attested by the correct grouping of clones that shared the same homology to *S. mansoni* or other organisms database sequences. Considering that one goal of the EST sequencing program is the discovery of new genes, the diversity in the non-*Sm* match and in the unknown categories are particularly relevant. In this respect, in all libraries with exception of the Adult 1 and Adult 2 libraries, more than 70% of the transcripts are distinct in these two categories. This fact counterbalances the low

efficiency in obtaining useful clones from the Egg, Lung stage and Adult 3 libraries. An intermediate degree of diversity is observed for the Adult 1 library, while a very low diversity of transcripts is seen in the Adult 2 library. A tendency of decreasing the variety of transcripts in the *Sm* match category is also observed, which can be explained by the presence of very abundant transcripts already characterized in *S. mansoni*. That is the case for the Cercariae, Adult 1 and Adult 2 libraries due to the enrichment of CaBP-, GAPDH- and eggshell protein-encoding<sup>15</sup> transcripts, respectively.

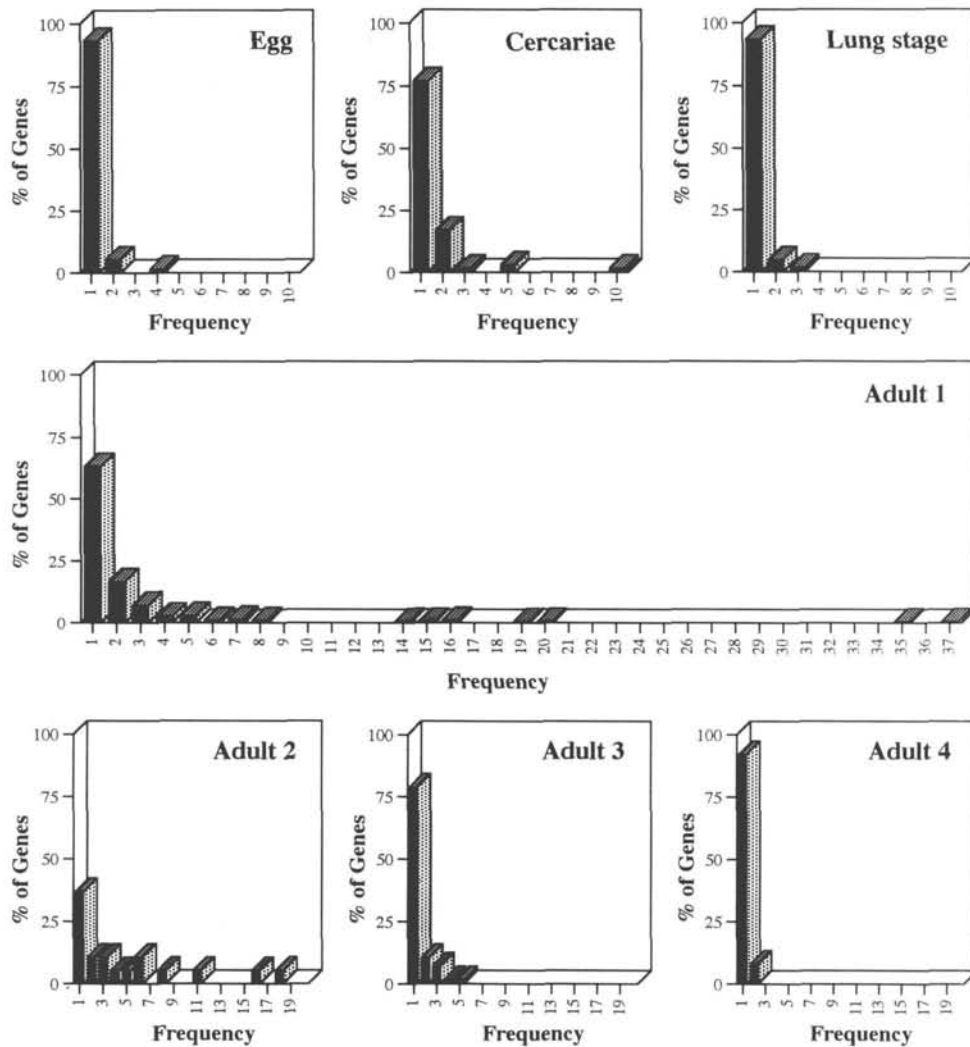
### 3.2. Gene content and redundancy analysis

The strategy of random-sampling of cDNA libraries always produces a series of clones corresponding to a single transcript; either because abundant mRNA will be more represented in the library, or because each library has an inherent bias that was introduced during its construction. Thus, clones obtained from a such library will reflect its cDNA composition. For this reason, we decided to analyze each library according to its gene content and to evaluate its quality based on the extent of redundancy. This was only possible after performing clustering analysis by ICATOOLS.

Table 2 shows the number of distinct genes, as well as the number of new genes obtained from each library. This last class includes genes homologous to genes from other organisms (non-*Sm* match category) and genes either partially homologous to genes from other organisms or non-database match genes (unknown category). A total of 522 distinct genes were obtained from the seven libraries, 458 of which (88%) were newly identified in *S. mansoni*. This corresponds to three times the number of new genes obtained in the beginning of the sequencing program.<sup>2</sup>

Considering the effort to get distinct or new genes from random selection of clones in each library, it is important to consider the percentage of genes in the total of sequenced clones. This is a measure of the library quality regarding both its redundancy and content of useless clones. It can be seen that, in all libraries with the exception of the Adult 1 and Adult 2 libraries, more than 50% of the sequenced clones were found to be distinct genes. It is important to note that the Adult 1 library was se-





**Figure 2.** Redundancy in EST sequencing of the *S. mansoni* cDNA libraries. On the abscissa we show the number of times that each gene was sampled and on the ordinate we depict the fraction of genes sharing a given sampling frequency.

quenced close to six times more than the other libraries (Table 1), and this might explain the rate of 32% of new genes. The same tendency was seen for the ratio of new genes per total of sequenced clones. Again, the Adult 1 and Adult 2 libraries provided the lowest efficiencies. Rates of 50% in acquirement of new genes as observed for the *S. mansoni* libraries met the criteria established for the human EST program.<sup>10</sup>

A direct representation of the extent of redundancy in each library is seen in Fig. 2, that shows the percentage of genes that appear in the library under a given frequency. As random sampling of a cDNA library should follow a Poisson distribution for rare events, the unex-

pected presence of genes under classes of high frequency of isolation reveals a bias in the library. This is evident for the Adult 2 library, where the profile of frequency distribution clearly escapes a typical Poisson distribution, which strongly supports our decision not to use this library for large-scale EST production. The high proportion of redundant genes in this library might have resulted from errors introduced during library construction and amplification, "en masse" excision or clone sampling for EST generation. The occurrence of genes under classes of high frequency of isolation is also seen in the Cercariae and Adult 1 libraries. Nevertheless, it would be possible to eliminate the most redundant genes (8 genes

**Table 3.** Putatively identified genes homologous to *S. mansoni* genes.<sup>a)</sup>

Gene	Library	EST accession <sup>b)</sup>
<b>Enzyme</b>		
Aspartic proteinase	Adult 4	AA169900
Carbonyl reductase	Egg	AA140589
Cathepsin B	Cercariae	AA143823
Cyclophilin B	Lung stage	AA125705
Enolase	Adult 1	T14396
ER-luminal cysteine protease (ER-60)	Adult 4	AA169915
Fructose-1,6-bisphosphate aldolase	Lung stage	AA125670
Glutathione peroxidase	Cercariae	AA143892
Glutathione S-transferase	Adult 1	T14549
Glyceraldehyde-3-phosphate dehydrogenase	Adult 1	T14434
Hemoglobinase (Sm32)	Adult 1	T14348
Hexokinase	Adult 1	T14603
Triose phosphate isomerase	Egg	AA140583
<b>Cytoskeletal/structural protein</b>		
Actin	Cercariae	AA143846
Alpha-tubulin	Egg	AA140633
Eggshell protein	Adult 4	AA169905
Female-specific polypeptide	Adult 3	AA218489
Myosin heavy chain	Lung stage	AA125688
P48 eggshell protein	Adult 3	AA218479
Sm23 integral membrane protein	Adult 1	T14382
Tropomyosin (GB:SCMTPM)	Adult 1	W06805
Tropomyosin (GB:SCMTROPO)	Adult 1	W06761
<b>Antigen</b>		
Antigen 10-3	Adult 1	T14386
Antigen Sm21.7	Adult 3	AA218508
Major egg Antigen (P40)	Egg	AA140559
Sm13 tegumental antigen	Adult 4	AA169901
<b>Transport/storage protein</b>		
Calcium binding protein (CaBP)	Cercariae	AA143886
Calcium-calmodulin binding protein	Cercariae	AA143883
Calreticulin	Adult 1	W06720
Fatty-acid binding protein (Sm14)	Adult 1	T14374
Ferritin	Adult 3	AA218482
Glucose transporter	Adult 1	T14364
<b>Other</b>		
Breast basic conserved protein/ribosomal protein L13	Adult 1	T14585
Calnexin homolog SmIrV1	Adult 3	AA218511
Elongation factor 1 alpha	Lung stage	AA125724
Heat shock protein 86	Adult 1	T14407
<i>S. mansoni</i> mRNA for tandem repeat	Egg	AA140585
<i>S. mansoni</i> (Liberia) zinc finger protein	Lung stage	AA125700
Y-box-binding protein	Adult 1	T14571

a) Genes putatively identified by homology with *S. mansoni* database sequences. Only one representative EST matching the respective gene is shown, together with the name of the library it was isolated from.

b) EST accession corresponds to the GenBank accession number.

for the Adult 1 library and 3 genes for the Cercariae library) from these libraries by filter screening, using the abundant transcripts as probes, and this should result in a profile compatible with a Poisson distribution.

Although some libraries presented problems detected by quality analysis, they all contributed to the list of putatively identified genes, as well as 333 distinct unknown genes (see below). Genes identified by homology with previously described *S. mansoni* genes are distributed amongst various classes, such as genes coding for enzymes, structural proteins, antigens, proteins involved in transport and storage, etc (Table 3). Two genes in this list have been the subject of a more extensive study and were characterized in detail in our laboratory. They are

the *S. mansoni* homologues of the Y-box-binding protein (Franco et al., submitted) and the breast basic conserved protein, or the 60S ribosomal protein L13 (Franco et al., submitted). Table 4 lists distinct genes putatively identified by homology with genes from other organisms. They code for enzymes of different metabolic pathways, a great variety of ribosomal proteins, several constituents of transcriptional/translational machinery, and regulatory cytoplasmic and membrane proteins, among others. Three of these genes were selected for further studies. One is the homologue of mago nashi gene from *Drosophila*. This gene is necessary for proper germ plasm assembly and mutations in it result in sterility of F1 progeny.<sup>16</sup> The *S. mansoni* purine nucleoside phosphorylase was selected for presenting a high similarity with the human counterpart, the 3D structure of which has already been resolved and deposited in the Protein Data Bank. Modeling studies with this protein have led to the identification of powerful inhibitors of this enzyme, whose activity is crucial in T cell guanosine metabolism.<sup>17</sup> The third gene is the homologue of the human HLA-DR-associated protein I, a protein which may be involved in signal transduction in B cells.<sup>18</sup> We are interested in the selection of proteins that can interact with it, which may help to define its biological function in the parasite.

### 3.3. Gene expression profile in *S. mansoni*

To obtain an initial profile of gene diversity in the parasite and a preliminary pattern of gene expression in distinct stages of the development of *S. mansoni*, we performed a clustering analysis, joining sequences from all libraries. This resulted in a total of 466 unique genes (considering only once the genes present in more than one library), corresponding to 427 new *S. mansoni* genes. From the total of unique genes, 39 (8.3%) matched previously characterized *S. mansoni* genes, 94 (20.2%) matched genes from other organisms and 333 (71.5%) represent unknown genes. From the clustering analysis, most genes (433 of 466) were present only in a single library (e.g. CaBP was found only in the Cercariae library). Other genes were expressed in more than one developmental stage and are listed in Table 5. They may represent housekeeping genes in the parasite and, curiously, ten of them were unknown. The antigenic potential of such genes should be investigated, since they might be specific to this parasite.

At this point of the sequencing program, only three genes were found to be expressed in all developmental stages analyzed: the cytochrome oxidase chain I, the fructose-1,6-bisphosphate aldolase and unknown gene 10. Somewhat unexpectedly, actin and GAPDH, the most frequent genes in the collection, were not isolated from all stages, perhaps because the number of transcripts sequenced in each library was not very large. Five genes

Table 4. Identified genes homologous to non-*S. mansoni* genes.<sup>a)</sup>

Gene	Library	EST accession <sup>b)</sup>
<b>Enzymes</b>		
Alcohol dehydrogenase class III	Adult 3	AA218449
Aldehyde dehydrogenase	Adult 1	T24129
Aldose reductase	Adult 1	W06782
ATP synthase, vacuolar	Adult 1	T24142
Cytochrome Oxidase chain I	Egg	AA140564
Cytochrome oxidase II	Adult 3	AA218486
Dakt1 serine/threonine protein kinase	Adult 4	AA169931
Dihydroipoamide acetyltransferase	Adult 1	W06795
Enoyl-CoA Hydratase	Lung stage	AA125690
Glutamine Synthetase	Adult 1	W06821
Glyceral 3-phosphate dehydrogenase	Cercariae	AA143842
H <sup>+</sup> -transporting ATP synthase alpha-chain	Adult 1	W06794
Lactate dehydrogenase	Adult 1	W06744
Oligosaccharyl transferase 48 KD	Adult 3	AA218463
Ornithine aminotransferase	Adult 1	T14588
Phosphoenolpyruvate Carboxykinase	Egg	AA140576
Phosphoglycerate kinase	Adult 1	T14620
Phosphoglycerate mutase	Adult 1	W06743
20S proteasoma subunit RC7-I=PRE1 homolog	Lung stage	AA125733
Proteasome zeta chain	Adult 1	T14568
Purine nucleoside phosphorylase	Adult 1	W06714
Pyruvate kinase	Adult 1	T24140
Ribonuclease- phosphate 3-epimerase (pentose-5-phosphate 3-epimerase)	Adult 3	AA218494
Vacuolar ATP synthase subunit B	Adult 1	W06824
<b>Transcriptional/ Translational Machinery</b>		
40S ribosomal protein S3	Adult 3	AA218471
40S ribosomal protein S4	Adult 1	W06814
40S ribosomal protein S7	Egg	AA140626
40S ribosomal protein S11	Adult 4	AA169892
40S ribosomal protein S12	Lung stage	AA125707
40S ribosomal protein S14	Adult 1	T14564
40S ribosomal protein S17	Lung stage	AA125727
40S ribosomal protein S20	Egg	AA140581
40S ribosomal protein S21	Egg	AA140582
40S ribosomal protein S26	Lung stage	AA125695
60S ribosomal protein L5	Lung stage	AA125687
60S ribosomal protein L7	Egg	AA140600
60S ribosomal protein L7a	Adult 1	T14431
60S ribosomal protein L10a	Egg	AA140612
60S ribosomal protein L25	Lung stage	AA125723
60S ribosomal protein L30	Adult 3	AA218468
Asp-tRNA synthetase	Adult 1	W06768
Elongation factor 1 gamma	Adult 1	T14422
<i>Homo sapiens</i> 9G8 splicing factor	Adult 1	W06725
Jun-binding protein	Adult 4	AA125664
Lys-tRNA synthetase	Adult 1	T14484
Polyadenylate binding protein	Adult 1	W06727
Putative transcriptional regulator	Lung stage	AA125694
Reverse transcriptase	Egg	AA140605
Rho-GDP dissociation inhibitor	Adult 1	W06723
RNA polymerase II subunit	Lung stage	AA125704
RNA-binding protein X-16	Adult 1	T14358
Small nuclear ribonucleoprotein	Adult 1	T14459

were present in two or more adult libraries, but absent in other stages. This is the case for the eggshell protein, that is recognized to be expressed in mature females, and also unknown gene 2.

Clustering analysis also included formation of contigs of sequences. As an example, the cDNA sequence of an unknown gene, that is abundant in the Adult 1 library, was obtained after assembling ESTs from both cDNA ends that were clustered together by ICATOOLS. This gene is currently being characterized in more detail. We

Table 4. Continued.

Gene	Library	EST accession <sup>b)</sup>
<b>Membrane/cytoplasm</b>		
ADP/ATP carrier protein	Adult 1	T14447
Annexin family	Adult 1	T14511
Beta-1 tubulin	Egg	AA140634
Chaperonin-like protein	Adult 1	T14632
Cytochrome c	Cercariae	AA143872
DNAJ homolog	Adult 1	W06722
GTP-binding protein	Adult 3	AA218450
Heat shock protein 108	Adult 1	T18621
HLA-DR associated protein I	Adult 3	AA218460
Polyubiquitin	Egg	AA140632
Possible membrane protein	Lung stage	AA125728
Protein kinase C inhibitor protein	Adult 1	T14595
Nonerythroid alpha-spectrin	Adult 1	T14622
UDP-galactose translocador	Adult 3	AA218519
<b>Other</b>		
52k active chromatin boundary protein	Adult 1	W06740
Alpha-collagen	Adult 1	T14493
Apoptosis-inducible	Cercariae	AA143880
Arginine-rich gene	Adult 1	T14555
<i>C. elegans</i> hypothetical 272 KD protein C50C3.6 in chromosome III	Adult 3	AA218465
<i>C. elegans</i> clone C16C10.10	Adult 1	W06746
Coded for by <i>C. elegans</i> cDNA	Adult 3	AA218495
Cysteine-rich intestinal protein	Adult 3	AA218481
<i>E. coli</i> hypothetical 53.1 KD protein in LYSU-CADA intergenic region	Lung stage	AA125683
Fibrillin 2	Cercariae	AA143891
GATA-3 gene	Egg	AA140598
Golden Syrian Hamster repetitive DNA	Egg	AA140590
Histone H3.3	Cercariae	AA143814
<i>H. sapiens</i> mRNA for <i>Sm</i> protein F	Lung stage	AA125673
Human Alu subfamily	Adult 1	W06750
Hypothetical protein - <i>D. melanogaster</i>	Cercariae	AA143820
Hypothetical protein 5 <i>Xanthobacter</i> sp	Adult 1	W06771
Hypothetical 30.5 KD protein of <i>C. elegans</i>	Egg	AA140628
Liver regeneration factor augmenter	Adult 2	AA185826
Mago nashi protein	Lung stage	AA125719
MER5 Protein	Adult 1	AA140649
NIFS-like 54.5 KD protein	Adult 1	W06757
Proliferation-associated protein	Lung stage	AA125729
Retrovirus-related GAG polyprotein	Egg	AA140602
Synaptophysin	Adult 1	W06818
Yeast hypothetical 103.7 KD Protein	Adult 1	W06819
Valosin-containing protein homologue	Adult 1	T14640

a) Genes putatively identified by homology with genes from other organisms. Only one representative EST matching the respective gene is shown, together with the name of the library it was isolated from.

b) EST accession corresponds to GenBank accession number.

expect that, with the advance of the sequencing program, a higher number of partial cDNA sequences will be assembled as full-length contigs, increasing the ability to identify unknown genes and more precisely define the real number of distinct genes in each library and in each developmental stage.

**Acknowledgments:** The authors thank Kátia Barroso for carrying out automated DNA sequencing. This investigation received financial support from the following sources: PADCT, CNPq, UNDP/WORLD BANK/WHO Special Program for Research and Training in Tropical Diseases (TDR N°: 940325 and 940751), USAID/HOH (N° 264.01.01.04), FAPEMIG, PAPES/ FIOCRUZ.

**Table 5.** Frequency of genes present in multiple *S. mansoni* cDNA libraries.<sup>a)</sup>

Genes	Egg (%)	Cercariae (%)	Lung stage (%)	Adult 1 (%)	Adult 2 (%)	Adult 3 (%)	Adult 4 (%)	Total (%)
1- Actin	—	1.0	—	6.9	—	3.9	1.9	4.1
2- Alpha tubulin	2.5	—	1.5	0.8	—	2.6	—	0.9
3- ATP synthase	—	—	1.5	0.2	—	—	—	0.2
4- Beta tubulin	1.3	—	—	0.4	—	—	—	0.3
5- Carbonyl reductase	1.3	—	—	0.2	—	—	—	0.2
6- Cathepsin	—	1.0	1.5	0.4	—	—	—	0.4
7- Cyclophilin B	—	—	1.5	—	—	2.6	—	0.3
8- Cysteine-rich intestinal protein	—	1.0	—	—	—	1.3	—	0.2
9- Cytochrome oxidase chain I	1.3	2.0	1.5	0.4	8.8	—	—	1.4
10- EF1alpha	—	—	1.5	3.0	—	—	—	1.6
11- Eggshell protein	—	—	—	0.2	19.8	—	1.9	2.1
12- Enolase	—	—	—	0.8	—	—	1.9	0.5
13- ER-luminal cysteine protease	—	—	—	0.4	—	—	1.9	0.3
14- Fibrillin	—	1.0	—	0.6	—	—	—	0.4
15- Fructose-1,6-BP aldolase	1.3	1.0	1.5	3.2	—	—	3.8	2.2
16- GAPDH	—	2.0	4.5	7.3	—	1.3	—	4.4
17- Major egg Antigen (P40)	1.3	—	—	0.2	—	—	—	0.2
18- Myosin heavy chain	—	—	1.5	0.4	—	—	—	0.3
19- Oligosaccharyl transferase 48 KD	—	—	—	0.2	—	1.3	—	0.2
20- Triose phosphate isomerase	1.3	—	—	0.4	—	—	—	0.3
21- Ubiquitin	1.3	—	—	1.0	—	—	—	0.6
22- 60S ribosomal protein L5	—	—	1.5	1.0	—	—	—	0.6
23- 60S ribosomal protein L30	—	—	1.5	—	—	1.3	—	0.2
24- Gene 1 <sup>b)</sup>	—	2.0	—	—	—	3.8	—	0.5
25- Gene 2	—	—	—	0.4	2.2	5.1	—	0.8
26- Gene 3	—	—	1.5	—	—	1.3	—	0.2
27- Gene 4	—	—	1.5	0.2	—	—	—	0.2
28- Gene 5	—	1.0	—	0.2	—	—	—	0.2
29- Gene 6	—	1.0	—	0.2	—	—	—	0.2
30- Gene 7	—	—	1.5	—	—	—	1.9	0.2
31- Gene 8	—	—	1.5	0.2	—	—	—	0.2
32- Gene 9	1.3	2.0	—	0.2	4.4	6.4	1.9	1.4
33- Gene 10	1.3	5.1	1.5	1.6	—	1.3	1.9	1.8

<sup>a)</sup> Percentage of clones matching the corresponding gene in the total of usable clones analyzed by ICATTOOLS. For the total of usable clones see Table 1. <sup>b)</sup> unknown genes are numbered 1–10.

## References

- Adams, M. D., Kelley, J. M., Gocayne, J. D. et al. 1991, Complementary DNA sequencing: expressed sequence tags and human genome project, *Science*, **252**, 1651–1656.
- Franco, G. R., Adams, M. D., Soares, M. B., Simpson, A. J. G., Venter, J. C., and Pena, S. D. J. 1995, Sequencing and Identification of expressed *Schistosoma mansoni* genes by random selection of cDNA clones from a directional library, *Gene*, **152**, 141–147.
- Smithers, S. and Terry, R. J. 1965, The infection of laboratory hosts with cercariae of *S. mansoni* and the recovery of adult worms, *Parasitology*, **55**, 695–700.
- Chomczynski, P. and Sacchi, N. 1987, Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction, *Anal. Biochem.*, **162**, 156–159.
- Aviv, H. and Leder, P. 1972, Purification of biologically active globin messenger RNA by chromatography on oligo-thymidylic acid-cellulose, *Proc. Natl. Acad. Sci. USA*, **69**, 1408.
- Short, J. M., Fernandez, J. M., Sorge, J. A., and Huse, W. D. 1988, λZAP: A bacteriophage λ expression vector with *in vivo* excision properties, *Nucleic Acids Res.*, **16**, 7583–7600.
- Sanger, F. 1981, Determination of nucleotide sequences in DNA, *Science*, **214**, 1205–1210.
- Altschul, S. F., Gish, W. Miller, W. Myers, E. W., and Lipman, D. 1990, Basic local alignment search tool, *J. Molec. Biol.*, **215**, 403–410.
- Parsons, J. D., Brenner, S., and Bishop, M. J. 1992, Clustering cDNA sequences, *Comput. Appl. Biosci.*, **8**, 461–466.
- Adams, M. D., Kerlavage, A. R., Fleischmann, R. D. et al. 1995, Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence, *Nature*, **377** (supp), 3–174.
- Pena, H. B., Souza, C. P., Simpson, A. J. G., and Pena, S. D. J. 1995, Intracellular promiscuity in *Schistosoma mansoni*: nuclear transcribed DNA sequences are part of a mitochondrial minisatellite region, *Proc. Natl. Acad. Sci. USA*, **92**, 915–919.
- Spotila, L. D., Rekosh, D. M., and LoVerde, P. T. 1991, Polymorphic repeated DNA element in the genome of *Schistosoma mansoni*, *Mol. Biochem. Parasitol.*, **48**,



- 117–120.
13. Goudot-Crouzel, V., Caillol, D., Djabali, M., and Dessein, A. J. 1989, The major parasite surface antigen associated with human resistance to schistosomiasis is a 37 kDa glyceraldehyde-3P-dehydrogenase, *J. Exp. Med.*, **170**, 2065–2080.
  14. Ram, D., Grossman, Z., Markovics, A. et al. 1989, Rapid changes in the expression of a gene encoding a calcium-binding protein in *Schistosoma mansoni*, *Mol. Biochem. Parasitol.*, **34**, 167–175.
  15. Menrath, M., Michel, A., and Kunz, W. 1995, A female-specific sequence of *Schistosoma mansoni* encoding a mucin-like protein that is expressed in the epithelial cells of the reproductive duct, *Parasitology*, **111**, 477–483.
  16. Boswell, R. E., Prout, M. E., and Steichen, J. C. 1991, Mutations in a newly identified *Drosophila melanogaster* gene, *mago nashi*, disrupt germ cell formation and result in the formation of mirror-image symmetrical double abdomen embryos, *Development*, **113**, 373–384.
  17. Ealick, S. E., Babu, Y. S., Bugg, C. E. et al. 1991, Application of the crystallographic and modeling methods in the design of purine nucleoside phosphorylase inhibitors, *Proc. Natl. Acad. Sci. USA*, **88**, 11540–11544.
  18. Vaesen, M., Barnikol-Watanable, S., Gotz, H. et al. 1994, Purification and characterization of two putative HLA class II associated proteins: PHAPI and PHAPII, *Biol. Chem. Hoppe-Seyler*, **375**, 113–126.

## **I.5. ARTIGO 3**

**Analysis of the gene expression profile of *Schistosoma mansoni* cercariae using  
the Expressed Sequence Tag approach**

## Analysis of the gene expression profile of *Schistosoma mansoni* cercariae using the expressed sequence tag approach<sup>☆</sup>

Túlio M. Santos<sup>a</sup>, David A. Johnston<sup>d</sup>, Vasco Azevedo<sup>b</sup>, Ian L. Ridgers<sup>d</sup>, Mercedes F. Martinez<sup>d</sup>, Gláucia B. Marotta<sup>b</sup>, Rogério L. Santos<sup>b</sup>, Sérgio J. Fonseca<sup>b</sup>, J. Miguel Ortega<sup>a</sup>, Élide M.L. Rabelo<sup>c</sup>, Mohamed Saber<sup>e</sup>, Hanem M. Ahmed<sup>e</sup>, Mahmoud H. Romeih<sup>e</sup>, Glória R. Franco<sup>a</sup>, David Rollinson<sup>d</sup>, Sérgio D.J. Pena<sup>a,\*</sup>

<sup>a</sup> Departamento de Bioquímica e Imunologia, ICB-UFMG, CP 486, 31270-010, Belo Horizonte, MG, Brazil

<sup>b</sup> Departamento de Biologia Geral, ICB-UFMG, CP 486, 31270-010, Belo Horizonte, MG, Brazil

<sup>c</sup> Departamento de Parasitologia, ICB-UFMG, CP 486, 31270-010, Belo Horizonte, MG, Brazil

<sup>d</sup> Department of Zoology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

<sup>e</sup> Biochemistry Department, Theodore Bilharz Research Institute, P.O Box 30, Imbaba, Cairo, 12411, Egypt

Received 17 February 1999; received in revised form 10 May 1999; accepted 13 May 1999

### Abstract

ESTs constitute rapid and informative tools with which to study gene-expression profiles of the diverse stages of the schistosome life cycle. Following a comprehensive EST study of adult worms, analysis has now targeted the cercaria, the parasite larval form responsible for infection of the vertebrate host. Two *Schistosoma mansoni* cercarial cDNA libraries were examined and partial sequence obtained from 957 randomly selected clones. On the basis of database searches, 551 (57.6%) ESTs generated had no homologs in the public databases whilst 308 (32.2%) were putatively identified, totaling 859 informative ESTs. The remaining 98 (10.2%) were uninformative ESTs (ribosomal RNA and non-coding mitochondrial sequences). By clustering analysis we have identified 453 different genes. The most common sequences in both libraries represented Sm8 calcium binding protein (8% of ESTs), fructose-1,6-bisphosphate aldolase, glyceraldehyde-3-phosphate dehydrogenase, cytochrome oxidase subunit 1, ATP guanidine kinase and triose phosphate isomerase. One hundred and nineteen identified genes were sorted into 11 functional categories,

*Abbreviations:* CaBP, calcium binding protein; ESTs, expressed sequence tags; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; HMG-CoA reductase, 3-hydroxymethylglutaryl-coenzyme A reductase; NCBI, National Center for Biotechnology Information; PBR, peripheral-type benzodiazepine receptor; PDE, phosphodiesterase.

<sup>☆</sup> Note: Nucleotide sequence data reported in this paper have been submitted to the GenBank™ database with the following accession numbers: cercarial library 1, AA999208 → AA999647, AI264972; cercarial library 2, AA559353 → AA559843.

\* Corresponding author. Tel.: + 55-31-4992628; fax: + 55-31-4992984.

E-mail address: spena@dcc.ufmg.br (S.D.J. Pena)

with genes associated with energy metabolism being the most abundant (13%) and diverse. The diversity and abundance of genes associated with the transcription/translation machinery and with regulatory/signaling functions were also marked. A paramyosin transcript was identified, indicating that this gene is not exclusively expressed in adult worms and sporocysts (as had been suggested previously). The possible physiological relevance to cercariae of the presence of transcripts with homology to calcium binding proteins of the EF-hand superfamily, Gq-coupled rhodopsin photoreceptor, rod phosphodiesterase  $\delta$  subunit and peripheral-type benzodiazepine receptor is discussed. © 1999 Elsevier Science B.V. All rights reserved.

**Keywords:** *Schistosoma mansoni*; Cercariae; Gene expression; Genomics; Benzodiazepine receptor; Photoreceptor

---

## 1. Introduction

Schistosome development occurs as a result of the coordinated expression of distinct sets of genes such that the biochemical and morphological alterations required for completion of the life cycle occurs in the correct order and at the appropriate time. Analysis of the gene expression profiles of different parasite stages should allow a better understanding of schistosome biology and host relationships, facilitating the discovery of novel gene products that could represent targets for the development of new drugs and vaccines to control schistosomiasis. Some stage-specific schistosome genes have already been identified and characterized [1–4], but the number is still small.

The production of ESTs, partial cDNA sequences (150–600 bp) obtained from the ends of cDNA molecules, has been shown to be a rapid and simple method for discovering new genes [5,6], for examining gene expression patterns [7–9] and for identifying differentially regulated genes [10,11].

Since we started the *Schistosoma mansoni* genome project, EST techniques have greatly increased the known gene catalogue for this parasite [12–15]. Until now, EST production has concentrated primarily on adult worm cDNA libraries [12,13]. However, it is also important to analyze the other stages of the developmental cycle: egg, miracidium, sporocyst, cercariae and schistosomulum. Thus, we began a program to explore stage-specific cDNA libraries [14]. A preliminary study of the gene expression profile in different developmental stages of *S. mansoni* was performed using data from 1401 ESTs, generated from seven different cDNA libraries constructed from four stages of

the parasite life cycle [15]. Here we report a comprehensive analysis of the gene expression profile of cercariae, the larval form responsible for infection of the vertebrate host, based on the sequencing of 957 new templates.

## 2. Materials and methods

### 2.1. cDNA libraries and sequencing

*S. mansoni* cercariae cDNA libraries 1 and 2, constructed in  $\lambda$ ZapII vector (Stratagene), were produced at the Theodore Bilharz Research Institute, Cairo, Egypt. Production of ESTs from the 5' or 3' ends of cDNA library 1 by in vivo excision of pBluescript SK + phagemids, purification of plasmid DNA from randomly selected clones and single pass sequencing on an A.L.F. automated DNA sequencer (Amersham Pharmacia Biotech) at Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, was as previously described [14]. Production of ESTs from the 5' ends of clones from cDNA library 2 at The Natural History Museum, London, was by isolation of individual phage plaques, PCR amplification of cloned inserts and direct sequencing of purified PCR product using *Taq* dye terminator cycle sequencing chemistry using an ABI automated DNA sequencer (Applied Biosystems). The full protocols are available from [http://www.nhm.ac.uk/hosted\\_sites/schisto/protocols/nhm\\_est\\_method.html](http://www.nhm.ac.uk/hosted_sites/schisto/protocols/nhm_est_method.html).

### 2.2. Data analysis

Raw sequence data was edited to remove vector sequence and low quality 3' sequence. Edited se-



quences were subjected to protein and nucleic acid homology searches using the Basic Local Alignment Search Tool (BLAST) service, at the NCBI. BLAST and clustering analysis using ICATOOLS [16] were performed as previously described [14].

### 3. Results

In total, 957 EST sequences were generated from the 5' or 3' ends of clones selected from the *S. mansoni* cercarial cDNA libraries (518 clones from library 1; 439 clones from library 2). Table 1 shows the composition of the two cercarial libraries based on the database match categories of ESTs. Of the 957 EST sequences analyzed by homology search, 551 (57.6%) showed no homology to any protein or DNA sequence in the public databases, while 308 (32.2%) ESTs were putatively identified. The remaining 98 (10.2%) corresponded to uninformative ESTs, i.e. sequences matching ribosomal RNA and non-coding mitochondrial DNA. This last category was not included in the clustering analysis.

To determine the number of different genes represented by the cercarial EST dataset, the 859 informative sequences were clustered using ICATOOLS. Each ICATOOLS cluster was treated as an independent gene. Library 1 produced 268 clusters, library 2 produced 229 (48% redundancy in each dataset). Combining the datasets, 44 clusters contained ESTs from both libraries. In total, 453 different genes were detected. The source of

the redundancy in each library was then analyzed by determining the redundancy within each database match category (Fig. 1). Proportionally, clusters showing homology to known *S. mansoni* genes were the most redundant in both libraries (Fig. 1(A)). However, these represented the minor part of the total gene 'catalogue' (Fig. 1(B)). Fifty-seven clusters formed by non-identified ESTs from one or both libraries were also detected. Some of them were highly redundant such as the clusters headed by ESTs AA559726 (3.8% of informative ESTs), AA999330 (3.6%), AA999248 (1.6%), AA559702 (1.3%) and AA559786 (1.2%). These clusters are of special interest since they probably represent *S. mansoni* specific genes.

Fig. 2 gives a picture of the major transcriptional activity in *S. mansoni* cercariae as assessed by the frequency with which individual genes were detected. Twenty-eight identified genes were represented by more than one EST. Some of them were found in both libraries at a similar level of redundancy. These included the highly redundant genes for Sm8 calcium binding protein (7.8% of informative ESTs), fructose-1,6-bisphosphate aldolase (2.9%), GAPDH (1.8%), cytochrome oxidase subunit 1 (1.4%), ATP guanidino kinase (1.2%) and triose phosphate isomerase (1.0%). Other genes frequently seen were actin and the 23-kD integral membrane protein (library 1) and calmodulin NB-1 related protein, cytochrome C, glutathione peroxidase, ATP synthase  $\delta$  chain and tegumental antigen Sm20.8 (library 2). The major-

Table 1  
Database match categories of ESTs generated by sequencing random clones from two *S. mansoni* (Sm) cercarial cDNA libraries<sup>a</sup>

Match category	Library 1	(%)	Library 2	(%)
No database match	298	57.5	253	57.6
Sm match	104	20.1	68	15.5
Sm partial match	3	0.6	9	2.1
Other organisms match	40	7.7	84	19.1
Mitochondrial	16	3.1	5	1.1
rRNA	57	11.0	20	4.6
Total	518	100.0	439	100.0

<sup>a</sup> 'Sm match' includes ESTs with >95% nucleotide identity to *S. mansoni* sequences in GenBank. 'Other organism matches' and 'Sm partial match' include significant matches to other organisms and matches with <95% nucleotide identity to *S. mansoni* sequences, respectively. ESTs with no significant database matches are shown as 'No database match'. Clones containing ribosomal RNA and non-coding mitochondrial sequences were considered 'uninformative' clones.

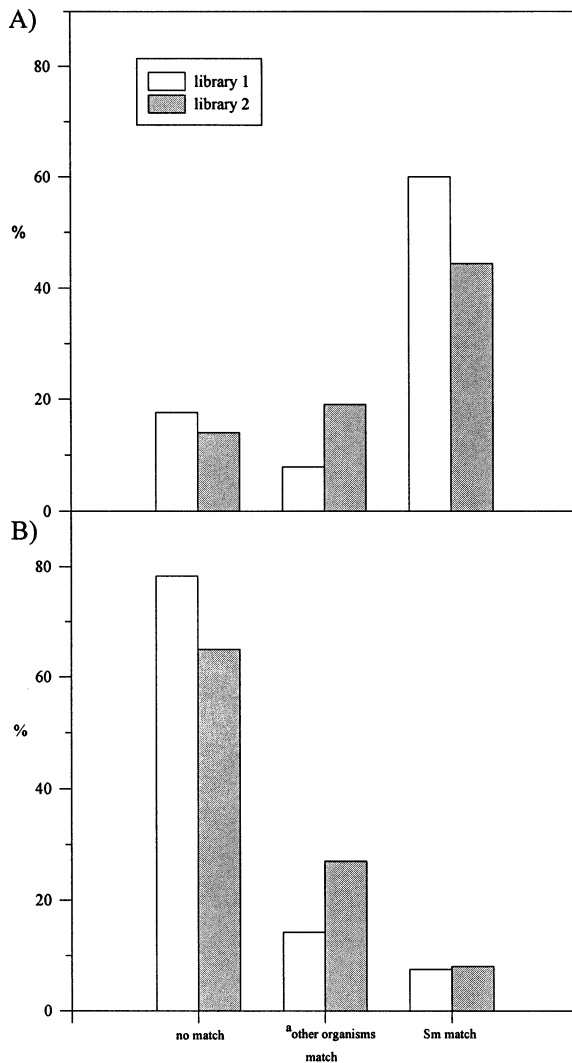


Fig. 1. Analysis of gene redundancy with respect to library and database match category. (A) % of genes within each database match category that displayed redundancy (i.e. were represented > 1 time in the dataset). (B) Distribution of observed clusters (= genes) with respect to database match category (totals sum to 100%). <sup>a</sup>The Sm partial match category is included within the 'other organism match' category

ity of redundant genes are *S. mansoni* housekeeping genes, suggesting that their redundancy is not an artefact of library construction, but a reflection of the fact that they are expressed at high levels in diverse tissues.

A profile of the physiological activity of cercariae can be outlined by grouping the identified genes by functional category, as described previously [7]. Where identification was assigned on the basis of homology with characterized genes from species other than schistosomes, similarity was always less than 95% at the nucleotide or peptide level. Table 2 lists 119 identified genes, sorted into 11 functional categories. Genes related to energy metabolism were the most abundant, comprising 13.1% of informative ESTs, with 30 different genes detected that belong to the glycolytic pathway, Krebs cycle or oxidative phosphorylation. Amongst the other functional categories identified, were genes encoding proteins involved with transcription and translation and those associated with regulatory and signaling functions.

#### 4. Discussion

According to the library quality parameters previously established for large scale EST sequencing projects, a useful library should comprise at least 50% new genes, a broad variety of transcripts and no more than 20% uninformative sequences [9,14]. Both of the cercarial cDNA libraries examined in this study pass these criteria, as shown by Table 1. However, the redundancy rate observed in both libraries was relatively high, especially with respect to the category of Sm matched genes (Fig. 1). This can be explained by two factors: firstly, abundant transcripts are more likely to have been discovered previously; and secondly, this category is composed primarily of housekeeping genes that can be found by gene discovery programs for any developmental stage of the parasite [14].

Similarity between libraries in the levels of redundancy of Sm8 calcium binding protein, fructose-1,6-bisphosphate aldolase, GAPDH, cytochrome oxidase subunit 1, ATP guanidino kinase and triose phosphate isomerase genes suggests that the results obtained are representative and that the libraries accurately reflect the physiological expression in cercariae (Fig. 2). However, the observation of some differences in redundancy between the two libraries (e.g. actin in library 1

Table 2

*S. mansoni* cercariae ESTs identified by matches to nucleotide and protein databases sequences

Clone identification	EST number <sup>a</sup>	Putative homologue (organism)	Accession number <sup>b</sup>	Number of clones	Similarity (%)	Identity (%)	Length	Program	Score	Probability
<i>Energy metabolism</i>										
MC260487NA3	AA559480	ATP synthase $\delta$ chain ( <i>Caenorhabditis elegans</i> )	SP Q09544	4	76	62	81	BLASTX	257	5.4e-027
MC031033NA3	AA559696	ATP synthase $\beta$ chain ( <i>Saccharomyces cerevisiae</i> )	SP P00830	2	86	82	52	BLASTX	217	4.5e-034
MC260422NA3	AA559439	ATP synthase c chain (lipid-binding protein P2 precursor-protein 9) – ( <i>Bos taurus</i> )	SP P07926	2	73	60	88	BLASTX	260	9.0e-031
MC260314NA3	AA559382	atp synthase d chain ( <i>Drosophila melanogaster</i> )	GNL X99667	1	56	36	65	BLASTX	128	1.1e-015
SMC03524F	AA999577	ATP:guanidino kinase—Smc 74 ( <i>Schistosoma mansoni</i> )	GB J05410	10	95	95	287	Blastn	1327	9.0e-126
SMC03228F	AA999384	Cytochrome B reductase ( <i>Nyctinomops lacticaudatus</i> )	GI 309611	1	63	46	79	BLASTX	210	1.1e-030
MC210217NA3	AA559772	Cytochrome C ( <i>Boettcherisca peregrina</i> )	SP P12831	4	84	73	107	BLASTX	445	2.9e-056
SMC03406F	AA999503	Cytochrome oxidase subunit I ( <i>Fasciola hepatica</i> )	GI 471038	12	84	78	57	BLASTX	231	1.4e-050
MC070579NA3	AA559572	Cytochrome oxidase subunit II ( <i>Schlechtendalia chinensis</i> )	GI 1036829	3	74	49	63	BLASTX	181	9.1e-020
SMC03092F	AA999283	Enolase ( <i>Schistosoma mansoni</i> )	GB U33177	3	100	100	186	BLASTN	930	2.7e-110
SMC03319F	AA999456	Fructose-1,6-bisphosphate aldolase ( <i>Schistosoma mansoni</i> )	GB L38658	25	100	100	267	BLASTN	1335	1.5e-103
MC090719NA3	AA559629	Glucose-6-phosphate isomerase ( <i>Drosophila melanogaster</i> )	SP P52031	1	86	80	89	BLASTX	417	8.2e-052
SMC03404F	AA999501	Glyceraldehyde 3-phosphate dehydrogenase ( <i>Schistosoma mansoni</i> )	GB M92359	16	100	100	149	blastn	745	1.9e-081

Table 2 (Continued)

Clone identification	EST number <sup>a</sup>	Putative homologue (organism)	Accession number <sup>b</sup>	Number of clones	Similarity (%)	Identity (%)	Length	Program	Score	Probability
SMC03140F	AA999325	Glycerol-3-phosphate dehydrogenase—NAD <sup>+</sup> ( <i>Drosophila melanogaster</i> )	PIR A32937	2	70	56	55	BLASTX	177	2.0e-017
MC240711NA3	AA559650	Lactato dehydrogenase H4 and S-lac-NAD <sup>+</sup> complex ( <i>Sus scrofa</i> )	PDB 5LDH	1	77	67	103	BLASTX	356	9.0e-043
MC250466NA3	AA559800	Muscle glycogen phosphorylase ( <i>Homo sapiens</i> )	GI 187595	1	75	49	61	BLASTX	154	3.8e-033
MC070522NA3	AA559531	NADH dehydrogenase subunit 3 ( <i>Fasciola hepatica</i> )	GI 471037	1	50	35	76	BLASTX	127	4.4e-011
SMC03126F	AA999312	NADH-ubiquinone oxidoreductase 18 kD subunit precursor ( <i>Bos taurus</i> )	SP Q02375	1	71	61	77	BLASTX	257	1.7e-030
SMC03541F	AA999583	NADH-ubiquinone oxidoreductase 20 kD subunit precursor—complex I ( <i>Bos taurus</i> )	SP P42026	1	87	77	40	BLASTX	176	1.7e-017
MC031038NA3	AA559699	NADH-ubiquinone oxidoreductase acyl-carrier subunit ( <i>Drosophila melanogaster</i> )	GNL Y09068	1	76	50	46	BLASTX	136	4.3e-012
MC080553NA3	AA559613	Phosphofructo kinase ( <i>Schistosoma mansoni</i> )	GB L31531	1	86	86	223	BLASTN	876	4.4e-065
SMC03105F	AA999293	Phosphoglycerate kinase—clone C5-1B3 ( <i>Schistosoma mansoni</i> )	GB L36833	2	99	99	190	BLASTN	943	2.2e-124
MC080524NA3	AA559596	Phosphoglycerate mutase ( <i>Homo sapiens</i> )	SP P15259	3	80	65	26	BLASTX	103	2.8e-023
MC260465NA3	AA559468	Pyruvate dehydrogenase E1 component, $\alpha$ subunit, type I precursor ( <i>Ascaris suum</i> )	SP P26267	1	73	51	49	BLASTX	127	2.9e-010

Table 2 (Continued)

Clone identification	EST number <sup>a</sup>	Putative homologue (organism)	Accession number <sup>b</sup>	Number of clones	Similarity (%)	Identity (%)	Length	Program	Score	Probability
MC270409NA3	AA559488	Pyruvate kinase (EC 2.7.1.40) isozyme M1 ( <i>Homo sapiens</i> )	PIR S64635	1	94	88	18	BLASTX	255	7.2e-027
SMC03302F	AA999444	Succinate dehydrogenase—ubiquinone ( <i>Homo sapiens</i> )	PIR A34045	2	78	59	32	BLASTX	110	2.4e-015
SMC03169F	AA999344	Triose phosphate isomerase ( <i>Schistosoma mansoni</i> )	GB L07286	6	99	99	275	BLASTN	1457	4.2e-116
MC031067NA3	AA559712	Ubiquinol cytochrome C reductase ( <i>Homo sapiens</i> )	PIR S00680	1	52	65	44	BLASTX	125	6.1e-009
SMC03160F	AA999337	Uncoupling protein homolog ( <i>Homo sapiens</i> )	GB U94592	1	91	91	79	BLASTN	332	1.3e-017
MC250410NA3	AA559402	Vacuolar ATP synthase 16 kD proteolipid subunit—ductin ( <i>Drosophila melanogaster</i> )	SP P23380	1	86	77	75	BLASTX	293	3.5e-040
<i>Other metabolism</i>										
SMC0233R	AA999215	Cytochrome P-450 P450CR10 ( <i>Catharanthus roseus</i> )	EMB X69782	1	74	74	75	BLASTN	208	3.1e-007
MC031064NA3	AA559710	Glutamine synthetase—glutamate—ammonia ligase ( <i>Xenopus laevis</i> )	SP P51121	3	83	73	67	BLASTX	263	2.3e-029
SMC03477F	AA999552	Glutathione peroxidase ( <i>Schistosoma mansoni</i> )	GB M86510	5	100	100	70	BLASTN	350	1.7e-020
SMC03400F	AA999498	Glutathione S-transferase gene ( <i>Schistosoma mansoni</i> )	GB M73624	2	98	98	368	BLASTN	1804	5.6e-166
SMC03074F	AA999273	Hydroxymethylglutaryl CoA reductase 1 ( <i>Hevea brasiliensis</i> )	EMB X54657	2	83	83	62	BLASTN	215	2.4e-008
MC180922NA3	AA559673	O-linked N-acetylglucosamine transferase, p110 subunit ( <i>Rattus norvegicus</i> )	GI 1931579	1	73	57	96	BLASTX	288	1.3e-032

Table 2 (Continued)

Clone identification	EST number <sup>a</sup>	Putative homologue (organism)	Accession number <sup>b</sup>	Number of clones	Similarity (%)	Identity (%)	Length	Program	Score	Probability
SMC03137F	AA999323	Phosphatidylinositol transfer protein $\alpha$ (Pitpn) ( <i>Mus musculus</i> )	GB U96726	1	68	68	111	BLASTN	210	1.5e-007
<i>Transporters</i>										
MC250481NA3	AA559810	ADP, ATP carrier protein 3 (ADP/ATP translocase 3)—( <i>Saccharomyces cerevisiae</i> )	SP P18238	1	85	71	88	BLASTX	345	2.1e-041
SMC0119R	AA999213	Coat protein gamma-cop ( <i>Bos taurus</i> )	PIR S33957	1	71	57	42	BLASTX	139	1.4e-031
MC070517NA3	AA559528	Epsilon subunit of mitochondrial F1-ATPase ( <i>Arabidopsis thaliana</i> )	GNL D88377	1	51	60	43	BLASTX	116	1.6e-010
MC140336NA3	AA559756	Natural resistance-associated macrophage protein 2 (NRAMP2)—( <i>Homo sapiens</i> )	SP P49281	1	76	58	34	BLASTX	105	1.7e-022
MC250457NA3	AA559795	Peripheral-type benzodiazepine receptor (PBR)—( <i>Homo sapiens</i> )	SP P30536	1	64	52	42	BLASTX	135	2.3e-015
MC031026NA3	AA559694	Similar to clathrin coat assemble protein ( <i>Caenorhabditis elegans</i> )	GI 2315354	1	75	62	85	BLASTX	277	3.5e-033
<i>Proteases and inhibitors</i>										
SMC03244F	AA999395	$\alpha$ -Proteasome ( <i>Entamoeba histolytica</i> )	EMB x99382	1	80	80	65	BLASTN	208	2.6e-007
MC260451NA3	AA559458	26 S protease regulatory subunit 7—MSS1 protein homolog ( <i>Mus musculus</i> )	PIR S39349	1	88	70	94	BLASTX	363	2.0e-041
MC080539NA3	AA559606	26S protease regulatory subunit 4 ( <i>Homo sapiens</i> )	SP Q03527	1	83	68	103	BLASTX	371	3.6e-044

Table 2 (Continued)

Clone identification	EST number <sup>a</sup>	Putative homologue (organism)	Accession number <sup>b</sup>	Number of clones	Similarity (%)	Identity (%)	Length	Program	Score	Probability
MC250418NA3	AA559406	ATP-dependent C1p proteinase (EC 3.4.21.92) chain P homolog ( <i>Homo sapiens</i> )	PIR 568421	1	75	47	48	BLASTX	137	1.9e-019
SMC03520F	AA999575	Cathepsin B (Sm 31) ( <i>Schistosoma mansoni</i> )	GB M21309	2	91	91	289	BLASTN	1224	2.2e-094
MC070564NA3	AA559560	Plasminogen activator inhibitor-2, type A—PAI2A ( <i>Rattus norvegicus</i> )	SP P29524	1	53	34	66	BLASTX	125	4.7e-010
SMC03356F	AA999481	X-Pro dipeptidase-EC 3.4.13.9 ( <i>Homo sapiens</i> )	PIR A61015	1	77	52	57	BLASTX	153	5.8e-014
<i>DNA scaffold</i>										
MC070518NA3	AA559529	52k active chromatin boundary protein ( <i>Drosophila melanogaster</i> )	PIR A37282	1	47	30	84	BLASTX	100	1.1e-008
MC031008NA3	AA559689	Histone H1 ( <i>Schistosoma mansoni</i> )	GB U54584	2	100	100	161	BLASTN	805	9.2e-066
SMC03261F	AA999409	Histone H3.3 (H3.B) (H3.3Q)—( <i>Homo sapiens</i> )	SP P06351	1	100	98	91	BLASTX	452	5.7e-071
<i>Structural and cytoskeletal</i>										
MC090727NA3	AA559633	$\alpha$ Actinin-actin binding protein ( <i>Caenorhabditis elegans</i> )	PIR A48403	1	59	40	101	BLASTX	207	7.8e-021
SMC03205F	AA999365	$\alpha$ -Tubulin ( <i>Schistosoma mansoni</i> )	GB M80214	2	100	100	400	BLASTN	2000	3.6e-164
SMC03047F	AA999252	Actin ( <i>Schistosoma mansoni</i> )	GB U19945	5	99	99	480	BLASTN	2391	1.7e-226
SMC03201F	AA999361	Fibrillin 2 ( <i>Schistosoma mansoni</i> )	GB U54590	2	89	89	167	BLASTN	673	5.9e-070
MC260444NA3	AA559453	Nuclear movement protein (MNUDC protein)—( <i>Mus musculus</i> )	GNL Y15522	1	91	81	49	BLASTX	214	1.9e-026

Table 2 (Continued)

Clone identification	EST number <sup>a</sup>	Putative homologue (organism)	Accession number <sup>b</sup>	Number of clones	Similarity (%)	Identity (%)	Length	Program	Score	Probability
SMC03235F	AA999390	Paramyosin ( <i>Schistosoma mansoni</i> )	GB m35499	1	97	97	157	BLASTN	749	4.3e-058
MC090746NA3	AA559641	Tropomyosin ( <i>Schistosoma mansoni</i> )	GB M27512	1	98	98	296	BLASTN	1459	5.0e-118
<i>Tegumental/membrane proteins</i>										
SMC03225F	AA999381	23 kDa integral membrane protein ( <i>Schistosoma mansoni</i> )	GB M34453	5	100	100	132	BLASTN	660	3.9e-057
SMC03135F	AA999321	Major tegumental antigen Sm 15 ( <i>Schistosoma mansoni</i> )	GB L07513	1	97	177	177	BLASTN	849	4.0e-090
MC080562NA3	AA559617	Tegumental antigen Sm 22.6 ( <i>Schistosoma mansoni</i> partial match)	GB M37003	1	61	61	150	BLASTN	228	6.4e-009
SMC03283F	AA999427	Tegumental antigen Sm20.8 ( <i>Schistosoma mansoni</i> partial match)	GB U86673	5	67	67	180	BLASTN	369	1.3e-020
<i>Kinases and phosphatases</i>										
MC260427NA3	AA559442	5'-AMP-activated protein kinase, gamma-1-subunit ( <i>Homo sapiens</i> )	SP P54619	1	81	59	37	BLASTX	117	5.8e-009
MC250423NA3	AA559411	Adenylate kinase AK2 ( <i>Bos taurus</i> )	PRF 1008165A	2	78	55	90	BLASTX	281	1.1e-033
MC210271NA3	AA559783	Casein kinase I $\alpha$ (CKI- $\alpha$ ) ( <i>Drosophila melanogaster</i> )	GB U55848	1	95	95	46	BLASTN	212	1.3e-007
SMC03129F	AA999315	Nucleoside monophosphate kinase ( <i>Schistosoma mansoni</i> )	GB M80542	2	91	91	239	BLASTN	1006	4.0e-076
SMC03549F	AA999589	Protein kinase—Dd GPK2 ( <i>Dictyostelium discoideum</i> )	GB M38703	1	73	73	102	BLASTN	267	1.4e-015
MC080555NA3	AA559615	Serine/threonine protein phosphatase pp2a- $\alpha$ catalytic subunit ( <i>Homo sapiens</i> )	SP P05323	1	80	71	56	BLASTX	212	1.7e-036



Table 2 (Continued)

Clone identification	EST number <sup>a</sup>	Putative homologue (organism)	Accession number <sup>b</sup>	Number of clones	Similarity (%)	Identity (%)	Length	Program	Score	Probability
<i>Regulatory and signaling proteins</i>										
SMC03015F	AA999233	Calcium binding protein Sm8 ( <i>Schistosoma mansoni</i> )	GB J04641	67	100	100	289	BLASTN	1445	6.2e-115
MC280420NA3	AA559511	Calmodulin related protein ( <i>Homo sapiens</i> )	GI 167676	9	51	36	66	BLASTX	109	9.2e-007
SMC03078F	AA999276	Calponin homolog ( <i>Schistosoma mansoni</i> )	GB U86674	1	95	95	138	BLASTN	636	9.9e-057
SMC03096F	AA999287	Cyclin B ( <i>Cricetulus longicaudatus</i> )	EMB X64588	1	86	86	58	BLASTN	218	7.1e-008
SMC0257R	AA999217	Cysteine-rich protein 2—ESP1 protein ( <i>Rattus norvegicus</i> )	SP P36201	3	76	62	64	BLASTX	246	4.6e-028
MC250424NA3	AA559412	DNAJ protein homolog 2 ( <i>Homo sapiens</i> )	SP P31689	1	70	50	97	BLASTX	264	1.7e-029
MC180929NA3	AA559680	Gq-coupled rhodopsin ( <i>Patinopecten yessoensis</i> )	DBJ AB006454	1	67	45	55	BLASTX	141	6.1e-030
SMC03405F	AA999502	Heat shock protein 86 gene-clone Dd2 ( <i>Plasmodium falciparum</i> )	GB L34027	1	73	73	76	BLASTN	200	1.6e-009
MC090718NA3	AA559628	Human homolog of the fission yeast skb1 gene ( <i>Homo sapiens</i> )	GI 2323410	1	73	67	49	BLASTX	165	8.3e-035
SMC03648F	AA999635	Polyubiquitin ( <i>Cricetulus griseus</i> )	GI 49477	1	100	100	100	BLASTX	496	6.5e-064
SMC0220R	AI264972	RAS-like GTP binding protein—Rho ( <i>Aplysia californica</i> )	SP P01122	1	94	89	96	BLASTX	458	1.8e-069
MC090712NA3	AA559625	Rod phosphodiesterase $\delta$ subunit ( <i>Bos taurus</i> )	GI 1565306	1	78	61	75	BLASTX	250	2.1e-029
MC100409NA3	AA559394	Sorcin ( <i>Schistosoma japonicum</i> )	GI 1655733	1	75	64	95	BLASTX	325	5.9e-047

Table 2 (Continued)

Clone identification	EST number <sup>a</sup>	Putative homologue (organism)	Accession number <sup>b</sup>	Number of clones	Similarity (%)	Identity (%)	Length	Program	Score	Probability
<i>Transcription factors</i>										
MC090722NA3	AA559630	ADR6 gene for transcription factor ( <i>Saccharomyces cerevisiae</i> )	EMBLX12493	1	72	72	81	BLASTN	209	2.8e-007
MC260470NA3	AA559472	OCP2 putative transcription factor expressed in the inner ear ( <i>Homo sapiens</i> )	GI 1017813	1	93	74	43	BLASTX	198	3.0e-043
SMC03296F	AA999439	Y-box binding protein ( <i>Schistosoma mansoni</i> )	GB U39883	3	99	99	294	BLASTN	1454	1.4e-113
<i>Transcription and translation machinery</i>										
SMC03055F	AA999258	40S ribosomal protein S4 ( <i>Drosophila melanogaster</i> )	SP P41042	1	84	65	89	BLASTX	336	1.3e-041
SMC03189F	AA999352	40S ribosomal protein S5 ( <i>Caenorhabditis elegans</i> )	SP P49041	1	100	95	62	BLASTX	297	4.1e-036
MC240706NA3	AA559648	40S ribosomal protein S9 ( <i>Rattus norvegicus</i> )	SP P29314	1	83	79	62	BLASTX	244	2.2e-034
MC210261NA3	AA559782	40S ribosomal protein S28 ( <i>Homo sapiens</i> )	SP P25112	1	86	72	59	BLASTX	223	2.4e-025
MC180904NA3	AA559661	60S ribosomal protein L36 ( <i>Rattus norvegicus</i> )	SP P39032	1	72	57	75	BLASTX	223	1.6e-025
MC180932NA3	AA559682	Cyclophilin ( <i>Schistosoma mansoni</i> )	GB L46884	1	94	94	274	BLASTN	1260	1.1e-099
SMC03526F	AA999578	Cyclophilin B-like mRNA ( <i>Schistosoma mansoni</i> )	GB U30874	1	98	98	185	BLASTN	898	7.6e-099
MC070530NA3	AA559536	Elongation factor 1- $\alpha$ ( <i>Schistosoma mansoni</i> )	EMBL Y08487	1	98	98	306	BLASTN	1476	3.0e-115
MC180948NA3	AA559687	Ganglioside expression factor-2 (GEF-2)—( <i>Rattus norvegicus</i> )	GNL AB003515	1	78	68	41	BLASTX	140	1.1e-036
MC260462NA3	AA559466	Nucleolin ( <i>Xenopus laevis</i> )	PIR S32644	1	48	48	52	BLASTX	136	1.0e-008

Table 2 (Continued)

Clone identification	EST number <sup>a</sup>	Putative homologue (organism)	Accession number <sup>b</sup>	Number of clones	Similarity (%)	Identity (%)	Length	Program	Score	Probability
SMC03097F	AA999288	P43 mitochondrial elongation factor homolog ( <i>Homo sapiens</i> )	GI 833999	1	61	48	49	BLASTX	113	2.1e-019
SMC03198F	AA999359	Polypirimidine tract-binding protein ( <i>Sus scrofa</i> )	GNL X93009	1	80	65	41	BLASTX	130	9.2e-011
MC250434NA3	AA559420	Ribosomal protein S2 ( <i>Homo sapiens</i> )	PIR S08228	1	80	71	87	BLASTX	299	2.3e-036
MC280405NA3	AA559507	RNA polymerase II chain hRPB17 ( <i>Homo sapiens</i> )	PIR S55370	1	80	66	60	BLASTX	209	1.5e-020
SMC03087F	AA999281	RNA polymerase iii large subunit gene ( <i>Plasmodium falciparum</i> )	GB M73770	1	81	81	78	BLASTN	222	3.5e-012
SMC0217R	AA999214	Transcriptional regulator protein HCNGP ( <i>Mus musculus</i> )	SP Q02614	1	61	44	63	BLASTX	141	1.4e-010
<i>Others</i>										
MC250461NA3	AA559797	Antigen Sm 21.7 ( <i>Schistosoma mansoni</i> )	GB U30663	1	97	97	269	BLASTN	1292	3.4e-117
MC031061NA3	AA559709	C43G2.1 gene product ( <i>Caenorhabditis elegans</i> )	GI 1572756	1	75	58	48	BLASTX	165	1.3e-015
MC270424NA3	AA559497	Contains two c2-like domains, exon 7 and exons 15 and 16 ( <i>Caenorhabditis elegans</i> )	GI 529355	1	56	43	51	BLASTX	101	9.3e-008
MC080518NA3	AA559590	Cosmid T20G5 ( <i>Caenorhabditis elegans</i> )	EMB Z30423	1	68	68	123	BLASTN	264	7.0e-012
SMC03166F	AA999342	Hypothetical 17.2 kDa protein F44E2.6 in chromosome III ( <i>Caenorhabditis elegans</i> )	GI 1788077	1	68	50	83	BLASTX	257	2.2e-030
SMC03414F	AA999508	Hypothetical 63.5 kDa protein ZK 353.1 in chromosome III ( <i>Caenorhabditis elegans</i> )	SP P34624	1	57	43	71	BLASTX	151	1.1e-013

Table 2 (Continued)

Clone identification	EST number <sup>a</sup>	Putative homologue (organism)	Accession number <sup>b</sup>	Number of clones	Similarity (%)	Identity (%)	Length	Program	Score	Probability
MC260460NA3	AA559465	Hypothetical 69.9 kD protein in MIC1-SRB5 intergenic region ( <i>Homo sapiens</i> )	GB U05340	1	75	59	44	BLASTX	156	3.1e-013
SMC03168F	AA999343	Hypothetical protein (z69385) zk593.7 ( <i>Caenorhabditis elegans</i> )	GNL Z69385	1	69	49	59	BLASTX	155	3.8e-016
SMC03586F	AA999614	Knob associated histidine-rich protein ( <i>Plasmodium falciparum</i> )	GB S39048	2	79	79	77	BLASTN	295	1.8e-016
SMC03021F	AA999239	mRNA for tandem repeat ( <i>Schistosoma mansoni</i> )	EMB Z29960	1	69	69	91	BLASTN	203	7.5e-008
MC260493NA3	AA559483	P125 protein (Pestivirus, bovine viral diarrhea virus type I)	GI 567767	1	98	98	101	BLASTX	494	4.4e-059
SMC03475F	AA999550	Putative ATPase n2b (HFN2B) —( <i>Haematobia irritans</i> )	SP P46441	1	79	55	34	BLASTX	112	5.9e-007
MC080549NA3	AA559610	Set $\alpha$ isoform ( <i>Rattus</i> sp.)	PIR I51908	1	87	72	55	BLASTX	217	9.5e-027
SMC03104F	AA999292	Synthetic mRNA for p28II protein (synthetic construct)	EMB A03660	1	97	97	136	BLASTN	653	1.4e-086
MC041007NA3	AA559731	Translationally controlled tumor protein—TCTP ( <i>Schistosoma japonicum</i> )	GI 1816592	1	72	53	65	BLASTX	184	1.2e-036
MC031091NA3	AA559723	Trans-spliced mRNA, clone SL cDNA-6 ( <i>Schistosoma mansoni</i> )	GB U30264	1	98	98	280	BLASTN	1357	7.7e-116
SMC03641F	AA999630	Uteroglobin—CC10 gene ( <i>Homo sapiens</i> )	EMB X59875	1	80	80	62	BLASTN	202	8.7e-007

<sup>a</sup> ESTs identified in the databases are listed with the description of the sequence matched, the length and percent of similarity and identity of the alignment. Matches were sorted into functional categories and only one representative match is given. Match lengths are in nucleotides for DNA matches and amino acid residues for peptide matches. DNA and peptide homologues are indicated, respectively, by BLASTN and BLASTX local alignment search programs. Percent identity refers to the percentage of identical nucleotides or amino acids residues in the alignment of the EST with the database sequence. For amino acid matches, percent similarity refers to the percentage of identical plus conservative amino acid substitutions in the alignment.

<sup>b</sup> Database abbreviations: DDBJ, DNA Data Bank of Japan; EMB, European Molecular Biology Laboratory Data Library; GB, GenBank™; GI, NCBI's general database identifier; GNL, NCBI's general database identifier; PDB, Brookhaven Protein Data Bank; PIR, Protein Information Resource; PRF, Protein Research Foundation; SP, SwissProt.

and calmodulin NB-1 related protein in library 2) is indicative of the existence of distinct biases, possibly arising from cloning procedures, and differences in the cercaria-biology. The possibility that the in vivo excision of library 1 also introduced bias cannot be discounted.

A more precise characterization of the diversity of gene expression and physiological activity in cercariae can be ascertained by sorting the identified genes into functional categories, as shown in Table 2. Genes related to energy metabolism formed the largest category, which might be expected considering that cercariae need much energy for rapid swimming and for body contraction, in order to find and penetrate the skin of their vertebrate host. This finding is in accordance with previous expression studies of energy metabolism genes using Northern blot techniques [17–19]. The detection of genes in-

volved in transcriptional and translational activities or with regulatory and signaling functions is also considered significant as the cercariae need to be fully prepared for the rapid morphological and physiological changes that will be necessary for their transformation into schistosomula. The EST approach provides evidence for the importance of these genes in the overall transcriptional activity of the cercariae.

New genes that were similar to, but not identical to, known genes of *S. mansoni* (i.e. <95% homology at nucleotide level) were found, suggesting that these may be new members of gene families that include several genes with similar sequences and probably similar functions. These genes are represented by the Sm partial match category (Table 1) and include sequences similar to Sm8 calcium binding protein, to Sm 20.8 tegumental antigen, to HMG-CoA reductase and to

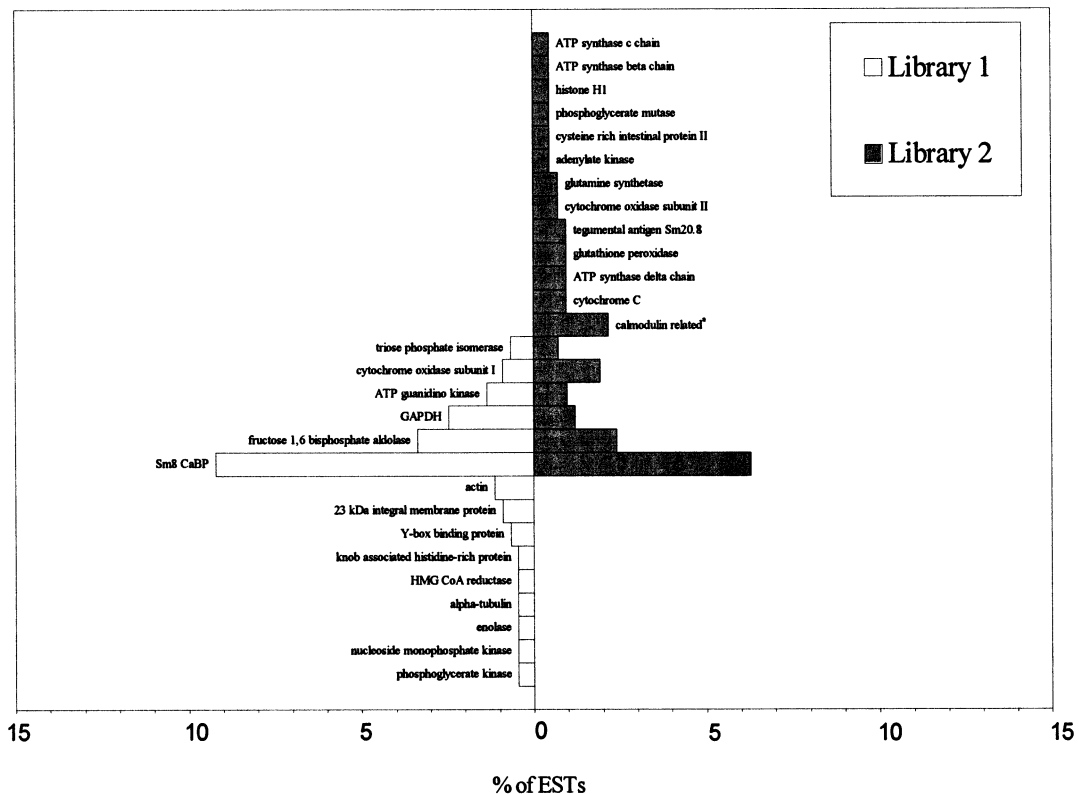


Fig. 2. Comparison of the gene expression profiles of the two cercarial cDNA libraries. The frequency of each redundant gene is shown as a percentage of the total 'informative' ESTs for each library. <sup>a</sup>Includes Sm8 CaBP partial match ESTs (see text).

Sm 22.6 antigen. ICATOOLS grouped the four ESTs showing partial homology with Sm8 calcium binding protein into the same cluster as ESTs matching human calmodulin NB-1. Therefore, we considered all ESTs within this cluster to be 'calmodulin related'.

The most abundant gene found in both cercarial cDNA libraries (nearly 8% of all ESTs (Fig. 2)) showed over 95% nucleotide identity with Sm8, an 8-kDa calcium binding protein (CaBP) from *S. mansoni* that was originally identified by anti-schistosome antibodies in assays to identify genes expressed in a stage-specific manner [1]. CaBPs are interesting molecules because most metabolic processes (e.g. activation of regulatory enzymes such as kinases) and physiological events (e.g. muscle contraction) that are triggered by  $\text{Ca}^{2+}$  are mediated through calcium binding proteins. Intriguingly, expression of Sm8 only occurs after release from the snail (within 1 h) and ceases within 24 h of transformation into the schistosomulum, thus showing a rapid change in gene expression characterized by high levels of specific mRNA [1]. The precise physiological significance of Sm8 expression in cercariae remains unclear. Ram et al. [20] have suggested that Sm8 CaBP is synthesized in the head gland and subtegumental cells and translocated to the tegument where it plays a role in modifications required for adaptation to parasitic life in the host. These authors also showed that it was not found in muscle mitochondria, suggesting that it is not involved in the rapid motility and aerobic metabolism that are characteristic of cercariae [20].

Another redundant member of the EF-hand CaBP superfamily identified in library 2 was a gene similar to the human calmodulin related protein NB-1. The identity between the cercarial EST and the human protein was about 36% at the amino acid level. In man this protein probably performs similar functions to authentic calmodulins and has been identified as a down-regulated gene during the transformation of human mammary epithelial cells [21]. An EST showing 64% amino acid identity to sorcin, another protein from the CaBP family, was also found in library 2. Sorcin, a 22-kDa  $\text{Ca}^{2+}$ -binding protein identified from multidrug-resistant cells has been

shown to be most abundant in cardiac tissue, but is also present in kidney, brain and skeletal muscle [22]. Its biological role has not been determined but there is evidence that a calcium-dependent translocation process moves the sorcin protein from the cytosol to cellular membranes in both cardiac tissue and multidrug-resistant cells. This translocation process may affect ionic flux through the membranes of sorcin rich tissues [22]. It is conceivable that sorcin plays a similar role as a calcium homeostasis regulator in cercariae.

Some interesting matches to known proteins from *S. mansoni* and other organisms were observed that may provide important insights into the physiology of cercariae and adult worms. Amongst the structural genes identified were those for the *S. mansoni* muscle proteins paramyosin and tropomyosin (Table 1). This is interesting because previous Northern blot analyses had identified paramyosin and myosin mRNAs in sporocysts and adult worms, but not in cercariae, despite the fact that the protein itself was found in cercariae. This suggested that the protein was synthesized at the sporocyst stage and retained throughout the life of the cercariae [23], leading to the proposal that genes for *S. mansoni* muscle proteins were subject to stage-specific regulation. However, the existence of paramyosin and tropomyosin transcripts in library 2 indicates that these muscle genes are not subject to stage-specific regulation, in accordance to previous findings for schistosome myosin [24].

Two ESTs showed significant homology with genes encoding important proteins involved in mammalian light perception mechanisms, Gq-coupled rhodopsin photoreceptor and rod phosphodiesterase  $\delta$  subunit. *S. mansoni* cercariae alternate between passive drifting and active swimming, the latter being stimulated by alterations in light intensity [25]. Passive cercariae show a pronounced response to shadow ('inverse photokinesis'—where stimulation of swimming activity is proportional to the amount of shadow) and a lesser response to increased illumination ('direct photokinesis'). These two responses may be mediated by separate photoreceptor systems in cercariae. It has been proposed [25] that direct

photokinesis could result from light absorption by nerves, tail muscle or other tissue whereas inverse photokinesis is mediated by ciliary receptors (as in other invertebrates [25]). Two ciliary structures discovered near the anterior end of *S. mansoni* cercariae have been interpreted as larval photoreceptors [26] and may be implicated in this response [25] although the mechanisms that control and modify cercarial swimming behavior are not yet understood at the neurophysiological level [27]. At the molecular level, invertebrate organisms employ a light perception mechanism similar to that of vertebrates, being based on 11-*cis*-retinal chromophore and rhodopsin [28]. EST AA559680 showed 67% homology to a molluscan G<sub>q</sub>-coupled rhodopsin, a serpentine photoreceptor protein containing seven transmembrane helices. The chromophore, which lies near the center of the protein, undergoes a conformational change after it has absorbed a photon and the photoexcited rhodopsin triggers the activation of a GTP binding protein (G protein) [28]. Downstream of the G protein, vertebrate and invertebrate phototransduction pathways are divergent [29]. In vertebrates, photoexcited rhodopsin activates the heterotrimeric G-protein, transducin (G<sub>t</sub>), by forming a complex with it and catalyzing the exchange of GTP for bound GDP. The binding of GTP to transducin stimulates cGMP phosphodiesterase, which very rapidly hydrolyzes cyclic GMP, leading to transmission of the excitation stimulus. In invertebrates, both inositol trisphosphate (IP<sub>3</sub>) and Ca<sup>2+</sup> may be photoexcitation transmitters rather than a cyclic GMP phosphodiesterase [28]. Interestingly, cercarial EST AA559625 showed 78% homology to an 8-kD  $\delta$  inhibitory subunit of bovine rod cGMP phosphodiesterase (i.e. a vertebrate light response protein). This observation is not without precedent as it has been reported that the hyperpolarizing response of scallop photoreceptor cells also involves a cGMP-sensitive potassium channel [29]. These EST matches may provide clues to an understanding of the biochemical mechanism by which cercariae can respond to light.

We found also a cercarial gene displaying significant (64%) homology to the human PBRs. These receptors are multimeric polypeptides that

are located mainly on the outer mitochondrial membrane and which bind with high affinity to 4'-chlorodiazepam [30]. They have been identified in virtually all mammalian peripheral tissues and may also exist on the plasma membrane. PBRs are believed to be involved in several functions, including cell growth, calcium channel modulation and mitochondrial respiration. In particular, they are directly involved in mitochondrial steroidogenesis [30]. In normal metabolism the rate of steroid formation is limited by the rate at which cholesterol is transported from intracellular stores to the inner mitochondrial membrane and loaded onto cytochrome P450 side-chain cleavage enzyme, where it is converted to pregnenolone steroid precursor [31]. It has been observed that the expression of PBRs is especially high in steroidogenic organs [32]. Recent *in vitro* studies suggested that a key element in the translocation of cholesterol from the outer to the inner mitochondrial membrane is the mitochondrial PBR [33]. Targeted disruption of the PBR gene in steroidogenic cells dramatically reduces the ability of the cells to transport cholesterol to the mitochondria and produce steroids [34]. Studies on the lipid composition of *S. mansoni* show that there are differences in sterol contents between cercariae, schistosomula, and adult worms. In particular, cercariae and schistosomula hold primarily cholesterol and desmosterol, as well as several minor sterols, whereas adults contain only cholesterol, which is absorbed from their hosts [35]. Hence, the existence of PBRs (and thus, cholesterol conversion pathways) could explain the presence of these sterols in cercariae.

A novel member of the RAS superfamily in *S. mansoni* matches the Rho GTP binding protein from *Aplysia californica* with 89% identity over 96 amino acid residues. Rho GTPases are involved in regulation of actin assembly. The molecular characterization and functional analysis of this gene are underway at the Universidade Federal de Minas Gerais, Brazil.

In conclusion, EST methodology has proved to be an effective tool for studying gene expression in cercariae. It has allowed analysis of gene expression in this important stage of the parasite life cycle and has revealed several interesting new

genes, increasing and complementing our knowledge of the biology of *S. mansoni*.

### Acknowledgements

The authors thank Kátia Barroso for carrying out automated DNA sequencing at UFMG. This investigation received financial support from the following sources: PADCT, CNPq, UNDP/WORLD BANK/WHO Special Program for Research and Training in Tropical Diseases (TDR No.: 940325 and 940751), USAID/HOH (No.: 264.01.01.04)

### References

- [1] Ram D, Grossman Z, Markovics A, et al. Rapid changes in the expression of a gene encoding a calcium-binding protein in *Schistosoma mansoni*. *Mol Biochem Parasitol* 1989;34:167–75.
- [2] Moser D, Doenhoff MJ, Klinkert MQ. A stage-specific calcium-binding protein expressed in eggs of *Schistosoma mansoni*. *Mol Biochem Parasitol* 1992;51:229–38.
- [3] Fishelson Z, Amiri P, Friend DS, et al. *Schistosoma mansoni*: cell-specific expression and secretion of a serine protease during development of cercariae. *Exp Parasitol* 1992;75:87–98.
- [4] Neumann S, Ziv E, Lantner F, Schechter I. Regulation of HSP70 gene expression during the life cycle of the parasitic helminth *Schistosoma mansoni*. *Eur J Biochem* 1993;212:589–96.
- [5] Adams MD, Soares MB, Kerlavage AR, Fields C, Venter JC. Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nat Genet* 1993;4:373–80.
- [6] Hillier LD, Lennon G, Becker M, et al. Generation and analysis of 280 000 human expressed sequence tags. *Genome Res* 1996;6:807–28.
- [7] Adams MD, Kerlavage AR, Fields C, Venter JC. 3400 new expressed sequence tags identify diversity of transcripts in human brain. *Nat Genet* 1993;4:256–67.
- [8] Blaxter ML, Raghavan N, Ghosh I, et al. Genes expressed in *Brugia malayi* infective third stage larvae. *Mol Biochem Parasitol* 1996;77:77–93.
- [9] Adams MD, Kerlavage AR, Fleischmann RD, et al. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* 1995;377:3–174.
- [10] Lee NH, Weinstock KG, Kirkness EF, et al. Comparative expressed-sequence-tag analysis of differential gene expression profiles in PC-12 cells before and after nerve growth factor treatment. *Proc Natl Acad Sci USA* 1995;92:8303–7.
- [11] Manger ID, Hehl A, Parmley S, et al. Expressed sequence tag analysis of the bradyzoite stage of *Toxoplasma gondii*: identification of developmentally regulated genes. *Infect Immun* 1998;66:1632–7.
- [12] Franco GR, Adams MD, Soares MB, Simpson AJ, Venter JC, Pena SD. Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. *Gene* 1995;152:141–7.
- [13] Rabelo EM, Franco GR, Azevedo VA, et al. Update of the gene discovery program in *Schistosoma mansoni* with the expressed sequence tag approach. *Mem Inst Oswaldo Cruz* 1997;92:625–9.
- [14] Franco GR, Rabelo EM, Azevedo V, et al. Evaluation of cDNA libraries from different developmental stages of *Schistosoma mansoni* for production of expressed sequence tags (ESTs). *DNA Res* 1997;4:231–40.
- [15] Williams SA, Johnston DA. Helminth genome analysis: the current status of the filarial and schistosome genome projects. *Parasitology* 1999;in press.
- [16] Parsons JD, Brenner S, Bishop MJ. Clustering cDNA sequences. *Comput Appl Biosci* 1992;8:461–6.
- [17] Skelly PJ, Stein LD, Shoemaker CB. Expression of *Schistosoma mansoni* genes involved in anaerobic and oxidative glucose metabolism during the cercaria to adult transformation. *Mol Biochem Parasitol* 1993;60:93–104.
- [18] Skelly PJ, Shoemaker CB. A molecular genetic study of the variations in metabolic function during schistosome development. *Mem Inst Oswaldo Cruz* 1995;90:281–4.
- [19] Stein LD, Harn DA, David JR. A cloned ATP: guanidino kinase in the trematode *Schistosoma mansoni* has a novel duplicated structure. *J Biol Chem* 1990;265:6582–8.
- [20] Ram D, Romano B, Schechter I. Immunochemical studies on the cercarial-specific calcium binding protein of *Schistosoma mansoni*. *Parasitology* 1994;108:289–300.
- [21] Yaswen P, Smoll A, Peehl DM, Trask DK, Sager R, Stampfer MR. Down-regulation of a calmodulin-related gene during transformation of human mammary epithelial cells. *Proc Natl Acad Sci USA* 1990;87:7360–4.
- [22] Meyers MB, Zamparelli C, Verzili D, Dicker AP, Blanck TJ, Chiancone E. Calcium-dependent translocation of sorcin to membranes: functional relevance in contractile tissue. *FEBS Lett* 1995;357:230–4.
- [23] Grossman Z, Ram D, Markovics A, et al. *Schistosoma mansoni*: stage-specific expression of muscle-specific genes. *Exp Parasitol* 1990;70:62–71.
- [24] Newport GR, Harrison RA, McKerrow J, Tarr P, Kallestad J, Agabian N. Molecular cloning of *Schistosoma mansoni* myosin. *Mol Biochem Parasitol* 1987;26:29–38.
- [25] Saladin KS. *Schistosoma mansoni*: cercarial responses to irradiance changes. *J Parasitol* 1982;68:120–4.
- [26] Short RB, Gagne HT. Fine structure of possible photoreceptor in cercariae of *Schistosoma mansoni*. *J Parasitol* 1975;61:69–74.



- [27] Haas W. Physiological analyses of host-finding behaviour in trematode cercariae: adaptations for transmission success. *Parasitology* 1994;109(Suppl.):S15–29.
- [28] Yarfitz S, Hurley JB. Transduction mechanisms of vertebrate and invertebrate photoreceptors. *J Biol Chem* 1994;269:14329–32.
- [29] del Pilar Gomez M, Nasi E. Activation of light-dependent K<sup>+</sup> channels in ciliary invertebrate photoreceptors involves cGMP but not the IP<sub>3</sub>/Ca<sup>2+</sup> cascade. *Neuron* 1995;15:607–18.
- [30] Woods MJ, Williams DC. Multiple forms and locations for the peripheral-type benzodiazepine receptor. *Biochem Pharmacol* 1996;52:1805–14.
- [31] Papadopoulos V, Amri H, Boujrad N, et al. Peripheral benzodiazepine receptor in cholesterol transport and steroidogenesis. *Steroids* 1997;62:21–8.
- [32] Gavish M, Weizman R. Role of peripheral-type benzodiazepine receptors in steroidogenesis. *Clin Neuropharmacol* 1997;20:473–81.
- [33] Papadopoulos V. Structure and function of the peripheral-type benzodiazepine receptor in steroidogenic cells. *Proc Soc Exp Biol Med* 1998;217:130–42.
- [34] Papadopoulos V, Amri H, Li H, Boujrad N, Vidic B, Garnier M. Targeted disruption of the peripheral-type benzodiazepine receptor gene inhibits steroidogenesis in the R2C Leydig tumor cell line. *J Biol Chem* 1997;272:32129–35.
- [35] Furlong ST, Caulfield JP. *Schistosoma mansoni*: sterol and phospholipid composition of cercariae schistosomula and adults. *Exp Parasitol* 1988;65:222–31.

## **PARTE II**

**Estudos de Genômica Funcional de uma Rho GTPase de**

***Schistosoma mansoni*:**

**Caracterização Molecular e Funcional de um Gene de *S. mansoni***

**Homólogo a uma Ras-Like GTP Binding Protein (Rho)**

## II.1. JUSTIFICATIVA

Esta segunda parte do trabalho se justifica em vista das importantes funções regulatórias desempenhadas pelas GTPases de baixo peso molecular da família Rho nos organismos eucarióticos e, sobretudo, pelo pouco conhecimento que se tem ainda hoje das funções desempenhadas por essas proteínas no desenvolvimento de *S. mansoni*, em especial, dos órgãos reprodutivos da fêmea do parasita e na postura de ovos.

## II.2. OBJETIVOS

### Objetivo geral:

Caracterizar molecular e funcionalmente um gene de *Schistosoma mansoni* homólogo a Rho GTPases que foi isolado de uma biblioteca de cDNA de cercárias e identificado por EST;

### Objetivos específicos:

- 1- Sequenciar por completo e caracterizar o clone de cDNA de um gene de *S. mansoni* homólogo a Rho GTPases que foi isolado de uma biblioteca de cDNA de cercárias e identificado por EST;
- 2- Clonar, sequenciar e caracterizar a sequência genômica completa do gene de *S. mansoni* homólogo a Rho GTPases;
- 3- Por meio de modelagem por homologia de estrutura de proteína, obter um modelo da estrutura tridimensional do polipeptídeo codificado pelo gene de *S. mansoni* homólogo a Rho GTPases e fazer a caracterização da molécula com base no modelo gerado *in silico*;
- 4- Estudar as funções do gene de *S. mansoni* que codifica uma Rho GTPase através da substituição do seu homólogo (*RHO1*) na levedura *Saccharomyces cerevisiae* pelo gene do parasita;

## **II.3. ARTIGO 4**

**Characterization and Complementation in Yeast of a Rho GTPase Gene From  
*Schistosoma mansoni* (em preparação)**

*Article for The Journal of Biological Chemistry*

(in preparation)

# **Characterization and Complementation in Yeast of a Rho GTPase Gene from *Schistosoma mansoni*\***

**Túlio M. Santos<sup>a</sup>, Carlos R. Machado<sup>a</sup>, Humberto M. Pereira<sup>b</sup>, Richard Garratt<sup>b</sup>, Sérgio D. J. Pena<sup>a#</sup>**

<sup>a</sup>*Departamento de Bioquímica e Imunologia, ICB - UFMG, CP 486, 31270-010, Belo Horizonte, MG, Brazil*

<sup>b</sup>*Departamento de Física e Informática, Instituto de Física de São Carlos, USP, CP 369, 13560-970, São Carlos, SP, Brazil*

## Summary

Low-molecular weight GTP-binding proteins (LMWGs) of the Ras superfamily play an important role in *Schistosoma mansoni* female development and sexual maturation. Therefore, the study of this large protein family in the blood fluke *S. mansoni* may open new perspectives on schistosomiasis control. Here we report the characterization of the first *S. mansoni* gene (*SMRHO*) that encodes a 193 amino acid polypeptide highly homologous to several Rho GTPase and Rho-type LMWGs from other eukaryotic organisms. We sequenced the cDNA and also the genomic sequence that showed three small introns. The protein sequence obtained by translation of the cDNA, has as peculiarities the insertion of a serine residue in the N-terminal end and the substitution of the first aliphatic residue from the CAAX prenylation motif for a polar residue. However, its molecular structure obtained by comparative computational modelling follows the conserved folding pattern of Rho-type GTPases. The *SMRHO* overexpression complements a *Sacharomyces cerevisiae rho1* null mutant strain even in restrictive conditions of Ca<sup>2+</sup> concentration and temperature. However, some phenotypical alterations such as cell cycle arrest, cell enlargement and random budding were observed, indicating that *SMRHO* not complement all *S. cerevisiae RHO1* functions. The *SMRHO* complementation results and sequence comparisons allowed evidence that the Pro<sup>96</sup> and Thr<sup>100</sup> amino acid residues in human RHOA  $\alpha$ 3-helix loop7 region are probably responsible for its conditional complementation of the *S. cerevisiae rho1* null mutant.

\*This investigation received financial support from the following sources: PRPq, PADCT, CNPq, UNDP/WORLD BANK/WHO Special Program for Research and Training in Tropical Diseases (TDR N<sup>o</sup>: 940325 and 940751), USAID/HOH (No 264.01.01.04).

*The nucleotide sequence reported in this paper has been submitted to the GenBank<sup>TM</sup>/EMBL Data Bank with the accession number AF140785.*

<sup>1</sup>The abbreviations used are: LMWGPs, low-molecular weight GTP-binding proteins; r.m.s.d., root-mean square deviation; *SMRHO*, *Schistosoma mansoni* Rho GTPase gene; SMRHO, *S. mansoni* Rho GTPase protein; UPGMA, unweighted pair-group method with arithmetic mean.

<sup>#</sup>To whom correspondence should be addressed: Tel.: +5531 499 2628; fax: +5531 499 2984; E-mail: spena@dcc.ufmg.br

## INTRODUCTION

The Rho GTPases are part of the large Ras superfamily of low-molecular weight GTP-binding proteins (LMWGPs<sup>1</sup>) (1). Like other Ras superfamily members, they are able, through the binding and hydrolysis of GTP, to create a switch between an active GTP-bound conformation and an inactive GDP-bound conformation (2). The “on-off” activity of Rho GTPases is controlled by several regulatory proteins: the guanine nucleotide exchange factors (GEFs) promote the exchange of the GDP for GTP; the GTPase-activating proteins (GAPs) accelerate the hydrolysis of the GTP to GDP; and the GDP dissociation inhibitors (GDIs) inhibit the dissociation of the GDP from the GTPase (3). Rho GTPases receive upstream signals through their regulators and transduce signals to a multitude of effector molecules, while remaining in the GTP-bound form (4,5,6). In this fashion, members of Rho GTPases family perform important regulatory roles, controlling a variety of cellular activities such as organization of the actin cytoskeleton (7), cell cycle progression (8,9), and intracellular transport (10).

Few LMWGPs and their regulatory counterparts have been identified and characterized in the blood fluke *Schistosoma mansoni*. This parasite assumes a diversity of forms during its life cycle, and LMWGPs may perform a pivotal regulatory role in the evolution of the different stages. Consequently, the characterization of this family of guanine nucleotide switcher proteins is important to understand the schistosome development and host relationships. There is evidence to LMWGPs are involvement in the *S. mansoni* female maturation process (induced by pairing with the male) and egg production (11,12). This is of great interest for schistosomiasis because the pathological lesions observed in affected humans are caused by the inflammatory reaction to the eggs produced by female parasites. A reduction of isoprenoid end-product synthesis by mevinolin, an inhibitor of hydroxymethylglutaryl-CoA reductase, was capable of decreasing egg production and blocking the schistosomiasis pathology in infected mice (12) as well as causing loss of parasite viability at higher concentrations (13). The same effects in schistosome females were observed as result of prenyltransferase activity inhibition by BZA-5B (14). This is important because blocking of



prenylation of a 25 kDa GTP-binding proteins group from schistosomes using BZA-5B led to the speculation that LMWGs could be playing a role in the egg laying mechanism (15). These observations stimulated the search for LMWGs associated with the regulation of signal transduction pathways in *S. mansoni*. A gene encoding a RAB-related GTP-binding protein was cloned from *S. mansoni* genomic DNA using degenerate oligonucleotide primers as an attempt to identify LMWGs involved in female egg production (16). The *in vitro* translation product of smrab gene is geranylgeranylated and the recombinant protein binds GTP. Using an immunochemical approach, Schussler *et al.* (11) showed the sexual and developmental regulation of Ras, GAP and MAP kinases in schistosome extracts. The involvement of these molecules in the female maturation process was expected once Ras, GAP and MAP kinases proteins interact in the same signal transduction pathway known to regulate cell proliferation events. More recently, a schistosome cDNA encoding a gene highly similar with K-Ras from various organisms was isolated and characterized (17). The *S. mansoni* K-Ras homologue gene is overexpressed in females in comparison to male worms had been detected in all developmental stages of the parasite. The recombinant smras protein is farnesylated *in vitro* and was identified by immunocytochemical localization associated with the worm subtegument.

As part of The *Schistosoma mansoni* Genome Project, we isolated from a cercarial library a cDNA clone (18) highly homologous to the Rho GTPase gene from several eukaryotic organisms. We report in this paper the complete cDNA and genomic sequences of this gene. The results with sequence alignments, molecular phylogenetic analysis, protein structural characterization by comparative computational molecular modeling, and complementation in a *Saccharomyces cerevisiae rho1* null mutant strain allowed us identify this gene as the first of *Schistosoma mansoni* that encodes a Rho-type LMWGP.

## EXPERIMENTAL PROCEDURES

*Cloning Procedures and DNA Sequencing*—The proofreading *Pfu* DNA polymerase (Stratagene) was used in PCR amplifications for subcloning. The amplification products used in subcloning or transformation procedures were previously purified using the Magic PCR-Prep kit (Promega). Cloning in pUC18 vector were done using SureClone Ligation Kit (A. P. Biotech). Plasmid purifications were by alkaline lysis using Wizard Plus SV Miniprep kit (Promega). The DNA sequencing reactions (19) were prepared using Thermo Sequenase Sequencing Kit (Amersham Pharmacia Biotech) with fluorescent M13 primers and carried out in an A.L.F. DNA sequencer (A. P. Biotech).

*cDNA Clone Sequencing*— The SMC0220R cDNA clone containing the *S. mansoni* Rho GTPase gene was obtained by random selection from a *S. mansoni* cercariae library as described previously (18). Initially we sequenced the 5' and 3' extremities of the SMC0220R cDNA clone using the M13 primers. Subsequently, using these flanking sequences, the PTV3 (5'-TGA TGT TCC CAT TGT CTT AG-3') and PTV4 (5'-AAA TCA AAT CTC AGA CAA TT-3') primers were designed to amplify the internal region of the insert. The 291 bp amplified internal fragment was subcloned in pUC18 vector and sequenced on both strands. In order to obtain a reliable full-length sequence, the extremities of the insert were amplified using the M13 reverse/PTV5 (5'-CTA AGA CAA TGG GAA CAT CA-3') and PTV6 (5'-TGT CTA GAT TTG ATT TTC CG-3')/M13 universal primer pairs. The generated fragments of 527 bp and 234 bp, respectively, were subcloned in pUC18 and sequenced on both strands.

*Genomic Sequencing*— We used also the SMRHOF/PTV4 primers to amplify the *S. mansoni* Rho GTPase genomic sequence. The PCR amplification generated a fragment of around 700 bp. This fragment was excised from the gel, purified, cloned in pUC18 vector and sequenced on both strands. In order to obtain a reliable sequence, the genomic clone was PCR amplified into two

distinct fragments using primer pairs SMRHOF/PTV5 and PTV3/PTV4. The generated fragments of 417 and 324 bp were subcloned in pUC18 and sequenced on both strands.

*Sequence Analyses*—Sequence alignments and open reading frame (ORF) search were done using the DNASIS program (Hitachi). Homology searches were performed using NCBI's BLAST programs for sequence similarity search. Alignments of amino acid sequences and UPGMA analysis were performed with the software DIALIGN 2.1 (20).

*Comparative Computational Molecular Modeling*—The molecular models of *S. mansoni* Rho GTPase protein associated with GDP and GTP $\gamma$ S were constructed using as templates the high-resolution crystal structures of human RHOA GTPase in association with GDP (21) and GTP $\gamma$ S (22) obtained from the Protein Data Bank. The MODELLER program was used for automatic construction of the models based on satisfaction of spatial restraints (23) on a Power Challenger workstation (Silicon Graphics). Forty models for schistosome Rho GTPase protein binding GDP and GTP $\gamma$ S were constructed and the models with best energy score were chosen.

*Yeast strains, Media, and Genetic Manipulation Methods*—The yeast strains employed are listed in Table 1. The diploid strain BY4727/36 was obtained crossing BY4727 with BY4736 by replica plating. Yeast media were as described (25) and 2% glucose or 2% galactose were included as carbon source when required.

*Cloning on Yeast Super-expression Vector*—The *S. mansoni* Rho GTPase gene was amplified from the SMC0220R cDNA clone using primers SMRHOF (5'-GGC AAC AAT GGC GAG TG-3') and SMRHOR (5'-TTA GAG AAC AGT GGA GAT CA-3'). The *S. cerevisiae* *RHO1* gene was obtained by direct amplification from yeast genomic DNA by colony PCR using the primers RHOSACF (5'-ACA TGA GAT CTA CGC AGA TCA-3') and RHOSACR (5'-CAT CGA TCA

TTC CTC TGA GTA-3'). The fragment containing the *RHO1* gene was digested with *DraI* in order to remove the promoter region. The fragments containing the schistosome Rho GTPase and *RHO1* genes were cloned in pUC18. Afterwards, the fragments were recovered by digestion with *BamHI/KpnI* and cloned into the same restriction sites of pYEDP60-2-ADE2-URA3 plasmid vector (a kind gift of Prof. Francisco Nóbrega, UNIVAP, São Paulo) containing the galactose inducible promoter *GAL10-CYCI* for overexpression in yeast (26).

*Yeast Genetic Manipulation Methods*—Transformations were done with 1 µg of purified linear DNA or 0.5 µg of plasmid DNA by LiAc/SS-DNA/PEG procedure (27). *S. cerevisiae rho1* null mutant strain TSY-D was obtained via PCR-mediated gene disruption method (24) using the diploid strain BY4727/36. The 60 mer primers RHO1HIS3F (5'-ATG TCA CAA CAA GTT GGT AAC AGT ATC AGA AGA AAG CTG GCT TCA TTC AAC GTT TCC CAT-3') and RHO1HIS3R (5'-GCT AGA AAT ATG AAC CTT CCA ACA AAA CTG AGG TTG GAG GAG TAT CAT ACT GTT CGT ATA-3') were used to amplify the *HIS3* auxotrophic marker gene cloned in pDIS plasmid (a kind gift of Prof. Francisco Nóbrega). The 1,243 bp amplified product contained at the 3' and 5' ends 40 bp of sequences to the left and right of the *S. cerevisiae RHO1* gene. This product was purified and transformed in BY4727/36 diploid strain for *RHO1* gene disruption through a homologous recombination event. Disrupted mutants were selected on plates of YNB-glucose containing the required nutrients except the histidine amino acid. PCR amplification of a 788 bp hybrid fragment containing part of *RHO1* and *HIS3* genes using RHOSACF and HIS311 (5'- AAC CCT ATA CCT GTG TGG A-3') primers, confirmed the insertion of auxotrophic marker in place of one of the *RHO1* genomic copies in transforming yeast clones (data not show). The *rho1* disrupted diploid strain TSY-D was transformed with the pYEDP, pYEDP(*RHO1*) and pYEDP(*SMRHO*) plasmids to produce the strains TSY-SE0-D, TSY-SE1-D, and TSY-SE2-D, respectively. Transforming clones were selected on YNB-glucose containing the required nutrients. The TSY-SE2-D strain was induced to sporulate in order to obtain Rho1 null mutant haploid cells

complemented with *SMRHO*. Sporulation of diploid cells in liquid medium was induced as described (24). The sporulated cultures were treated with NovoZym (Sigma) to digest the ascus wall, vortexed vigorously 2 min to release spores from the ascis and subsequently heated at 58°C for 15 min to kill any remaining vegetative cells (25). Spores dilutions were spread on YNB-galactose plates supplemented with the required nutrients. The *rho1* null mutant haploid clone, complemented by overexpression of *SMRHO* gene induced by *GAL10-CYCI* promoter (TSY-SE2 strain), was confirmed by negative PCR amplification of *RHO1* gene using RHOSACF/RHOSACR primers, amplification of *SMRHO* gene using SMRHOF/SMRHOR primers, and amplification of a 788 bp hybrid fragment containing part of *HIS3* and *RHO1* genes using RHOSACF/HIS311 primers (data not show).

## RESULTS

*cDNA Sequence Analyses*—The partial sequences obtained of the subclones were aligned and produced a 1,016 bp consensus sequence that represents the full-length clone. We observed, however, that the insert was a quimeric artifact of two cDNA molecules joined by adapters used in the cercariae library construction. The actual cDNA molecule encoding the *S. mansoni* Rho GTPase homologous gene represented 841 bp in this sequence (Fig. 1). The most significant ORF found had 579-nucleotide long and encoded a putative polypeptide of 193 amino acid with an estimated molecular mass of 21.8 kDa (Fig. 1). This cDNA contained a long 3' untranslated portion, which is a known feature of *S. mansoni* transcripts. In comparison with other Rho GTPases, the deduced amino acid sequence was found to have the insertion of a serine residue near to the N-terminal end. Apparently this insertion is peculiar to the *S. mansoni* Rho GTPase amino acid sequence since it has not been observed in any other Rho GTPase homologous sequence examined (Fig. 2A). Another interesting feature was the presence of the polar amino acid aspartate (position 191) in an unusual substitution of the first aliphatic amino acid of the C-terminal motif CAAX for prenylation. With this exception, the putative amino acid sequence of this ORF contained all others consensus

sequences conserved among the Rho-type GTPases (28,29,30), and was highly homologous to these LMWGs (Fig 2A), most of all to *Aplysia californica* RHO protein (78% of identity). UPGMA analysis revealed that the schistosome Rho GTPase although apparently more evolutionarily related to *Hemicentrotus pulcherrimus* echinoderm RHO1 was quite divergent, grouping as an isolated branch in the phylogenetic tree (Fig 2B). We named this gene “*SMRHO1*” for *Schistosoma mansoni* homologue of the Rho-type LMWGs.

*Genomic Sequence Analyses* —The partial sequences obtained of the subclones were aligned to produce a 722 bp consensus genomic sequence that contained an ORF identical to the *SMRHO* and 3 introns. Fig. 1 highlights the small intron sequences found in the genomic sequence. The first intron has 32 bp and is located between switch I and switch II domains. The second intron, of 33 bp, is in the middle of the ORF. The last intron has also 33 bp long and is located in the SAK motif that participates in interactions with guanine base of GTP molecule in the active site. All introns interrupt the protein frame. The canonical donor/acceptor splice sites are conserved at all exon/intron junctions.

*Comparative Computational Models of SMRHO Protein*—The existence of human RHOA refined crystal structures in association with  $GTP_{\gamma}S$  and GDP (21,22) allowed us to construct, by comparative computational molecular modeling, feasible models of the SMRHO tridimensional structures in its activated and inactivated states. The SMRHO structures in association with  $GTP_{\gamma}S$  (r.m.s.d. of 0.25 Å) and GDP (r.m.s.d. of 0.17 Å) follow the general folding pattern of the Rho GTPase structures already determined by crystallography (21,22,31). The superposition of  $C_{\alpha}$ -carbon backbone of SMRHO models bound to  $GTP_{\gamma}S$  and GDP revealed that the segments of largest structural displacements among the activated and inactivated forms corresponded to switch I

(28 to 46) and II (62 to 71) regions (Fig. 3). Except by these alterations, the C<sub>α</sub>-carbon atoms of SMRHO-GTP $\gamma$ S and SMRHO-GDP superimposed with a r.m.s.d. of 0.74 Å.

*Complementation of the S. cerevisiae rho1 Null Mutant by SMRHO*—The schistosome Rho GTPase gene complements the *rho1* null mutation on YNB-galactose even at 37°C but not on YPD plates containing glucose, where the overexpression of *SMRHO* shuts of (Fig 4B). Furthermore, the *rho1* null mutants harboring pYEDP(*SMRHO*) plasmid grew on YNB-galactose plate containing 300 mM CaCl<sub>2</sub> (Fig. 4A). However, the overproduction of SMRHO causes cell cycle arrest in null mutants in the same fashion as RHO1 overproduction (32) (Fig. 5). In addition, we verified alterations in the morphology and budding pattern of these cells. We observed the presence of large cells, sometimes with buds, and cells with the overall morphology apparently normal, but exhibiting a random budding process (Fig 6). Interestingly, the cell cycle (Fig. 5) and budding pattern (Fig. 6) were apparently normal when SMRHO was overproduced in TSY-SE2-D, although there was yet the presence of enlarged cells (Fig. 6).

## DISCUSSION

The first Rho GTPase homologue gene (*SMRHO*) isolated from *Schistosoma mansoni* encodes a protein that is highly homologous to several Rho-type LMWGs, with maximal identity (78%) to the *Aplysia californica* Rho GTPase. The Rho-type LMWGP characteristic motifs (28,29,30) are conserved in SMRHO, except for the CAAX prenylation domain that has a polar amino acid substituting the first aliphatic residue. Another peculiarity found in the *SMRHO* deduced amino acid sequence was the insertion of a serine residue at the N-terminal. Such sequence peculiarities and UPGMA analysis showed that schistosome Rho GTPase protein is evolutionarily more divergent than the Rho-type GTPases already identified in other organisms. However, the Rho-type GTPases general molecular folding (21,22,31) is conserved in SMRHO. The parasite molecule displays the same structural displacements in switch I (28-46 amino acids) and switch II (62-71 amino acids)

regions when its activated and inactivated forms are superimposed (Fig. 3). Additionally, we demonstrate in this work that the SMRHO homology (62%) to *S. cerevisiae* RHO1 is sufficient to complement the yeast GTPase functions when overexpressed in a null mutant strain. The complementation occurred even at 37°C (Fig. 4B) and in the presence of 300 mM CaCl<sub>2</sub> (Fig. 4A). In *S. cerevisiae* RHO1 is required for cell cycle progression (33) and mediates bud growth by controlling polarization of the actin cytoskeleton and cell wall synthesis (34). Therefore, SMRHO can perform similar biological functions. However, we observed some phenotypic alterations such as cell cycle arrest (Fig. 5), cell enlargement, and random budding caused by SMRHO overexpression in *rho1* null mutant (Fig. 6). In contrast, the cell cycle and budding pattern were apparently normal when SMRHO was overexpressed in wild-type cells, although enlarged cells were present yet. Such results suggest a RHO1 dominance effect over the SMRHO with respect to the regulatory functions of the cell cycle and budding site selection process but not in relation to the actin cytoskeleton regulatory functions. This could be explained by defective molecular interactions of the SMRHO with the distinct RHO1 regulatory and effector proteins in *S. cerevisiae* intracellular environment, due to structural differences that exist among the heterologue and native GTPases.

The functional effect caused by structural differences becomes still more evident when we compare the results of the human RHOA and SMRHO complementation in *S. cerevisiae rho1* null mutants. Human RHOA overexpression is lethal and its expression under control of the yeast RHO1 promoter produces temperature and high Ca<sup>+</sup> sensitive phenotypes (35). Nevertheless, SMRHO overexpression complements *S. cerevisiae rho1* null mutant even in restrictive conditions of temperature and Ca<sup>+</sup> concentration (Fig. 4). The  $\alpha$ 3-helix loop7 was previously defined as the region of the human RHOA molecule responsible for the complementation inability in such conditions (35). In order to identify the amino acid residues functionally significant for yeast growing on such restrictive conditions, we aligned the sequence of *S. cerevisiae* RHO1  $\alpha$ 3-helix-loop7 with the corresponding regions of *C. albicans* RHO1, human RHOA, and SMRHO GTPases. The *C. albicans* RHO1 protein was included in the alignments because it complements *S. cerevisiae rho1* null mutants



without any apparent phenotypic alterations (36). Fig. 7 shows the conservative and non-conservative amino acid changes among the  $\alpha$ 3-helix loop7 sequences. The *C. albicans* sequence showed few amino acid alterations in comparison with *S. cerevisiae* RHO1 which were apparently without any functional importance to complementation. Nevertheless, the amino acid modifications in relation to the human RHOA and SMRHO  $\alpha$ 3-helix loop7 regions were qualitative and in greater number (Fig 7), determining significant hydrophobic changes. The comparison among the amino acid constitution of the  $\alpha$ 3-helix loop7 regions with the *C. albicans*, human and schistosome Rho GTPases capacity to complement *S. cerevisiae* rho1 null mutants suggest that Pro<sup>96</sup> and Thr<sup>100</sup> are been the amino acids responsible by human RHOA conditional complementation. The presence of Thr<sup>100</sup> determined a notable hydrophobic modification in RHOA  $\alpha$ 3-helix to a more hydrophilic character (Fig 7), while Pro<sup>96</sup> can determine a conformational alteration in  $\alpha$ 3-helix reducing its capacity of interaction with the *S. cerevisiae* regulatory and effector molecules. The amino acid residues in RHOA loop7 are functionally equivalent to these in SMRHO, explaining why schistosome Rho GTPase complements rho1 null mutant even in restrictive conditions. Thus, although the Rho GTPases from different species share a high homology degree, subtle amino acid differences in certain regions determine the interaction specificity of them with its multiple regulatory and effector proteins in every organism. It is then conceivable to postulate the existence of a direct relationship between specific regions of Rho-type LMWGs and the different functions performed by these proteins, in accordance with previous observations done for *S. cerevisiae* RHO1 (33).

The Schistosome Genome Project has already produced sequence of many genes and demands at this time intensive efforts on functional genomics (37). Nevertheless, such endeavor can be restricted by the technical difficulties of parasite genetic manipulation. The use of *S. cerevisiae* and other model organisms will constitute a powerful tool for the schistosome gene function investigations, as was demonstrated here.

*Acknowledgments*-The authors thank Vasco Azevedo for initial directions on SmRho sequencing, Miguel Ortega by use of your laboratory facilities for yeast cultivation and Kátia Barroso for carrying out automated DNA sequencing.

## REFERENCES

1. Ridley A. J. (1997) *Int. J. Biochem. Cell. Biol.* **29**, 1225-9
2. Bourne H. R. , Sanders D. A., and McCormick F. (1991) *Nature* **349**, 117-27
3. Sasaki T., and Takai Y. (1998) *Biochem. Biophys. Res. Commun.* **245**, 641-5
4. Machesky, L. M., and Hall, A. (1996) *Trends in Cell Biology* **6**, 304-310.
5. Ridley, A. J. (1997) *Biochem. Soc. Trans.* **25**, 1005-10
6. Aspenstrom, P. (1999) *Curr. Opin. Cell Biol.* **11**, 95-102
7. Hall, A. (1990) *Science* **249**, 635-40
8. Hirai, A., Nakamura, S., Noguchi, Y., Yasuda, T., Kitagawa, M., Tatsuno, I., Oeda, T., Tahara, K., Terano, T., Narumiya, S., Kohn, L. D., and Saito, Y. (1997) *J. Biol. Chem.* **272**, 13-6
9. Hu, W., Bellone, C. J., and Baldassare, J. J. (1999) *J. Biol. Chem.* **274**, 3396-401
10. Ellis, S., and Mellor, H. (2000) *Trends. Cell Biol.* **10**, 85-8
11. Schussler, P., Grevelding, C. G., and Kunz, W. (1997) *Parasitology*, **115**, 629-634.
12. VandeWaa, E. A., Mills, G., Chen, G. Z., Foster, L. A., and Bennett J. L. (1989) *Am. J. Physiol.*, **257**, r618-25,
13. Chen, G. Z., Foster, L., Bennett, J. L. (1990) *Naunyn. Schmiedebergs. Arch. Pharmacol.* **342**, 477-82
14. Bennett, J. L. cited in Loeffler I. K., and Bennett J. L. (1996) *Mol. Biochem. Parasitol.* **77**, 31-40
15. Chen, G. Z., and Bennett, J. L. (1993) *Mol. Biochem. Parasitol.* **59**, 287-92,
16. Loeffler, I. K., and Bennett, J. L. (1996) *Mol. Biochem. Parasitol.* **77**, 31-40

17. Osman, A., Niles, E. G., and LoVerde, P. T. (1999) *Mol. Biochem. Parasitol.* **100**, 27-41
18. Santos, T. M., Johnston, D. A., Azevedo, V., Ridgers, I. L., Martinez, M. F., Marotta, G. B., Santos, R. L., Fonseca, S. J., Ortega, J. M., Rabelo, E. M., Saber, M., Ahmed, H. M., Romeih, M. H., Franco, G. R., Rollinson, D., and Pena, S. D. J. (1999) *Mol. Biochem. Parasitol.* **103**, 79-97
19. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463-7
20. Morgenstern, B. (1999) *Bioinformatics* **15**, 211-218.
21. Wei, Y., Zhang, Y., Derewenda, U., Liu, X., Minor, W., Nakamoto, R. K., Somlyo, A. V., Somlyo, A. P., and Derewenda, Z. S. (1997) *Nat. Struct. Biol.* **4**, 699-703
22. Ihara, K., Muraguchi, S., Kato, M., Shimizu, T., Shirakawa, M., Kuroda, S., Kaibuchi, K., and Hakoshima, T. (1998) *J. Biol. Chem.* **273**, 9656-66
23. Sanchéz, R., and Sali, A. (1997) *Proteins Suppl* **1**, 50-8
24. Brachmann, C. B., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P., and Boeke, J. D. (1998) *Yeast* **14**, 115-32
25. Rose, M. D., Winston, F. and Hieter, P. (1990) *Methods in yeast genetics* (Cold Spring Harbor Lab. Press, Plainview, NY)
26. Renaud, J. P., Cullin, C., Pompon, D., Beaune, P. and, Mansuy, D. (1990) *Eur. J. Biochem.* **194**, 889-896
27. Gietz, R. D., Schiestl, R. H., Willems, A. R., and Woods, R.A. (1995) *Yeast* **11**, 355-60
28. Bourne, H. R., Sanders, D. A., and McCormick, F. (1991) *Nature* **349**, 117-27
29. Kahn, R. A., Der, C. J., and Bokoch, G. M. (1992) *FASEB J.* **6**, 2512-3
30. Self, A. J., Paterson, H. F., and Hall, A. (1993) *Oncogene* **8**, 655-61
31. Hirshberg, M., Stockley, R. W., Dodson, G., and Webb, M. R.. (1997) *Nat. Struct. Biol.* **4**, 147-52
32. Espinet, C., de la Torre, M. A., Aldea, M., and Herrero, E. (1995) *Yeast* **11**, 25-32
33. Drgonova, J., Drgon, T., Roh, D. H., and Cabib, (1999) *E. J. Cell. Biol.* **146**, 373-387

34. Cabib, E., Drgonova, J., and Drgon, T. (1998) *Annu. Rev. Biochem.* **67**, 307-333
35. Qadota, H., Anraku, Y., Botstein, D., and Ohya, Y. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**,  
9317-21
36. Kondoh, O., Tachibana, Y., Ohya, Y., Arisawa, M., and Watanabe, T. (1997) *J. Bacteriol.* **179**,  
7734-41
37. Franco, G. R., Valadão, A. F., Azevedo, V., Rabelo, E. M. (2000) *Int. J. Parasitol.* **30**, 453-63

## Figure legends

**Fig. 1.** Nucleotide sequence and predicted amino acid sequence of *Schistosoma mansoni* Rho GTPase gene. Intron sequences are underlined above the ORF sequence and its respective insertion positions in the nucleotide sequence are highlighted in bold face.

**Fig. 2. (A)** Amino acid sequence alignment of *Schistosoma mansoni* Rho GTPase and other orthologue sequences. Identical amino acids among the sequences are marked by asterisks. The functional regions of the GTPase active site are highlighted in yellow. Geranylgeranylation site is in pink. Switch I and switch II regions are indicated above the sequences. **(B)** Dendogram analysis of *Schistosoma mansoni* Rho GTPase with other orthologue and non-orthologue sequences. %I, identity; %S, similarity; O, alignment overlap; *Ap*, *Aplysia californica*; *Xl*, *Xenopus laevis*; *Hs*, *Homo sapiens*; *Hp*, *Hemicentrotus pulcherrimus*; *Dm*, *Drosophila melanogaster*; *Mm*, *Mus musculus*; *Gg*, *Gallus gallus*; *Ce*, *Caenorhabditis elegans*; *Cf*, *Canis familiaris*; *Do*, *Discopyge ommata*; *Sp*, *Schizosaccharomyces pombe*; *Sc*, *Saccharomyces cerevisiae*; *Fn*, *Filobasidiella neoformans*; *Ca*, *Candida albicans*.

**Fig. 3.** Structural comparison between SMRHO-GTP $\gamma$ S and SMRHO-GDP models. Superposition of C $\alpha$ -carbon atom backbone of smrho bound to GTP $\gamma$ -S (red) and smrho bound to GDP (green). Segments displaying large displacements are highlighted in SMRHO-GTP $\gamma$ -S. These were switch I (residues 28-46) in light blue and switch II (residues 62-71) in pink.

**Fig. 4.** Complementation of a *Saccharomyces cerevisiae* rho1 null mutant by overexpression of *Schistosoma mansoni* Rho GTPase gene. **(A)** Growth of TSY-SE0-D, TSY-SE1-D, TSY-SE2-D, and TSY-SE2 strains at 30°C in YNB-GAL, YNB-GAL + 300 mM CaCl<sub>2</sub> and YNB agar plates. **(B)**

Growth of TSY-SE0-D, TSY-SE1-D, TSY-SE2-D, and TSY-SE2 strains at 37°C in YNB-GAL and YNB agar plates.

**Fig. 5.** Cell growth of *Saccharomyces cerevisiae* rho1 null mutants complemented with *Schistosoma mansoni* Rho GTPase. TSY-SE0-D, TSY-SE1-D, TSY-SE2-D, and TSY-SE2 cells were incubated at 30°C in YNB-Gal medium supplemented with methionine, tryptophane, leucine and lysine amino acids for the indicated times.

**Fig. 6.** Morphological aspect of *Saccharomyces cerevisiae* rho1 null mutants cells complemented with *Schistosoma mansoni* Rho GTPase. Asynchronous cultures of (A) TSY-SE0-D, (B) TSY-SE1-D, (C) TSY-SE2-D, and (D) TSY-SE2 were grown at 30°C in YNB-Gal medium supplemented with methionine, tryptophane, leucine and lysine amino acids.

**Fig. 7.** BLASTP alignment of *Saccharomyces cerevisiae* RHO1 and *Candida albicans* RHO1, human RHOA, and *Schistosoma mansoni* Rho GTPases  $\alpha$ 3-helix loop 7 region. Identical amino acids are highlighted in yellow, alterations for similar amino acids are in green and for non-similar amino acids are in red. %I, identity; %S, similarity.

**Table 1**

TABLE 1. Yeast strains used in this study

Strain	Genotype	Reference
BY4727	<i>MAT<math>\alpha</math> his3<math>\Delta</math>200 leu2<math>\Delta</math>0 lys2<math>\Delta</math>0 met15<math>\Delta</math>0 trp1<math>\Delta</math>63 ura3<math>\Delta</math>0</i>	24
BY4736	<i>MAT<math>\alpha</math> ade2<math>\Delta</math>::hisG his3<math>\Delta</math>200 met15<math>\Delta</math>0 trp1<math>\Delta</math>63 ura3<math>\Delta</math>0</i>	24
BY4727/36	<i>MAT<math>\alpha</math>/MAT<math>\alpha</math> his3<math>\Delta</math>200 met15<math>\Delta</math>0 trp1<math>\Delta</math>63 ura3<math>\Delta</math>0</i>	this study
TSY-D	<i>MAT<math>\alpha</math>/MAT<math>\alpha</math> his3<math>\Delta</math>200 met15<math>\Delta</math>0 trp1<math>\Delta</math>63 ura3<math>\Delta</math>0 rho1::HIS3</i>	this study
TSY-SE0-D	<i>MAT<math>\alpha</math>/MAT<math>\alpha</math> his3<math>\Delta</math>200 met15<math>\Delta</math>0 trp1<math>\Delta</math>63 ura3<math>\Delta</math>0 rho1::HIS3 pYEDP</i>	this study
TSY-SE1-D	<i>MAT<math>\alpha</math>/MAT<math>\alpha</math> his3<math>\Delta</math>200 met15<math>\Delta</math>0 trp1<math>\Delta</math>63 ura3<math>\Delta</math>0 rho1::HIS3 pYEDP(RHO1)</i>	this study
TSY-SE2-D	<i>MAT<math>\alpha</math>/MAT<math>\alpha</math> his3<math>\Delta</math>200 met15<math>\Delta</math>0 trp1<math>\Delta</math>63 ura3<math>\Delta</math>0 rho1::HIS3 pYEDP(SMRHO)</i>	this study
TSY-SE2	<i>MAT? ade2<math>\Delta</math>::hisG? his3<math>\Delta</math>200 met15<math>\Delta</math>0 trp1<math>\Delta</math>63 ura3<math>\Delta</math>0 leu2<math>\Delta</math>0 lys2<math>\Delta</math>0 rho1::HIS3 pYEDP(SMRHO)</i>	this study

**Figure 1**

-6 5' gcaaca

1 ATGGCGAGTGCGGTACGTAAGAACTTGTATTGTTGGAGATGGTGCATGCGGGAAAAC  
 1 M A S A V R K K L V I V G D G A C G K T

61 TGTTTACTTATTGTATTCAGCAAAGACCAGTTTCCCGAAGTTTACGTTCTACTGTATTC  
 21 C L L I V F S K D Q F P E V Y V P T V F

aggtgtataagtacttactatatgtgtaggt

121 GAAAACATGTTGCAGATATCGAAGTTGATAACAGACAAGTTGAATTAGCTCTCTGGGAC  
 41 E N Y V A D I E V D N R Q V E L A L W D

181 ACTGCTGGCCAAGAAGATTATGATAGGCTACGGCCACTTCTTATCCTGATACGGATGTA  
 61 T A G Q E D Y D R L R P L S Y P D T D V

241 GTTCTGTTATGTTATAGTATTGACTCTCCCGATAGTTTTCGAAACATTGAGGAAAAATGG  
 81 V L L C Y S I D S P D S F A N I E E K W

gtagataccgcgagttcataaatcctattatag

301 TTACCAGAAATCCGTCATTTTTGTCTGATGTTCCCATGTCTTAGTTGAAACAAAAA  
 101 L P E I R H F C P D V P I V L V G N K K

361 GACTTACGACATGACGAGGCTACAAAGAATGAGCTACACAGAACCAACAACCTTCTGTC  
 121 D L R H D E A T K N E L H R T K Q L P V

421 ACTTTTAACGAGGGTAAACAAGTAGCTGAAAAGATTTCTGCTTATGCCTTCTTTGAGTGT  
 141 T F N E G K Q V A E K I S A Y A F F E C

gtactgtctttacttatgtctagccttcttttag

481 TCAGCTAAGACCAAGGAAGGAGTCAGCGATGTTTTTCGTAGCAGCTACTCGAGCCGCTTG  
 161 S A K T K E G V S D V F V A A T R A A L

541 AATTCAGCGAAGAAGAAGAGGAGGTGTGATTTAATTTgatctccactgttctctaaa  
 181 N S A K K K R R R C D L I \*

601 ttgtctagatTTGATTTTCCGGTCTCCGAAACATACATTCCCCTCCTAATTCTTCTCAG  
 661 gaccttaaatTGTACTTCTATCGGCTGCGTGTCTTCTTTACAATCATATTTTGTGTTG  
 721 ccaacaagacaatccattccacggaatatTTTGCCATTTATTTTGCCTTGGTACTAAAA  
 781 tagggcatgcctcatgtaacttatatttattgagaataaattactataatcagtc 3'



Figure 2A

		Switch I					Switch II								
<i>Sm</i> Rho	1	MASAVRKKLV	IVGDGACGKT	CLLIVFSKDQ	FPEVYVPTVF	ENYVADIEVD	NRQVELALWD	TAGQEDYDRL	RPLSYPDTDV	VLLCYSIDSP	DSFANIEEKW	LPEIRHFPCPN			
<i>Ac</i> Rho	1	MA-AIRKKLV	IVGDGACGKT	CLLIVFSKDQ	FPEVYVPTVF	ENYVADIEVD	GKQVELALWD	TAGQEDYDRL	RPLSYPDTDV	IILMCFSIDSP	DSLENIPEKW	TPEVRHFPCPN			
<i>Xl</i> RHOA	1	MA-AIRKKLV	IVGDGACGKT	CLLIVFSKDQ	FPEVYVPTVF	ENYVADIEVD	GKQVELALWD	TAGQEDYDRL	RPLSYPDTDV	IILMCFSIDSP	DSLENIPEKW	TPEVKHFPCPN			
<i>Hs</i> RhoC	1	MA-AIRKKLV	IVGDGACGKT	CLLIVFSKDQ	FPEVYVPTVF	ENYIADIEVD	GKQVELALWD	TAGQEDYDRL	RPLSYPDTDV	IILMCFSIDSP	DSLENIPEKW	TPEVKHFPCPN			
<i>Hp</i> RHO1	1	MA-AIRKKLV	IVGDGACGKT	CLLIVFSKDQ	FPEVYVPTVF	ENYVADIEVD	GKQVELALWD	TAGQEDYDRL	RPLSYPDTDV	IILMCFSIDNP	DSLETIPEKW	TPEVKHFPCPN			
<i>Dm</i> RHO1	1	MT-TIRKKLV	IVGDGACGKT	CLLIVFSKDQ	FPEVYVPTVF	ENYVADIEVD	GKQVELALWD	TAGQEDYDRL	RPLSYPDTDV	IILMCFSDSP	DSLENIPEKW	TPEVKHFPCPN			
<i>Mm</i> RhoC	1	MA-AIRKKLV	IVGDGACGKT	CLLIVFSKDQ	FPEVYVPTVF	ENYIADIEVD	GKQVELALWD	TAGQEDYDRL	RPLSYPDTDV	IILMCFSIDSP	DSLENIPEKW	TPQVKHFPCPN			
<i>Gg</i> RhoC	1	MA-AIRKKLV	IVGDGACGKT	CLLIVFSKDQ	FPEVYVPTVF	ENYIADIEVD	GKQVELALWD	TAGQEDYDRL	RPLSYPDTDV	IILMCFSIDSP	DSLENIPEKW	TPEVKHFPCPN			
<i>Ce</i> RHOA	1	MA-AIRKKLV	IVGDGACGKT	CLLIVFSKDQ	FPEVYVPTVF	ENYVADIEVD	GKQVELALWD	TAGQEDYDRL	RPLSYPDTDV	IILMCFSIDSP	DSLENIPEKW	TPEVRHFPCPN			
<i>Gg</i> RHOA	1	MA-AIRKKLV	IVGDGACGKT	CLLIVFSKDQ	FPEVYVPTVF	ENYVADIEVD	GKQVELALWD	TAGQEDYDRL	RPLSYPDTDV	IILMCFSIDSP	DSLENIPEKW	TPEVKHFPCPN			
<i>Mm</i> RHOA	1	MA-AIRKKLV	IVGDGACGKT	CLLIVFSKDQ	FPEVYVPTVF	ENYVADIEVD	GKQVELALWD	TAGQEDYDRL	RPLSYPDTDV	IILMCFSIDSP	DSLENIPEKW	TPEVKHFPCPN			
<i>Hs</i> RHOA	1	MA-AIRKKLV	IVGDGACGKT	CLLIVFSKDQ	FPEVYVPTVF	ENYVADIEVD	GKQVELALWD	TAGQEDYDRL	RPLSYPDTDV	IILMCFSIDSP	DSLENIPEKW	TPEVKHFPCPN			
		*	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*	*	*
													%I	%S	O
<i>Sm</i> Rho	111	VPIVLVGNKK	DLRHDEATKN	ELHRTKQLPV	TFNEGKQVAE	KISAYAFFEC	SAKTKEGVSD	VFVAATRAAL	NSAKKKRRR-	CDLI	193				
<i>Ac</i> Rho	110	VPIILVGNKK	DLRNDDESTKR	ELMKMKQEPV	RPEDGRAMAE	KINAYSYLEC	SAKTKEGVRD	VFETATRAAL	QVKKKKKGG-	CVVL	192	78	88	188	
<i>Xl</i> RHOA	110	VPIILVGNKK	DLRNDDEHTRR	ELTKMKQEPV	KPEEGRDMAN	RISAYAYMEC	SAKTKEGVRE	VFELATRAAL	QARRGKKKPR	CLLI	193	76	87	192	
<i>Hs</i> RhoC	110	VPIILVGNKK	DLRQDEHTRR	ELAKMKQEPV	RSEEGRDMAN	RISAFGYLEC	SAKTKEGVRE	VFEMATRAGL	QVRKNKRRRG	CPIL	193	75	85	187	
<i>Hp</i> RHO1	110	VPVILVGNKK	DLRNDDATKR	ELSKMKQEPV	KYNDAQTMSD	KINAYKYLEC	SAKSNDGVRE	VFETATRAAL	QVKKKKSSK-	CSLL	192	72	86	191	
<i>Dm</i> RHO1	110	VPIILVGNKK	DLRNDPNTIR	DLAKMKQEPV	KPQEGRAMAE	KINAFAYLEC	SAKSKEGVRE	VFETATRAAL	QVKKKKKTR-	CLLL	192	74	86	191	
<i>Mm</i> RhoC	110	VPIILVGNKK	DLRQDEHTRR	ELAKMKQEPV	RSEEGRDMAN	RISAFGYLEC	SAKTKEGVRE	VFEMATRAGL	QVRKNKRRRG	CPIL	193	75	85	187	
<i>Gg</i> RhoC	110	VPIILVGNKK	DLRNDDEHTRR	ELAKMKQEPV	KPEEGRDMAN	RINAFGYLEC	SAKTKEGVRE	VFEMATRAGL	QVRKNKRRRG	CPIL	193	74	85	192	
<i>Ce</i> RHOA	110	VPIILVGNKR	DLRSDPQTVR	ELAKMKQEPV	KPEQGRAIAE	QIGAFAYLEC	SAKTKEGVRE	VFEMATRAGL	QVKKKKKSK-	CMIL	192	72	85	191	
<i>Gg</i> RHOA	110	VPIILVGNKK	DLRNDDEHTRR	ELAKMKQEPV	KPEEGRDMAN	RIGAFGYMEC	SAKTKEGVRE	VFEMATRAAL	QARRGKKKSG	CLLL	193	73	85	192	
<i>Mm</i> RHOA	110	VPIILVGNKK	DLRNDDEHTRR	ELAKMKQEPV	KPEEGRDMAN	RIGAFGYMEC	SAKTKEGVRE	VFEMATRAAL	QARRGKKKSG	CLLL	193	73	85	191	
<i>Hs</i> RHOA	110	VPIILVGNKK	DLRNDDEHTRR	ELAKMKQEPV	KPEEGRDMAN	RIGAFGYMEC	SAKTKEGVRE	VFEMATRAAL	QARRGKKKSG	CLVL	193	74	86	186	
		**	*****	***	*	*	**	**	**	*	*	*	*	*	*

Figure 2B

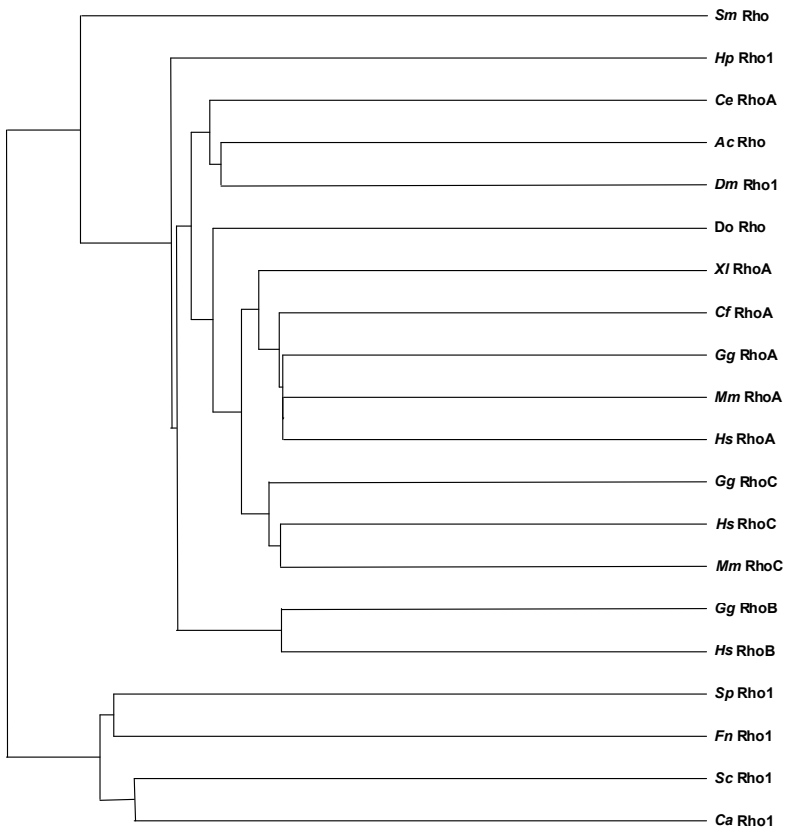


Figure 3.

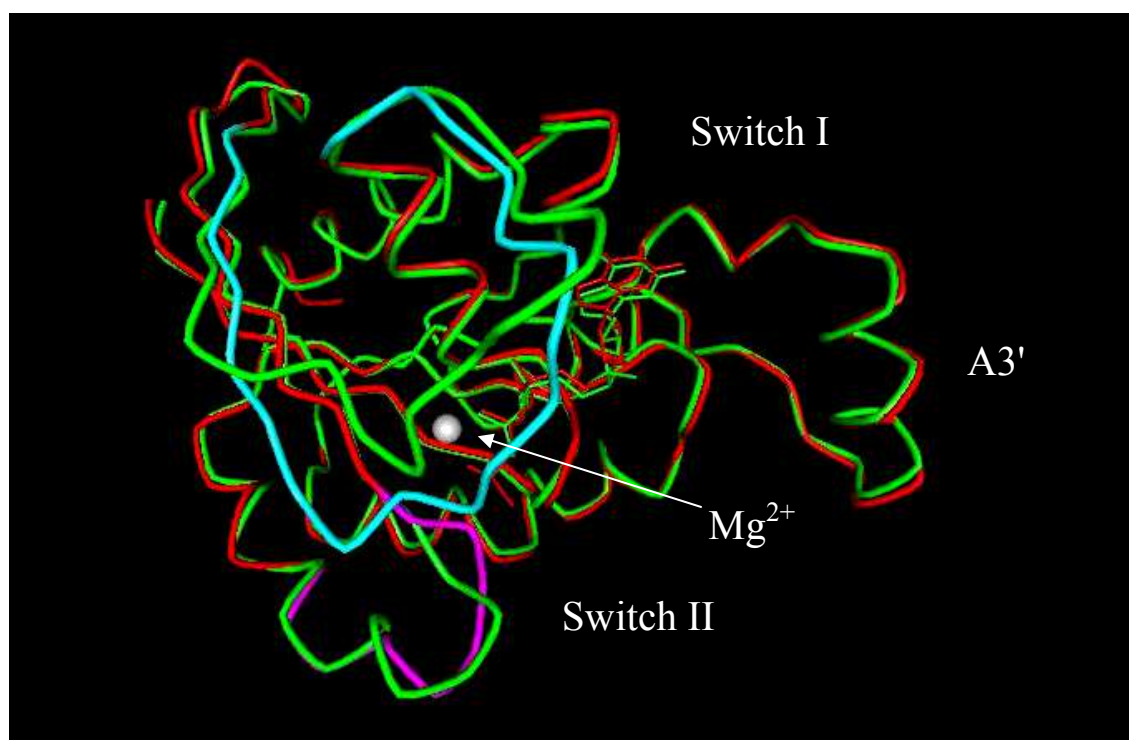


Figure 4

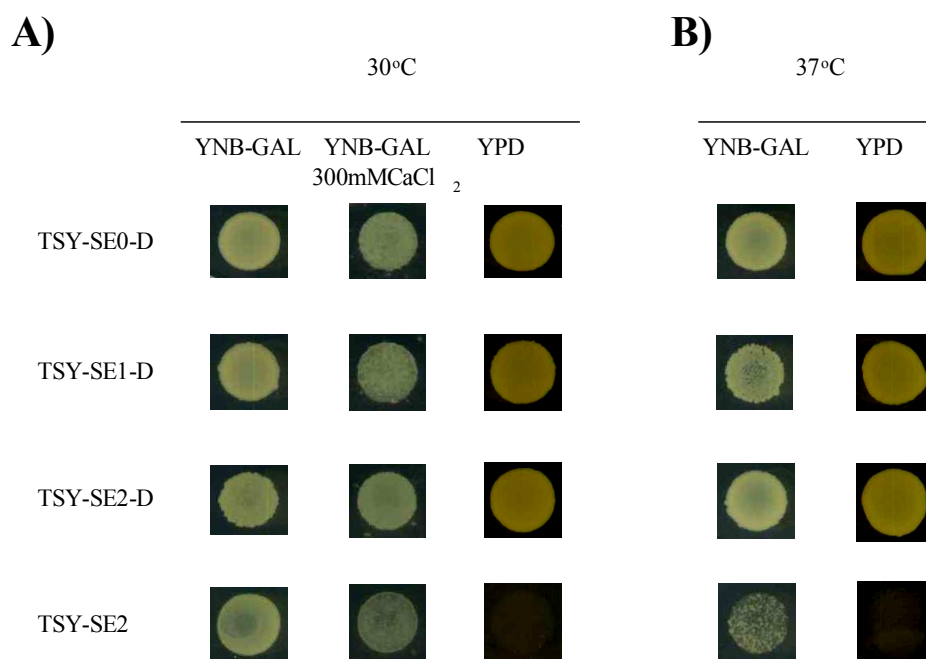


Figure 5

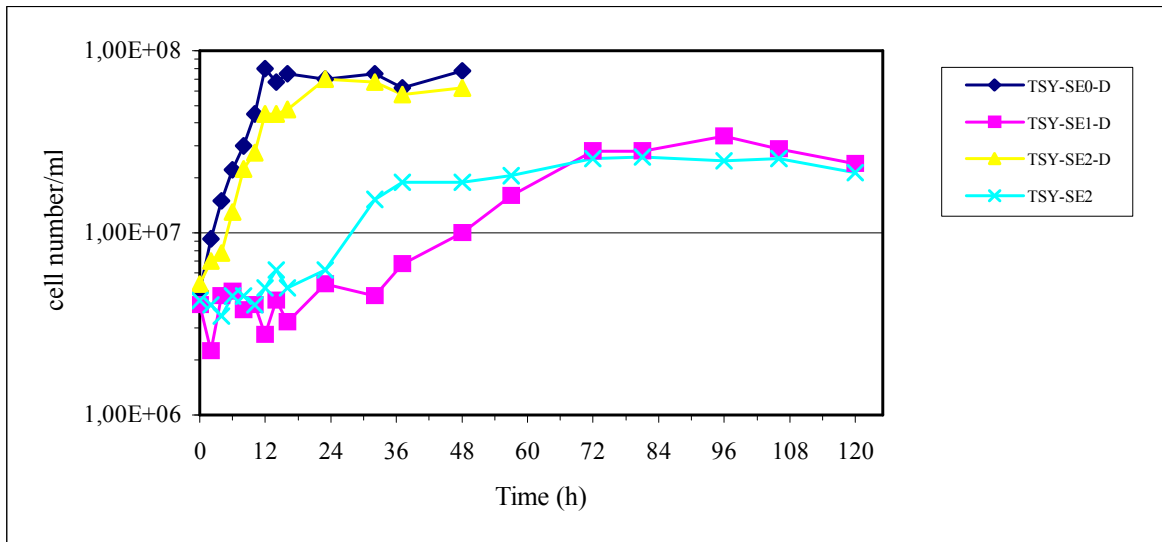
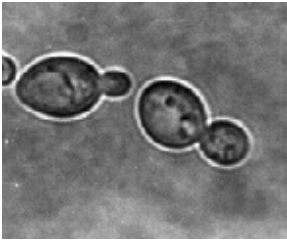
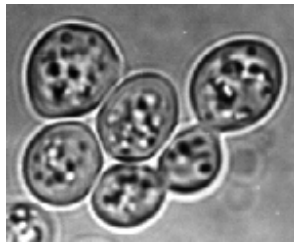


Figure 6

**A**



**B**



**C**



**D**

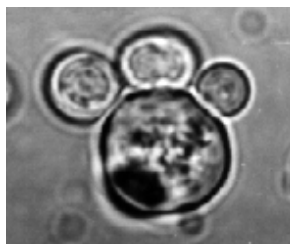


Fig 7

		$\alpha$ 3-helix	--L7--	I (%)	S (%)
( <i>S. cerevisiae</i> RHO1)	95	DSLENVQEKWIAEVLHFCQGVPI	117		
( <i>C. albicans</i> RHO1)	94	DSL <sup>green</sup> DNV <sup>red</sup> LEK <sup>green</sup> WIS <sup>green</sup> EVLHFCQGVPI	116	86	94
( <i>H. sapiens</i> RHOA)	90	DSLENI <sup>green</sup> PEK <sup>red</sup> WT <sup>red</sup> PEV <sup>red</sup> KHFC <sup>red</sup> PNVPI	112	69	73
( <i>S. mansoni</i> Rho)	91	DS <sup>red</sup> FANIE <sup>green</sup> EK <sup>green</sup> WL <sup>green</sup> PE <sup>red</sup> IR <sup>red</sup> HFC <sup>red</sup> PDVPI	113	56	73

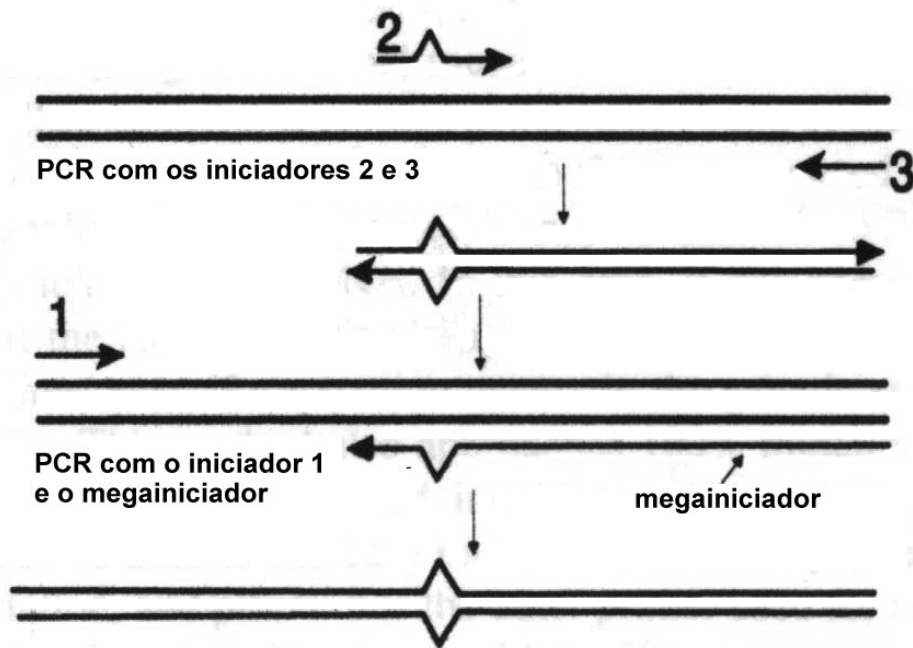
## 2. CONSIDERAÇÕES FINAIS

### 2.1. Experimentos Finais Para Complementação do Artigo 4

Conforme descrevemos no artigo 4, que está sendo preparado para ser submetido ao *Journal of Biological Chemistry*, o gene de *S. mansoni* que codifica uma proteína similar a Rho GTPases é capaz de complementar a ausência do gene Rho1 quando superexpresso em mutantes nulos de *Saccharomyces cerevisiae* obtidos artificialmente. A complementação é efetiva mesmo em condições restritivas de temperatura e concentração de Ca<sup>2+</sup>. Porém, verificamos alterações fenotípicas relacionadas ao ciclo celular, morfologia e brotamento. Duas hipóteses podem ser levantadas para explicar as alterações que nós observamos: que elas podem ter ocorrido simplesmente por efeito inespecífico da superexpressão do gene de *S. mansoni* na levedura ou por uma incapacidade do gene exógeno de complementar adequadamente as funções de Rho1 devido às diferenças existentes entre a proteína selvagem e heteróloga. Sendo assim, estamos clonando a smrho GTPase juntamente com o promotor natural de Rho1 para que dessa forma possamos verificar se numa condição normal de expressão a complementação das funções de Rho1 pela smrho é completa ou não.

Os nossos resultados descritos no artigo 4 indicam também que os aminoácidos Pro96 e Thr100 da RhoA GTPase humana sejam os responsáveis diretos pelo fato de que a expressão heteróloga do gene dessa proteína não complemente adequadamente mutantes nulos de *S. cerevisiae* em condições restritivas de temperatura e concentração de Ca<sup>2+</sup>. Para testarmos esta hipótese, estamos conduzindo experimentos de mutagênese sítio dirigida de smrho pela PCR utilizando a estratégia do megainiciador (SARKAR & SOMMER, 1990) mostrada na Fig. 3 para substituímos os aminoácidos Glu97 e Leu101 da sequência pelos resíduos Pro e Thr, respectivamente. Estão sendo preparados três mutantes: um primeiro substituindo apenas a Glu97 por uma Pro, um segundo substituindo apenas a Leu101 por uma Thr e um terceiro contendo as duas





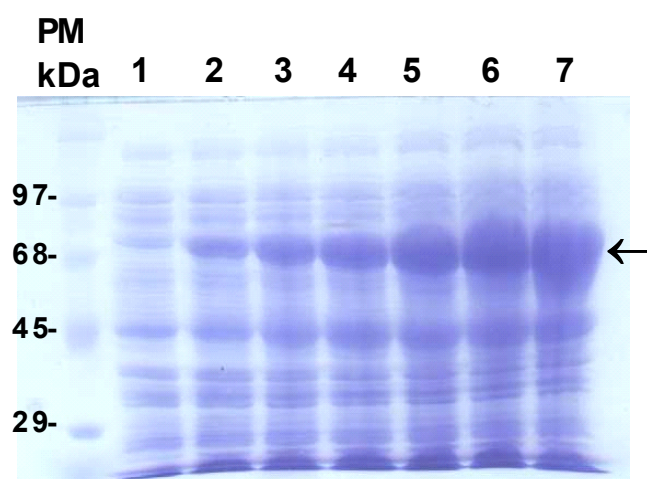
**Figura 3.** Esquema para geração de mutações sítio-dirigidas pela PCR usando megainiciadores. O método faz uso de apenas três iniciadores em duas rodadas de ampliações pela PCR. A primeira amplificação usa os iniciadores 2 e 3 para produzir um curto segmento de DNA sendo que o iniciador 2 é o que contém a mutação desejada. O produto desta primeira PCR é então purificado para remover os iniciadores originais e usado como "megainiciador" juntamente com o iniciador 1 para produzir o mutante desejado numa segunda amplificação pela PCR. Adaptado de REIKOFSKI & TAO (1992).

substituições ao mesmo tempo. Os genes de *smrho* mutados serão então clonados no vetor pYEDP 60-2 (POMPON, LOUERAT, BRONNE & URBAN, 1996) para que sejam superexpressos nos mutantes nulos de *S. cerevisiae*. Esperamos com isso não observar a complementação por pelo menos um dos genes mutados. Isso porque a superexpressão do gene da RhoA GTPase humana não é capaz de complementar mutantes nulos de *S. cerevisiae* (QADOTA & COLS., 1997) em nenhuma condição. Pretendemos clonar também os mesmos mutantes de *smrho* juntamente com o promotor natural de Rho1 no vetor pYEplac112 (GIETZ & SUGINO, 1988) e verificar também a complementação em condições normais de expressão. Através desses dois experimentos esperamos identificar indiretamente qual dos resíduos de aminoácidos da região da hélice  $\alpha 3$  de Rho1 são importantes para o crescimento da levedura em condições restritivas de temperatura e concentração de  $Ca^{2+}$ .

## **2.2. Expressão da proteína Rho GTPase de *S. mansoni* em *E. coli***

O gene que codifica SMRHO foi clonado no vetor pMAL-c2 (New England Biolabs) que provê um método para expressar e purificar a proteína produzida. O gene de *smrho* foi inserido no plasmídeo após a sequência do gene *malE* de *E. coli*, que codifica uma proteína ligadora de maltose (MBP). O vetor contém o promotor forte "tac" e os sinais iniciadores da tradução do gene *malE* que fazem com que as sequências clonadas em fase sejam expressas em grande quantidade. A proteína de fusão produzida pode ser purificada em um único passo por cromatografia de afinidade em colunas contendo resina de amilose com base na propriedade que MBP tem de se ligar a maltose. A proteína de fusão fica retida na coluna enquanto que o restante das proteínas da bactéria é eluído e desprezado até sua completa eliminação. A proteína de fusão purificada, que permanece ligada à resina, é coletada aplicando-se na coluna um tampão contendo maltose. O vetor contém o gene *lacIq*, que codifica o repressor Lac. Esse sistema mantém a expressão de Ptac baixa na ausência de indução por IPTG. O vetor contém ainda a sequência

codificadora do sítio de reconhecimento do fator Xa localizada logo antes da sequência do sítio múltiplo de clonagem. Isso permite que MBP (42,7 kDa) seja eliminada da proteína de interesse após a purificação. O fator Xa cliva a cadeia polipeptídica logo após o seu sítio de reconhecimento formado por 4 aminoácidos, de tal maneira que poucos ou nenhum resíduo de aminoácido derivado do vetor permaneça junto com a proteína de interesse, dependendo do sítio de restrição usado para clonagem. A Fig. 4 mostra um gel de poliacrilamida da cinética de indução da expressão da proteína SMRHO em fusão com MBP (~64 kDa).



**Figura 4.** Eletroforese em gel de poliacrilamida (SDS-PAGE) de extratos protéicos de *E. coli* expressando a proteína SMRHO clonada no vetor pMAL c2. Os extratos aplicados no gel foram de: **1.** cultura não induzida e de culturas onde a expressão de SMRHO foi induzida com IPTG por: **2.** 1 hora, **3.** 2 horas, **4.** 3 horas, **5.** 4 horas, **6.** 5 horas e **7.** 6 horas. As proteínas separadas foram coradas com azul de coomassie. A seta indica a posição da proteína SMRHO recombinante em fusão com MBP (~64 kDa).

### 3. PERSPECTIVAS

As possibilidades que surgem a partir dos resultados gerados por esse estudo são muitas e de grande relevância para compreensão de aspectos bioquímicos fundamentais de *S. mansoni*. Isso se aplica em especial à participação da proteína SMRHO GTPase no desenvolvimento do parasita, uma vez que avançamos bastante na caracterização molecular e funcional desse gene. Será interessante avaliar por Western blot a expressão de SMRHO em vermes adultos machos e fêmeas, acasalados ou não, assim como em todos os outros estágios do ciclo de vida do parasita, a partir de seus extratos protéicos. Tais experimentos poderão ser feitos utilizando soro policlonal de coelho imunizado com a proteína SMRHO recombinante já obtida e poderá dessa forma fornecer informações valiosas a respeito da regulação da expressão dessa GTPase de baixo peso molecular durante o desenvolvimento de *S. mansoni*. Assim será possível determinar, de uma maneira mais específica, o papel de SMRHO nos mecanismos envolvidos com a maturação sexual da fêmea e oviposição.

A identificação das moléculas efetoras e regulatórias de SMRHO no parasita, através de ensaios de duplo-híbrido em leveduras, poderá contribuir de maneira efetiva para que se possa entender melhor as funções e como são reguladas as GTPases de baixo peso molecular da família Rho em *S. mansoni*.

Contudo, pretendemos de modo mais imediato avançar nos estudos estruturais da SMRHO GTPase a partir da obtenção de cristais da proteína recombinante purificada. Poderemos assim determinar experimentalmente sua estrutura tridimensional por meio de cristalografia de raios-X e a partir dela desenhar inibidores específicos com vistas a produzir no futuro novas drogas contra o parasita. Os estudos de cristalografia deverão ser feitos em colaboração com a equipe do Prof. Richard Garratt do Instituto de Física de São Carlos da USP em São Carlos.

As análises da expressão do gene de *S. mansoni* homólogo a Rho GTPases nas diferentes fases do ciclo de vida do parasita serão feitas pela técnica de RT-PCR,

conforme FRANCO & COLS. (1997) e em colaboração com a Profa. Élide M. L. Rabelo, do Laboratório de Biologia Molecular de Parasitas do Depto. de Parasitologia do ICB-UFMG.

## 4. CONCLUSÕES

- Foram geradas mais 205 ESTs de uma biblioteca de verme adulto que permitiram a identificação de 91 genes diferentes dentre os quais havia 74 genes que ainda não tinham sido descritos em *S. mansoni*;
- A partir de uma amostra de 101 ESTs, a biblioteca 1 de cercária foi avaliada e aprovada para uso em larga escala no programa de descoberta de novos genes de *S. mansoni* com base em um número alto encontrado de sequências não identificadas e de sequências identificadas com homologia a genes de outros organismos, além de um baixo índice de redundância e de sequências com homologia a genes de *S. mansoni* e de sequências espúrias (rRNA e DNA mitocondrial não codificantes);
- Após a avaliação da qualidade da biblioteca 1 produzimos mais 518 sequências úteis. Estas sequências somaram-se a outras 439 produzidas a partir de uma segunda biblioteca de cercária pelo grupo do Dr David Johnston em Londres. Nós analisamos o conjunto dessas sequências com base em buscas por homologias em bancos de dados resultando em 957 ESTs de cercária. Desse total, 551 (57,6%) não foram identificadas enquanto que 308 (32,2%) foram identificadas por homologia com sequências de DNA ou proteínas depositadas nos bancos de dados e 98 (10,2%) eram ESTs não informativas (RNA ribossomal e sequências mitocondriais não codificantes);
- Dentre as 859 ESTs informativas identificamos 453 genes diferentes com base em análises de agrupamento de sequências homólogas;
- As sequências que codificavam calcium binding protein Sm8 (7,8%), frutose 1,6 bisfosfato aldolase (2,9%), GAPDH (1,8%), citocromo oxidase (1,4%), ATP guanidino

kinase (1.2%) e triose fosfato isomerase (1.0%) foram as mais comuns dentre as ESTs informativas;

- Foram detectados 57 grupos formados por ESTs não identificadas provindas de uma ou de ambas as bibliotecas. Alguns deles foram altamente redundantes como aquele encabeçado pela sequência AA559726 (3,8%);

- Cento e dezenove genes diferentes foram identificados e distribuídos em 11 categorias funcionais sendo que os genes associados com o metabolismo energético foram os mais abundantes (13%) e diversos. A diversidade e abundância de genes associados com as atividades de transcrição/tradução e com funções regulatórias e de sinalização também foi significativa;

- Um transcrito de paramiosina de *S. mansoni* foi identificado na biblioteca 1 indicando que este gene é expresso também em cercária e não apenas em esporocisto e verme adulto como havia sido sugerido anteriormente;

- Dentre os genes novos identificados destacamos a descoberta de genes que codificam transcritos com alto grau de homologia a proteínas ligadoras de cálcio da superfamília EF-hand (calmodulina NB1 e sorcina), que desempenham importantes funções regulatórias, proteínas relacionadas à percepção luminosa como o fotoreceptor rodopsina e a subunidade delta da fosfodiesterase de bastonete além de um receptor de benzodiazepínicos do tipo periférico, envolvido no metabolismo biosintético de esteróides;

- Identificamos também um transcrito que codifica uma proteína homóloga a uma GTPase de baixo peso molecular da família Rho. O sequenciamento completo desse gene revelou que a maior janela aberta de leitura encontrada codifica uma proteína de

193 aminoácidos que apresenta 78% de identidade com a Rho GTPase do molusco marinho *Aplysia californica* e tem uma massa molecular estimada em 21,8 kDa. A sequência genômica desse gene contém três pequenos introns. Com base em alinhamentos com sequências de proteínas homólogas já identificadas em diversos organismos, análises filogenéticas moleculares, caracterização estrutural da proteína por modelagem computacional comparativa e complementação funcional de mutantes nulos de *Saccharomyces cerevisiae*, preparados por nós para esse estudo, podemos afirmar que identificamos o primeiro gene que codifica uma GTPase de baixo peso molecular da família Rho de *S. mansoni* o qual denominamos smrho.



## **5. REFERÊNCIAS BIBLIOGRÁFICAS**

- ABATH F.G., HAGAN P., JEFFS S.A., SCHECHTER I., MEADOWS H.M., HOLDER A.A. & SIMPSON A.J. Structure of the gene encoding a putative *Schistosoma mansoni* tegumental antigen precursor. *Mol. Biochem. Parasitol.*, v. 60, p. 81-91, 1993.
- ADAMS M.D., CELNIKER S.E., HOLT R.A. & COLS. The genome sequence of *Drosophila melanogaster*. *Science*, v. 287, p. 2185-95, 2000.
- ADAMS M.D., KELLEY J.M., GOCAYNE J.D., DUBNICK M., POLYMEROPOULOS M.H., XIAO H., MERRIL C.R., WU A., OLDE B., MORENO R.F. & COLS. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science*, v. 252, p. 1651-6, 1991.
- ADAMS M.D., KERLAVAGE A.R., FLEISCHMANN R.D., FULDNER R.A., BULT C.J., LEE N.H., KIRKNESS E.F., WEINSTOCK K.G., GOCAYNE J.D., WHITE O & COLS. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature*, v. 377, p. 3-174, 1995
- ADAMS S.M., HELPS N.R., SHARP M.G., BRAMMAR W.J., WALKER R.A. & VARLEY J.M. Isolation and characterization of a novel gene with differential expression in benign and malignant human breast tumours. *Hum. Mol. Genet.*, v. 1, p. 91-6, 1992.
- ALI P.O., SIMPSON A.J., ALLEN R., WATERS A.P., HUMPHRIES C.J., JOHNSTON D.A. & ROLLINSON D. Sequence of a small subunit rRNA gene of *Schistosoma mansoni* and its use in phylogenetic analysis. *Mol. Biochem. Parasitol.*, v. 46, p. 201-8, 1991.
- ANDERSON L. & SEILHAMER J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis*, v. 18, p. 533-7, 1997.
- ANDERSON S., BANKIER A.T., BARRELL B.G., DE BRUIJN M.H., COULSON A.R., DROUIN J., EPERON I.C., NIERLICH D.P., ROE B.A., SANGER F., SCHREIER P.H., SMITH A.J., STADEN R. & YOUNG I.G. Sequence and organization of the human mitochondrial genome. *Nature*, v. 290, p. 457-65, 1981.
- ASHBURNER M. & GOODMAN N. Informatics--genome and genetic databases. *Curr. Opin. Genet. Dev.*, v. 7, p. 750-6, 1997.

- BAI G., JOHNSTON L.A., WATSON C.O & YOSHINO T.P. Phenoloxidase activity in the reproductive system of *Biomphalaria glabrata*: role in egg production and effect of schistosome infection. *J. Parasitol.*, v. 83, p. 852-8, 1997.
- BALLOUL J.M., GRZYCH J.M., PIERCE R.J. & CAPRON A. A purified 28,000 dalton protein from *Schistosoma mansoni* adult worms protects rats and mice against experimental schistosomiasis. *J. Immunol.*, v. 138, p. 3448-53, 1987.
- BECK S. & STERK P. Genome-scale DNA sequencing: where are we? *Curr. Opin. Biotechnol.*, v. 9, p. 116-21, 1998.
- BERTELSEN A.H. & VELCULESCU V.E. High-throughput gene expression analysis using SAGE. *Trends in Genetics*, v. 3, p. 152-159, 1998.
- BOGUSKI M.S., LOWE T.M. & TOLSTOSHEV C.M. dbEST--database for "expressed sequence tags". *Nat. Genet.*, v. 4, p. 332-3, 1993.
- BORK P. & KOONIN E.V. Predicting functions from protein sequences--where are the bottlenecks? *Nat. Genet.*, v. 18, p. 313-8, 1998.
- BOSWELL R.E., PROUT M.E. & STEICHEN J.C. Mutations in a newly identified *Drosophila melanogaster* gene, mago nashi, disrupt germ cell formation and result in the formation of mirror-image symmetrical double abdomen embryos. *Development*, v. 113, p. 373-84, 1991.
- BRACHMANN C.B., DAVIES A., COST G.J., CAPUTO E., LI J., HIETER P. & BOEKE J.D. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*, v. 14, p. 115-32, 1998.
- BRACHMANN R.K. & BOEKE J.D. Tag games in yeast: the two-hybrid system and beyond. *Curr. Opin. Biotechnol.*, v. 8, p. 561-8, 1997.
- BROWN P.O. & BOTSTEIN D. Exploring the new world of the genome with DNA microarrays. *Nat. Genet.*, v. 21, p. 33-7, 1999.

- BROWN S.D. & NOLAN P.M. Mouse mutagenesis-systematic studies of mammalian gene function. *Hum. Mol. Genet.*, v. 7, p. 1627-33, 1998
- BRUCE J.I., DIAS L.C., LIANG Y.S., COLES G.C. Drug resistance in schistosomiasis: a review. *Mem. Inst. Oswaldo Cruz*, v. 82, suplemento 4, p. 143-50, 1987.
- BUTTERWORTH, A.E. Immunological aspects of human schistosomiasis. *Br. Med. Bull.*, v. 54, p. 357-68, 1998
- CASTELLINO A.M. When the chips are down. *Genome Res.*, v. 7, p. 943-6, 1997.
- CHEEVER A.W. & YAP G.S. Immunologic basis of disease and disease regulation in schistosomiasis. *Chem. Immunol.*, v. 66, p. 159-76, 1997
- CHITSULO L, ENGELS D, SAVIOLI L & MONTRESOR A. The global status of schistosomiasis and its control. *Acta Tropica*, no prelo.
- COELHO, P.M., LIMA E SILVA, F.C. & NOGUEIRA-MACHADO, J.A. Resistance to oxamniquine of a *Schistosoma mansoni* strain isolated from patient submitted to repeated treatments. *Rev. Inst. Med. Trop. São Paulo*, v. 39, p. 101-6, 1997
- COLLINS F. & GALAS D. A new five-year plan for the U.S. Human Genome Project. *Science*, v. 262, p. 43-6, 1993
- COOKE R., MACHE R. & HOFTE H. EST and genomic sequence projects. In: Plant gene isolation: Principles and Practice, (Foster, G.D. and Twell, D., eds.). John Wiley & Sons Ltd, London, UK. pp. 401-419.
- DAVIES R.W., GALLAGHER E.J. & SAVIOZ A. Reverse genetics of the mouse central nervous system: targeted genetic analysis of neuropeptide function and reverse genetic screens for genes involved in human neurodegenerative disease. *Prog. Neurobiol.*, v. 42, p. 319-31, 1994.
- DAVIES T.W. & ERASMUS D.A. An ultrastructural study of the effect of parasitism by larval *Schistosoma mansoni* on the calcium reserves of the host, *Biomphalaria glabrata*. *Cell. Tissue Res.*, v. 236, p. 643-9, 1984.
- dbEST release 061600, [http://www.ncbi.nlm.nih.gov/dbEST/dbEST\\_summary.html](http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html), 2000.

- DE ANGELIS M.H., FLASWINKEL H., FUCHS H. & COLS. Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat. Genet.*, v. 25, p. 444-7, 2000.
- DE FRANCESCO L. Going beyond genomics with proteomics and two-dimensional gel technology. *The Scientist*, v. 13, n. 1, 1999.
- DERISI J.L., IYER V.R. & BROWN P.O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, v. 278, p. 680-6, 1997.
- DIAS NETO E., GARCIA CORREA R., VERJOVSKI-ALMEIDA S., BRIONES M.R., NAGAI M.A., DA SILVA W JR, ZAGO M.A., BORDIN S., COSTA F.F., GOLDMAN G.H., CARVALHO A.F., MATSUKUMA A., BAIA G.S., SIMPSON D.H., BRUNSTEIN A., DE OLIVEIRA P.S., BUCHER P., JONGENEEL C.V., O'HARE M.J., SOARES F., BRENTANI R.R., REIS L.F., DE SOUZA S.J. & SIMPSON A.J. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc. Natl. Acad. Sci. U S A*, v. 97, p. 3491-6, 2000.
- DIAS NETO E., HARROP R., CORREA-OLIVEIRA R., PENA S.D., WILSON R.A. & SIMPSON A.J. The schistosome genome project: RNA arbitrarily primed PCR allows the accelerated generation of expressed sequence tags. *Mem. Inst. Oswaldo Cruz*, v. 91, p. 655-7, 1996.
- DUNHAM I. & COLS. The DNA sequence of human chromosome 22. *Nature*, v. 402, p. 489-95, 1999. Published erratum appears in *Nature*, 2000 Apr 20;404(6780):904.
- DUNN J.J. & STUDIER F.W. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *Mol. Biol.*, v. 166, p. 477-535, 1983.
- DUNNE D.W., HAGAN, P & ABATH, F.G.C. Prospects for immunological control of schistosomiasis. *The Lancet*, v. 345, p. 1488-1492, 1995.
- ERICKSON, R.P. Antisense transgenics in animals. *Methods* v. 18, p. 304-10, 1999.
- EVANS M.J., CARLTON M.B. & RUSS A.P. Gene trapping and functional genomics. *Trends Genet.*, v. 13, p. 370-4, 1997.

- FAN J., MINCHELLA D.J., DAY S.R., McMANUS D.P., TIU W.U. & BRINDLEY P.J. Generation, identification, and evaluation of expressed sequence tags from different developmental stages of the Asian blood fluke *Schistosoma japonicum*. *Biochem. Biophys. Res. Commun.*, v. 252, p.348-56, 1998.
- FARIA M.S.C. Programa de descoberta gênica em *Schistosoma mansoni*. Geração e análise de 1335 Etiquetas de Sequências Transcritas (ESTs) de duas bibliotecas de ovo. Dissertação de Mestrado (2000) – Departamento de Biologia Geral do Instituto de Ciências Biológicas da UFMG, Belo Horizonte, MG, Brasil.
- FLEISHMANN R.D. & COLS. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, 269, p. 496-512, 1995
- FRANCO G.R. O uso das Etiquetas de Sequências Transcritas na identificação de genes de *Schistosoma mansoni*. Tese de Doutorado (1996) – Departamento de Bioquímica e Imunologia do Instituto de Ciências Biológicas da UFMG, Belo Horizonte, MG, Brasil.
- FRANCO G.R., ADAMS M.D., SOARES M.B., SIMPSON A.J., VENTER J.C. & PENA S.D. Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. *Gene*, v. 152, p. 141-147, 1995.
- FRANCO G.R., GARRATT R.C., TANAKA M., SIMPSON A.J. & PENA S.D. Characterization of a *Schistosoma mansoni* gene encoding a homologue of the Y-box binding protein. *Gene*, v. 198, p. 5-16, 1997.
- FRANCO G.R., RABELO E.M., AZEVEDO V., PENA H.B., ORTEGA J.M., SANTOS T.M., & COLS. Evaluation of cDNA libraries from different developmental stages of *Schistosoma mansoni* for production of expressed sequence tags (ESTs). *DNA Res.*, v. 4, p. 231-40, 1997a.
- FRANCO G.R., TANAKA M., SIMPSON A.J. & PENA S.D. Characterization of a *Schistosoma mansoni* homologue of the gene encoding the breast basic conserved

- protein 1/L13 ribosomal protein. *Comp. Biochem. Physiol. B. Biochem. Mol. Biol.*, v. 120, p. 701-8, 1998.
- FRANCO G.R., VALADAO A.F., AZEVEDO V. & RABELO E.M. The Schistosoma gene discovery program: state of the art. *Int. J. Parasitol.*, v. 30, p. 453-63, 2000.
- GIETZ RD & SUGINO A. New yeast-Escherichia coli shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. *Gene*, v. 74, p. 527-34, 1988.
- GOFFEAU A., BARRELL B.G., BUSSEY H., DAVIS R.W., DUJON B., FELDMANN H. & COLS. Life with 6000 genes. *Science*. v. 274, p. 546, 563-7, 1996.
- GRANJEAUD S., BERTUCCI F. & JORDAN B.R. Expression profiling: DNA arrays in many guises. *Bioessays*. v. 21, p. 781-90, 1999.
- GUYER M.S. & COLLINS F.S. How is the Human Genome Project doing, and what have we learned so far? *Proc. Natl. Acad. Sci. U. S. A.*, v. 92, p. 10841-8, 1995
- HATTORI M. & COLS. The DNA sequence of human chromosome 21. *Nature*, v. 405, p. 311-319, 2000.
- HELPS N.R., ADAMS S.M., BRAMMAR W.J. & VARLEY J.M. The Drosophila melanogaster homologue of the human BBC1 gene is highly expressed during embryogenesis. *Gene*, v. 162, n. 245-8, 1995.
- HILLIER L.D., LENNON G., BECKER M. & COLS. Generation and analysis of 280,000 human expressed sequence tags. *Genome. Res.*, v. 6 p. 807-828, 1996.
- HILLYER G.V. Buoyant density and thermal denaturation profiles of schistosome DNA. *J. Parasitol.*, v. 60, p. 725-7, 1974.
- HIRAI H. & LOVERDE P.T. Identification of the telomeres on *Schistosoma mansoni* chromosomes by FISH. *J. Parasitol.*, v. 82, p. 511-2, 1996.
- ISMAIL M., METWALLY A., FARGHALY A., BRUCE J., TAO L.F. & BENNETT J.L. Characterization of isolates of *Schistosoma mansoni* from Egyptian villagers that tolerate high doses of praziquantel. *Am. J. Trop. Med. Hyg.*, v. 55, p. 214-8, 1996

- JOHNSTON D.A. & COLS. Opening the can worms: molecular analysis of Schistosoma population. *Parasitology Today*, v. 9, p. 286-91, 1993.
- JOHNSTON D.A. The WHO/UNDP/World Bank Schistosoma Genome initiative: Current status. *Parasitology Today*, v. 13, p. 45-46, 1997.
- KATZ N. Schistosomiasis control in Brazil. *Mem. Inst. Oswaldo Cruz*, v. 93, suplemento 1, p. 33-5, 1998.
- KATZ N., COELHO P.M., PELLEGRINO J. Evaluation of Kato's quantitative method through the recovery of *Schistosoma mansoni* eggs added to human feces. *J. Parasitol.*, v. 56, p. 1032-3, 1970.
- KING, C.L. & HIGASHI, G.I. *Schistosoma mansoni*: silver ions (Ag<sup>+</sup>) stimulates and reversibly inhibits lipids-induced cercarial penetration. *Exp. Parasit.*, v. 75, p. 31-39, 1992.
- KLOSE J. & KOBALZ U. Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. *Electrophoresis*, v. 16, p. 1034-59, 1995.
- KUSTER B. & MANN M. Identifying proteins and post-translational modifications by mass spectrometry. *Curr. Opin. Struct. Biol.*, v. 8, p. 393-400, 1998.
- KUWABARA P.E. Worming your way through the genome. *Trends. Genet.*, v. 13, p. 455-60, 1997.
- KYTE J. & DOOLITTLE R.F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.*, v. 157, p. 105-32, 1982.
- LE PASLIER M.C., PIERCE R.J., MERLIN F., HIRAI H., WU W., WILLIAMS D.L., JOHNSTON D., LOVERDE P.T., LE PASLIER D. Construction and characterization of a *Schistosoma mansoni* bacterial artificial chromosome library. *Genomics*, v. 65, p. 87-94, 2000.
- MARRA M.A., HILLIER L. & WATERSON R.H. Expressed sequence tags – ESTablishing bridges between genomes. *Trends Genet.*, v. 14, p. 4-7, 1998.



- MATSUMURA, K., MITSUI, Y., SATO, K., SAKAMOTO, M. & AOKI, Y. *Schistosoma mansoni*: possible involvement of protein kinase C in linoleic acid-induced proteolytic enzyme release from cercariae. *Exp. Parasit.*, v. 72, p. 311-320, 1991.
- MAXAM A.M. & GILBERT W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA*, v. 74, p. 560-4, 1977.
- MEADOWS H.M. & SIMPSON A.J. Codon usage in *Schistosoma*. *Mol. Biochem. Parasitol.*, v. 36, p. 291-3, 1989.
- MEIRA W.S., FRANCO G.R., RABELO E.M. & PENA S.D. Characterization of an abundant *Schistosoma mansoni* transcript with no homologs in the databases. *Mem. Inst. Oswaldo Cruz*, v. 93, suplemento 1, p. 211-3, 1998.
- MILHON J.L. & TRACY J.W. Updated codon usage in *Schistosoma*. *Exp. Parasitol.* v. 80, p. 353-6, 1995.
- MUSTO H., ROMERO H. & RODRIGUEZ-MASEDA H. Heterogeneity in codon usage in the flatworm *Schistosoma mansoni*. *J. Mol. Evol.*, v. 46, p. 159-67, 1998.
- NAKAMURA Y., GOJOBORI T. & IKEMURA T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic acids Research*, 28(1), 292-292. 2000.
- NEVES, D.P. Parasitologia Humana. 10a edição, Editora Atheneu. Rio de Janeiro, RJ, 2000.
- NEVES, J. Esquistossomose mansoni. Clínica da forma aguda ou toxêmica. Rio de Janeiro, Medsi, 1986.
- NOLAN PM, PETERS J, STRIVENS M. & COLS. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat. Genet.*, v. 25, p. 440-3, 2000.
- O'Farrell P.H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, v. 250, p. 4007-21, 1975.
- OLIVER S. Guilt-by-association goes global. *Nature*, v. 403, p. 601-3, 2000.

- OLIVER S.G., WINSON M.K., KELL D.B. & BAGANZ F. Systematic functional analysis of the yeast genome. *Trends Biotechnol.*, v. 16, p. 373-8, 1998. Published erratum appears in *Trends Biotechnol.*, 1998 Oct;16(10):447
- OSMAN A., NILES E.G. & LOVERDE P.T. Characterization of the Ras homologue of *Schistosoma mansoni*. *Mol. Biochem. Parasitol.*, v. 100, p. 27-41, 1999.
- PEARCE E.J. & SHER A. Three major surface antigens of *Schistosoma mansoni* are linked to the membrane by glycosylphosphatidylinositol. *J. Immunol.*, v. 142, p. 979-84, 1989.
- PEARCE E.J., JAMES S.L., HIENY S., LANAR D.E. & SHER A. Induction of protective immunity against *Schistosoma mansoni* by vaccination with schistosome paramyosin (Sm97), a nonsurface parasite antigen. *Proc. Nat. Acad. Sci. U S A*, v. 85, p. 5678-82, 1988.
- PEARCE E.J., MAGEE A.I., SMITHERS S.R. & SIMPSON A.J. Sm25, a major schistosome tegumental glycoprotein, is dependent on palmitic acid for membrane attachment. *EMBO J.*, v. 10, p. 2741-6, 1991.
- PENA S.D. Third World participation in genome projects. *Trends Biotechnol.*, v. 14, p. 74-7, 1996.
- PEREIRA H.M. Caracterização do gene codificador da enzima purina nucleosídeo fosforilase de *Schistosoma mansoni*. Dissertação de Mestrado (1999) – Departamento de Física e Informática do Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP, Brasil.
- PESSÔA S.B. & MARTINS A.V. Parasitologia Médica, 11.ed. Rio de Janeiro: Guanabara Koogan, 1982. 872p.
- POMPON D, LOUERAT B., BRONNE A. & URBAN P. Yeast expression of animal and plant P540s in optimized redox experiments. *Methods Enzymol.*, 272: 51–64. 1996.

- QADOTA, H., ANRAKU, Y., BOTSTEIN, D. & OHYA, Y. Conditional lethality of a yeast strain expressing human RHOA in place of RHO1. *Proc. Natl. Acad. Sci. U S A*, v. 91, p. 9317-21, 1994.
- RABELO E.M., FRANCO G.R., AZEVEDO V.A., PENA H.B., SANTOS T.M. & COLS. Update of the gene discovery program in *Schistosoma mansoni* with the expressed sequence tag approach. *Mem. Inst. Oswaldo Cruz*, v. 92, p. 625-9, 1997
- RABELO E.M., HOBAIKA A.B., COELHO P.M. & PENA S.D. Sequencing and expression analysis of a *Schistosoma mansoni* gene homologue to a *Drosophila* gene involved in germ plasm assembly. *Mem. Inst. Oswaldo Cruz*, v. 93, suplemento 1, p. 207-9, 1998.
- RASO P., PEDROSO E.R.P. & NEVES, J. Patologia da forma aguda, toxêmica, da esquistossomose mansoni. *Rev. Soc. Bras. Med. Trop.*, v.19, p.45-55, 1986.
- RASTAN S. & BEELEY L.J. Functional genomics: going forwards from the databases. *Curr. Opin. Genet. Dev.*, v. 7, p. 777-83, 1997.
- REIKOFFSKI J. & TAO B.Y. Polymerase chain reaction (PCR) techniques for site-directed mutagenesis. *Biotech. Adv.*, v. 10, p. 535, 1992.
- REYNOLDS S.R., SHOEMAKER C.B. & HARN D.A. T and B cell epitope mapping of SM23, an integral membrane protein of *Schistosoma mansoni*. *J. Immunol.*, v. 149, p. 3995-4001, 1992.
- ROCHA M.O., PEDROSO E.R., NEVES J., ROCHA R.S., GRECO D.B., LAMBERTUCCI J.R., ROCHA R.L. & KATZ N. Characterization of the non-apparent clinical form in the initial phase of schistosomiasis mansoni. *Rev. Inst. Med. Trop. São Paulo*, v. 35, p. 47-51, 1993.
- SALAFSKY, B., WANG, S., KEVIN, M.B., HILL, H. & FUSCO, A.C. The role of prostaglandins in cercarial (*Schistosoma mansoni*) response to free fat acids. *J. Parasitol.*, v. 70, p. 584-591, 1984.

- SANGER F., NICKLEN S. & COULSON A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.*, v. 74, p. 5463-7, 1977.
- SANTOS F.R., PENA S.D. & EPPLEN J.T. Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. *Hum. Genet.*, v. 90, p. 655-6, 1993.
- SARKAR G. & SOMMER S.S. The "megaprimer" method of site-directed mutagenesis. *Biotechniques*, v. 8, p. 404-7, 1990.
- SCHIFF, C.J., CMELIK, S.H.W., LEY, H.E. & KRIEL, F.L. The influence of human skin lipids on the cercarial penetration responses of *Schistosoma haematobium* and *Schistosoma mansoni*. *J. Parasit.*, v. 58, p. 476-480, 1972.
- SHAW J.R., MARSHALL I. & ERASMUS D.A. *Schistosoma mansoni*: in vitro stimulation of vitelline cell development by extracts of male worms. *Exp. Parasitol.*, v. 42, p. 14-20, 1977.
- SHOEMAKER C., GROSS A., GEBREMICHAEL A., HARN D. cDNA cloning and functional expression of the *Schistosoma mansoni* protective antigen triose-phosphate isomerase. *Proc. Natl. Acad. Sci. U S A*, v. 89, p. 1842-6, 1992.
- SIMPSON A.J., SHER A. & MCCUTCHAN T.F. The genome of *Schistosoma mansoni*: isolation of DNA, its size, bases and repetitive sequences. *Mol. Biochem. Parasitol.*, v. 6, p. 125-37, 1982.
- SIMPSON, A.J.G. & COLS. The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature*, v. 406, p. 151-159, 2000.
- SMITHERS S.R. & TERRY R.J. The infection of laboratory hosts with cercariae of *Schistosoma mansoni* and the recovery of the adult worms. *Parasitology*, v. 55, p. 695-700, 1965,
- SOISSON L.M., MASTERSON C.P., TOM T.D., MCNALLY M.T., LOWELL G.H. & STRAND M. Induction of protective immunity in mice using a 62-kDa recombinant

- fragment of a *Schistosoma mansoni* surface antigen. *J. Immunol.*, v. 149, p. 3612-20, 1992.
- SPOTILA L.D., REKOSH D.M. & LoVERDE, P.T. Polymorphic repeated DNA element in the genome of *Schistosoma mansoni*. *Mol. Biochem. Parasitol.*, v. 48, p. 117-120, 1991.
- STADECKER M.J. & HERNANDEZ H.J. The immune response and immunopathology in infection with *Schistosoma mansoni*: a key role of major egg antigen Sm-p40. *Parasite Immunol.*, V. 20, p. 217-21, 1998
- STIREWALT, M.A. & FREGEAU, W.A. An invasive enzyme system present in cercariae but absent in schistosomules of *Schistosoma mansoni*. *Exp. Parasitol.*, v. 19, n. 2, p.206-215, 1966.
- STIREWALT, M.A. *Schistosoma mansoni*: cercaria to schistosomule. *Adv. Parasitol.*, v. 12, p. 115-182, 1974.
- STRAUSBERG R.L., DAHL C.A. & KLAUSNER R.D. New opportunities for uncovering the molecular basis of cancer. *Nat. Genet.*, v. 16, p. 415-6, 1997.
- TANAKA M., HIRAI H., LOVERDE P.T., NAGAFUCHI S., FRANCO G.R., SIMPSON A.J. & PENA S.D. Yeast artificial chromosome (YAC)-based genome mapping of *Schistosoma mansoni*. *Mol. Biochem. Parasitol.*, v. 69, p. 41-51, 1995.
- TANAKA M., TANAKA T., INAZAWA J., NAGAFUCHI S., MUTSUI Y., KAUKAS A., JOHNSTON D.A. & ROLLINSON D. Proceedings of the schistosome genome project. *Mem. Inst. Oswaldo Cruz*, v. 92, p. 829-34, 1997.
- TEIXEIRA, M. Brasil se afirma no seleto clube da genômica mundial. *Pesquisa FAPESP*, novembro, p. 19-24, 2000.
- TENDLER M., BRITO C.A., VILAR M.M., SERRA-FREIRE N., DIOGO C.M., ALMEIDA M.S. & COLS. A *Schistosoma mansoni* fatty acid-binding protein, Sm14, is the potential basis of a dual-purpose anti-helminth vaccine. *Proc. Natl. Acad. Sci. U S A*, v. 93, p. 269-73, 1996.

The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science.*, v. 282, p. 2012-8, 1998.

TROTTEIN F., GODIN C., PIERCE R.J., SELLIN B., TAYLOR M.G., GORILLOT I., SILVA M.S., LECOCQ J.P. & CAPRON A. Inter-species variation of schistosome 28-kDa glutathione S-transferases. *Mol. Biochem. Parasitol.*, v. 54, p. 63-72, 1992.

U.S. Human Genome Project 5-Year Research Goals 1998-2003. [http://www.ornl.gov/TechResources/Human\\_Genome/hg5yp/index.html](http://www.ornl.gov/TechResources/Human_Genome/hg5yp/index.html), 1998.

UniGene Human Release Statistics. <http://www.ncbi.nlm.nih.gov/UniGene/Hs.stats.shtml>, 2000.

VALADÃO A.F. Estudos funcionais da proteína SMYB1 de *Schistosoma mansoni*. Dissertação de Mestrado (1998) – Departamento de Bioquímica e Imunologia do Instituto de Ciências Biológicas da UFMG, Belo Horizonte, MG, Brasil.

WEBSTER P., MANSOUR T.E. & BIEBER D. Isolation of a female specific, highly repeated *Schistosoma mansoni* DNA probe and its use in an assay of cercarial sex. *Mol Biochem. Parasitol.*, v. 36, p. 217-22, 1989.

WHO, Fact Sheet 115, <http://www.who.int/inf-fs/en/fact115.html>, 1996.

WHO, Workplan 1997, [http://www.who.int/tdr/workplan/1997/schi\\_vac.htm](http://www.who.int/tdr/workplan/1997/schi_vac.htm), 1997

WHO, Report of the WHO Informal Consultation on schistosomiasis control. WHO/CDS/CPC/SIP/99.2, <http://www.who.int/ctd/schisto/strategies.htm>, 1998a.

WHO, Report of the WHO informal consultation on monitoring of drug efficacy in the control of schistosomiasis and intestinal nematodes. WHO/CDS/CPC/SIP/99.1, [http://www.who.int/emc/WHO\\_docs/schisto.html](http://www.who.int/emc/WHO_docs/schisto.html), 1998b.

WHO/TDR REPORT, Scientific working group on the utilization of genomic information for tropical disease drug and vaccine discovery, 1998c

WHO, <http://www.who.int/ctd/schisto/>, 2000.

- WILSON R.A., DRASKAU T., MILLER P. & LAWSON J.R. *Schistosoma mansoni*: the activity and development of the schistosomulum during migration from the skin to the hepatic portal system. *Parasitology*, v. 77, p. 57-73, 1978.
- WINZELER E.A. & DAVIS R.W. Functional analysis of the yeast genome. *Curr. Opin. Genet. Dev.*, v. 7, p. 771-6, 1997.
- YATES J.R. 3RD Mass spectrometry. From genomics to proteomics. *Trends Genet.*, v. 16, p. 5-8, 2000.
- ZOUAIN C.S., AZEVEDO V.A., FRANCO G.R., PENA S.D. & GOES A.M. Identification of genes encoding *Schistosoma mansoni* antigens using an antigenic sequence tag strategy. *J. Parasitol.*, v. 84, p. 1307-10, 1998.
- ZWEIGER G. & SCOTT R.W. From expressed sequence tags to 'epigenomics': an understanding of disease processes. *Curr. Opin. Biotechnol.*, v. 8, p. 684-7, 1997.

## APÊNDICES

### A.1. Aminoácidos e suas propriedades

Aminoácidos	Nomes abreviados		Massa molecular	Índice de hidropaticidade KYTE & DOOLITTLE (1982)
<i>Não-polares, alifáticos</i>				
Glicina	Gly	G	75	-0,4
Alanina	Ala	A	89	1,8
Valina	Val	V	117	4,2
Leucina	Leu	L	131	3,8
Isoleucina	Ile	I	131	4,5
Prolina	Pro	P	115	-1,6
<i>Aromáticos</i>				
Fenilalanina	Phe	F	165	2,8
Triptofano	Trp	W	204	-0,9
Tirosina	Tyr	Y	181	-1,3
<i>Polares, não carregados</i>				
Serina	Ser	S	105	-0,8
Treonina	Thr	T	119	-0,7
Cisteína	Cys	C	121	2,5
Metionina	Met	M	149	1,9
Asparagina	Asn	N	132	-3,5
Glutamina	Gln	Q	146	-3,5
<i>Carregados negativamente</i>				
Ác. Aspártico	Asp	D	133	-3,5
Ác. Glutâmico	Glu	E	147	-3,5
<i>Carregados positivamente</i>				
Lisina	Lys	K	146	-3,9
Arginina	Arg	R	174	-4,5
Histidina	His	H	155	-3,2



## A.2. Codon Usage de *Schistosoma mansoni*

179 genes (63970 codons)

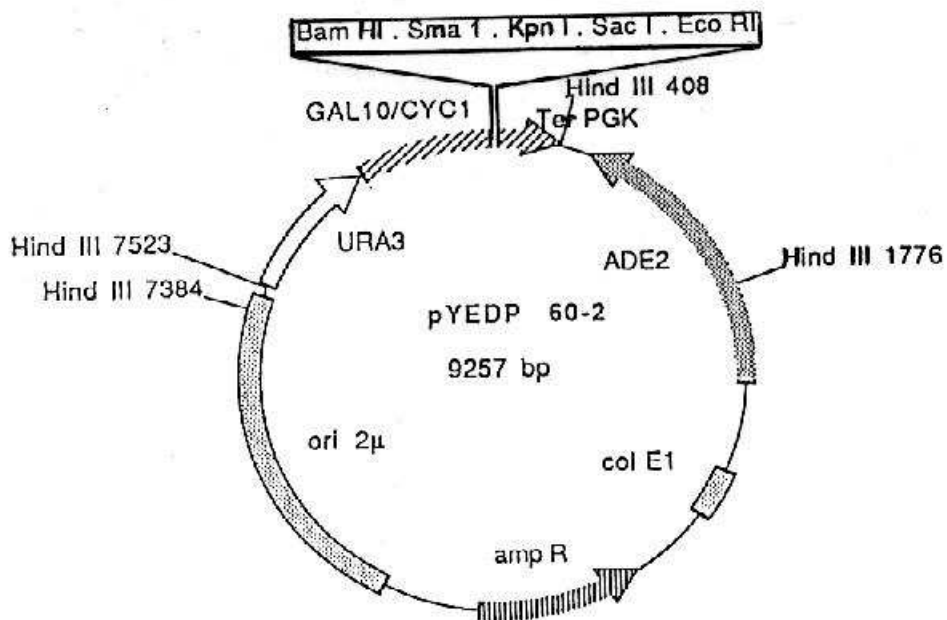
campos: [tripleto] [aminoácido] [fração] [frequência/1000]

<b>UUU</b> F 0.60 24.4 (1563)	<b>UCU</b> S 0.23 17.3 (1108)	<b>UAU</b> Y 0.71 23.5 (1503)	<b>UGU</b> C 0.76 16.0 (1023)
<b>UUC</b> F 0.40 16.4 (1048)	<b>UCC</b> S 0.09 7.1 (452)	<b>UAC</b> Y 0.29 9.8 (626)	<b>UGC</b> C 0.24 5.2 (331)
<b>UUA</b> L 0.35 31.6 (2021)	<b>UCA</b> S 0.29 21.6 (1382)	<b>UAA</b> * 0.46 1.6 (102)	<b>UGA</b> * 0.32 1.1 (71)
<b>UUG</b> L 0.18 16.3 (1043)	<b>UCG</b> S 0.07 5.3 (338)	<b>UAG</b> * 0.21 0.7 (47)	<b>UGG</b> W 1.00 10.1 (648)
<b>CUU</b> L 0.19 17.5 (1118)	<b>CCU</b> P 0.34 13.8 (881)	<b>CAU</b> H 0.76 17.9 (1145)	<b>CGU</b> R 0.38 17.9 (1148)
<b>CUC</b> L 0.06 5.5 (354)	<b>CCC</b> P 0.09 3.7 (239)	<b>CAC</b> H 0.24 5.6 (360)	<b>CGC</b> R 0.08 3.9 (249)
<b>CUA</b> L 0.12 10.5 (671)	<b>CCA</b> P 0.46 18.6 (1188)	<b>CAA</b> Q 0.74 27.9 (1786)	<b>CGA</b> R 0.20 9.3 (595)
<b>CUG</b> L 0.09 8.3 (532)	<b>CCG</b> P 0.10 4.2 (266)	<b>CAG</b> Q 0.26 9.9 (635)	<b>CGG</b> R 0.05 2.5 (158)
<b>AUU</b> I 0.49 32.0 (2046)	<b>ACU</b> T 0.37 21.4 (1366)	<b>AAU</b> N 0.76 40.6 (2597)	<b>AGU</b> S 0.25 18.9 (1208)
<b>AUC</b> I 0.20 12.9 (823)	<b>ACC</b> T 0.11 6.4 (410)	<b>AAC</b> N 0.24 13.0 (832)	<b>AGC</b> S 0.07 5.1 (329)
<b>AUA</b> I 0.31 19.9 (1271)	<b>ACA</b> T 0.42 24.1 (1542)	<b>AAA</b> K 0.74 49.3 (3152)	<b>AGA</b> R 0.23 10.7 (684)
<b>AUG</b> M 1.00 25.3 (1617)	<b>ACG</b> T 0.10 5.9 (377)	<b>AAG</b> K 0.26 17.4 (1110)	<b>AGG</b> R 0.07 3.2 (206)
<b>GUU</b> V 0.42 27.0 (1727)	<b>GCU</b> A 0.47 27.9 (1785)	<b>GAU</b> D 0.78 43.8 (2803)	<b>GGU</b> G 0.50 31.8 (2034)
<b>GUC</b> V 0.15 9.6 (617)	<b>GCC</b> A 0.14 8.0 (512)	<b>GAC</b> D 0.22 12.6 (808)	<b>GGC</b> G 0.13 8.2 (526)
<b>GUA</b> V 0.25 16.3 (1044)	<b>GCA</b> A 0.32 18.9 (1207)	<b>GAA</b> E 0.77 50.2 (3214)	<b>GGA</b> G 0.30 19.0 (1217)
<b>GUG</b> V 0.18 11.5 (737)	<b>GCG</b> A 0.07 4.2 (270)	<b>GAG</b> E 0.23 14.7 (939)	<b>GGG</b> G 0.08 5.1 (329)

Fonte: Codon Usage Database (NAKAMURA, GOJOBORI & IKEMURA, 2000).

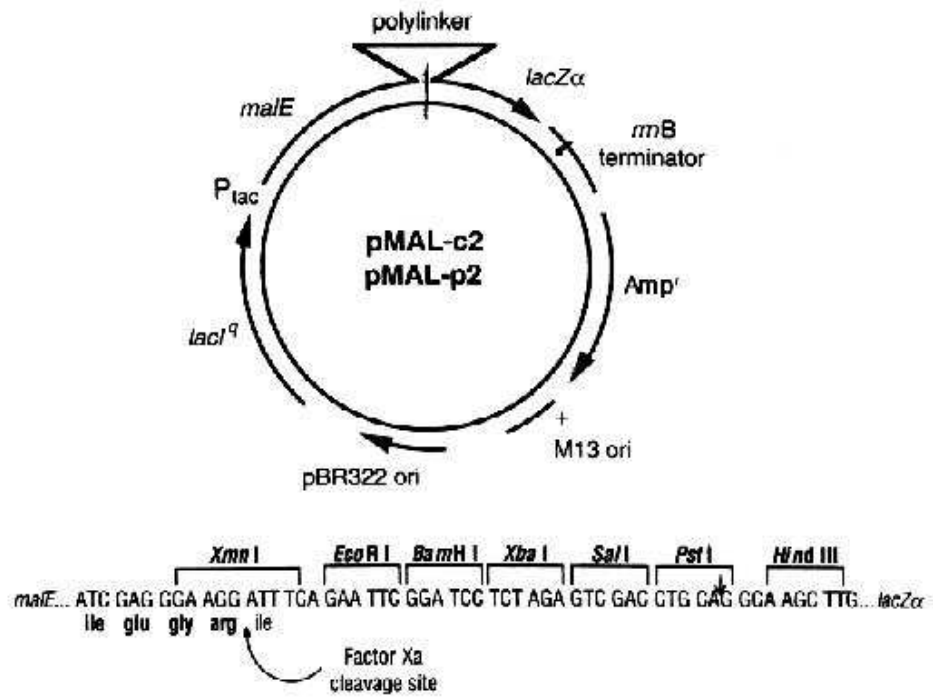
<http://www.kazusa.or.jp/codon/>

### A.3. Mapa do vetor pYEDP 60-2



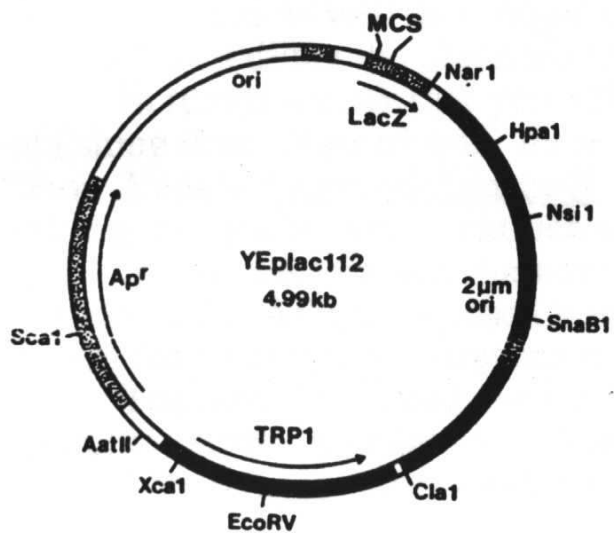
Fonte: POMPON, LOUERAT, BRONNE & URBAN (1996).

#### A.4. Mapa do vetor pMAL



Fonte: New England Biolabs, Ipswich - MA, EUA.

### A.5. Mapa do vetor pYEplac112



Fonte: GIETZ & SUGINO (1988)