UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA

**CONSTRUÇÃO E CARACTERIZAÇÃO DE MINI-
BIBLIOTECAS DE EST GERADAS COM RT-PCR DE BAIXA
ESTRINGENCIA E CLONAGEM DE UMA APIRASE DE**
Schistosoma mansoni.

RICARDO DE MARCO
TESE DE DOUTORADO

ORIENTADOR: PROF. DR. SERGIO VERJOVSKI-ALMEIDA

SÃO PAULO
24 de Fevereiro de 2003

28
11/03

Ficha Catalográfica

Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP.

De Marco, Ricardo

D372c Construção e caracterização de mini-bibliotecas de EST
geradas com RT-PCR de baixa estringencia e clonagem de
uma apirase de *Schistosoma mansoni* / Ricardo De Marco.
-- São Paulo, 2003.
121p

Tese (doutorado) - Instituto de Química da Universidade
de São Paulo. Departamento de Bioquímica.
Orientador : Verjovski-Almeida, Sergio

I. Biologia molecular I. T. II. Verjovski-Almeida,
Sergio, orientador.

574.88 CDD

“Construção e Caracterização de Mini- Bibliotecas de EST Geradas com RT- PCR de Baixa Estringencia e Clonagem de uma Apirase de *Schistosoma mansoni*”

RICARDO DE MARCO

Tese de Doutorado submetida ao Instituto de Química da Universidade de São Paulo como parte dos requisitos necessários à obtenção do grau de Doutor em Ciências – Área: Bioquímica.

Aprovada por:

Prof. Dr. SERGIO VERJOVSKI DE ALMEIDA
IQ – USP
(Orientador e Presidente)

Prof. Dr. HAMZA FAHMI ALI EL DORRY
IQ – USP

Prof. Dr. EMMANUEL DIAS NETO
FM – USP

Prof. Dr. HERNANDO ANTONIO DEL PORTILLO OBANDO
ICB – USP

Prof. Dr. WIM MAURITS SYLVAIN DEGRAVE
FIOCRUZ – RJ

SÃO PAULO
24 DE FEVEREIRO 2003.

“Não devemos acreditar nos muitos que dizem que só as pessoas livres devem ser educadas, deveríamos antes acreditar nos filósofos que dizem que apenas as pessoas educadas são livres”.

Epicteto, filósofo romano.

Dedico esta tese aos meus pais,
Mario e Francisca,
cujo amor e dedicação
me permitiriam chegar ate aqui.

AGRADECIMENTOS

Ao Prof. Dr. Sergio Verjovski-Almeida pela orientação e oportunidade de desenvolver este trabalho.

Aos professores e funcionários do Instituto de Química da USP.

À Dra. Cybele Gargioni e o setor de endoparasitoses do Instituto Adolfo Lutz pelo fornecimento dos parasitas utilizados neste trabalho.

À Juliana Fietto e André Kowaltowski que participaram ativamente de parte deste trabalho.

Ao amigo Fabio Siviero pelo empréstimo de materiais e protocolos que ajudaram a realização deste trabalho.

Aos amigos da graduação César, Gláucia, Herbert, Rodrigo e Sheila pela amizade e apoio.

Aos amigos, colegas e ex-colegas de laboratório: Abimael, Adriana, Ana Claudia, Ana Paula, Apuã, Camila, Camille, Carol, Cristielle, Daniel, Denise, Flavio, Eduardo, Gilson, Giulliana, Glauber, Helder, Ivone, Luciana, Jefferson, Milton, Renato e Rodrigo.

A todos os amigos dos diferentes laboratórios do Instituto de Química.

A minha família: Eduardo, Francisca, Gabriella, Joanna, Marcio, Mariella e Mario, pelo incentivo e carinho dados durante todo este período.

À FAPESP pela bolsa concedida.

Índice

1-Introdução.....	1
1.1-O <i>Schistosoma mansoni</i>	1
1.2-A esquistossomose.....	5
1.3-Aspectos imunológicos da esquistossomose.....	7
1.4-O tegumento do <i>Schistosoma mansoni</i>	9
1.5-As ATP-difosfohidrolases.....	11
1.6-A ATP-difosfohidrolase do <i>Schistosoma mansoni</i>	12
1.7-A estratégia de EST.....	15
2-Objetivos.....	17
3-Materiais e Métodos.....	19
3.1- Materiais.....	19
3.2- Obtenção de parasitas.....	19
3.3-Extração de RNA mensageiro de <i>Schistosoma mansoni</i>	20
3.4- Transcrição reversa.....	21
3.5- Reações de PCR para construção das bibliotecas.....	21
3.6 – Análise das reações de RT-PCR utilizando gel de poliacrilamida corado por prata.....	22
3.7- Seleção e Ligação de Fragmentos de DNA.....	22
3.8- Preparação de células eletrocompetentes.....	23
3.9 Transformação de bactérias eletrocompetentes.....	24
3.10- Extração de plasmídeos pelo método de lise por fervura.....	24
3.11- Seqüenciamento dos clones	25
3.12- Análise do seqüenciamento.....	25
3.13- Amplificação rápida de pontas de cDNA (RACE).....	26
3.14- Expressão da proteína recombinante em <i>E. coli</i>	26

3.15- Solubilização dos corpos de inclusão.....	27
3.16- Coluna de níquel para purificação de proteína com hexahistidina....	27
3.17- Produção de anticorpo anti-apirase de <i>S. mansoni</i>	28
3.18- Extração de tegumento de verme adulto.....	28
3.19- Solubilização do tegumento para ensaio de Western-blot.....	29
3.20-Western-blots.....	29
3.21- Extração de DNA genômico de <i>S. mansoni</i>	31
3.22-Southern-blots.....	31
4- Resultados e discussão.....	33
4.1-Produção de minibibliotecas de EST de <i>S. mansoni</i>	33
4.1.1- Verificação da integridade do RNA mensageiro.....	33
4.1.2- Determinação dos parâmetros da reação de PCR.....	33
4.1.3-Estudo da influência dos primers na construção de bibliotecas.....	45
4.1.4- Análise de parâmetros globais da técnica.....	59
4.1.5- Avaliação da contribuição das seqüências geradas para o conhecimento do transcriptoma de <i>S. mansoni</i>	62
4.1.6- Avaliação da homologia entre genes de <i>S. mansoni</i> e de organismos totalmente seqüenciados.....	77
4.2- Clonagem de uma ATP-difosfohidrolase (Apirase) de <i>S. mansoni</i>	79
4.2.1- Amplificação do gene da Apirase de <i>S. mansoni</i> utilizando a técnica de RACE.....	79
4.2.2- Análise da seqüência do gene da apirase.....	82
4.2.3- Detecção do gene da apirase no genoma do parasita.....	95
4.2.4- Expressão heteróloga da apirase de <i>S. mansoni</i>	96
4.2.5- Purificação em coluna de níquel e imunização para produção de soro policlonal.....	102
5-Conclusões e perspectivas.....	106

5.1-Construção de bibliotecas de cDNA.....	106
5.2-Clonagem da apirase de <i>S. mansoni</i>	108
6-Bibliografia.....	110
7-Curriculum vitae.....	114

Resumo

Este trabalho demonstrou a construção de minibibliotecas de “Expressed Sequence Tags” (EST) com o uso de RT-PCR de baixa estringência e “primer” consenso-degenerado. Através do estudo de parâmetros críticos como concentração de sais, temperatura e velocidade de ciclagem, composição dos “primers” e qualidade do RNA mensageiro, foi possível padronizar um protocolo. Tal protocolo permitiu um aumento do número de seqüências por minibiblioteca em relação a protocolos similares existentes na literatura (Dias *et al.*, 1997 e 2000) sem perda de características desejáveis como amplificação preferencial do centro dos genes e normalização das mensagens.

As seqüências produzidas levaram a um significativo aumento das seqüências de EST de *S. mansoni* disponíveis publicamente e permitiu a detecção da existência de novos fragmentos de genes expressos na fase adulta do parasita.

Neste trabalho também clonamos uma apirase de *S. mansoni* cuja a seqüência não havia sido descrita anteriormente. O gene possui cerca de 2,7 mil pares de bases e codifica para uma proteína de 544 aminoácidos. Esta foi expressa em sistema heterólogo e utilizada para obtenção de anticorpos policlonais contra esta proteína. Utilizando estes anticorpos foi possível detectar a proteína no tegumento do parasita.

Summary

This work demonstrates the construction of EST minilibraries employing low stringency RT-PCR and consensus-degenerate primers. Through the study of critical parameters such as salt concentration, cycling temperature and ramp speed, composition of primers and quality of messenger RNA a standard protocol was obtained. Such protocol allowed an increase in the number of sequences per minilibrary in relation to similar protocols in the literature (Dias Neto et al., 1997 and 2000) without no loss in desirable characteristics such as preferential amplification of the central portion of messages and normalization of messages.

The sequences produced allowed a significant increase in the publicly available EST sequences of *S. mansoni* and permitted discovery of new gene fragments expressed in adult stage of the life cycle of the parasite.

In this work we also cloned an apyrase of *S. mansoni* whose sequence had never been described previously. The gene has about 2.7 kilobases and codes for a protein of 544 aminoacids. This protein was expressed in a heterologous system and was used for production of a polyclonal antibody. This antibody was used for detection of this protein in the tegument of the parasite.

Abreviações

BAC- “Bacterial Artificial Chromosome”(Cromossomo artificial de bactéria).

BLAST- “Basic Local Alignment Search Tool” (Ferramenta básica de procura de alinhamentos locais).

cDNA- DNA complementar ao RNA mensageiro.

DNA- Acido deoxiribonucleico

EST- “Expressed Sequence Tag”(Etiqueta de sequência expressa).

IPTG- Isopropiltio- β -D-galactosideo.

kDa- kiloDaltons

ORF- “Open Reading Frame” (Quadro de leitura aberto)

PBS- “Phosphate Buffered Saline”(Tampão salina fosfato)

PCR- “Polymerase Chain Reaction” (reação em cadeia da polimerase).

RACE- “Rapid Amplification of cDNA Ends”(Amplificação rápida de pontas de cDNA)

RNA- Acido ribonucléico.

RT-PCR- Reação sucessiva de transcrição reversa e reação em cadeia da polimerase.

SDS- Dodecilsulfato de sódio.

SDS-PAGE- Eletroforese em gel de poliacrilamida com Dodecilsulfato de sódio.

SNP- “Single Nucleotide Polymorphism” (Polimorfismo de uma única base).

1-Introdução

1.1-O *Schistosoma mansoni*

O *Schistosoma mansoni* é membro da família Schistosomatidae, classe Trematoda, subclasse Digenea. Ele é o principal agente etiológico da esquistossomose, uma doença crônica que atinge humanos. O *Schistosoma* difere de outros tremátodos pelo fato de serem dióicos e os dois sexos apresentarem diferenças morfológicas marcantes.

Das espécies de *Schistosoma* que habitualmente parasitam o homem somente a *mansoni* existe na América. Ele é encontrado também na África e no Oriente Médio.(figura 1) As outras principais espécies são o *Schistosoma haematobium* que se concentra em países da África e Leste do Mediterrâneo e o *Schistosoma japonicum*, que se concentra no Sudeste Asiático e no Pacífico Ocidental.

Figura 1- Distribuição geográfica do *Schistosoma mansoni* (Segundo a WHO, 1985).

A introdução do *S. mansoni* na América é um fato recente visto que o nível de diferença de polimorfismos do DNA mitocondrial entre isolados africanos e americanos é o mesmo do que entre isolados americanos comparados entre si (Deprès et al., 1993). A hipótese mais provável é a introdução do *Schistosoma* há algumas centenas de anos atrás, devido ao tráfico de escravos que trouxe um grande número de populações nativas da África que deveriam ter infecções pelo *S. mansoni*.

O parasita possui um ciclo de vida que inclui a passagem por dois hospedeiros e formas intermediárias (figura 2)

Figura 2- O ciclo de vida do *Schistosoma* (segundo Cummings and Nyame, 1999).

A forma adulta tem como hospedeiro o ser humano. A fêmea adulta é delgada e mais longa que o macho. O macho é mais robusto e tem em sua parte ventral o canal ginecóforo no qual se alojada a fêmea. Eles se fixam nas paredes dos vasos por meio de ventosas. O *Schistosoma* encontra condições favoráveis para seu desenvolvimento no sistema porta intra-

hepático (Veronesi, 1972). Ele obtém aminoácidos para seu crescimento através da ingestão de eritrócitos, e é estimado que macho e a fêmea consumam aproximadamente 39.000 e 330.000 eritrócitos por hora, respectivamente. Presume-se que a maior necessidade de alimentação pela fêmea seja devido a maior demanda por proteínas devido a produção de ovos (Halton, 1997). Os eritrócitos são lisados e a hemoglobina é proteolisada por enzimas como as catepsinas D, L, B e C para dar origem a di-peptídeos ou aminoácidos que são absorvidos pelas células gastrodermais (Brindley et al., 1997).

Existe um conhecimento muito restrito de quais os mecanismos das interações sexuais do *Schistosoma* a nível molecular. É conhecido que infecções unissexuais de *Schistosoma* dão origem a adultos imaturos (Khalil and Mansour, 1995); foi demonstrado também que quando machos e fêmeas são mantidos separados *in vitro* há um padrão diferente de expressão gênica do que quando eles são mantidos em pares nas mesmas condições (Grevelding et al., 1997).

A fêmea produz de 300 a 3000 ovos por dia. Os ovos apresentam espinhos que assistem na retenção dos ovos nos vasos sanguíneos. O destino principal dos ovos é romper a barreira para o intestino e serem eliminados pelas fezes. Aparentemente os ovos conseguem cruzar o endotélio para chegar ao intestino via um mecanismo de transporte que envolve ativamente as células endoteliais (File, 1995).

Parte destes ovos acabam sendo depositados no fígado, onde realizam uma contínua liberação de antígenos que estimulam uma intensa resposta imune. Isto leva a uma acumulação de células do sistema imune ao redor do ovo formando um granuloma. Estes granulomas são responsáveis por uma fibrose devido ao dano tecidual. Em casos mais agudos ocorre uma hepatosplenomegalia devido a um impedimento do fluxo sanguíneo pelo fígado.

Os ovos que conseguem ser eliminados pelo hospedeiro em condições apropriadas de luz, temperatura e umidade se rompem liberando o miracídio que penetra em um molusco do gênero

Biomphalaria, que é o hospedeiro intermediário. O miracídio utiliza o reconhecimento de um glicoconjugado de alto peso molecular para reconhecer o hospedeiro intermediário (Halber et al., 1995; Kalbe et al., 1998).

Neste hospedeiro se forma um esporocisto mãe que dará origem a larvas de vida livre, as cercárias, que serão eliminadas na água. Estas possuem um fototropismo que causa sua acumulação na superfície de águas rasas onde é mais freqüente o contato com humanos. Sob o contato da pele humana a cercaria responde a sinais químicos, principalmente de ácidos graxos de cadeia media, como um sinal para invasão. Elas se fixam sobre a pele ou mucosas e penetram com o auxílio de substâncias líticas secretadas pela glândula acetabular em um mecanismo ativo de penetração.

A cercária então perde sua cauda e o glicocálix, um revestimento complexo de carboidratos, e começa se diferenciar com esquistossômulo. Ele se aprofunda na epiderme até a derme onde precisa achar uma veia ou um vaso linfático para entrar na circulação. Uma vez dentro do espaço vascular ocorre a parada de secreções da glândula acetabular (McKerrow and Salter, 2002).

O esquistossômulo é carregado pelo fluxo sanguíneo, e de quatro a seis dias após a infecção chega até o pulmão onde sofre uma série de mudanças morfológicas. O parasita começa a se alimentar de plasma sanguíneo e migra então para a veia porta hepática para completar seu desenvolvimento e se tornar sexualmente maduro.

O *Schistosoma mansoni* possui um genoma estimado em 270 milhões de pares de bases organizados em oito pares de cromossomos. O genoma apresenta de 4 a 8 % de seqüências altamente repetitivas e 32 a 36 % de seqüências medianamente repetitivas. O 60% restante do genoma é composto de seqüências de cópia única (Simpson et al., 1982).

Postula-se que o *Schistosoma mansoni* possua um numero total de genes proximo ao valor de organismos com grau similar de complexidade tais como *D. melanogaster* e *C. elegans* que possuem preditos 13,600 e 19,099 genes, respectivamente (Adams et al.,2000; The *C. elegans* Sequencing Consortium, 1998).

O *Schistosoma mansoni* apresenta certas peculiaridades no processamento de suas moleculas de mRNA tais como a presença de introns muito pequenos (de 31 a 40 pares de bases) em seus genes (Craig et al, 1989) e *trans-splicing* de sequencias de um *spliced leader* em algumas de suas menssagens (Davis et al., 1995).

Existem atualmente depositadas no Genbank cerca de 480 seqüências de mRNA parciais ou totais, 17.000 ESTs e 27.000 seqüências genômicas sendo a maioria pontas de BACs. Esta informação ainda representa uma pequena fração em relação a toda informação existente no genoma do parasita.

1.2- A esquistossomose

A esquistossomose constitui hoje um dos maiores problemas de saúde pública em vários países tropicais. A doença é endêmica em 74 países em desenvolvimento, infectando cerca de 200 milhões de pessoas. Há uma estimativa de que há entre 500 a 600 milhões de pessoas com risco de infecção (WHO, 1996).

A esquistossomose tem ampla distribuição geográfica no Brasil, com maior intensidade de transmissão na região Nordeste do país e norte de Minas Gerais. Entretanto, todos os estados apresentam áreas de transmissão. (FUNASA, 2002).

A tendência histórica para estes indicadores aponta para uma redução na morbidade e mortalidade por formas graves pela doença, para o país como um todo. Entretanto, na áreas

endêmicas da região Nordeste do Brasil, a ocorrência da forma hepato-esplênica, que pode levar ao morte por hemorragia digestiva, ainda representa um importante problema de saúde (FUNASA,2002). No estado de São paulo foram notificados 5.239 casos no ano de 1998 (SUCEN, 2000).

Os sintomas da doença começam logo após a penetração das cercárias, ocorrendo um edema localizado e prurido que podem durar até 3 dias. Após três semanas da infecção podem haver manifestações transientes de alergia e intoxicação. A fase aguda da doença está associada ao início da postura de ovos pelo *Schistosoma*, que normalmente ocorre de 1 a 3 meses após a infecção; os sintomas desta fase incluem febre, urticária, dor abdominal e diarreia. Após este período a doença gradualmente assume seu estado crônico. O crescimento de granulomas no fígado e intestino tem um pico na fase aguda da doença e apresenta um declínio espontâneo no estágio crônico (Boros, 1999). A formação de granulomas podem levar a hepatosplenomegalia após 18 meses em infecções severas ou vários anos em infecções leves (Markell et al., 1992).

O diagnóstico da esquistossomose é feito através da contagem de ovos nas fezes utilizando técnicas como sedimentação, filtração e flotação. É possível estimar o numero de vermes infectantes a partir das contagens de ovos, entretanto infecções mais leves não são detectadas. Em alguns casos é possível realizar uma biopsia retal que apresenta maior sensibilidade, entretanto ela é uma técnica mais invasiva e não indica necessariamente uma infecção ativa, visto ser difícil diferenciar ovos vivos de mortos (Marshall, 1995).

O diagnóstico também pode ser feito com base em antígenos que são liberados na circulação do hospedeiro pelo parasita, originários dos adultos (tegumento e estomago) e dos ovos. A maior parte da pesquisa se concentra em dois antígenos, o antígeno catódico circulante (CCA) e o antígeno anódico circulante (CAA). Os dois são glico-conjugados que são detectados

na circulação e que através de testes sorológicos do tipo ELISA proporcionam contagens mais estáveis e com maior sensibilidade que a contagem de ovos (Van Lieshout et al., 2000).

Atualmente existem medicamentos efetivos contra o parasita, como é o caso do praziquantel, cuja ação está associada à indução do influxo de cálcio através da membrana do tegumento, que causa uma imediata contração muscular e vacuolização do tegumento (Köhler, 2001). O praziquantel é administrado via dose única oral de 40 mg/kg e afeta todas as espécies de *Schistosoma* que infectam humanos (Marshall, 1995). Um dos problemas é que a esquistossomose tem uma progressão lenta e muitas vezes o indivíduo só procura assistência quando a condição de hepatosplenomegalia se torna evidente; neste ponto o uso de medicamentos já não é mais eficaz (Wilson and Coulson, 1999). Além disso, o uso de medicamentos não impede a reinfeção dos pacientes e já foram detectadas cepas resistentes ao praziquantel em regiões onde se realizaram tratamentos em larga escala (Ismail et al., 1996; Liang et al., 2001), embora o mecanismo de seleção de vermes resistentes no campo esteja sob discussão (Doenhoff et al., 2002).

1.3-Aspectos imunológicos da esquistossomose

O parasita possui um sofisticado sistema de escape de seus hospedeiros envolvendo expressão de antígenos de superfície e modificação da resposta imune do parasita (Salzet et al., 2000). Já foi determinado que um dos fatores de modificação desta resposta se dá via um mecanismo ativo de interferência do parasita na síntese de citocinas através da secreção de substâncias que interferem nesta síntese (Angeli et al., 2001).

Foi demonstrado que o parasita secreta em diferentes fases do seu ciclo substâncias que induzem a apoptose de células T do hospedeiro. Aparentemente o mecanismo efetor deste processo seria a ativação de caspases destas células causadas por um aumento na expressão de

Fas e FasL. É postulado que tal mecanismo permitiria um ganho de tempo para estágios iniciais do parasita se adaptar ao novo hospedeiro e induziria a dominância de um tipo de resposta favorável ao parasita (Fallon et al., 1998; Chen et al., 2002).

A formação de granulomas nos tecidos está associada ao recrutamento celular realizado por células linfóides através da secreção de mediadores como o fator inibitório de migração de macrófagos (MIF) e promotor de estimulação de eosinófilos (ESP). IL-2 e IL-4 parecem estar relacionadas também a formação de granulomas, pois quando injetadas em camundongos no estado crônico da doença causaram a expansão destes (Boros, 1999).

Já foram descritos casos em que ocorre o desenvolvimento de resistência a infecção (Butterworth et al., 1985). O tipo de resposta imune associada ao desenvolvimento de resistência foi do tipo Th2 com a presença de IL- 4 e 5, anticorpos do tipo IgE e eosinófilos (Hagan et al, 1998). Este dado é reforçado pelo fato de que ratos tratados com anticorpos anti-IgE demonstraram uma inibição da resistência a infecção (Khalife et al., 2000).

Outros experimentos sugerem que na verdade tanto uma resposta Th1 quanto Th2 desde que polarizada é eficaz para combater uma infecção (Hoffmann et al., 1999). Foi proposta uma teoria chamada de “Happy valley” segundo a qual o *Schistosoma* se adaptaria melhor a uma resposta Th0, uma resposta intermediária que ficaria entre os picos de uma resposta polarizada Th1 e de uma Th2 (Wilson and Coulson, 1999).

Um maior nível de resistência a uma reinfeção foi verificado em indivíduos a partir da puberdade (Fulford et al., 1998), este fato seria indicativo que haveria uma influência do níveis de hormônios na corrente sanguínea no desenvolvimento da infecção. De fato foi demonstrado que altos níveis de testosterona (Nakazawa et al. 1997) e hormônios hipotalâmico-pituitário-adrenal (Morales-Montor, 2001) exercem uma influência negativa no desenvolvimento do *Schistosoma*.

Entretanto, apesar destas evidências não se desenvolveu ainda nenhuma vacina efetiva contra a esquistossomose. Existem atualmente vários candidatos a vacina, como a Gliceraldeído fosfato desidrogenase (GAPDH) (Argiro et al., 2000) , Glutathione-S-transferase (GST) (Capron et al., 2001), a triose fosfato isomerase (TPI) (Reynolds et al., 1994) e a Sm 23 (Da'dara et al., 2001). Apesar destes candidatos terem mostrado resultados positivos em desafios realizados em testes animais nenhum ainda foi demonstrado como eficaz em humanos.

1.4- O tegumento do *Schistosoma mansoni*

O tegumento é formado logo após a penetração através da pele do hospedeiro vertebrado; o esquistossômulo começa a desenvolver a partir de pequenos vacúolos originários das células sub-tegumentais. Ele consiste de um sincício citoplasmático encapsulado por duas membranas tegumentais. Os principais tipos de inclusões membranosas presentes no tegumento são as mitocôndrias, corpos membranosos e corpos discóides. O sincício anuclear está conectado às células subtegumentais que se encontram abaixo do tegumento, entre as camadas musculares. Estas células são responsáveis pela síntese dos corpos de inclusão que tem um papel na formação e substituição da dupla membrana externa (Abath and Werkhauser, 1996).

O tegumento tem um papel importante no escape do parasita das defesas do hospedeiro. Foi demonstrado que parasitas que sob ação de drogas que causam distúrbios no tegumento, como o praziquantel, apresentavam sensibilidade maior a soros imunes (Fallon et al., 1996).

Antígenos do próprio hospedeiro, tais com antígenos A, B e H e MHCs, são incorporados ao tegumento para mascaramento de possíveis antígenos de superfície do parasita. Além disso, o parasita é capaz de sintetizar de forma endógena moléculas semelhantes as presentes no

hospedeiro que apresentam imunoreatividade cruzada com antígenos do hospedeiro (Abath and Werkhauser, 1996).

No tegumento se localizam também receptores do parasita que irão interagir com moléculas do sangue como é o caso do SmRK, um receptor serina-treonina kinase que pertence a família dos receptores de TGF β , um fator de crescimento celular. Postula-se que este possa interagir com o fator do próprio hospedeiro facilitando a adaptação do parasita (Davies et al., 1998).

Foi demonstrado que em esquistossômulos ocorre uma renovação constante da membrana do tegumento e que este mecanismo esta associado ao escape do parasita do sistema imune (Pearce et al., 1996). Nos vermes adultos ocorre também a renovação dos fosfolipídios constituintes da membrana do tegumento, entretanto esta renovação se dá via uma acetilação/desacetilação dos fosfolipídios e não devido à perda da membrana (Brouwers et al., 1999).

O tegumento também tem papel ativo na alimentação do parasita absorvendo glicose do sangue do hospedeiro, o equivalente ao seu peso seco a cada 5 horas. Este processo ocorre através de um transportador (SGTP4) localizado na superfície externa do tegumento e outro (SPTG1) localizado na membrana do tegumento adjacente aos outros tecidos do parasita (Skelly e Shoemaker, 1996). Ocorre também a absorção de aminoácidos via permeases. Foi descrita uma permease (SPRM11c) que funciona associada a outra subunidade, constituindo um transportador heterodimérico que se localiza na membrana externa do tegumento (Skelly et al., 1999).

Devido a esta absorção de nutrientes pelo tegumento, o *Schistosoma* pode sobreviver sem um intestino funcional por um prolongado período de tempo (Popiel and Basch, 1984).

1.5- As ATP-difosfohidrolases

As ATP-difosfohidrolases (E.C. 3.6.1.5.), comumente chamadas de apirases, têm sido descritas em tecidos de plantas e animais. Elas hidrolisam ligações pirofosfato de nucleosídeos di- e trifosfatados, na presença de cátions divalentes, com liberação de ortofosfato, em um mecanismo seqüencial onde NDP aparece como intermediário (Laliberte and Beaudoin 1983). Elas são insensíveis a inibidores específicos de ATPases tipo-P, tipo-F e tipo-V. Esta família de proteínas apresenta 5 pequenas regiões conservadas, as ACRs (apirase conserved regions) (Figura 3), espalhadas ao longo da proteína (Handa and Gudotti, 1996; Vasconcelos et al., 1996). Duas das regiões tem motivos conservados em relação a superfamília da Actina/HSP70/carbohidrato kinase (Smith and Kirley, 1999).

Já foi demonstrado que as apirases possuem variados papéis fisiológicos. Um deles é a hidrólise de ATP até adenosina em conjunto com a 5'nucleotidase nas junções sinápticas. Esta ação tem como objetivo a terminação da ação do ATP que é um co-transmissor na junção sináptica (Sarkis and Salto, 1991).

A apirase tem um papel importante no controle da agregação plaquetária pois hidrolisa o ADP que é um potente agregador plaquetário. O ação do ADP ocorre através da interação com três receptores, um associado à inibição da adenosil ciclase, inibindo a produção de AMP cíclico, e os outros dois à mobilização de cálcio intracelular (Daniel et al., 1998). Foi demonstrado também que ele potencializa a ação de outros agentes de agregação (Gear et al., 2001). A apirase bloqueia esta ação *in vitro* via conversão de ADP a AMP (Gayle et al., 1998). Entretanto a disrupção do gene da apirase *in vivo* leva a uma perda de capacidade de agregação plaquetária (Enjyoji et al., 1999), sugerindo um papel mais complexo da enzima neste sistema.

A apirase também tem um papel na inibição de permeabilização celular. O ATP extracelular permeabiliza a membrana a íons e outros compostos não permeáveis (Saribas et al., 1993). Foi demonstrado que algumas células humanas epidermais eram resistentes a este efeito à presença da apirase em sua própria superfície, ao contrário de outras células que não possuíam a proteína em sua membrana (Girolomoni et al., 1993). Este efeito estaria associado portanto à quebra do ATP extracelular pela apirase que impediria a formação de poros.

Algumas apirases se localizam na membrana do complexo de Golgi onde hidrolisam o NDP do lúmen a NMP, que é um antiporter da translocação de açúcares ligados a nucleotídeos, que são transportados do citoplasma para o lúmen do Golgi. A presença da apirase no Golgi já foi demonstrada como necessária para a N- e O-glicosilação em *Saccharomyces cerevisiae*, cujos mutantes nulos para o gene YND1 apresentavam defeitos na glicosilação de proteínas (Gao et al., 1999).

Já foram realizados vários estudos para tentar avaliar a mudança na atividade da apirase através de modificações da sua sequência, para inferir aminoácidos ou regiões importantes para a sua atividade. Tais ensaios incluem mutações sítio-dirigidas (Smith et al., 1999; Grinthal and Guidotti, 2000; Kirley et al., 2001), construções de quimeras com regiões de diferentes apirases (Heine et al., 2001; Grinthal and Guidotti, 2002) e expressão de fragmentos da proteína (Esch et al., 1999). Estes estudos ajudaram a mapear algumas regiões ou aminoácidos que quando modificados levavam à perda de atividade ou modificação da razão de hidrólise de ATP/ADP característica da enzima.

1.6- A ATP-difosfohidrolase do *Schistosoma mansoni*.

A ATP-difosfohidrolase de *Schistosoma mansoni* é uma proteína de membrana que se localiza na superfície externa do parasita. A atividade desta foi primeiramente caracterizada por

Vasconcelos et al. em 1993 a partir de experimentos com extratos de tegumento que demonstraram haver uma atividade hidrolítica de ATP e ADP.

Esta atividade não era inibida por levamisole e tetramisole, que são inibidores de fosfatase alcalina. Ap_5A , um inibidor da adenilato quinase, também não diminuía a atividade hidrolítica. Mostraram-se inalteradas as medidas de hidrólise de ADP com um sistema com glicose e hexoquinase, que consumiria o ATP mantendo o grupo fosfato ligado à glicose, excluindo a possibilidade de que a atividade ADPásica fosse devido a uma ação combinada de uma adenilato quinase e uma ATPase. Vanadato, um inibidor de ATPases do tipo P, não inibiu a hidrólise em concentrações três ordens de grandeza maiores que as requeridas para inibição de ATPases do tipo P (Vasconcelos et al., 1993).

Experimentos demonstraram que amostras de ADP incubadas durante mais tempo com o extrato de tegumento demonstravam menor capacidade de agregar o plasma sanguíneo quando eram adicionadas a este. Tal fato poderia refletir uma função postulada da enzima que seria de escape do parasita de uma das defesas do hospedeiro via agregação plaquetária. Devido a localização exterior da apirase ela seria capaz de hidrolisar o ADP existente no sangue em sua proximidade, impedindo assim uma estimulação da agregação plaquetária (Vasconcelos et al., 1993).

Posteriormente foi verificada uma imunoreatividade cruzada entre a apirase de *S. mansoni* com anticorpos gerados a partir da apirase de batata. Tal fato permitiu que fosse determinado o tamanho de 63 kDa para proteína e imunolocaliza-la na superfície externa do tegumento do parasita (Vasconcelos et al., 1996).

Além disso, foi sugerido que existiriam duas isoformas da apirase de *Schistosoma mansoni*. Isto foi proposto a partir do experimento em que o tegumento solubilizado em Triton X-100 e Deoxicolato de Sódio era submetido a uma eletroforese não desnaturante. Quando o gel era incubado com ATP ou ADP em meio com presença de cálcio se verificava a formação de um precipitado de fosfato de cálcio em duas regiões diferentes do gel. A partir disso foi determinado que existiam duas proteínas com carga total diferente que apresentavam características cinéticas de quebra de nucleotídeo diferentes mas um peso molecular idêntico (Vasconcelos et al., 1996).

1.7- A estratégia de EST.

Seqüências geradas através da uma leitura de uma parte do cDNA são ditas Expressed Sequence Tags (EST), e constituem uma ferramenta importante no esforço de sequenciamento. Com esta técnica são produzidas bibliotecas de cDNA que representam os genes expressos em um dado estágio do organismo estudado.

Inicialmente as seqüências eram geradas a partir das extremidades poli A das moléculas de mRNA. Deste modo seria possível se utilizar os ESTs como marcadores para o mapeamento físico do genoma, da mesma maneira pela qual os STS (Sequence-tagged-sites) são usados, com a vantagem de que os EST apontam para um gene expresso (Adams et al., 1991). Recentemente com a finalização de genomas completos de vários organismos, as EST vem sendo utilizadas em larga escala para mapear genes e determinação de “splicings” (de Souza et al., 2000; Kan et al., 2001; Xie et al., 2002). Adicionalmente pode-se usar estes trechos como sondas para o isolamento de genes completos, através de uma busca em bibliotecas de cDNA, ou utilizar a técnica de amplificação rápida de pontas de cDNA (RACE) com primers desenhados a partir da seqüência de EST.

Devido à diferente abundância relativa dos mRNAs, o sequenciamento de ESTs apresentava uma alta redundância de certas seqüências e baixa possibilidade de identificação de seqüências raras. Devido a isso foram introduzidas técnicas para a normalização dos níveis das mensagens das bibliotecas, que se baseiam em cinéticas de reassociação de seqüências mais abundantes acopladas a colunas de ligação de DNA (Soares et al., 1994).

Já foram construídas várias bibliotecas de cDNA de *S. mansoni* de diferentes fases do parasita para gerar ESTs, sendo uma delas normalizada (Franco et al., 1995; Franco et al., 1997; Santos et al., 1999).

Posteriormente foram construídas minibibliotecas direcionadas de EST em *S. mansoni* com o uso de RT-PCR com baixa estringência (Dias Neto et al, 1997). A técnica utilizada tem como característica o uso de primers com seqüências arbitrárias e um anelamento inicial dos primers a baixas temperaturas (usualmente a 37°C). Um problema apresentado foi a redundância das seqüências, com apenas cerca de 20 seqüências diferentes entra si por reação, o que é um fator limitante no caso de produção em larga escala de seqüências.

Foi demonstrado que a utilização deste tipo de abordagem permitia a amplificação de genes raros, em um processo de normalização das mensagens, e a amplificação preferencial do centro das mensagens (Dias Neto et al., 2000). Tais características se mostram desejáveis uma vez que além de gerar um maior número de mensagens raras, estas conteriam informação complementar às ESTs obtidas pelo método tradicional a partir das extremidades poliA (que tende a amplificar as pontas dos genes).

2- Objetivos

Os objetivos deste trabalho podem ser divididos em dois subitens:

1- Construção de bibliotecas de EST em *S. mansoni*.

- Caracterização de um método de construção de bibliotecas de EST utilizando a combinação de primers híbridos consenso-degenerado (Rose et al., 1998) e touchdown-PCR.
- Verificação da influência de fatores da reação de PCR na diversidade e normalização da biblioteca construída.
- Análise da influência do primer em fatores na diversidade e direcionamento da biblioteca.
- Análise do papel das diferentes bases componentes do primer (degeneradas e não degeneradas) na seleção dos alvos e estabilização do anelamento do primer.
- Análise da validade dos princípios descritos para o primer CODEHOP (Rose et al., 1998) no contexto da construção de minibiblioteca.
- Verificação da manutenção de parâmetros anteriormente descritos para bibliotecas construídas com PCR de baixa estrigência (Dias Neto et al. 2000) nas bibliotecas construídas com primers CODEHOP.
- Avaliação da contribuição das ESTs de *S. mansoni* geradas em nosso trabalho para a descrição do transcriptoma do parasita.

2-Apirase de *S. mansoni*.

- Clonagem de um gene codificante para uma apirase em *S. mansoni*
- Detecção do gene no genoma do parasita.

- Caracterização de domínios conservados e topologia da proteína codificada pelo gene clonado.
- Expressão e purificação da proteína em sistema heterologo.
- Geração de um soro policlonal para a proteína recombinante produzida.
- Detecção da proteína no tegumento do parasita.

3-Materiais e métodos

3.1- Materiais

Os Kits- ECL, Rediprime II e GFX foram obtidos da Amersham-Pharmacia. Anti-IgG de coelho conjugada à peroxidase e o filme Hyperfilm também foram obtidos da Amersham-Pharmacia. Leupeptina, pepstatina, inibidor de tripsina, fluoreto de fenilmetilsulfonil (PMSF) e dodecil sulfato de sódio (SDS) foram obtidos da Sigma (St. Louis, MO). O kit Oligotex mRNA foi obtido da QIAGEN. O kit μ MACS mRNA Isolation foi obtido de Miltenyi Biotec. Trizol e os kits Superscript II, RACE 3'e RACE 5' foram obtidos de GIBCO BRL. Todos os demais reagentes foram do mais alto grau analítico.

3.2- Obtenção de parasitas

Caramujos da espécie *Biomphalaria glabrata* foram infectados com 15 miracídeos/caramujo através de exposição direta dentro de placas de Petri, por duas horas. Ao final de 6 a 7 semanas, as cercárias liberadas pelos caramujos foram utilizadas para a infecção de camundongos ou de hamsters, inoculando-se ~ 150 cercárias/animal por via subcutânea. O tipo de infecção usada foi a bissexuada, resultando em vermes adultos machos e fêmeas. Os animais foram sacrificados ao final de aproximadamente 45 dias. Os vermes adultos foram obtidos a partir de perfusão do sistema porta hepático dos camundongos, usando-se solução salina fisiológica acrescida de EDTA 0,15 %. Os parasitas foram lavados com salina para descartar possíveis contaminantes do hospedeiro e mantidos em solução salina a 4 °C até o processamento de extração do tegumento.

O ciclo biológico do *S. mansoni* é mantido no Instituto Adolfo Lutz, na Seção de Enteroparasitoses, e todos os detalhes pertinentes ao desenvolvimento, manutenção, infecção e coleta dos materiais foram realizados sob a supervisão da Dra. Cybele Gargioni.

3.3-Extração de RNA mensageiro de *Schistosoma mansoni*.

Os vermes adultos recém perfundidos eram adicionados a uma solução de RNAlater (Ambion) em uma proporção de cerca de 1.000 vermes para 15mL de solução. Esta solução então era acondicionada a 4°C durante 24 horas, de acordo com a recomendação do fabricante. Após este período a amostra era conservada a -20°C.

No momento da extração de RNA os vermes eram retirados da solução com o auxílio de uma espátula e colocados em um tubo vazio previamente tarado, sendo feita então a pesagem do tubo com vermes para determinação da massa total utilizada.

Dois protocolos foram utilizados para a obtenção do RNA mensageiro. No primeiro os vermes eram colocados em Trizol (Gibco), na proporção recomendada pelo fabricante, e a lise dos tecidos era feito com o auxílio de um disruptor celular DIAx900 (Heidolph), que processava a amostra a velocidade mínima durante 30 segundos. Após isso era seguido processo de extração era seguido de acordo com as recomendações do fabricante para a obtenção do RNA total. O RNA mensageiro era então extraído do RNA total utilizando o kit Oligotex mRNA (Qiagen) que utiliza oligonucleotídeos de timina acoplados a resina de poliestireno para separar as moléculas poliadeniladas de RNA mensageiro dos outros RNAs.

O segundo protocolo utilizava o kit “ μ MACS mRNA Isolation” (Miltenyi Biotec) que utiliza oligonucleotídeos de timina acoplados a uma resina que esta fixada a grânulos metálicos. A resina é misturada com o lisado celular, que em seguida é aplicado em uma coluna plástica

submetida a um campo magnético. A resina é retida pelo campo magnético, já estando ligado ao RNA magnético pelas seqüências poliT. A extração do RNA mensageiro era feita diretamente dos vermes que eram homogeneizados na solução de lise do kit utilizando o disruptor celular DIAX900 (Heidolph), a velocidade mínima durante 30 segundos. Após isso o RNA mensageiro era extraído seguindo as instruções do fabricante com a exceção do tampão de eluição que era substituído por água.

3.4- Transcrição reversa

O RNA mensageiro foi submetido à tratamento com DNase livre de RNase (Promega) na concentração de 1U/10 µl por 15 minutos à temperatura ambiente. A enzima foi então inativada por aquecimento a 65°C por 10 minutos. Cerca de 250 ng do RNA eram submetidos à transcrição reversa a 42°C, por 60 minutos com 200 U de Superscript II (Gibco) e 20 pmol do primer desejado em um volume de 20 µl como recomendado pelo fabricante do kit. As amostras foram então incubadas com RNase H por 20 min à 37°C e o cDNA era estocado a -70 °C.

3.5- Reações de PCR para construção das bibliotecas.

Aliquotas de 0,5 µl do cDNA obtido foram submetidas a reações de PCR utilizando-se a enzima AmpliTaq-Gold (Perkin Elmer), conforme instrução do fabricante, com 1 µM do primer escolhido, 200 nM de dNTP e concentrações variáveis de Magnésio e tampão de reação. Todas as reações foram feitas no termociclador 9700 da Perkin Elmer.

3.6 – Análise das reações de RT-PCR utilizando gel de poliacrilamida corado por prata.

Aliquotas de 5µl da reação de RT-PCR era adicionadas a 1µ do tampão de corrida de DNA 6x concentrado contendo bromofenolblue 0,25%; xilenocianol FF 0,25% e glicerol 30%. As amostras foram submetidas à eletroforese em gel de acrilamida 6% em tampão TAE. Após a eletroforese o gel foi submerso em solução de fixação (etanol 10% e ácido acético glacial 0,5%) por 3 minutos, sendo então lavado rapidamente com água deionizada e incubado em solução de coloração (etanol 10%, ácido acético glacial 0,5% e nitrato de prata 12mM) por 5 minutos. Em seguida foi feita uma lavagem de 2 minutos com água deionizada e o gel foi submerso em solução de revelação (NaOH 0,75M e formaldeído 38% a 2,6 µl/ml) até que a impregnação desejada fosse alcançada. A coloração foi interrompida pela adição da solução de fixação por 5 minutos e posterior lavagem com água deionizada.

3.7- Seleção e Ligação de Fragmentos de DNA.

As amostras de RT-PCR que apresentavam um perfil desejável, após inspeção do gel corado por prata, eram precipitadas com 2,5 volumes de etanol e centrifugação a 12.000g por 20 minutos. A amostra foi ressuspensa em 50 µl de tampão 1X para EcoRI. Foram adicionados 20 unidades de EcoRI (New England Biolabs) e a amostra era incubada a 37°C por 1 hora.

O produto da digestão foi submetido a uma eletroforese em gel de agarose 1,2 % e foi feita uma seleção de tamanho cortando o gel em trechos contendo produtos do peso molecular desejado. O DNA foi então extraído do gel com o kit GFX (Amersham-Pharmacia) sendo finalmente eluido em um volume de 25µl.

As ligações foram feitas utilizando-se 10µl do DNA obtido e 25 ng do plasmídeo Bluescript II Sk (Stratagene) previamente digerido com a enzima EcoRI. As reações foram feitas sob o tampão da ligase 1X (Tris.Cl 200mM pH 7,5; MgCl₂ 50mM; DTT 50mM e BSA 500:g/ml) e 5U de T4 DNA ligase eram adicionadas. As amostras eram incubadas a 16 °C por 16 horas.

Após este período as amostras foram dialisadas sendo colocadas no topo de um filtro de nitrocelulose de 0,025µm de porosidade (Millipore) que se encontrava flutuando sobre água deionizada. A amostra era dialisada por um período de duas horas sendo estocadas a -20°C para transformação após este período.

3.8- Preparação de células eletrocompetentes.

Um pré-inóculo era feito a partir de uma colônia de *E. coli* XL1-Blue MRF' em 2 ml de meio LB (bacto-triptona 1%; extrato de levedura 0,5%) com tetraciclina (20µg/ml) e incubada a 37°C por 16 horas sob agitação. Este pré-inóculo foi adicionado a 500 ml de meio LB e incubado a 37°C com 200 rpm de agitação. até atingir uma OD_{600nm} de 0,6-0,8 . As células foram coletadas por centrifugação a 5.800 g por 10 minutos a 4°C, e o meio de cultura descartado. As células foram lavadas com 500 ml de água deionizada estéril gelada. Em seguida as células foram recolhidas por nova centrifugação sob as mesmas condições e ressuspensas em 100 ml de glicerol 10% estéril gelado. Esta suspensão foi novamente centrifugada por 10 minutos, e o precipitado foi ressuspensão em 1 ml de glicerol 10% gelado . Aliquotas de 40:1 foram congeladas à -80°C até utilização em transformações.

3.9 Transformação de bactérias eletrocompetentes.

As células eletrocompetentes descongeladas em gelo eram misturadas com o volume desejado das ligações e em seguida submetidas a eletroporação de 2,5 volts, 25 μ F e 200ohms em um aparelho Gene Pulser (Bio Rad). As células foram recuperadas por 40 minutos a 37°C em meio SOC (bacto-triptona 2%; extrato de levedura 0,5%; NaCl 0,01M; glicose 20 mM) . Em seguida eram plaqueadas em meio sólido LB (bacto-triptona 1%; extrato de levedura 0,5%) suplementado com ampicilina 100: g/ml, sendo adicionados 30:1 de IPTG 100 mM (isopropil-tiogalactopianosídeo) e 120 μ l de X-gal 2% (5-bromo-4cloro-3indolil-2-D-galactosídeo) por placa. As placas foram incubadas por 12-16 horas a 37°C em estufa. As colônias brancas eram então usadas como inóculo para placas de 96 poços contendo 100 μ l de TB/ampicilina (bacto-triptona 1,2%; extrato de levedura 2,4%; glicerol 0,4%; KH₂PO₄ 0,017M e K₂HPO₄ 0,072M) sendo incubados a 37°C por 16 horas. Em seguida eram adicionados 100 μ l de glicerol 30% e as amostras eram estocadas a -80°C.

3.10- Extração de plasmídeos pelo método de lise por fervura.

Os clones bacterianos transformados estocados em placas de fundo redondo de 96 poços. foram replicados em 1 ml de TB suplementado com ampicilina 100 μ g/ml com auxílio de replicador tipo ouriço. O inóculo foi crescido por 16 horas e as células recolhidas por centrifugação a 3000 rpm por 8 minutos sendo o meio de cultura descartado. Ressuspendeu-se as células em 25 μ l de água deionizada estéril por poço com a ajuda de um aparelho de vórtex. Em seguida foi adicionado a cada poço 70 μ l de MW/STET/TWEEN (Tween-20 5%; NaCl 100mM; Tris.Cl 16mM pH8 e EDTA 1mM pH 8) seguido de homogeneização e incubação à temperatura

ambiente por 5 minutos. As células foram lisadas por fervura em microondas e eram adicionados 300 µl de água deionizada estéril seguida por uma incubação por 10 minutos no gelo. A amostra foi submetida a centrifugação por 20 minutos à 6000 rpm e foi retirada uma alíquota de 100µl do sobrenadante que foi estocada a -20°C.

3.11- Sequenciamento dos clones .

As reações de sequenciamento foram feitas de acordo com o método de dideoxinucleotídeos (Sanger, 1977). Foram utilizados 2µl do mix de sequenciamento Big Dye (Perkin Elmer), 8 pmol do primer desejado e 4µl de uma solução 5mM de MgCl e 200mM de Tris pH 9 em cada reação de 10 µl. O programa utilizado nas máquinas 9700 da Perkin Elmer foi: 94°C por 30 seg, 56°C por 30 seg e 60°C por 3 minutos, repetindo-se o ciclo 35 vezes.

O DNA foi precipitado com a adição 25µl de etanol, 1 µl de acetato de sódio 3M pH 5,2; e 1µl de solução de glicogênio (1g/l). A amostra era centrifugada a 12.000g por 20 minutos. O sobrenadante era descartado. Era feita uma lavagem com a adição de 50µl etanol 70% seguida por uma centrifugação a 12.000 g por 8 minutos. Após precipitação o DNA foi ressuspenso em tampão de formamida e desnaturado por 5 minutos à 95°C. Imediatamente após a desnaturação as amostras foram acondicionadas no gelo.

As amostras eram então submetidas a eletroforese em gel de acrilamida em meio TBE durante 7 horas como descrito no manual do ABI377 (Perkin Elmer).

3.12- Análise do sequenciamento

O imagem do gel de sequenciamento era manualmente inspecionada para verificação de problemas no programa de detecção das amostras. Caso fosse detectado algum problema a

determinação de cada amostra no gel era feita manualmente. Após a extração do cromatograma a partir do dado do gel bruto, cada um deles era inspecionado manualmente e trechos de baixa qualidade eram retirados do arquivo. Os arquivos eram então enviados para outro computador onde as seqüências eram analisadas com o programa PhredPhrap (Ewing et al., 1998). Seqüências com “Phred score” acima de 20 foram utilizadas.

3.13- Amplificação rápida de pontas de cDNA (RACE)

Aproximadamente 200ng de RNA mensageiro foram utilizadas para a transcrição reversa seguindo o protocolo dos kits 3'RACE system for rapid amplification of cDNA ends e 5'RACE system for rapid amplification of cDNA ends (Gibco). No caso do 5' ainda foram feitas purificações e tratamentos adicionais para adição do sítio de anelamento do primer na ponta 5'seguindo as orientações do kit. O PCR para amplificação destes produtos foi feito com a enzima Advantage 2 (Clontech) com o tampão fornecido pelo fabricante e dNTPs em uma concentração de 200 nM. O programa de PCR utilizado foi: 95°C-1min + 35 ciclos de 95°C-30s+55°C-30s+ 68°C-3min + extensão final de 68°C-3min. Os produtos da reação foram analisados por eletroforese em gel de agarose 1,2%.

3.14- Expressão da proteína recombinante em *E. coli*.

Uma colônia da bactéria BL21(DE3) transformada com o plasmídeo de expressão pET-21b contendo como inserto a seqüência da proteína de interesse era inoculada em um volume de 3 ml de LB contendo 50µg/ml de ampicilina e era incubada a 37°C sob agitação até alcançar uma OD₆₀₀de aproximadamente 0,5. Esta suspensão era usada como inóculo para 100ml de

LB/ampicilina, que era incubado a 37°C sob agitação até uma OD₆₀₀ de 0,5. Era adicionado IPTG para uma concentração final de 1mM e a amostra era incubada por mais 3 horas.

Após este período as células eram recolhidas através de centrifugação a 10.000g por 10 minutos a 4°C e o meio de cultura era descartado. As células eram então ressuspensas em 10ml de Tris 20 mM pH 7,5 e eram realizadas duas passagens a 20.000 psi na “French press” para lise das células. As amostras eram centrifugadas a 14.000g para a separação das frações solúvel e insolúvel.

3.15- Solubilização dos corpos de inclusão.

Os precipitados contendo os corpos de inclusão da bactéria foram lavados duas vezes com tampão RIPA (SDS 0,1%, Triton X-100 1% e deoxicolato de sódio 1% em tampão fosfato) para retirada de contaminantes do sobrenadante. Foi adicionado 2 volumes em relação ao volume do precipitados de SDS 10% e ferveu-se a amostra durante 1 hora. A amostra foi dialisada durante 16 horas contra 100 volumes de SDS 0,05%. Dialisou-se a amostra 8 horas adicionais contra 100 volumes de SDS 0,01%.

3.16- Coluna de níquel para purificação de proteína com hexahistidina.

Foi adicionado às amostras um tampão de ligação da coluna 5X concentrado (NaCl 1.5 M, NaH₂PO₄ 250 mM pH 8 e imidazol 50mM) para concentração final 1X. Para cada 4 ml de amostra foram adicionados 1 ml de Ni-NTA agarose (QIAGEN) e realizada uma incubação por 1 hora a 4°C. A suspensão foi transferida para uma coluna onde o polímero de agarose ficou retido e o sobrenadante escoou. A resina foi lavada duas vezes com 4 volumes de solução de lavagem

(NaCl 300 mM, NaH₂PO₄ 50 mM pH 8 e imidazol 20mM). A eluição da proteína foi feita com 4 volumes de solução de eluição (NaCl 300 mM, NaH₂PO₄ 50 mM pH 8 e imidazol 250mM).

3.17- Produção de anticorpo anti-apirase de *S. mansoni*.

O fragmento da apirase que foi expresso em *E. coli* foi usado para imunização de um coelho fêmea de aproximadamente 3 meses de idade. Dez dias antes da imunização, cerca de 5 ml de sangue foram retirados pela veia marginal da orelha, incubado a 37°C por 30 minutos e posteriormente incubado por 16 horas a 4°C para o encolhimento do coágulo. A amostra foi centrifugada 10 minutos a 2500rpm para deposição do coágulo sendo o soro pré-imune sobrenadante recolhido e congelado em alíquotas a -20°C.

A primeira injeção foi feita com cerca de 50 µg de proteína em 2 ml de PBS adicionado de 2 ml de adjuvante de Freund completo (Difco), subcutaneamente no dorso do animal em quatro diferentes pontos. Após 15 dias a imunização foi repetida nas mesmas condições, exceto pelo uso de adjuvante de Freund incompleto. Um outro reforço foi realizado após mais 15 dias. Cerca de 5 ml de sangue foram retirados após dez dias do último reforço e tratados como descrito para o soro pré-imune.

3.18- Extração de tegumento de verme adulto

Com o objetivo de induzir a liberação do tegumento, cerca de 2.000 parasitas vivos machos e fêmeas, foram incubados por 10 min a 37 °C em aproximadamente 10 ml de tampão salina-fosfato em um meio contendo KCl 2,7 mM, CaCl₂ 0,9 mM, MgCl₂ 0,5 mM e inibidores de proteases: leupeptina (0,5 µg/ml), pepstatina (0,07 µg/ml), inibidor de tripsina (50 µg/l) e fluoreto de fenilmetilsulfonila (2 µg/ml). O material foi submetido a agitação em vórtex por 10 segundos

(Payares et al., 1985). Os vermes inteiros, cujo tegumento havia sido liberado, foram removidos por decantação. Qualquer partícula macroscópica em suspensão foi retirada do sobrenadante com auxílio de pipeta Pasteur. O sobrenadante da decantação foi submetido à ultra-centrifugação a 100.000 x g. O precipitado foi homogeneizado em 300 a 400µl de Tris-HCl 5 mM, pH 7,4, contendo sacarose 8 % e os já indicados inibidores de proteases, com o auxílio de um Potter-Elvehjem. As amostras foram estocadas em nitrogênio líquido.

3.19- Solubilização do tegumento para ensaio de Western-blot

Cerca de 600µl de tegumentos estocados (correspondentes a cerca de 3000 vermes) eram descongelados e a eles foi adicionado SDS para uma concentração final de 0,1%. A amostra foi homogeneizada com auxílio de um Potter-Elvehjem e a amostra foi primeiramente centrifugada a 5.200g por 2 minutos, sendo o sobrenadante transferido para um tubo novo e sofrendo uma nova centrifugação a 14.000g por 2 minutos. O sobrenadante foi aplicado em uma coluna Ultrafree com membrana biomax de 30 K (Millipore) e centrifugado durante 15 minutos a 7.500g. Após a centrifugação foram adicionados 500µl de SDS 0,1% no topo da coluna e foi feita outra centrifugação a 7.500g. Este procedimento de lavagem foi repetido mais uma vez e no final 50µl que se concentraram no topo da coluna foram recolhidos e centrifugados a 14.000g onde ocorreu a formação de um precipitado. O sobrenadante foi utilizado para ensaio de *Western-blot*.

3.20- Western-blots

Para realização de *Western blots*, as amostras foram solubilizadas em tampão de amostra (Tris 62,5mM pH 6,8; glicerol 20%; SDS 2% (p/v); ditiotretitol 42mM; corante azul de

bromofenol), fervidas durante cinco minutos e submetidas a uma eletroforese em gel de poliacrilamida 11% com dodecil sulfato de sódio 0,1% (SDS-PAGE).

As proteínas do gel foram transferidas para membranas de nitrocelulose Hybond C+(Amersham-Pharmacia) com a utilização do aparelho Multiphor II (Pharmacia). O sistema foi montado fazendo-se um sanduíche com 6 folhas de papel eletrodo novablot (Amersham-Pharmacia), 1 folha de membrana de nitrocelulose, o gel contendo amostras a serem transferidas e mais 6 folhas de papel eletrodo, todos foram embebidos em tampão de transferência contínua. A transferência era feita sob uma corrente de 0,8mA por cm² de membrana durante 1 hora e trinta minutos.

As membranas de nitrocelulose foram bloqueadas em solução tampão PBS (tampão fosfato salino) acrescida de leite desnatado 5% e azida sódica 0,05% durante 16 horas a 4°C. Após o bloqueio foram incubadas com soro policlonal anti-apirase de batata nativa de *S. tuberosum* var. Desirée em PBS (diluição 1:1.000), por 15 horas a temperatura ambiente ou 4 horas com o soro policlonal anti-apirase recombinante de *S. mansoni* (diluição 1:1000). Em seguida foram lavadas quatro vezes em solução tampão PBS com 0,1 % de tween-20 por 5 minutos. As membranas foram incubadas durante 1 hora com anticorpo secundário anti-IgG de coelho acoplado à peroxidase (Amersham-Pharmacia) na diluição de 1:3000 em PBS. Foi feita então novamente uma lavagem com PBS-tween 20 sob as mesmas condições.

As membranas foram colocadas sobre película de PVC e incubadas durante 1 minuto com o substrato do Kit ECL (Amersham-Pharmacia). O excesso de solução foi retirado e as membranas foram envolvidas outra camada de filme de PVC. Filmes de autoradiografia ECL (Amersham-Pharmacia) foram expostos por diferentes tempos a membrana e revelados.

3.21- Extração de DNA genômico de *S. mansoni*

A extração do DNA genômico foi feita a partir de aproximadamente 0,4 g de vermes adultos. Os vermes foram macerados com nitrogênio líquido em cadinho de porcelana e ressuspensos em tampão de digestão (Tris-Cl 10mM pH 8,0; NaCl 100mM; SDS 0,5% e proteinase K 0,1mg/ml). A digestão foi feita por 18 horas à 50°C, sob agitação. A solução foi submetida à extração de proteínas pela adição de 1 volume de fenol:clorofórmio e centrifugação de 10 minutos à 1700 x g. A fase aquosa superior foi recuperada e o DNA precipitado com adição de 0,5 volume de acetato de amônio 7,5M e 2 volumes de etanol 100%. A recuperação do DNA genômico foi feita por centrifugação de 10 minutos a 1700 x g, este foi lavado com etanol 70% e recolhido por centrifugação por 10 minutos a 5000g. O precipitado foi ressuspensão através de incubação em 0,5 ml de TE a 55°C por 2 horas.

3.22- Southern-blots

Para a realização dos Southern-blots cerca de 10µg de DNA genômico foram submetidos à digestão por diferentes enzimas de restrição em 50µl de meio de reação apropriado durante 16 horas a 37°C. Em seguida as amostras eram precipitadas com a adição de 2,5 volumes de etanol e 0,1 volumes de acetato de sódio 3M pH 5,2 e uma centrifugação a 12.000g por 20 minutos. A amostra foi ressuspensa em 10µl de tampão de corrida de DNA e submetida a uma eletroforese de 10V em gel de agarose 0,8%. Após isso o DNA era transferido para uma membrana de nylon Hybond N+ (Amersham-Pharmacia) utilizando o aparato de transferência Posiblot (Stratagene). O DNA era fixado na membrana através de *crosslink* com UV utilizando o aparelho GS Gene Linker (Bio-rad).

Esta membrana foi pré-hibridizada por 1 hora em 40 ml de solução 6X SSC (cloreto de sódio 0,9M, citrato de sódio 0,09M), 5X Solução de Denhardtts (Ficoll tipo 400 1mg/ml, polivinilpirrolidona 1mg/ml e soro albumina bovina 1mg/ml), 0,5 % de SDS e 2 µg de DNA de esperma de salmão desnaturado a 65°C.

A este foi adicionada sonda com fósforo radioativo preparada a partir de 50ng de DNA. A marcação desta DNA foi feita com o “kit” Rediprime II (Amersham-Pharmacia) segundo as instruções do fabricante. A membrana era incubada por 16 horas a 65°C.

Foi feita uma lavagem com 1X SSC e 0,2% SDS por 30 minutos a 65°C seguida por duas lavagens com 0,5 X SSC e 0,2% SDS por 30 minutos a 65°C. A membrana foi então utilizada para a sensibilização de um filme Hyperfilm MP (Amersham-Pharmacia) a -70°C por um período de até uma semana.

4- Resultados e discussão

4.1-Produção de minibibliotecas de EST de *S. mansoni*

4.1.1- Verificação da integridade do RNA mensageiro.

Para a verificação da integridade do RNA mensageiro utilizado decidiu-se realizar um controle com a amplificação da região 5' de genes conhecidos de *S. mansoni*. Foram escolhidos os genes da Actina (número de acesso do GenBank U19945), Catepsina C (número de acesso do GenBank Z32531) e do Receptor-x-retinóico (número de acesso do GenBank AF158102) de *Schistosoma mansoni*. Estes genes foram escolhidos por representarem genes com diferentes níveis de expressão e tamanho da mensagem, sendo assim possível verificar o grau de degradação do mRNA. Os *primers* foram desenhados com o auxílio do programa *Primer 3* disponível no endereço: <http://www-genome.wi.mit.edu>.

Foram realizadas transcrições reversas com cada um dos *primers* reversos para estes genes. Em seguida o cDNA gerado foi amplificado por uma reação de PCR com os *primers* para cada gene. O resultado foi a amplificação de um produto com 500 pb (Figura 4) para cada um dos genes tomados como indicadores, conforme era esperado. Este resultado sugeriu que o mRNA estava livre de degradações.

4.1.2- Determinação dos parâmetros da reação de PCR.

O primeiro *primer* utilizado em nossos experimentos foi o APYBAF1, que originalmente foi desenhado como parte da estratégia para clonagem do gene da apirase utilizando a metodologia CODEHOP (Rose et al., 1998), para clonagem de genes a partir do alinhamento de proteínas distantemente relacionadas (Figura 5). Decidimos utilizá-lo para a transcrição reversa, e como único *primer* no PCR, para tentarmos obter uma boa normalização. Havia o risco de que, com a

utilização de dois *primers* degenerados em um PCR com baixa estringência, ocorresse um predomínio da amplificação de seqüências abundantes.

Figura 4- Caracterização da integridade do RNA usado nos experimentos- Produtos da reação de RT-PCR dos Controles para os genes da Actina, Catepsina C e Receptor-x-retinóico com o uso dos *primers* desenhados para estes genes, tomados como indicadores da integridade das mensagens. As reações foram feitas tanto com o RNA mensageiro quanto com o RNA total. O tamanho dos produtos é o esperado para a amplificação de cada um dos três produtos. **PM-** 100 bp ladder. **Controle-** Produto da reação de PCR usando-se mRNA incubado na ausência de transcriptase reversa.

Realizando uma primeira reação com o uso deste *primer* e um protocolo de touchdown PCR (94 °C- 8min, 30 ciclos de 94°C- 1min; 60 a 45°C (-0,5°C por ciclo)- 1min; 72°C- 1 min + 10 ciclos de 94°C- 1min; 45°C- 1min; 72°C- 1min + 7min a 72°C de extensão final) obtivemos um perfil que apresentava alguma variabilidade, mas possuía um nítido padrão de bandejamento. Entretanto, o grau de amplificação já era mais intenso que utilizando o protocolo descrito em Dias Neto et al., 1997 para este mesmo *primer* (Figura 6 painel A).

Devido ao forte padrão de bandejamento de baixo peso molecular verificado, decidiu-se experimentar uma variação na concentração do tampão de PCR. Como descrito por Henerariu et

al. em 1997, uma diminuição na concentração de tampão sem variação da concentração de magnésio favorece a amplificação de fragmentos maiores em PCRs, com a amplificação de múltiplas bandas.

Como era do nosso interesse obter fragmentos grandes, foi testada uma variação com um alcance maior do que o descrito no artigo. Observou-se uma amplificação mais intensa com uma diminuição da concentração de tampão para 0,7X da concentração original (figura 6 painel B). Entretanto o sequenciamento dos clones obtidos a partir deste perfil demonstrou que ela ainda apresentava um alto índice de redundância (biblioteca fl9- tabela I), provavelmente devido a amplificação específica de algumas mensagens, que se reflete na presença de bandas nitidamente visíveis (Figura 6 painel B)

Figura 5- Esquematização da estratégia de desenho de *primers* utilizada para a clonagem da apirase de *S. mansoni*. As caixas escuras representam as regiões contendo motivos conservados que foram utilizados para a construção dos *primers*. Os números abaixo representam a distância dos blocos em relação ao primeiro bloco.

Figura 6- Otimização dos parâmetros para a geração de perfis para a construção de minibibliotecas. Eletroforese em gel de acrilamida 6% corado com prata. **A-** Comparação do perfil seguindo o protocolo descrito em Dias Neto et al., 1997 (LS) com o protocolo de touchdown PCR (LS+TD). **B-** Efeito de diluição de sal combinado com touchdown PCR. **C-** Efeito de diluição de sal combinado com ciclo com rampa lenta e touchdown PCR. **D-** Comparação do perfil produzido com o protocolo de touchdown PCR com 0,4X de sal e rampa lenta (LS+MTD) e com o protocolo descrito em Dias Neto et al., 1997. **0.4X, 0.7X e 1X-** concentrações de tampão utilizadas. **Controle-** Reação controle sem a adição de transcriptase reversa.

Decidiu-se modificar ainda mais um parâmetro com a finalidade de aumentar a variabilidade das seqüências. Introduziu-se a utilização da rampa lenta (15 % da velocidade normal de transição) entre as etapas de anelamento e extensão dos ciclos de PCR. Este procedimento já havia sido descrito por Ellinghaus et al.(1999) para um aumento no rendimento das reações de PCR. Apesar disso, não existia nenhuma descrição do efeito da combinação da variação de tampão combinada com a rampa lenta.

Conseguiu-se uma melhoria na diversidade das mensagens amplificadas, em especial em baixas concentrações de tampão (0,4 X) (Figura 6 painel C), que apresenta um padrão de amplificação arrastado sem nenhuma banda proeminente. Devido a esta característica gerou-se uma biblioteca com o perfil produzido com baixo tampão. A figura 6 painel D mostra a comparação direta entre a amplificação obtida usando-se uma combinação dos 3 parâmetros otimizados (touchdown, diluição do tampão para 0,4X e rampa lenta) e a amplificação utilizando o protocolo original. Observa-se uma grande diferença em relação ao perfil produzido com o protocolo original da técnica.

Esta biblioteca foi construída utilizando o procedimento padrão na construção de bibliotecas, que envolvia primeiro a seleção de tamanho através da separação em um gel de agarose, e posterior tratamento com a enzima EcoRI (biblioteca f23 - tabela II). Percebeu-se que o tratamento nesta ordem não era conveniente, pois quando trata-se com a endonuclease algumas seqüências geram fragmentos muito pequenos devido a um sítio de clivagem interno, principalmente na biblioteca construída a partir da seleção de tamanho de peso molecular mais alto. Devido a seu baixo peso molecular estes fragmentos eram preferencialmente clonados, gerando uma grande número de mensagens com baixo nível de informação.

Tabela I- Tabelamento dos resultados do sequenciamento da biblioteca f19.

Biblioteca f19 - seleção de tamanho de 400 a 700 pb (24 seqüências)

Classificação das seqüências ^a	Numero de seqüências
Seqüências conhecidas (79 %)	
EST <i>S. mansoni</i>	16 (67%)
rRNA	3 (12%)
Seqüências novas (21%)	
Sem similaridade	5 (21%)
Seqüências não redundantes	15 (62%)
Seqüências redundantes	9 (38%)

^a A classificação das seqüências foi feita de acordo com a seqüência de maior similaridade indicada pelo programa BLASTn e BLASTx com uma exigência de e^{-10} . A classificação das seqüências foi feita nas seguintes categorias: **EST *S. mansoni***- possuem similaridade com seqüências de nucleotídeos de *S. mansoni* do banco de dados dbEST do NCBI. **Sem similaridade**- não possuem nenhuma similaridade com nenhum dos bancos de dados pesquisados. **rRNA**- - possuem similaridade com seqüências de RNA ribossomal.

Para excluir fragmentos pequenos que são gerados a partir da clivagem em sítios internos muito próximos, decidiu-se primeiro digerir os produtos de RT-PCR com a enzima EcoRI e em seguida fazer a eletroforese e seleção de tamanhos. A partir deste procedimento gerou-se a biblioteca f24 (tabela III), que passou a apresentar seqüências com maior tamanho.

Apesar disso, pode-se notar uma grande quantidade de RNA ribossomal contaminante (Tabela II e III) que diminuía sensivelmente o rendimento da técnica. Aparentemente o protocolo de extração com trizol (Gibco) seguido do “kit” oligotex (Qiagen) gerava uma alta contaminação, provavelmente causada pela adsorção de RNA ribossomal na resina de poliestireno do oligotex.

Optou-se então por modificar o procedimento de extração de RNA utilizando o “kit” μ MACs (Miltenyi) para uma nova extração. Ele tem por princípio a purificação com uma resina

Tabela II- tabelamento dos Resultados do sequenciamento da biblioteca f23.

Biblioteca f23-I - seleção de tamanho de 400 a 700 pb (119 seqüências)

Classificação das seqüências ^a	Numero de seqüências
Seqüências conhecidas (63,9 %)	
EST <i>S. mansoni</i>	40 (33,6%)
Gene <i>S. mansoni</i>	18 (15,1%)
rRNA	17 (14,2%)
DNA	1 (1%)
Seqüências novas(36 %)	
Proteína $\neq$ <i>S. mansoni</i>	4 (3,3%)
Sem similaridade	39 (32,7%)
Seqüências não redundantes	83 (69,7%)
Seqüências redundantes	36 (30,3%)

Biblioteca f23-S - seleção de tamanho de 700 a 1200 pb (36 seqüências)

Classificação das seqüências ^a	Numero de seqüências
Seqüências conhecidas (75 %)	
EST <i>S. mansoni</i>	20 (55,6 %)
Gene <i>S. mansoni</i>	1 (2,7 %)
rRNA	6 (16,7 %)
Seqüências novas (25 %)	
Sem similaridade	5 (13,9 %)
Proteína $\neq$ <i>S. mansoni</i>	4 (11,1 %)
Seqüências não redundantes	19 (53%)
Seqüências redundantes	17 (47%)

^a A classificação das seqüências foi feita de acordo com a seqüência de maior similaridade indicada pelo programa BLASTn e BLASTx com uma exigência de e^{-10} . A classificação das seqüências foi feita nas seguintes categorias: **EST *S. mansoni***- possuem similaridade com seqüências de nucleotídeos de *S. mansoni* do banco de dados dbEST do NCBI. **Gene *S. mansoni***- possuem similaridade com seqüências de nucleotídeos de *S. mansoni* do banco de dados nr do NCBI. **Proteína $\neq$ *S. mansoni***- após tradução em diferentes fases de leitura possuem similaridade com seqüências de proteínas de organismos diferentes do *S. mansoni* do banco de dados nr do NCBI. **Sem similaridade**- não possuem nenhuma similaridade com nenhum dos bancos de dados pesquisados. **rRNA**- possuem similaridade com seqüências de RNA ribossomal. **DNA**- possuem similaridade com seqüências de DNA não expressas.

Tabela III- tabelamento dos resultados do sequenciamento da biblioteca f24.

Biblioteca f24-I - seleção de tamanho de 400 a 700 pb (401 seqüências)

Classificação das seqüências ^a	Numero de seqüências
Seqüências conhecidas (71,6 %)	
EST <i>S. mansoni</i>	132 (34,6%)
Gene <i>S. mansoni</i>	44 (11%)
rRNA	100 (19,1%)
DNA	11 (3,4%)
Seqüências novas (28,4%)	
Sem similaridade	63 (15,7 %)
Proteína $\neq$ <i>S. mansoni</i>	37 (9,2%)
EST <i>S. japonicum</i>	7 (1,7%)
Gene $\neq$ <i>S. mansoni</i>	7 (1,7%)
Seqüências não redundantes	219 (53,8%)
Seqüências redundantes	192 (46,2%)

Biblioteca f24-S - seleção de tamanho de 700 a 1200 pb (34 seqüências)

Classificação das seqüências ^a	Numero de seqüências
Seqüências conhecidas (76,5 %)	
EST <i>S. mansoni</i>	10 (29,4 %)
Gene <i>S. mansoni</i>	3 (8,8 %)
rRNA	13 (38,2 %)
Seqüências novas (23,5 %)	
Sem similaridade	8 (23,5 %)
Seqüências não redundantes	23 (67,6%)
Seqüências redundantes	11 (32,4%)

^a A classificação das seqüências foi feita de acordo com a seqüência de maior similaridade indicada pelo programa BLASTn e BLASTx com uma exigência de e^{-10} . A classificação das seqüências foi feita nas seguintes categorias: **EST *S. mansoni***- possuem similaridade com seqüências de nucleotídeos de *S. mansoni* do banco de dados dbEST do NCBI. **Gene *S. mansoni***- possuem similaridade com seqüências de nucleotídeos de *S. mansoni* do banco de dados nr do NCBI. **Proteína $\neq$ *S. mansoni***- após tradução em diferentes fases de leitura possuem similaridade com seqüências de Proteínas de organismos diferentes de *S. mansoni* do banco de dados nr do NCBI. **Sem similaridade**- não possuem nenhuma similaridade com nenhum dos bancos de dados pesquisados. **rRNA**- possuem similaridade com seqüências de RNA ribossomal. **DNA**- possuem similaridade com seqüências de DNA não expressas.

acoplada a esferas metálicas que ficam presas em um campo magnético, não havendo um empacotamento da resina como no caso do oligotex. A ausência de empacotamento e a maior dispersão das esferas poderia facilitar a lavagem e diminuir a contaminação pela adsorção de moléculas contaminantes na coluna, quando comparado com a resina oligotex. Além disso o protocolo de extração com o kit μ MACs requer bem menos tempo do que o protocolo combinado dos dois outros “kits”, permitindo que haja menor degradação do RNA, e preservando eventuais mensagens raras para a amplificação.

Foi montada uma nova biblioteca seguindo o mesmo protocolo da biblioteca f24-I, mas com o uso de RNA mensageiro obtido a partir deste novo protocolo (Biblioteca f33- tabela IV). Ocorreu uma diminuição do nível de mensagens contaminantes de RNA ribossomal, o que permitiu que a biblioteca tivesse um nível menor de redundância.

Uma análise global das bibliotecas construídas ao longo deste processo de otimização do protocolo quanto a sua não redundância (Figura 7) permite-nos dizer a introdução de modificações em nosso protocolo favoreceu um acúmulo de um maior número de mensagens com menor índice de redundância. Podemos ver que a biblioteca f33 apresenta um índice menor de redundância mesmo após a aquisição de cerca de 640 seqüências, comparado com a biblioteca f19 após a aquisição de apenas 24 seqüências.

Além disso, comparamos o tipo de seqüência produzida por cada biblioteca (Figura 8). É possível perceber que a biblioteca f33 apresenta um nível bem maior de seqüências novas (Cerca de 45%) do que as outras bibliotecas (variando de 20% a 35%). Devido ao fato de que as seqüências existentes no banco público são, de modo geral, seqüências abundantes, podemos dizer que com a mudança de protocolo estamos conseguindo amostrar seqüências mais raras.

Tabela IV- tabelamento dos resultados do sequenciamento da biblioteca f33.

Biblioteca f33 - seleção de tamanho de 400 a 700 pb (640 sequências)

Classificação das seqüências ^a	Numero de seqüências
Seqüências conhecidas (53,8 %)	
EST <i>S. mansoni</i>	221 (34,5%)
Gene <i>S. mansoni</i>	55 (8,6%)
rRNA	66 (10,3%)
DNA	2 (0,3%)
Seqüências novas (46,2 %)	
Sem similaridade	221 (34,5 %)
Proteína ≠ <i>S. mansoni</i>	65 (10,1%)
EST <i>S. japonicum</i>	1 (0,2%)
Gene ≠ <i>S. mansoni</i>	9 (1,4%)
Seqüências não redundantes	413 (64,6%)
Seqüências redundantes	227 (35,4%)

^a A classificação das seqüências foi feita de acordo com a seqüência de maior similaridade indicada pelo programa BLASTn e BLASTx com uma exigência de e^{-10} . A classificação da seqüências foi feita nas seguintes categorias: **EST *S. mansoni***- possuem similaridade com seqüências de nucleotídeos de *S. mansoni* do banco de dados dbEST do NCBI. **Gene *S. mansoni***- possuem similaridade com seqüências de nucleotídeos de *S. mansoni* do banco de dados nr do NCBI. **Proteína ≠ *S. mansoni***- após tradução em diferentes fases de leitura possuem similaridade com seqüências de proteínas de organismos diferentes do *S. mansoni* do banco de dados nr do NCBI. **Sem similaridade**- não possuem nenhuma similaridade com nenhum dos bancos de dados pesquisados. **rRNA**- - possuem similaridade com seqüências de RNA ribossomal. **DNA**- possuem similaridade com seqüências de DNA não expressas.

Figura 7- Análise da não redundância das bibliotecas em função do número cumulativo de seqüências de cada biblioteca.

Figura 8- - Classificação dos EST gerados por biblioteca de acordo com sua similaridade com seqüências já depositadas no Genbank. Consideraram-se duas seqüências como similares quando seu E value fosse menor que $e-10$ no BLASTn ou BLASTx. **ESTSm**- EST similar a outros ESTs de *S. mansoni* já depositados no Genbank. **GeneSm**- EST similar a seqüências de nucleotídeos full lenght de *S. mansoni* depositadas no Genbank. **rRNASm**- EST similar a seqüências de rRNA de *S. mansoni* depositadas no Genbank. **DNASm**- EST similar a seqüências de DNA genômico de *S. mansoni* depositadas no genbank. **Prot-não-SM**- EST com um frame de tradução similar a uma proteína de um organismo diferente do *S. mansoni* depositada no Genbank. **Sem similaridade**- EST sem homologia a qualquer seqüência no Genbank. **ESTSj**- EST similar a outros EST de *S. japonico* já depositados no genbank. **Gene-não-Sm**- EST similar a seqüências de nucleotídeo de um gene completo de um organismo diferente de *S. mansoni* depositadas no Genbank.

4.1.3-Estudo da influência dos *primers* na construção de bibliotecas.

A metodologia CODEHOP se utiliza um *primer* longo (de até 45 pares de bases) contendo um núcleo degenerado (*degenerate core*) na extremidade 3' que possui cerca de 12 bases. As degenerações desta região cobrem todos os códons possíveis para os aminoácidos correspondente a esta região do alinhamento.

O *primer* possui também um grampo não degenerado (non-degenerated Clamp) de até 30 bases na extremidade 5'. A função desta região seria de estabilizar a anelamento de *primers* após o ciclo inicial de amplificação, pois todos os *primers* possuem esta região. Deste modo seria possível um dado *primer* se anelar em um sítio de um amplicon gerado a partir de *primer* com bases completamente diferentes da sua nas posições degeneradas, teoricamente estabilizado pela região comum do grampo. Isto geraria um consumo por igual de todos os *primers*, não importando quais bases este possuísse no 3', ao contrário do que ocorre com *primers* degenerados tradicionais (Figura 9).

Para um estudo mais aprofundado da influencia das degenerações na seleção das seqüências desenhamos novos *primers* com modificações nas duas bases da extremidade 3' (hib 8 e 9) ou com adição de degenerações nas bases 4 e/ou 5 da extremidade 3' (hib 10,11 e 12) em relação ao desenho já utilizado (APYAF1) (Figura 10).

Realizamos as reações de RT-PCR com estes *primers* nas mesmas condições que já haviam sido utilizadas com o *primer* APYAF1 na montagem da biblioteca F24I. Deste modo poderíamos comparar os resultados da bibliotecas tendo como única variante os *primers* utilizados.

A) PCR degenerado

B) Primer CODEHOP

Figura 9- Representação esquemática do anelamentos de *primers* degenerados tradicionais e CODEHOP. A flecha indica o *primer* que se anela com a sequência alvo, representada pela linha horizontal abaixo deste. As linhas verticais representam bases com pareamento correto (adaptado de Rose et al., 1998).

APYBAF1

CGGAATTCCGTGTTAAATATGGTATTGTGATNGAYKSNGGNTC

Hib8

CGGAATTCCGTGTTAAATATGGTATTGTGATNGAYKSNGGNAG

Hib9

CGGAATTCCGTGTTAAATATGGTATTGTGATNGAYKSNGGNCT

Hib10

CGGAATTCCGTGTTAAATATGGTATTGTGATNGAYKSNWWNTC

hib11

CGGAATTCCGTGTTAAATATGGTATTGTGATNGAYKSNGWNTC

hib 12

CGGAATTCCGTGTTAAATATGGTATTGTGATNGAYKSNWGNTC

Figura 10- *Primers* arbitrários consenso-degenerado desenhados a partir do *primer* APYAF1. As letras sublinhadas representam as bases modificadas em relação ao *primer* original.

O resultado foi um padrão de amplificação arrastado, indicativo de uma amplificação bem diversa de mensagens, com os *primers* hib 8 e 9 (Figura 11-a) e praticamente nenhuma amplificação com os *primers* hib 10, 11 e 12 (dados não mostrados). Decidiu-se então modificar as condições da reação de PCR destes últimos três *primers* a fim de se obter uma amplificação que permitisse a clonagem destes perfis. Para isso realizamos um abaixamento de 5°C tanto na temperatura máxima quanto na mínima do touchdown-PCR que passou de 60 a 45 °C para 55 a 40 °C. Isso proporcionou um aumento da intensidade de amplificação que permitiu a clonagem destes perfis (figura 11-b).

Figura 11- Eletroforese em gel de acrilamida 6% corado com prata de perfis produzidos a partir de PCR de baixa estrigência .. **A-** Reações com touchdown de 60 a 45 °C. **B-** Reações com touchdown de 55 a 40°C. **H8-1, H9-1, H10-2, H11-2 e H12-2-** reações utilizando os *primers* hib8 a hib12. **Controle** Reação controle sem a adição de transcriptase reversa.

Foram geradas novas bibliotecas a partir destas reações, sendo feita uma análise de não redundância de cada uma (Figura 12) e a classificação da similaridade destas seqüências com seqüências do GenBank utilizando-se dos programas BLASTN e BLASTX (Figura 13). Pode-se ver a partir dos dados de não redundância que o perfil das cinco bibliotecas é bem semelhante, não havendo grandes diferenças entre elas.

Quando comparamos o gráfico de similaridade via BLASTn e BLASTx entretanto, existem diferenças marcantes quanto ao tipo de seqüências que estão sendo amostradas. Isso demonstra que pequenas modificações no *primer* podem alterar seu direcionamento permitindo a amplificação de seqüências diferentes. Apesar disso o nível de seqüências novas (isto é, sem

Figura 12- Análise da não-redundâncias das bibliotecas em relação ao número cumulativo de seqüências de cada biblioteca.

Figura 13- classificação dos EST gerados por biblioteca de acordo com sua similaridade com seqüências já depositadas no Genbank. Consideraram-se duas seqüências como similar quando seu E value fosse menor que $e-10$ no BLASTn ou BLASTx. **EST-** EST similar a outros EST de *S.mansoni* já depositados no Genbank. **Gene SM-** EST similar a seqüências de nucleotídeo completas de genes de *S. mansoni* depositadas no Genbank. **rRNA-** EST similar a seqüências de rRNA de *S. mansoni* depositadas no genbank. **DNA-** EST similar a seqüências de DNA genômico de *S. mansoni* depositadas no Genbank. **Pot. Non SM-** EST com um frame de tradução similar a uma proteína de um organismo diferente do *S. mansoni* depositada no Genbank. **Sem similaridade-** EST sem homologia a qualquer seqüência no Genbank. **ESTSJ-** EST similar a outros EST de *S. japonico* já depositados no Genbank. **Gene nonSM-** EST similar a seqüências de nucleotídeo “full lenght” de um organismo diferente de *S. mansoni* depositadas no Genbank

similaridade via nucleotídeo com seqüências de *S. mansoni* já depositadas no Genbank) é praticamente a mesma para todas as bibliotecas, girando em torno de um índice de 30%.

Selecionamos seqüências que possuíam alta similaridade com seqüências de *S. mansoni* no GenBank (e value menor que e^{-20}) em um alinhamento com o programa BLASTn e um alinhamento sem erros na região adjacente ao *primer*. Selecionamos a região correspondente ao *primer* de nossa seqüência e verificamos o alinhamento desta região com a seqüência depositada por outros autores no banco de dados (assumimos assim que a seqüência depositada corresponderia a seqüência alvo verdadeira de *S. mansoni* onde o *primer* pareou).

Com isso aferimos o índice de identidade entre o *primer* e a seqüência alvo, ao longo das diferentes bases do *primer* (Figura 14). A partir deste gráfico podemos ver que de modo sistemático são apenas as primeiras bases a 3' que determinam a escolha do alvo; podemos perceber também que apesar de terem suas reações de PCR a temperaturas mais baixas os *primers* hib 10, 11 e 12 apresentam um nível de seleção maior que os *primers* hib 8 e 9.

Devido ao fato das bases 3, 6 e 12 do *primer* serem totalmente degeneradas estudamos a proporção em que alvos com bases citosina (C) e guanina (G) foram selecionadas em relação as bases adenina (A) e timina (T) nestas posições. Verificamos que existe um desproporção entre alvos com bases G e C nas bases equivalentes a regiões de pareamento próximas a ponta 3' do *primer* (Figura 15). Este fato pode estar relacionado com o fato de a ligação G com C ser mais estável que A com T, o que proporcionaria maior estabilidade aos anelamentos dos *primers* contendo estas bases.

Como já vimos anteriormente que as bases mais a 3' tem um papel muito importante na seleção do alvo (Figura 14) não é de se estranhar que a estabilidade gerada por ligações GC é mais crítica para seleção de alvos quando está situada próxima do 3'.

Figura 14- Gráfico de coincidências entre as bases dos *primers* utilizados para a construção das diferentes bibliotecas e os supostos trechos de anelamento das seqüências alvo com alta similaridade. As bases 3, 6 e 12 não foram computadas por serem totalmente degeneradas. No caso de posições com degenerações parciais qualquer bases cujo anelamento fosse possível dentro do espectro da degeneração foi considerada correta. Após a 15ª base como a maioria de bases tende a ficar no valor próximo de 25% (indicando uma coincidência por mero acaso) e por esta razão estas bases não foram mostradas.

Figura 15- Análise da porcentagem de bases C ou G na suposta seqüência alvo nas posições de anelamento com as bases totalmente degeneradas dos *primers* utilizados para a construção das diferentes bibliotecas. A linha vermelha indica a porcentagem esperada se a seleção fosse completamente randômica.

Além disso, foi feita uma contagem de bases CG em relação as bases AT presentes nas posições totalmente degeneradas do *primer* seqüenciadas (Figura 16). Podemos ver através deste dado que existe um consumo preferencial de *primers* com bases CG nas posições degeneradas, o que está em desacordo com o que foi descrito para o uso do codehop por Rose et al (1997).

Foi feito um estudo para se verificar se de fato existe uma liberdade de anelamento de um *primer* com qualquer base na posição degenerada após o primeiro ciclo de amplificação, ou se existe uma preferência por *primers* com bases idênticas aquelas a seqüência alvo. Para isso computamos a base original na posição degenerada da seqüência alvo e a base final seqüenciada do *primer*. A partir da base conhecida da seqüência alvo verificamos se nesta mesma posição no *primer* seqüenciado havia uma base que fazia um pareamento ideal. Caso haja estabilização por parte do grampo, a porcentagem de bases com pareamento ideal seria próxima de 25%, visto que já não haveria preferência por um anelamento ideal após o anelamento inicial. Os resultados foram classificados a partir da base original da seqüência alvo, CG ou AT (Figura 17).

Os resultados mostram que nos pareamentos CG há uma tendência à manutenção da base existente no anelamento inicial, ao passo que com o pareamento AT ocorre justamente o contrário. Portanto, os pontos de degeneração do *primer* não funcionaram de maneira aleatória com era esperado. Devido à maior força da ligação CG do que a AT, obtivemos um favorecimento da ligação GC ao longo da reação.

Em conjunto estes dados nos mostravam que em nossas reações não estava havendo uma estabilização pelo grampo, como estava previsto inicialmente pela técnica do CODEHOP. Uma

Figura 16- Análise do consumo dos *primers* utilizados para a construção das diferentes bibliotecas em relação ao consumo preferencial dos *primers* com bases C ou G nas posições totalmente degeneradas em relação ao total de *primer*. A linha vermelha indica o valor esperado caso tivéssemos um consumo igual dos diferentes *primers*.

Figura 17- Análise da estabilização do grampo. Foi analisada a variação de base na posição degenerada em relação à base inicial que fazia um pareamento com a base da sequência alvo. Um índice próximo da linha vermelha indica uma variação randômica desta base após o ciclo inicial e, portanto uma boa estabilização do grampo não-degenerado do *primer*. **b#** refere à posição da base degenerada em relação à ponta 3^o do *primer*. **Tot** é referente ao computo total nesta posição do *primer* e **AT** ou **CG** se refere à estatística considerando apenas os casos com sequências alvo com as referidas bases na posição pesquisada.

hipótese levantada seria que devido ao alto conteúdo de A e T na região do grampo este não conseguiria gerar a estabilização necessária ao longo do *primer*, dada a baixa energia desta ligação.

Foram desenhados novos *primers* que possuíam alto conteúdo de C e G na região do grampo (Figura 18). Entretanto as reações de PCR utilizando-se destes *primers* falharam, não produzindo nenhum perfil de amplificação (dados não mostrados).

Desenhamos também *primers* sem a presença do grampo (Figura 18). As reações de amplificação utilizando-se estes *primers* se tornaram mais robustas existindo inclusive o problema de amplificação em algumas das reações de controle (dados não mostrados).

As bibliotecas geradas a partir deste *primer* sem grampo mostraram-se com alto índice de redundância. Um exemplo é da biblioteca h19-1, que apresentou um rápido esgotamento, se comparada com a biblioteca construída com o *primer* de mesma região 3' com o grampo (h11-2) (Figura 19).

Estes dados indicam que um *primer* com uma possibilidade de pareamento forte na região 5' inibe a amplificação de alvo, provavelmente devido a um pareamento em alvos que possuíam apenas a região 5' similar ao *primer*, o que não permitia que a polimerase iniciasse a extensão a partir deste ponto. A ausência do grampo eliminou este problema permitindo amplificações robustas que no entanto apresentavam baixa variabilidade de mensagens.

Podemos concluir que apesar do grampo não exercer o papel descrito originalmente (Rose et al. 1998), ele é importante para a obtenção de um número variado de seqüências. Foi possível perceber, pela análise feita pelo computo da influência de cada base individual na seleção da seqüência alvo, que individualmente as bases do grampo não têm influência na escolha do alvo (Figura 14). Entretanto coletivamente as bases do grampo exercem uma influência, pois sua retirada leva a seleção de um número bem mais restrito de alvos.


```

hib13 (apyaf1)
CGGAATTCCGTGTCGACGATGGTACGGTGATNGAYKSNGGNTC

hib14 (Hib9)
CGGAATTCCGTGTCGACGATGGTACGGTGATNGAYKSNGGNCT

hib15 (Hib11)
CGGAATTCCGTGTCGACGATGGTACGGTGATNGAYKSNGWNTC

hib16 (Hib10)
CGGAATTCCGTGTCGACGATGGTACGGTGATNGAYKSNWWNTC

hib17 (apyaf1)
CGGAATTCCGGATNGAYKSNGGNTC

hib 18 (Hib9)
CGGAATTCCGGATNGAYKSNGGNCT

hib19 (Hib11)
CGGAATTCCGGATNGAYKSNGWNTC

hib 20 (Hib10)
CGGAATTCCGGATNGAYKSNWWNTC

```

Figura 18- Primers desenhados com alto conteúdo de GC no grupo (Hib 13 a 16) ou sem grampo (Hib 17 a 20). A região 3' é idêntica aos *primers* indicados entre parênteses.

Figura 19- Análise da não-redundâncias das bibliotecas H11-2 e H19-1 em relação ao número cumulativo de seqüências de cada biblioteca.

4.1.4- Análise de parâmetros globais da técnica.

Mapeamos a posição das EST em genes descritos de *S. mansoni*. Sessenta e um dos genes totalmente seqüenciados (full-length) apresentavam identidade com pelo menos uma EST do nosso banco de dados. Foram mapeadas apenas as coordenadas de uma EST com o maior índice de similaridade por cada gene. Uma relativização destas coordenadas foi realizada dividindo-as pelo comprimento total do respectivo gene. O cômputo de todas estas coordenadas gerou um gráfico do posicionamento relativo das ESTs em relação ao gene (Figura 20-A). Este gráfico demonstra que, apesar de haver um direcionamento das ESTs para o centro dos genes, há um pequeno deslocamento do pico para a região 3'. Este fato poderia ser proveniente de uma degradação da região 5' do RNA utilizado.

Levantou-se a hipótese que este desvio poderia ser causado devido ao baixo número de casos analisados. Para obter uma amostragem maior fizemos o mesmo tipo de análise mas desta vez mapeando a tradução das 6 fases de leitura dos ESTs em todas as proteínas do GenBank. Obtivemos 137 casos onde toda EST se mapeava na proteína e estes foram computados do mesmo modo descrito na análise anterior (Figura 20-B).

Neste caso temos um gráfico com um pico centralizado, indicando um direcionamento do RT-PCR para o centro das mensagens. Além disso, o segundo gráfico apresenta um desenho menos irregular, provavelmente devido ao maior nível de amostragem.

Figura 20- Análise do direcionamento das ESTs produzidas. Todos os genes e proteínas tiveram seu tamanho normalizado de modo que seu comprimento total representado pelo eixo x seja 1. O gráfico mostra o mapeamento de nossas ESTs ao longo destas seqüências. O eixo y indica a fração das ESTs alinhadas que se mapeiam no determinado ponto do gene. **A.-** Análise realizada com 61 seqüências completas de RNA de *S. mansoni* existentes no GenBank que apresentaram similaridade com nossas ESTs. **B.-** Análise realizada com 137 seqüências completas de proteínas existentes no GenBank que apresentaram similaridade com nossas ESTs traduzidas (em 1 dos 6 possíveis frames).

Um método de verificação da normalização das bibliotecas foi descrito por Dias Neto et al. em 2000 através da utilização de um artifício de atribuir a abundância de um gene pelo número de seqüências que o agrupamento do Unigene (Schuler et al., 1996) daquela seqüência possui. Foi possível utilizar-se deste artifício, pois há um número enorme (aproximadamente 4 milhões) de seqüências humanas no GenBank, com isso a amostragem de seqüências é randômica, e portanto genes mais abundantes produzirão mais seqüências gerando agrupamentos maiores. Entretanto o *Schistosoma* não possui um número tão grande de seqüências de EST públicas e portanto esta abordagem não é aplicável.

Para contornar este problema produzimos algumas seqüências com nossa metodologia a partir de RNA extraído de amostras de câncer de mama humano (Gentilmente cedido pelo Dr. A. Simpson). As seqüências produzidas foram então relacionadas com os agrupamentos do Unigene através de BLASTn e classificadas de acordo com o tamanho do destes (Figura 21).

É possível perceber que nossa biblioteca apresenta maior número de seqüências similares a agrupamentos com baixo número de membros (de 2 a 64), o que indicaria genes raros. Comparando nossos resultados com os obtidos por Macke et al. (1999), com uma biblioteca construída com *primers* totalmente degenerados e RNA de retina humana, é possível ver que conseguimos uma fração significativamente maior de genes raros (normalização), ao passo que o uso de *primers* totalmente degenerados por Macke et al. levou a um favorecimento da amplificação de seqüências mais abundantes (Figura 21), isto é, perda de normalização.

A partir destes dados verificamos que a introdução de alterações no protocolo de construção das bibliotecas não modificou características de direcionamento para o centro do gene e normalização das mensagens, que haviam sido descritos anteriormente para técnica por Dias Neto et al (2000).

Figura 21- Capacidade de normalização das bibliotecas. Comparação do tamanho dos agrupamentos do Unigene com os quais as seqüências de uma biblioteca apresentam maior similaridade (através do programa BLASTn). **Barras pretas-** Seqüências produzidas com nosso protocolo a partir de RNA de câncer de mama. **Barras brancas-** Seqüências produzidas com *primers* totalmente degenerados por Macke et al. (1996) a partir de RNA de retina humana.

4.1.5- Avaliação da contribuição das seqüências geradas para o conhecimento do transcriptoma de *S. mansoni*.

Foram obtidas 2771 seqüências de boa qualidade a partir do seqüenciamento de todas as bibliotecas produzidas a partir de RT-PCR de baixa estringência. A classificação destas seqüências mostrou que aproximadamente um terço das seqüências eram novas em relação aos bancos de dados públicos existentes (Tabela V).

Tabela V- Tabela dos Resultados do sequenciamento total de bibliotecas

Classificação das seqüências ^a	
Seqüências conhecidas (68,4%)	
ESTs <i>S. mansoni</i>	1003 (36,2%)
Genes <i>S. mansoni</i>	396 (14,3%)
rRNA	463 (16,7%)
DNA	33 (1,2 %)
Seqüências novas (31,6%)	
Sem similaridade	643 (23,2%)
Proteínas $\neq$ <i>S. mansoni</i>	189 (6,8%)
Genes $\neq$ <i>S. mansoni</i>	24 (0,9%)
ESTs <i>S. japonicum</i>	15 (0,5%)
Contaminantes	5 (0,2%)

..^a A classificação das seqüências foi feita de acordo com a seqüência de maior similaridade indicada pelo programa BLASTn e BLASTx com uma exigência de e^{-10} . A classificação das seqüências foi feita nas seguintes categorias: **EST *S. mansoni***- possuem similaridade com seqüências de nucleotídeos de *S. mansoni* do banco de dados dbEST do NCBI. **Genes *S. mansoni***- possuem similaridade com seqüências de nucleotídeos de *S. mansoni* do banco de dados nr do NCBI. **Proteína $\neq$ *S. mansoni***- após tradução em diferentes fases de leitura possuem similaridade com seqüências de Proteínas de organismos diferentes do *S. mansoni* do banco de dados nr do NCBI. **Sem similaridade**- não possuem nenhuma similaridade com nenhum dos bancos de dados pesquisados. **rRNA**- possuem similaridade com seqüências de RNA ribossomal. **DNA**- possuem similaridade com seqüências de DNA não expressas. **Contaminantes** – possuem similaridade acima de e^{-50} no programa BLASTn com seqüências de camundongo.

Para avaliar mais detalhadamente a contribuição de nossas seqüências em relação ao número de novas seqüências de *S. mansoni* geradas decidimos realizar um agrupamento (utilizando o programa Phrap) das seqüências existentes no banco público em conjunto com as nossas seqüências. Como não possuíamos o cromatograma das seqüências públicas um valor de qualidade arbitrário foi atribuído a elas. Foram feitas 4 montagens: somente com as nossas seqüências, somente com as ESTs do banco público, com as ESTs e outras seqüências codificantes do banco público e, finalmente, com todas as seqüências (Tabela VI).

Tabela VI- Resultados das montagens de seqüências

Montagem ^a	# de seqüências ^b	# de agrupamentos ^c	# de seqüências órfãs ^d	Média de tamanho dos agrupamentos ^e	Média de seqüência por agrupamento ^f
GenBank ESTs + CDS	14.278	1.884	4.976	667	4,9
ESTs GenBank +CDS + nossas ESTs	17.049	2.161	5.650	639	5,3
ESTs GenBank	13.967	1.858	4.988	581	4,8
Nossas ESTs	2.771	328	1.003	387	5,4

^aAs seqüência de EST de *S. mansoni* do banco público (EST Genbank), seqüências codificantes não redundantes do banco publico (CDS) e seqüências produzidas neste trabalho (Nossas ESTs) foram agrupadas utilizando o programa Phrap. ^b**# de seqüências**- número total de seqüências que participaram da montagem. ^c**# de agrupamentos**- número de seqüências geradas a partir do agrupamento de pelo menos duas seqüências. ^d**# de seqüências órfãs** – número de seqüências não agrupadas. ^e**Média do tamanho dos agrupamentos**- média em número de bases do conjunto de agrupamentos gerados na montagem. ^f**Média de seqüências por agrupamento**- número médio de seqüências componentes de cada agrupamento.

Quando montadas sozinhas nossas seqüências geram um total de 1331 seqüências únicas (agrupamentos+órfãs) a partir de 2771 seqüências, indicando uma não redundância geral de 48%.

Percebe-se que nossos agrupamentos possuem um tamanho médio menor do que o agrupamento das ESTs do banco público. Isto se deve ao fato de que nossas ESTs são menores por serem produtos de PCR e representarem uma fração delimitada do gene.

Nossas seqüências adicionam 951 seqüências únicas à montagem que corresponderia a todas as seqüências expressas de *S. mansoni* conhecidas (EST Genbank+CDS). Isso corresponde um acréscimo de cerca de 15% ao número de seqüências conhecidas.

A adição de nossas seqüências ao conjunto leva a uma diminuição do tamanho médio dos agrupamentos, apesar do aumento do número de seqüências por alinhamento. Isto se deve ao fato dos agrupamentos novos gerados apenas com nossas seqüências apresentarem tamanho menor que a média.

Cento e sete de nossas EST são novos fragmentos de genes de *S. mansoni* que codificam para proteínas ortólogas a proteínas de outros organismos (Tabela VII). Estes ortólogos foram classificados de acordo com a função inferida para eles. As classes que tiveram maior número de fragmentos foram para proteínas de processos celulares, tais como transdução de sinal e tráfego de proteínas, e enzimas envolvidas em síntese e processamento de proteínas, ambas com 21 seqüências.

Muitas das ESTs foram preditas como tendo potencial codificante pelo programa ESTscan. A seqüência protéica predita por esta programa foi utilizada para busca de domínios conservados nos bancos Smart e PFAM. Em 47 dos nossos fragmentos foi detectado um domínio conservado. A presença destes domínios conservados nestes fragmentos reforça a predição de função.

Entre estes novos ortólogos descobertos podemos destacar a synembryn (ric-8), uma enzima chave da cascata de sinalização $G_o\alpha$ - $G_q\alpha$, descrita recentemente em *C. elegans*, que regula a secreção de neurotransmissores (Miller et al., 2000) e a embriogênese (Miller and Rand, 2000).

Tabela VII- Fragmentos de genes ortólogos gerados em nosso trabalho cujas seqüências codificadas eram similares a proteínas de outros organismos. Análise com o programa BLASTx usando-se a matriz de comparação Blosum 62, com um corte de e^{-5} no parâmetro e-value do programa. **Identificação da seqüência-** identificação das ESTs ou agrupamentos (contig) componentes do fragmento de gene com similaridade. **Proteína com maior similaridade no banco não redundante do GenBank-** Proteína que possuía alinhamento com uma maior pontuação em relação à tradução da seqüência identificada. No campo também são fornecidos o número de acesso da seqüência protéica no GenBank e o organismo do qual a proteína é proveniente. **BLAST e value-** parâmetro de expectativa do programa BLASTx. **Identidade-** número de aminoácidos idênticos em relação ao total entre a proteína e a tradução do fragmento no alinhamento. **Similaridade-** número de aminoácidos semelhantes em relação ao total entre a proteína e a tradução do fragmento no alinhamento. **Cobertura da EST-** porção do fragmento que é alinhado com a proteína. **Cobertura proteína-** Porção da proteína que é alinhada com o fragmento. **Escore do ESTscan-** escore do programa ESTscan em relação ao potencial de codificação da proteína. Fragmentos sem escore são considerados com baixo potencial codificante pelo programa. **Domínio conservado-** Resultado da busca nos bancos PFAM e Smart de domínios conservados da seqüência protéica predita pelo ESTscan. Seqüências que não tiveram escore do ESTscan não puderam ser analisadas. **CD e-value-** parâmetro de expectativa da busca no banco de domínios conservados.

Identificação da sequência	Proteína com maior similaridade no banco não redundante do GenBank	BLAST e value	identidade	similaridade	Cobertura da EST	Cobertura da proteína	escore do EST scan	Domínio conservado(CD)	CD E value
Processos celulares (21 Sequencias)									
f33-69	pir A54694 MAP kinase activator XMEK2, long form - African clawed frog (fragment) [Xenopus laevis]	2e-053	94/130 (72%)	114/130 (87%)	390/397 (98%)	130/457 (28%)	79.2	smart00220 S_TKc, Serine/Threonine protein kinases, catalytic domain; Phosphotransferases.	4e-29
h102-3	gi 2642620 vacuolar sorting protein 28 [Homo sapiens]	4e-040	78/113 (69%)	93/113 (82%)	339/351 (97%)	113/221 (51%)	105.7	Similaridade não encontrada	-
r39-201	gi 8163739 SNAP-24 protein [Drosophila melanogaster]	5e-039	70/114 (61%)	88/114 (76%)	342/445 (77%)	114/212 (53%)	127.5	pfam00835 SNAP-25, SNAP-25 family	1e-29
f24i-103	sp P30676 guanine nucleotide-binding protein g(i), alpha subunit (adenylate cyclase-inhibiting g alpha protein) [Asterina pectinifera]	6e-030	73/112 (65%)	88/112 (78%)	336/348 (96%)	133/354 (39%)	48.8	smart00275, G_alpha, G protein alpha subunit;	6e-07
h72-153	ref XP_040842.1 Vasopressin-activated calcium-mobilizing receptor-1 [Homo sapiens]	4e-028	63/128 (49%)	93/128 (72%)	384/396 (97%)	130/494 (26%)	86.7	pfam00888, Cullin, Cullin family.	4e-23
f33-166	NDR kinase [Drosophila melanogaster]	1e-026	52/82 (63%)	63/82 (76%)	246/272 (91%)	82/463 (18%)	14.3	Similaridade não encontrada	-
Contig59	ref NP_036377.1 ski-interacting protein [Homo sapiens]	8e-026	49/76 (64%)	63/76 (82%)	228/304 (75%)	76/536 (14%)	-	-	-
Contig71	gb AAG29810.1 phosphoinositide phosphatase SAC1 [Rattus norvegicus]	2e-024	56/132 (42%)	88/132 (66%)	396/411 (96%)	133/587 (23%)	39.4	pfam02383 Syja_N, SacI homology domain	4e-09
h112-80	ref XP_007906.3 calcium-regulated heat-stable protein (24kD) [Homo sapiens]	7e-022	51/102 (50%)	67/102 (65%)	306/323 (95%)	102/147 (69%)	51.8	LOAD_cold, cold, Cold Shock RNA binding domain of the OB fold	5e-10
h81-379	gi 12652807 G protein pathway suppressor 1 [Homo sapiens]	5e-21	45/77 (58%)	60/77 (77%)	231/245 (94%)	77/475 (16%)	11.2	Similaridade não encontrada	-
h81-182	ref XP_008830.4 protein phosphatase 2, regulatory subunit A (PR 65) [Homo sapiens]	3e-019	43/58 (74%)	52/58 (89%)	174/240 (82%)	58/589 (10%)	-	-	-
h81-154	dbj BAB00640.1 haspin [Mus musculus]	6e-019	41/90 (45%)	62/90 (68%)	270/366 (74%)	90/754 (12%)	19.2	Similaridade não encontrada	-
Contig306	dbj BAA03338.1 Max [Rattus norvegicus]	3e-015	42/88 (47%)	60/88 (67%)	264/356 (74%)	87/151 (58%)	74.4	Similaridade não encontrada	-
h112-134	gb AAG10199.1 synembryn	1e-013	35/82 (42%)	56/82 (67%)	246/371	82/566	104.1	Similaridade não encontrada	-

h122-17	[Caenorhabditis elegans] dbj BAB14917.1 NADPH- cytochrome P-450 reductase [Hydra vulgaris]	3e-013	29/62 (46%)	49/62 (78%)	186/204 (91%)	62/222 (28%)	17.4	encontrada pfam00667, FAD binding, FAD binding domain	1e-14
Contig124	gi 1703205 arfapin 2 (partner of rac1) (por1 protein) [Homo sapiens]	2e-012	37/81 (45%)	54/81 (65%)	243/376 (65%)	81/224 (36%)	38.2	Similaridade não encontrada	-
h81-339	ref NP_071520.1 calcium/calmodulin-dependent serine protein kinase [Rattus norvegicus]	3e-011	51/133 (38%)	73/133 (54%)	399/452 (88%)	138/909 (15%)	116.5	smart00072 GuKc, Guanylate kinase homologues.	
f24i-196	BAA25269.1 tFZR1 [Oncorhynchus mykiss]	1e-009	25/46 (54%)	34/46 (73%)	138/144 (93%)	46/447 (10%)	-	-	-
f24i-387	NP_032650.1 muscle and microspikes RAS [Mus musculus]	2e-009	32/55 (58%)	44/55 (79%)	165/321 (51%)	66/208 (32%)	-	-	-
f33-772	gi 6136914 nicotinic acetylcholine receptor [Gallus gallus]	6e-009	26/46 (56%)	34/46 (73%)	137/261 (52%)	46/528 (9%)	27.8	pfam02931 Neur_chan_LBD, Neurotransmitter-gated ion- channel ligand binding domain.	7e-12
f33-341	gb AAC33902.1 (AF071315) COP9 complex subunit 6 [Mus musculus]	1e-004	19/48 (39%)	32/48 (66%)	144/209 (69%)	48/324 (15%)	36.6	Similaridade não encontrada	-
Estrutura celular (10 sequencias)									
h81-19	ref XP_032781.1 tubulin, gamma 1 [Homo sapiens]	3e-044	87/102 (85%)	94/102 (91%)	306/317 (97%)	102/451(23 %)	51.7	pfam00091 tubulin, Tubulin/FtsZ family	3e-41
Contig117	ref NP_114105.1 golgi periphera membrane protein p65 [Homo sapiens]	2e-037	74/147 (50%)	96/147 (64%)	441/524 (84%)	141/440 (32%)	76.1	Similaridade não encontrada	-
h122-147	emb CAB66001.1 laminin gamma 1 precursor [Anopheles gambiae]	5e-024	43/87 (49%)	59/87 (67%)	261/339 (77%)	98/1623 (6%)	20.1	pfam00053, laminin_EGF, Laminin EGF-like (Domains III and V)	0.009
f33-114	gi 4803678 ankyrin (brank-2) [Homo sapiens]	2e-020	47/83 (56%)	62/83 (74%)	249/257 (96%)	83/1872 (4%)	52.3	Similaridade não encontrada	-
f33-452	gi 10947027 type IIA procollagen [Canis familiaris]	2e-020	47/90 (52%)	56/90 (62%)	261/271 (96%)	90/1487 (6%)	2.9	Similaridade não encontrada	-
f33-1035	gi 3337431 projectin [Drosophila melanogaster]	6e-011	25/73 (34%)	51/73 (69%)	219/228 (96%)	73/6658 (1%)	4.4	Similaridade não encontrada	-
f33-847	sp p20689 myosin light chain kinase, skeletal muscle (mlck) [Rattus norvegicus]	1e-010	43/135 (31%)	68/135 (49%)	405/439 (92%)	113/610 (18%)	18.9	S_TKc, Serine/Threonine protein kinases, catalytic domain;	7e-11
f33-550	gb AAK82364.1 prominin [Rattus norvegicus]	2e-009	27/74 (36%)	39/74 (52%)	222/371(60)	74/857	-	-	-

	norvegicus							(7%)				
f33-773	ref NP_115268.1 protocadherin gamma subfamily B, 3, isoform 2 precursor [Homo sapiens]	4e-008	29/89 (32%)	53/89 (58%)	267/437 (61%)	87/814 (10%)	62.4	pfam00028 cadherin, Cadherin domain				2e-08
f33-743	gb AAD31715.1 fat-spondin [Drosophila melanogaster]	4e-007	25/86 (29%)	44/86 (51%)	258/369 (69%)	90/763 (12%)	-	-				-
Divisão celular (9 sequências)												
f33-256	gi 2500771 septin 7 (cdc10 protein homolog) [homo sapiens]	1e-035	65/110 (59%)	87/110 (79%)	330/331 (100%)	108/418 (29%)	57.4	pfam00735, GTP_CDC, Cell division protein				9e-25
f33-369	gi 13431705 wd-repeat protein mip1 [Schizosaccharomyces pombe]	2e-25	53/95 (55%)	72/95 (75%)	285/311 (91%)	95/1313 (7%)	27.1	Similaridade não encontrada				-
Contig249	sp P48608 diaphanous protein [Drosophila melanogaster]	2e-022	54/131 (41%)	83/131 (63%)	393/406 (97%)	131/1091 (12%)	99.2	pfam02181 FH2, Formin Homology 2 Domain				1e-18
f33-51	>gi 214647 proliferating cell nuclear antigen (PCNA) [Xenopus laevis]	2e-020	45/79 (56%)	63/79 (78%)	237/247 (96%)	79/261 (30%)	27.0	PCNA_C, Proliferating cell nuclear antigen, C-terminal domain				9e-21
f33-128	gi 4099609 cell division control-related protein 2b [Homo sapiens]	3e-020	53/110 (48%),	73/110 (66%)	330/363 (91%)	110/459 (24%)	94.6	Similaridade não encontrada				-
Contig303	gi 6491856 Microtubule Associated Protein 215 kDa (XMAP215) [Xenopus laevis]	2e-018	54/145 (37%)	88/145 (60%)	429/455 (94%)	139/2065 (7%)	92.2	Similaridade não encontrada				-
f33-1131	gi 6118531 cyclin T1 [Equus caballus]	1e-011	37/112 (33%)	66/112 (58%)	330/335 (99%)	111/727 (15%)	35.6	Similaridade não encontrada				-
h91-131	NP_073153.1 meiotic checkpoint regulator [Homo sapiens]	3e-007	24/64 (37%)	40/64 (62%)	192/264 (73%)	64/2030 (3%)	22.2	Similaridade não encontrada				-
f24i-373	NP_002364.2 microtubule-associated protein 1 A [Homo sapiens]	2e-006	40/111 (36%)	59/111 (53%)	333/361 (92%)	104/2805 (4%)	36.4	Similaridade não encontrada				-
Metabolismo intermediário (7 Sequências)												
h102-86	pir S37300 glycogen phosphorylase (EC 2.4.1.1), brain [Rattus norvegicus]	1e-53	105/152 (69%)	116/152 (76%)	441/451 (98%)	151/843 (18%)	124.5	pfam00343 phosphorylase, Carbohydrate phosphorylase.				4e-51
h122-90	ref NP_065956.1 PAN2 protein [Homo sapiens]	3e-25	59/117 (50%)	81/117 (68%)	351/423(82 %)	215/336 (64%)	32.8	pfam00106 adh. short, short chain dehydrogenase.				1e-05
h122-179	gi 3241933 phosphoglucosyltransferase [Agrobacterium tumefaciens]	3e-23	58/120 (48%)	79/120 (65%)	360/372 (97%)	108/542 (20%)	63.5	pfam02878, PGM_PMM_1, Phosphoglucosyltransferase/homannomutase, alpha/beta/alpha domain I.				2e-08
f33-986	ref NP_000447.1 sulfite oxidase	4e-021	49/102 (48%)	65/102 (63%)	306/319	103/488	49.2	pfam00174,				5e-08

	[Homo sapiens]					(96%)	(21%)		oxidored_molyb, Oxidoreductase molybdopterin binding domain.	
f33-903	gi 7490104 aminomethyltransferase precursor- fission yeast [Schizosaccharomyces pombe]	8e-013	33/48 (68%)	41/48 (84%)	144/247 (58%)	48/387 (12%)	-	-	-	-
h81-210	gi 2995391 dihydrolipoamide S-acetyltransferase [Zymomonas mobilis]	2e-011	32/59 (54%)	44/59 (74%)	177/193 (92%)	59/440 (13%)	-	-	-	-
r39-79	ref NP_127072.1 2-hydroxyhepta-2,4-diene-1,7-dioate isomerase. [Pyrococcus abyssi]	9e-007	21/45 (46%)	30/45 (66%)	135/211 (64%)	45/225 (20%)	-	-	-	-
Biosíntese de pequenas moléculas (6 Sequências)										
h102-38	gi 13277380 lipoic acid synthetase [Mus musculus]	6e-41	81/135 (60%)	100/135 (74%)	405/426 (95%)	135/373 (36%)	85.9	pfam02546, Lipocate synth, Lipocate synthase.	2e-28	
f241-68	ref XP_046481.1 phosphate cytidyltransferase 2, ethanolamine [Homo sapiens]	1e-023	47/80 (58%)	64/80 (79%)	240/292 (82%)	80/358 (22,3%)	76.2	pfam01467, Cytidyltransferase.	9e-08	
h91-160	gb AAD50592.1 phytochelatin synthase [Triticum aestivum]	2e-019	39/80 (48%)	56/80 (69%)	240/365 (65%)	80/500 (16%)	-	-	-	
h81-34	ref NP_010742.1 guanylate kinase; Guk1p [Saccharomyces cerevisiae]	4e-017	44/87 (50%)	65/87 (74%)	261/420 (62%)	82/187 (44%)	62.3	smart00072 GuKc, Guanylate kinase homologues	4e-10	
f231-15	para-hydroxy benzoate polyphenyl diphosphate transferase [Arabidopsis thaliana]	2e-016	38/75 (50%)	51/75 (67%)	224/261 (85%)	74/355 (21%)	45.6	pfam01040, UbiA, UbiA prenyltransferase family.	1e-04	
h81-235	gb AAK25383.1 (AE006002) 4-hydroxybenzoate octaprenyltransferase [Caulobacter crescentus]	7e-013	31/54 (57%)	40/54 (73%)	162/185 (88%)	54/328 (16%)	-	-	-	
Síntese e processamento de proteínas (21 Sequências)										
Contig212	>gi6434956 26S proteasome regulatory complex subunit p42B [Drosophila melanogaster]	2e-047	90/156 (57%)	120/156 (76%)	468/477 (98%)	156/422 (37%)	102.9	pfam01399, PCI, PCI domain.	2e-08	
f33-218	dbj BAA11226.1 human 26S proteasome subunit p97 [Homo sapiens]	5e-043	84/113 (74%)	101/113 (89%)	339/378 (90%)	113/908 (12%)	80.9	Similaridade não encontrada	-	
f33-364	gb AAK12347.1 elongation factor-2 [Machiloides banksi]	1e-042	83/103 (80%)	94/103 (90%)	309/358 (86%)	103/633 (16%)	3.7	Similaridade não encontrada	-	

t24i-226	refNP_201189.1 glucosidase II alpha subunit [Arabidopsis thaliana]	1e-037	66/111 (59%)	83/111 (74%)	333/349(95%)	111/921 (12%)	82.7	pfam01055, Glyco_hydro_31, Glycosyl hydrolases family 31	8e-32
Contig106	>gi6683668 alpha 4 subunit of 20S proteasome [Carassius auratus]	1e-032	67/88 (76%)	79/88 (89%)	264/278 (95%)	88/251 (35%)	26.5	pfam00227, proteasome, Proteasome A-type and B-type.	3e-14
Contig154	refNP_005691.1 dipeptidylpeptidase III [Homo sapiens]	3e-031	61/87 (70%)	75/87 (86%)	261/280 (93%)	87/737 (12%)	71.1	Similaridade não encontrada	-
f23i-147	gb AAB02656.1 ubiquitin-conjugating enzyme E2-32k[Oryctolagus cuniculus]	3e-031	46/61 (75%) + 20/34 (58%)	52/61 (84%) + 27/34 (78%)	288/335 (81%)	96/250(38%)	8.3	pfam00179, UQ_con, Ubiquitin-conjugating enzyme.	9e-05
f33-149	gi86406 procollagen-lysine 5-dioxygenase (EC 1.14.11.4) precursor [Gallus gallus]	2e-029	52/113 (46%)	75/113 (66%)	339/407 (83%)	109/730 (15%)	71.7	pfam02062, Lysyl_hydro, Lysyl hydrolase	1e-37
h142-41	pir S66261 X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - [Flavobacterium meningosepticum]	4e-025	52/101 (51%)	74/101 (72%)	303/366(83%)	101/711 (14%)	53.1	pfam00561 abhydrolase, alpha/beta hydrolase fold.	0.001
t24i-436	refXP_010886.1 eukaryotic translation initiation factor 3, subunit 5 (epsilon, 47kD) [Homo sapiens]	2e-022	31/60 (51%) + 26/37 (70%)	43/60 (71%) + 31/37 (83%)	291/329 (89%)	97/361 (27%)	25.7	pfam01398 Mov34, Mov34/MPN/PAD-1 family	3e-12
h72-68	refNP_032313.1 heat-responsive protein 12 [Mus musculus]	7e-021	49/98 (50%)	68/98 (69%)	294/360 (81%)	98/173 (57%)	28.8	pfam01042, UPF0076, YigF family.	2e-19
f33-251	gi6671670 calpain 6 [Mus musculus]	5e-15	45/102 (44%)	59/102 (57%)	291/313 (93%)	100/641 (16%)	94.8	smart00230, CysPc, Calpain-like thiol protease family	9e-11
h81-246	>gi11558402 SIL1 protein [Homo sapiens]	7e-013	41/114 (35%)	68/114 (58%)	342/440 (78%)	114/461 (25%)	71.3	-	-
f33-1144	pir T50223 rs6/17a ribosomal protein homolog [imported] - fission yeast [Schizosaccharomyces pombe]	4e-012	34/45 (75%)	40/45 (88%)	135/179 (75%)	45/125 (36%)	25.2	pfam01248, Ribosomal_L7Ae, Ribosomal protein L7Ae/L30e/S12e/Gadd45 family.	3e-05
h72-296	gi15234339 ubiquitin-protein ligase UBC9 [Arabidopsis thaliana]	4e-012	31/76 (40%)	50/76 (65%)	228/342 (66%)	76/148 (51%)	31.8	Similaridade não encontrada	-
f33-637	emb CAA04539.1 E6-AP ubiquitin-protein ligase [Homo sapiens]	2e-008	27/41 (65%)	33/41 (79%)	123/152 (81%)	41/76 (54%)	10.5	smart00119, HECTc, Domain Homologous to E6-AP Carboxyl Terminus with ; E3 ubiquitin-protein	3e-04

h112-137	emb CAB40910.1 (AJ238236) ribosome associated membrane protein RAMP4 [Rattus norvegicus]	7e-008	27/61 (44%)	44/61 (71%)	183/309 (59%)	61/75 (81%)	-	ligases	-
Contig172	dbj BAA33466.1 Fly Fas-associated factor (FFAF) [Drosophila melanogaster]	2e-007	28/85 (32%)	49/85 (56%)	255/296 (86%)	85/464 (18%)	40.1	Similaridade não encontrada	-
f33-1137	ref XP_027716.1 mitochondrial ribosomal protein L20 [Homo sapiens]	7e-007	27/77 (35%)	43/77 (55%)	231/233 (99%)	76/149 (51%)	10.6	Similaridade não encontrada	-
f33-60	sp P16255 signal recognition particle 14 kd protein (SRP14) [Canis familiaris]	1e-005	21/30 (70%)	27/30 (90%)	90/144 (63%)	30/110 (27%)	38.6	Similaridade não encontrada	-
h185-40	gb AAK95145.1 ribosomal protein L18a [Ictalurus punctatus]	1e-005	23/39 (58%)	28/39 (70%)	117/137(85 %)	39/176 (22%)	-	-	-
Síntese e processamento de RNA (13 Sequências)									
h122-124	emb CAB56791.1 spliceosomal protein SAP 130 [Homo sapiens]	3e-060	111/129 (86%)	122/129 (94%)	387/392 (99%)	129/1217 (11%)	137.2	Similaridade não encontrada	-
h81-188	ref NP_008937.1 cleavage and polyadenylation specific factor 5, 25 kD subunit; pre-mRNA cleavage factor Im(25kD) [Homo sapiens]	4e-45	79/103 (76%)	93/103 (89%)	306/317 (97%)	103/227 (45%)	14.3	Similaridade não encontrada	-
h72-83	gi 6755594 U5 small nuclear ribonucleoprotein 116 kDa [Mus musculus]	4e-39	74/106 (69%)	89/106 (83%)	318/329 (97%)	106/971 (11%)	53.7	Similaridade não encontrada	-
f33-1157	gb AAF73861.1 p68 RNA helicase [Xenopus laevis]	1e-033	52/89 (58%)+ 23/54 (42%)	69/89 (77%)+ 38/54 (69%)	424/439 (97%)	144/608 (24%)	41.0	pfam00271, helicase C, Helicase conserved C- terminal domain.	6e-18
f33-32	gi 9087150 lim domain transcription factor lmo4 (lim-only protein 4) (lmo-4) -Mus musculus]	3e-032	56/112 (50%)	78/112 (69%)	336/344 (97%)	112/165 (68%)	25.6	Similaridade não encontrada	-
h102-126	ref NP_035265.1 catabolite repressor protein (CCR4)- associative factor 1 [Mus musculus]	6e-029	61/138 (44%)	90/138 (65%)	414/425 (98%)	137/285 (48%)	106.7	Similaridade não encontrada	-
f33-521	>gb AAF06330.1 vitamin D response element binding protein [Saguinus oedipus]	6e-024	48/86 (55%)	65/86 (74%)	258/278 (93%)	86/341(25 %)	52.8	pfam00076 rrm, RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain)	9e-11
Contig48	emb CAB41262.1 CSTF2 (cleavage	2e-022	46/68 (67%)	58/68 (84%)	204/242	68/102	23.3	pfam00076 rrm, RNA	1e-13

	stimulation factor, subunit 2, 64kD) [Homo sapiens]						(84%)	(67%)		recognition motif. (a.k.a. RRM, RBD, or RNP domain)	
f33-286	ref XP_008699.4 CpG binding protein [Homo sapiens]	3e-016	31/64 (48%)	45/64 (69%)			192/399 (48%)	64/656 (10%)	80.9	pfam00628, PHD, PHD- finger	4e-07
f33-1037	ref XP_011304.2 transcription factor CA150 [Homo sapiens]	6e-013	34/44 (77%)	38/44 (86%)			132/231 (57%)	44/588 (7%)	27.9	pfam01846, FF, FF domain.	5e-05
h81-82	gb AAK53461.1 calcium-responsive transcription coactivator [Rattus norvegicus]	1e-010	39/119 (32%)	64/119 (53%)			363/382 (95%)	115/418 (28%)	7.0	Similaridade não encontrada	-
f23s-51	NP_006441.1 splicing factor (45kD) [Homo sapiens]	2e-007	26/40 (65%)	33/40 (82%)			120/132 (91%)	40/405 (10%)	73.7	Similaridade não encontrada	-
f33-1106	dbj BA92375.1 zinc finger transcription factor ZNF207 [Mus musculus]	2e-007	30/59 (50%)	39/59 (65%)			177/355 (50%)	69/495 (14%)	-	-	-
Síntese e processamento de DNA (7 Sequências)											
f33-1057	gb AAC39923.1 zinc-finger helicase [Homo sapiens]	2e-059	107/145 (73%)	129/145 (88%)			435/440 (99%)	145/2000 (7%)	89.9	pfam00176, SNF2_N, SNF2 and others N- terminal domain.	3e-26
h102-133	gi 182179 DNA repair helicase [Homo sapiens]	1e-048	91/116 (78%)	103/116 (88%)			348/366 (95%)	116/782 (15%)	40.1	pfam00271, helicase_C, Helicase conserved C- terminal domain	4e-08
f24i-471	gb AAF52023.1 HDAC3 gene product [Drosophila melanogaster]	2e-039	72/89 (80%)	82/89 (91%)			267/275 (97%)	89/438 (20%)	80.6	pfam00850, Hist_deacetyl, Histone deacetylase family	3e-27
h112-120	gi 7085 histone H2A [Chironomus thummi]	1e-033	69/76 (90%)	74/76 (96%)			238/248 (96%)	76/125 (60%)	-	-	-
f33-363	gi 1786132 DNA topoisomerase I [Gallus gallus]	1e-029	60/83 (72%)	75/83 (90%)			249/258 (97%)	83/766 (11%)	123.0	smart00435, TOPEUc, DNA Topoisomerase I (eukaryota);	1e-21
Contig182	gb AAB87384.1 CHD3 [Drosophila melanogaster]	2e-026	48/63 (76%)	55/63 (87%)			189/272 (69%)	63/1518 (4%)	-	-	-
f24i-82	dbj BAB15807.1 XAB2 [Homo sapiens]	8e-021	44/71 (61%)	60/71 (83%)			213/236 (58%)	71/855 (8%)	93.7	similaridade não encontrada	-
Transporte (8 Sequências)											
h81-342	ref NP_055066.1 solute carrier family 25, member 13 (citrin) [Homo sapiens]	1e-027	64/140 (45%)	93/140 (65%)			420/438 (96%)	146/675 (22%)	35.3	Similaridade não encontrada	-
h102-155	ref NP_068380.1 deoxynucleotide carrier, mitochondrial; mitochondrial uncoupling protein 1	3e-026	65/139 (46%)	89/139 (63%)			417/424 (98%)	129/320 (40%)	18.0	pfam00153, mito_carr, Mitochondrial carrier protein.	4e-13

h102-68	[Homo sapiens] gb AAC25708.1 transportin; TRN [Drosophila melanogaster]	3e-025	49/79 (62%)	62/79 (78%)	237/303 (78%)	79/893 (9%)	7.9	Similaridade não encontrada	-
Contig239	ref NP_004913.1 transport protein Sec24C [Homo sapiens]	1e-023	49/81 (60%)	61/81 (74%)	243/253 (96%)	81/1125 (7%)	48.6	Similaridade não encontrada	-
f33-352	gb AAG49526.1 CLC-type chloride channel CLH-6 [Caenorhabditis elegans]	2e-22	50/94 (53%)	68/94 (72%)	279/311 (90%)	93/796 (12%)	46.1	Similaridade não encontrada	-
h91-38	gi 285987 mitochondrial outer membrane protein 19 [Homo sapiens]	6e-020	44/91 (48%)	63/91 (68%)	273/310 (88%)	91/145 (62%)	-	-	-
f24s-81	gb AAG25703.1 monocarboxylate transporter 4 [Gallus gallus]	1e-018	43/105 (40%)	64/105 (59%)	315/320 (98%)	105/473 (22%)	-	-	-
f33-915	sp P51905 sodium-dependent serotonin transporter (5ht transporter)(cocaine-sensitive serotonin transporter) [Drosophila melanogaster]	2e-008	22/58 (37%)	38/58 (64%)	174/211 (82%)	56/622 (9%)	-	-	-
Outros (5 Sequencias)									
Contig267	gb AAC62955.1 reverse transcriptase SjR1 [Schistosoma japonicum]	7e-044	82/134 (61%)	104/134 (77%)	406/415 (98%)	134/339 (36%)	-	-	-
h102-153	gb AAF21983.1 SM-11044 binding protein [Homo sapiens]	3e-029	54/73 (73%)	64/73 (86%)	217/287 (76%)	73/578 (13%)	16.7	Similaridade não encontrada	-
f24i-653	gb AAC80005.1 myotubularin related protein 2 [Danio rerio]	6e-021	47/95 (49%)	64/95 (66%)	285/303 (94%)	92/393 (23%)	23.1	Similaridade não encontrada	-
f24i-406	gb AAH11294.1 ectonucleotide pyrophosphatase /phosphodiesterase 5 [Mus musculus]	9e-011	37/92 (40%)	51/92 (55%)	276/295 (93%)	86/477 (18%)	61.9	pfam01663, Phosphodiesterase / Type I phosphodiesterase / nucleotide pyrophosphatase.	0.002
h122-45	gb AAB60512.1 Cys2/His2 zinc finger protein [Rattus norvegicus]	1e-007	35/101 (34%)	54/101 (52%)	303/358 (85%)	103/744 (14%)	50.2	Similaridade não encontrada	-

Obtivemos uma EST cuja proteína codificada é similar a um receptor nicotínico para acetilcolina. Esta proteína apresenta interesse, pois drogas antinematódicas agem neste receptor produzindo despolarização e paralisia em músculos de nematodos, devido à disfunção de um canal de cálcio acoplado ao receptor (Köler, 2001).

Obtivemos duas seqüências do proteassomo e Três relacionadas a ubiquitinização de proteínas. Estas proteínas são parte de uma maquinaria proteolítica específica que pode ser de interesse visto participar da regulação de diferenciação celular, crescimento e resposta a estresse.

Três ESTs apresentam similaridade com proteínas do sistema de microtúbulos que tem participação na segregação de cromossomos. Duas delas são proteínas associadas ao microtúbulo (XMAP215 e 1A) e uma tubulina gama, que apesar de ser uma proteína estrutural, foi mostrada como sendo essencial para a iniciação da montagem dos microtubulos (Zheng et al., 1995). O sistema de microtubulos é de interesse pois é alvo de compostos à base de Benzimidazole, um anti-helmintico (Köler, 2001). Apesar deste compostos não apresentarem eficiência contra a esquistosomose como a demonstrada em helmintos (Olds et al., 1999) alguns compostos derivados de Benzimidazole causaram diminuição da contagem de vermes adultos (Nessim et al., 2000) e de ovos (Pacera et al., 1997) em camundongos infectados.

Quatro ESTs foram mapeadas em um BAC construído a partir de DNA genômico do cromossomo 0 de *S. mansoni* (# acesso GenBank AC093457). Esta região codifica para um gene com alta similaridade a um fator de iniciação de *C. elegans*. As ESTs apresentaram descontinuidades em relação à seqüência genômica indicando a presença de íntrons. Apesar de já existirem outras ESTs públicas mapeadas na mesma região das nossas seqüências, uma de nossas seqüências mapeia em uma região para qual não havia nenhuma EST (Figura 22).

A

B

442	tgtataagtgaaaattatatcatataataggaacagacctgtgaatatatgtctcacgat	501
AI394805	tgt	ctgtgaatatatgtctcacaat
AI026192	tgt	ctgtgaatatatgtctcacgat
502	tactaggtaggtcatagttaacgactaatgagacttgcgtgacatcgataccacgcgcga	561
AI394805	tactagg	
AI026192	tactaggtaggtcatagttaacgactaatgagacggggg-catcgataccacgcgcga	272
562	gaatatcagtgacagcaaaaattcgtgatcttccactacggaattcctccataactgcat	621
AI026192	gaatatcagtgacagcaaaaattcgggatcttccactacggaattcctccataactgcat	
BG931893		gaattcctccataactgcat
622	cacgtttttcttgatccatatcactatgctagaggatagaacatttagtaggtttaatat	681
AI026192	cacgtttttcttgatccatatcactatg	
BG931893	cacgtgttgcttgatccatatcactatg	
682	gacaaattacaatgcatgaaacttggaaactttttgagcttaagctcattacaaatattgtg	741
AI026192	aatgcatgaaacttggaaactttttgagcttaagctcatt	
BG931893	aatgcatgaaacttggaaactgtttgagcttaagctcattacaaatattgtg	
742	tggacttgtttacagaattaacaaaaatgacgacttgacttaaattcattattccatata	801
BG931893	tggacttatttacagaattaacaaaaatgacgacttgacttaaattcattattccatata	
802	gatccattaaggtctctaatttgtagtcttctttggatacattaatataaaactgtttga	861
BG931893	gatccattaaggtctctaatttgtagtcttctttggatacattaatataaaactgtttga	
862	tcccatccaatgtcaattcttctttcttaatcagatttgaacaggatcatgcattatac	921
AI223465		ggatcatgcattatac
BG931893	tcccatccaatgtcaattcttctttcttaatcagatttgaacaggatcatgcattatac	
922	tccgtgcgatggaagtattgatttcggcatagtggcagacagtagcatgatctagtatc	981
AI223465	tccgtgcgatggaagtattgatttcggcatagtggcagacagtagcatgat	
BG931893	tccgtgcgatggaagtattgatttcggcatagtggcagacagtagcatgat	
982	aaatgaaatcatacagaagtgtaatcttctgagccgtttcaggtagaaatctatatattt	1041
AI223465		ttgancggtttcaggtagaaatctatatattt
BG931893		ttgagccgtttcaggtagaaatctatatattt
1042	ctttaatttggggttcaaaaacacgatcaagcatttgatcagcttcgtccaacacgaaaa	1101
AI223465	ctttaatttggggttcaaaaacacgatcaagcatttgatcagnttngtccaacncgaaaa	
BG931893	ctttaatttggggttcaaaaacacgatcaagcatttga	
BG931463		aagcatttgatcagcttcgtccaacacgaaaa
BG931416		aaaagcacgagcaagcatttgatcagcttcgtccaacacgaaaa
1102	tcttgatgtagcagttgccagagtattacgattcatcatatcaaaaacacggccgggag	1161
AI977513		gagtattacgattcatcatatcaaacacggccgggag
AI223465	tnttgatggtancagttgccanagtattacgattcatcatatcaaaaacacggccgggag	

Figura 22-A. Diagrama do alinhamento de seqüências em um fragmento de BAC. A numeração apresentada é arbitrária. As barras vermelhas representam seqüências geradas neste trabalho. As barras azuis representam seqüências existentes em bancos públicos. Trechos amarelos representam descontinuidades no alinhamento. B. Alinhamento detalhado do trecho central que inclui o trecho coberto apenas pela seqüência produzida em nosso trabalho.

4.1.6- Avaliação da homologia entre genes de *S. mansoni* e de organismos totalmente seqüenciados.

Com a finalidade de analisar o nível de similaridade entre as proteínas de *S. mansoni* em relação a outros organismos, decidimos realizar uma análise envolvendo as ESTs produzidas e conjuntos de proteínas de organismos totalmente seqüenciados.

Para isso foram utilizados conjuntos de ORFs (open reading frames ou quadro de leitura abertos) preditos a partir do genoma completo de três organismos: *Saccharomyces cerevisiae* (Goffeau *et al.*, 1997), *Caenorhabditis elegans* (The *C. elegans* Sequencing Consortium, 1998) e *Drosophila melanogaster* (Adams *et al.* 2000). As ORFs preditas para cada organismo foram utilizadas como banco de dados para uma busca utilizando as seqüências únicas obtidas na montagem de nossas seqüências (descrito no item 4.1.5) com o programa BLASTx com um parâmetro de exclusão de E-value e-4. Foi contado o numero de seqüências do nosso banco que possuía similaridade com alguma ORF para cada organismo (Figura 23 - barras pretas).

É possível notar que cerca de 30% de nossas seqüências apresentam similaridade com alguma proteína de *D. melanogaster*, valor ligeiramente maior do que o apresentado em relação ao nematodo *C. elegans*. Já em relação ao fungo *S. cerevisiae* há um valor de similaridade bem menor (20%), provavelmente devido a menor complexidade deste organismo unicelular em relação ao *S. mansoni*.

Adicionalmente fizemos o mesmo procedimento para as seqüências de EST existentes no banco público (Figura 23-barras brancas). Podemos ver que novamente os fragmentos apresentam uma similaridade semelhante entre *D. melanogaster* e *C. elegans* e um nível bem menor com *S. cerevisiae*.

Comparando nossos dados com os do banco público, podemos notar que, em todos os casos, uma maior porcentagem de nossos fragmentos apresentam similaridade com as ORFs de outros organismos.

Esta diferença provavelmente é devido ao fato que nossas seqüências tem como alvo preferencial o centro dos genes, que contém a região codificante das mensagens. Deste modo os nossos dados seriam mais adequados para estimar porcentagem de proteínas similares compartilhadas entre os organismos.

Figura 23- Nossa base dados de 1331 seqüências únicas (barras pretas) e a base de dados de 6844 seqüências únicas do banco de EST do GenBank (barras brancas) foram comparados contra uma banco de dados de ORFs preditas de organismos totalmente sequenciados utilizando o programa BLASTx. Aqueles com e-value menor que e^{-4} foram considerados similares. As barras indicam a porcentagem de seqüências que tinham similaridade com pelo menos uma ORF do organismo indicado.

4.2- Clonagem de uma ATP-difosfohidrolase (Apirase) de *S. mansoni*

4.2.1- Amplificação do gene da Apirase de *S. mansoni* utilizando a técnica de RACE.

Quatro meses após o início do projeto “Genoma EST de *Schistosoma mansoni*”, já se tendo acumulado um total de aproximadamente 16.000 seqüências, encontramos entre as ESTs produzidas pelo nosso laboratório para o projeto uma seqüência de 410 pares de bases que possuía uma razoável similaridade com a seqüência protéica de uma apirase de camundongo. Esta similaridade foi detectada utilizando-se o programa de alinhamento local tBLASTN (Figura 24). O que mais chamava a atenção para esta seqüência era a presença de três regiões conservadas da família das apirases, o que fortalecia em muito a hipótese de que esta seqüência seria do gene da apirase.

Entretanto, esta seqüência apresentava o problema que era necessário uma mudança no quadro de leitura para que fosse codificada a proteína que possuísse as três regiões conservadas. Isto indicava que provavelmente a seqüência possuía trechos com baixa qualidade, por isso decidiu-se sequenciar novamente este clone para deste modo obter a seqüência completa sem erros.

O resultado foi uma seqüência de 472 pb que possuía a maior similaridade com uma apirase humana (GenBank Accession number CAB41886) utilizando-se o programa BLASTx (Figura 25) em apenas um quadro de leitura. Este fato reforçava ainda mais a caracterização deste fragmento como pertencente a uma apirase.

Decidiu-se desenhar *primers* internos a esta seqüência com o objetivo de obter-se as seqüências de ambas as pontas, utilizando-se a técnica de RACE (rapid amplification of cDNA ends). Os *primers* foram desenhados utilizando o programa *primer 3* disponível na internet (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) (Figura 26).

Score = 48.9 bits (115), Expect(2) = 4e-11
 Identities = 25/82 (30%), Positives = 43/82 (51%), Gaps = 4/82 (4%)
 Frame = +2

Query: 110 ELSTELIPTSKHHQTPVYLGATAGMRLLRMESEQSADEVLAADVSTSLKSYPFDFQGA-- 166
 E++ L K H ++ TAGMRL +E + + + + LK F +
 Sbjct: 11 EITNSLTFEQKQHTPSLFGLOTAGMRLKLIEDPLGSLDLFSVIRQYLKQSGFQIETRNER 190

Query: 167 -KIITGQEEGAYGWITINYLIG 187
 +++ G EEC YGW+++NY+LG
 Sbjct: 191 IRLLYGSEECLYGWVSVNYILG 256

Score = 35.4 bits (80), Expect(2) = 4e-11
 Identities = 20/37 (54%), Positives = 25/37 (67%), Gaps = 3/37 (8%)
 Frame = +3

Query: 209 GALDIGGASTQITFVPQNSTIESPENSLQF---RLYG 242
 G+LDIGGASTQI F+P+ + P+ L F RLYG
 Sbjct: 300 GSLDIGGASTQIAFIPKVYS-NIPKEKLDFYPLRLYG 407

Figura 24- Resultado com maior escore do alinhamento local da sequência protéica da apirase de camundongo (GenBank accession number P55772, linha superior do alinhamento) contra o banco local de ESTs (linha inferior do alinhamento) gerado no projeto genoma de *S. mansoni* utilizando o programa tBLASTn. As caixas coloridas representam as regiões conservadas da apirase presentes na EST, e descritas como regiões número 2 (azul), 3 (verde) e 4 (vermelho).

```

>gi|4741545|emb|CAB41886.1| (AJ133133) ecto-ATP diphosphohydrolase I [Homo sapiens]
      Length = 517

Score = 89.0 bits (219), Expect = 1e-17
Identities = 52/135 (38%), Positives = 75/135 (55%), Gaps = 9/135 (6%)
Frame = +2

Query: 95  QKQHTPIYLAATAGMRILKLIEDPLGSLDLFSVIRQYLKQSGFQIETPNERIRLLYGSFEG 274
           Q Q TP+YL ATAGMRI +E + + V+ + L F + R++ G EEG
Sbjct: 127 QHQETPVYLGATAGMRILRMESEELADRVLDVVERSLSNYPFDFQGA----RIITGDEEG 182

Query: 275 LYGWVSVNYILGIIKEGK-----QTNPSDTVGSGLDIGCASTQIAFIPKVYS-NIPKE 427
           YGW++N+LG + +TN +T G+LDIGCASTQ F+P+ + P
Sbjct: 183 AYGWITINLLGKFSQKTRWFSIVPYETNNQETFGALDIGCASTQTFVPQNQTIESPDN 242

Query: 428 KLDFYPLRLYGNDFFS 472
           L F RLYG D++
Sbjct: 243 ALQF---RLYGKDYN 254

```

Figura 25- Resultado com maior escore do alinhamento local da sequência do fragmento da apirase de *S. mansoni* (linha superior do alinhamento) contra o banco não redundante de proteínas do GenBank (linha inferior do alinhamento) utilizando o programa BLASTx . As caixas coloridas representam as regiões conservadas da apirase descritas como número 2 (azul), 3 (verde) e 4 (vermelho).

```

APY3-1
GGCTGCAACTGCTGGTATG (19MER)

APY5-1
TACTAGCTCCGCCAAGATCC (20MER)

APY5-2
CTGATACCCAACCATAAAGTCC (22 MER)

```

Figura 26- Primers desenhados para a amplificação das pontas do cDNA. O primeiro primer se destina à amplificação da ponta 3' e os dois últimos a da ponta 5'.

Foi feito um RT-PCR para a ponta 3' (utilizando o *primer* apy3-1 e os *primers* do “kit” de RACE) e um “nested RT-PCR” para a ponta 5' (utilizando os *primers* apy5-1 e apy5-2 e o *primer* do “kit” de RACE) utilizando-se RNA mensageiro de vermes adultos, obtidos através da perfusão de camundongos infectados.

O resultado foi a amplificação de um fragmento em cada um dos experimentos, sendo que o fragmento 3' possuía aproximadamente 2 kb, e o 5' aproximadamente 800 pb (figura 27), que foram então clonados e seqüenciados. Devido ao tamanho do fragmento 3' foi necessária a clivagem deste com a enzima de restrição SspI (Sitio identificado no sequenciamento inicial do clone) e posterior subclonagem para se obter a seqüência completa.

O fato dos vermes usados para extração do RNA utilizado virem da perfusão de camundongos contribuiu para a certeza de que esta seqüência não seria contaminante do hospedeiro. O clone inicialmente seqüenciado no projeto genoma originou-se de vermes obtidos através da perfusão de hamsters. Portanto se fosse uma seqüência de hamster não seria de se esperar que uma seqüência com trechos idênticos fosse amplificada neste segundo experimento, onde os vermes provém de camundongos. Além disso, diversas isoformas da apirase de camundongo já estão descritas e disponíveis no GenBank, e não possuem similaridade com a seqüência aqui clonada.

4.2.2- Análise da seqüência do gene da apirase.

O seqüenciamento revelou que cada um dos dois fragmentos possuía regiões complementares ao clone inicial nas suas pontas correspondentes, de modo que foi possível conseguir o gene completo (Figura 28). A partir desta seqüência foi possível deduzir a proteína gerada, com o programa ORFfinder disponível na internet

(<http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi>). O resultado foi uma ORF com início na base 189 e término na base 1823. Há também uma segunda metionina na base 279 que também é candidata a ser o códon de início da tradução.

Figura 27- Produtos da reação de RT-PCR para a amplificação das pontas do cDNA. Gel de agarose 1,2% corado com brometo de etideo. A reação de transcriptase reversa foi feita durante 30 minutos a 37°C. A reação de PCR foi feita com o seguinte programa: 95°C, 1 minuto;+ 35 ciclos de 95°C, 30 segundos; 55°C, 30 segundos; 68°C, 3 minutos + 68°C, 3 minutos. **A.** Reação para a ponta 3'. **B.** Reação para ponta 5'. A seta indica a banda que foi retirada do gel para clonagem.

Utilizando o programa de predição de topologia transmembrânica MEMSAT (<http://bioinf.cs.ucl.ac.uk/psiform.html>) foi predita uma topologia com duas regiões transmembrânicas localizadas próximas as pontas amino e carboxi terminal. Foi predito também que a região hidrofílica, localizada entre estas duas regiões transmembranares, está voltada para o lado externo da membrana (Figura 29).

Ela foi comparada com outras seqüências de apirase via o programa de alinhamento global Pile up do pacote de programas GCG (Figura 30). A partir deste alinhamento é possível notar que todos os domínios conservados da apirase estão presentes em nossa proteína, e

Figura 28-A) Esquematisação da obtenção da sequência completa da apirase de *S. mansoni*. O retângulo representa o gene completo da apirase, incluindo a região não codificante (verde) e codificante (amarelo). O retângulo roxo representa o *primer* reverso utilizado na amplificação do fragmento 3' e os retângulos azuis representam os *primers* utilizados na clonagem do fragmento 5'. O número ao lado destes retângulos representa a posição da base da extremidade 5' do *primer* em relação à sequência completa. Na parte inferior está indicado o clone inicialmente obtido em nosso laboratório, como parte do projeto genoma de *S. mansoni* (Clone MA1-0037P-V022-A05). As barras marrons indicam 3 ESTs existentes no banco público que possuem identidade com nosso gene, os números adjacentes a estes indicam o início e o final da EST em nossa sequência. B) Sequência completa do gene da apirase de *S. mansoni*. Os códons das duas possíveis metioninas iniciais prevista estão indicado em verde. O códon de parada esta indicado em vermelho. As regiões em azul são os adaptadores do kit de race.

A

B

```

1  TGGCCACGCG TCGACTAGTA CGGGGGGGGG GGTGAAAACG GCAAAAAAATT
51  CCTCGATTCT ACTGATCATA TTAATTTTCAG TCAATTAACA ACTATTGAAC
101 AAGATACTCA ATATTACAGT GTATTACACA TAATATTTCC AATCTGATAT
151 ATGTACATAA TATAACAGT TGAAATATTC CGATCAACAT GGTTTTGGGA
201 AAACGCAACT TATCTCAATT ATTTGATTTT ACCAAACTTG TTAATAAAG
251 TAACATTTCA TTAACCTCTC GAAGAATCAT GGAACGTAAC ACTAATCTTC
301 ATAAGATTTT GACAATATTA GCTTTAATTA TTTTGGTGAT CAGTTTATT
351 GGATTATTAA TTATCATTTG TTTATATATT CTAAGTAGTA ACATTCCAGA
401 TGTGTATTCA GTAGTGATTG ATGCAGGATC AACTTCTTCA AAACCTCATT
451 TATATAAATG GATCGATGAA CCATTTTCGAT CAAATGGAAA AGTTGATGAA
501 GTGACTAATG AAAAAGCTTC TCCTGGCATA TCAGATTATA TTAATGATAC
551 AATTAAAGCT TATGATACAT TAAAACCGAA ACTTTTGAAA TTAACGAATA
601 GTTTAAGATT TGAACAAAAA CAGCATACAC CAATTATTTT GGCTGCAACT
651 GCTGGTATGC GATTAAAGCT AATTGAAGAT CCATTAGGAA GTTTAGATTT
701 ATCTCTAGTA ATTCGTCAAT ATTTAAACAA GTCTGGTTTT CAAATTGAGA
751 CACCAATGTA GCGTATTCGT TTATTGTATG GATCAGAAGA AGGACTTTAT
801 GGTTCGGTAT CAGTGAATTA TATTTTAGGT ATTATTAAG AAGGTAAACA
851 AACAAATCCT TCTGATACAG TTGGTTCTTT GGATCTGGCC GGAGCTAGTA
901 CACAAATCGC ATTTATTTCC AAAGTATATT CAAATATACC AAAAGAAAAA
951 TTAGACTTTT ATCCACTCCG ATTATATGGA AATGATTTTT CAGTATATAG
1001 TCATAGTTTC TTATGCTATG GAAAATCGGA ATTTGAAAGA CGTGTAATAA
1051 CTTCAATTGC AGCAGCATCT TTGTTACAAA GTAAAATTCC TAATCCATGC
1101 TTTCTTCAAG GCTATAAATC GGATTGTGAC AATGCATTTG AATGGTTCTC
1151 TGGAAGTTGT TTATCTGGAA CTTATGTAAG GAAAACATTT GCTGAGGAAA
1201 ATTTCCGACC TCCTAATATG AATAGTTTTT CATTTAATGG TACAGGACAA
1251 CCTAATGAAT GTGTCAATTA TATTCTAAAA CATTTTCAAA CGAAATGCGA
1301 TCATTCTATC TGCTCATTTA ATAATGTTTT TCAACAGGCT CCATTGGAAG
1351 AGTTTATGGC ATATTCTGGA TTTTCGTATG TGATGCGTTA TTTATTTTCT
1401 AATAAGAAC CAGGATTTAC ACTAACAGAA GTCAGTATG CCGTTATGAA
1451 GTTTTGCAAA AAACCATGGA AAGATGTGGC TAAAATTACA AAATTATCTG
1501 ACCAAGGGTT TACTGCTAAA TATTGTTTGT ATGGTTTATA CATTATTACT
1551 TTATTGAAAA TGTATGGTTT TACTACAGAT GAATCATGGA AAACAATTAC
1601 ATTTGAATCG AAAGTTAATG GTAAATCAGT TAGCTGGGCA CTAGGTACAA
1651 TGTTAGATCA ATCTGGTCAT TTACCAAGTG AAAGTCCAAA AGTTTCTGTC
1701 AGTACACAAT TATTTGTTTG TCTGTTTATT CTACTCCTTC TGTTAATCAT
1751 CGGTTCAATT ATTGGATTAT TACTAACTAG AAATTTCCTG TCGAATTCGA
1801 AAAAAATCAAG TAGTCAGGTA TAAATTATGT GTAAACAACA TAAACAACCA
1851 TCACAAATCC ATAAATTCCA TTTTACAATC TGTTATGTAA TACATCAACT
1901 AATCAATATC TTTGATCTTA TTCTCTTTAA TTCCACATTT TCTTCTTTTT
1951 GTGTATAGTA GTAAAACAAA AAAGCTAAAA GGAAATAAAA TCATAATTAC
2001 AGTTTAAATG CTTAATTGGA TTTTGAAAAA AAATAACTAA CAACAACAAA
2051 TAAATAGGCA ATATTTTGTG TTGCCAAGTA AGCAGCACAT TTAACAACAA
2101 AACGCACAAG AATTCACAAA AATCATTAGA AATGATTTTT TTTTCTTAAA
2151 GTTTTCACAA TTCTATTTTT TTGCATTTTA TCTGTCATAC AGACAATTCT
2201 ATATAAACAA AGTAACTGCA AAATAATAAT CTTGCTATTT TTGTATACCT
2251 AACAGCGTGT CTTCAAAACA ATTAATAATA TAAACAATGG TAGTTAATTT
2301 TCCTCCCTTT TTGTTTGTGT AATTTAGGAT ATTTTGTGTC TGTTAAATTG
2351 TGTAATTTTT TTTTCCAATA TCATGTCCTT ACTGTTTATT CTTTTTTCCC
2401 AGACTTAAGT GTCTAAATCA GCCAATCATA GTTATGTATT TGGGTGTGTG
2451 TGTTTTATTT TTTTATGTGT GTTTATCTAA CCATAAGTTG TGTACTCATC
2501 TTTTCTTCTG ATTGGTTTGT TCATACATTT ATTTTCGAGT TTTTATTGG
2551 TTATCGTTGT ATCTGCACTT TACGTTGTCA GAAATAAGGG GAAAATAATG
2601 AATTACTAAA TGTAATTTTA ATAATAATAA TTCTTTTAAC TCAACACAAA
2651 AAATCATTAC ACTTATCTCT ACATCAAATT TTCAAAAAAA AATTNTCAAA
2701 AATTTTTTAC TCAACGCGGG GCCAATCCCC

```

```

FINAL PREDICTION

=====
1: (in) 40-64 (5.58)
2: 505-529 (5.68)

      10      20      30      40      50      60
+++++IIIIIXXXXXXXXXXXXXXXXXX
MVFGKRNLSQLFDFTKLVTISNISLTTPRIMERNTNLHKILTILALIILVISFIGLLIII

      70      80      90     100     110     120
OOOO-----
VLYILSSNIPDVYSVVIDAGSTSSKLHLYKWIDEFPRSNGKVDEVTNEKLSPGISDYIND

      130     140     150     160     170     180
-----
TIKAYDTLKPKLLKLTNSLTFEQKQHTPIYLAATAGMRLKLIEDPLGSLDLFSVIRQYLK

      190     200     210     220     230     240
-----
QSGFQIETPNERIRLLYGSEEGLYGWVSVNYILGIIKEGKQTNPSDTVGSLLDGGASTQI

      250     260     270     280     290     300
-----
AFIPKVYSNIPKEKLDIFYPLRLYGNDFSVYSHSFLCYGKSEFERRVITSIAAASLLQSKI

      310     320     330     340     350     360
-----
PNPCFLQGYKSDLYNAFEWFSGSCLSGTYVKKTFAEEIFRPPNMNSFSFNGTGQPNECVN

      370     380     390     400     410     420
-----
YILKHFQTKCDHSSCSFNNVFQAPFGKFMAYSGFSYVMRYLFPNKNTGFTLTTEVTDVAM

      430     440     450     460     470     480
-----
KFCKKPWKDVAKITKLSQGTAKYCFDGLYIIITLLKMYGFTTDESWKTITFDSKVNGKS

      490     500     510     520     530     540
-----OOOXXXXXXXXXXXXXXXXXXIIII+++++
VSWALGYMLDQSGHLPSESPKVSSTQLFVCLFILLLLLIIGSIIGLLLTRNFLSKSKKS

+++++
TSQVVV

```

Figura 29- Predição de dois segmentos transmembrânicos na apirase de *S. mansoni* obtida pelo programa MEMSAT. (+) acima da letra do aminoácido representa localização interna a membrana, (-) representa localização externa, (O) representa região adjacente à hélice transmembrantar na parte externa da membrana, (I) representa região adjacente à hélice transmembrantar na parte interna da membrana e (X) representa a região da hélice transmembrantar.

Figura 30- Alinhamento global de diferentes apirases, incluindo a apirase de *S. mansoni* clonada em nosso trabalho utilizando o programa Pile up. Empregamos a matriz Blosum 35 com gapweight 10 e gaplengthweight 1. Os retângulos coloridos representam as diferentes regiões conservadas da família das apirases. As seqüências incluídas no alinhamento são as seguintes (# indica o GenBank accession number): NTP1_TOXGO- nucleoside-triphosphatase I precursor (ntpase-I) [*Toxoplasma gondii*] #Q27893; NTP2_TOXGO- nucleoside-triphosphatase II precursor (ntpase-II) [*Toxoplasma gondii*] #Q27895; P_sativum_NTPase- *P.sativum* (Alaska) NTPase translated product [[*Pisum sativum*] #CAA83655.1; apyrase_GS52_Glycine_soja- apyrase GS52 [*Glycine soja*] # AAG32960.1; apyrase_Arabidopsis- apyrase [*Arabidopsis thaliana*] # NP_197329.1; apyrase_potato- apyrase (EC 3.6.1.5) precursor [*Solanum tuberosum*] # JC4616; etdiphosp5_homo_sapiens- ectonucleoside triphosphate diphosphohydrolase 5; CD39-like 4 [*Homo sapiens*] # NP_001240.1; nucleoside_triphosphatase_Mus_musculus- nucleoside triphosphatase [*Mus musculus*] # AAC05181.1; etdiphosphohydrolase_6_homo_sapiens- ectonucleoside triphosphate diphosphohydrolase 6 (putative function); CD39-like 2 [*Homo sapiens*] # NP_001238.1; CD39_MOUSE- ectonucleoside triphosphate diphosphohydrolase 1 [*Mus musculus*] # P55772; CD39_RAT- ectonucleoside triphosphate diphosphohydrolase 1 [*Rattus norvegicus*] # P97687; CD39_HUMAN- ectonucleoside triphosphate diphosphohydrolase 1 [*Homo sapiens*] # P49961; CD39_BOVIN- ectonucleoside triphosphate diphosphohydrolase 1 [*Bos taurus*] # O18956, CD39_PIG- ectonucleoside triphosphate diphosphohydrolase 1 [*Sus scrofa*] # Q9MYU4; CD39_CHICK- ectonucleoside triphosphate diphosphohydrolase 1 [*Gallus gallus*] # O93295; etdiphosphohydrolase3_Homo_sapiens- ectonucleoside triphosphate diphosphohydrolase 3; CD39-like 3 [*Homo sapiens*] # NP_001239.1; Yeast_Nucleoside_Diphosphatase- Yeast Nucleoside Diphosphatase; Apyschistofull- sequencia obtida neste trabalho. Ynd1p [*Saccharomyces cerevisiae*] # NP_010920.1; apyrase_lyosomal_Homo_sapiens- apyrase, lysosomal [*Homo sapiens*] # NP_004892.1. As barras vermelha sob o alinhamento representam a região transmembrânica predita para a apirase de *S. mansoni*.

Resultados e discussão

!!AA_MULTIPLE_ALIGNMENT 1.0

PileUp of: @/tmp/31525.list

Symbol comparison table: genmoredata:blosum35.cmp CompCheck: 5355

GapWeight: 10
GapLengthWeight: 1

CD39_BOVIN_pileup_31525.txt MSF: 860 Type: P October 2, 19101 15:00 Check: 7414 ..

Name: NTP1_TOXGO Len: 860 Check: 409 Weight: 1.00
Name: NTP2_TOXGO Len: 860 Check: 1519 Weight: 1.00
Name: P_sativum_NTPase Len: 860 Check: 4820 Weight: 1.00
Name: apyrase_GS52_Glycine_soja Len: 860 Check: 6119 Weight: 1.00
Name: apyrase_Arabidopsis Len: 860 Check: 1368 Weight: 1.00
Name: apyrase_potato Len: 860 Check: 8631 Weight: 1.00
Name: etdiphosp5_homo_sapiens Len: 860 Check: 6072 Weight: 1.00
Name: nucleoside_triphosphatase_Mus_musculus Len: 860 Check: 5377 Weight: 1.00
Name: etdiphosphohydrolase_6_homo_sapiens Len: 860 Check: 3217 Weight: 1.00
Name: CD39_MOUSE Len: 860 Check: 7839 Weight: 1.00
Name: CD39_RAT Len: 860 Check: 477 Weight: 1.00
Name: CD39_HUMAN Len: 860 Check: 7478 Weight: 1.00
Name: CD39_BOVIN_ Len: 860 Check: 9110 Weight: 1.00
Name: CD39_PIG Len: 860 Check: 9800 Weight: 1.00
Name: CD39_CHICK Len: 860 Check: 7904 Weight: 1.00
Name: etdiphosphohydrolase3_Homo_sapiens Len: 860 Check: 6996 Weight: 1.00
Name: apyschistofull Len: 860 Check: 3963 Weight: 1.00
Name: Yeast_Nucleoside_Diphosphatase Len: 860 Check: 6677 Weight: 1.00
Name: apyrase_lyosomal_Homo_sapiens Len: 860 Check: 9638 Weight: 1.00

//

```

                                1                                50
      NTP1_TOXGO              ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~MWLPVY
      NTP2_TOXGO              ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~MWLPVY
      P_sativum_NTPase        ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~
      apyrase_GS52_Glycine_soja ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~
      apyrase_Arabidopsis     ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~
      apyrase_potato          ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~
      etdiphosp5_homo_sapiens ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~
nucleoside_triphosphatase_Mus_musculus ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~
      etdiphosphohydrolase_6_homo_sapiens M KKGI RYETS RKTSYIFQQP QHGPWQTRMR KISNHGSLRV AKVAYPLGLC
      CD39_MOUSE              ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~
      CD39_RAT                ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~
      CD39_HUMAN              ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~
      CD39_BOVIN              ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~
      CD39_PIG                ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~
      CD39_CHICK              ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~
      etdiphosphohydrolase3_Homo_sapiens ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~
      apyschistofull          ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~M
      Yeast_Nucleoside_Diphosphatase ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~
      apyrase_lyosomal_Homo_sapiens ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~MGRIG ISCLFPASWH FSISPVGCPR ILNTNLRQIM

                                51                                100
      NTP1_TOXGO              V P L L L V F G V S L S L P Q G S L G T D S S S L R G V D A D T E K R I N V G K K H L Q T L R N L E
      NTP2_TOXGO              V P L L L V F G V S L S L P H G S L G T D S S S L R G V D A D T E K R I N V G K T H L Q T L R N L E
      P_sativum_NTPase        ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~M E L L . I K L I T F L L F S M P A I T S S Q Y . L G N N L L T S R K I F L
      apyrase_GS52_Glycine_soja M V L V L W S R G T K N M D F L T L F T L L L L F I H T A L S S T Q Y H D G N I L L T H R K I F P
      apyrase_Arabidopsis     A D K I Q R H R G I I L V I S V P I V L I G L V L L L M P G R S I S D S V V E E Y S V H N R K G G P
      apyrase_potato          ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~M L N Q N S H F I F I I L A I F L V L P L S L L S K N V N A Q I P L . R R H L L
      etdiphosp5_homo_sapiens V C S A V S . . . . . H R . . N Q Q T W F . . . . . E G I F L S . . S M C P I N V S A . . . S .
nucleoside_triphosphatase_Mus_musculus V G S T V F . . . . . Y R . . E Q Q T W F . . . . . E G V F L S . . S M C P I N V S A . . . S .
      etdiphosphohydrolase_6_homo_sapiens V G V F I Y V A Y I K W H R A T A T Q A F F S I T R A A P G A R W G Q Q A H S P L G T A A D G H E .
      CD39_MOUSE              ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~M E D I K D S K V K R F C S K N I L I I L G F T S I L A V I A L I A V G L T Q N K
      CD39_RAT                ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~M E D I K D S K V K R F C S K N I L I I L G F S S V L A V I A L I A V G L T H N K
      CD39_HUMAN              ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~M E D T K E S N V K T F C S K N I L A I L G F S S I I A V I A L L A V G L T Q N K
      CD39_BOVIN              ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~M E D R R E S E L K V F C S K N I L S I L G F S C I I A V I A L L A L G L T Q N K
      CD39_PIG                ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~M E D R R E S E L K T F C S K N I L V I L G F S S I I A V I A L L A L G L T Q N K
      CD39_CHICK              ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~ ~~~~~~M E Y K G K V V A G L L T A T C V F S I I A L I L S A V D V . K D V
      etdiphosphohydrolase3_Homo_sapiens F T V L T R Q P C E Q A G L K A L Y R T P T I A L V . . V L L V S I V L V S I T V I Q I H K Q E
      apyschistofull          N I S L T P R R I M E . . R N T N L H K I L T I L A L I I L V I S F I G L L I I I V L Y I L S S N .

```

Transmembrana

Yeast_Nucleoside_Diphosphatase	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~MLIEN.
apyrase_lyosomal_Homo_sapiens	VISVLA AAVS	LLYFSVVIIR	NKYGR LTRDK	KFORYLARVT	DIEATDTNN.
				Transmembrana	
	101				150
NTP1_TOXGO	TRCHDSLQAL	VVIDAGSSST	RTNVFLAKTR	SCPNKGRSID	PDSIQLIGAG
NTP2_TOXGO	TRCHDSLQAL	VVIDAGSSST	RTNVFLAKTR	SCPNKGRSID	PDSIQLIREG
P_sativum_NTPase	KQEEISSYA	VVEDAGSTGS	RIHVY.HFNQ	NLDLLHIGKG	V.....
apyrase_GS52_Glycine_soja	KQEAITSYA	VIEDAGSTGS	RVHVF.HFDQ	NLDLLRIGNE	L.....
apyrase_Arabidopsis	NSRGPKNYA	VIEDAGSSGS	RVHVF.CFDQ	NLDLIPLGNE	L.....
apyrase_potato	SHES.EHYA	VIEDAGSTGS	RVHVF.RFDE	KLGLLPIGNN	I.....
etdiphosp5_homo_sapiens	TLYG	IMEDAGSTGT	RIHVY.TFVQ	KMPGQLPILE	G.....
nucleoside_triphosphatase_Mus_musculus	TFYG	IMEDAGSTGT	RIHVY.TFVQ	KTAGQLPFLE	G.....
etdiphosphohydrolase_6_homo_sapiens	VFYG	IMEDAGSTGT	RVHVF.QFTR	P.PRETPTLT	H.....
CD39_MOUSE	PLPENVKYG	IVIDAGSSHT	NLYIY.KWPA	EKE.....	
CD39_RAT	PLPENVKYG	IVIDAGSSHT	NLYIY.KWPA	EKE.....	
CD39_HUMAN	ALPENVKYG	IVIDAGSSHT	SLYIY.KWPA	EKE.....	
CD39_BOVIN	ALPENVKYG	IVIDAGSSHT	SLYIY.RWPA	EKE.....	
CD39_PIG	PLPENVKYG	IVIDAGSSHT	SLYIY.KWPA	EKE.....	
CD39_CHICK	FLPPGTYKG	LVEDAGSTHT	ALYVY.QWPA	DKE.....	
etdiphosphohydrolase3_Homo_sapiens	VLPPGLKYG	IVIDAGSSRT	TVYVY.QWPA	EKE.....	
apyschistofull	..IPD..VYS	VVIDAGSTSS	KLHLY.KWID	EPF.....	
Yeast_Nucleoside_Diphosphatase	...TNDREFG	IVIDAGSSGS	RIHVF.KWQD	TESLLHATNQ	DSQSILQSVP
apyrase_lyosomal_Homo_sapiens	...PNVNYG	IVVDCGSSGS	RVFVY.CWPR	HNGN.	P
		ACR1			
	151				200
NTP1_TOXGO	KRFAGLRVVL	EEWLDTYAGK	DWESRPVDAR	LLFQYVPMH	EGAKKLMQLL
NTP2_TOXGO	KRFTGLRVVL	EEWLDTYAGK	DWESRPVDAR	LLFQYVPMH	EGAKKLMQLL
P_sativum_NTPase		EYNN	ITPG.....	..LSSYANNPE	QAAS.LLIPL
apyrase_GS52_Glycine_soja		EFYDK	VTPG.....	..LSAYADNPQ	QAAS.LLIPL
apyrase_Arabidopsis		ELFLQ	LKPG.....	..LSAYPTDPR	QAANS.LVSL
apyrase_potato		EYFMA	TEPG.....	..LSSYADDPK	AAANS.LEPL
etdiphosp5_homo_sapiens		EVFDS	VKPG.....	..LSAFVDQPK	QGAET.VQGL
nucleoside_triphosphatase_Mus_musculus		EIFDS	VKPG.....	..LSAFVDQPK	QGAET.VQEL
etdiphosphohydrolase_6_homo_sapiens		ETFFKA	VKPG.....	..LSAYADDVE	KSAQG.IREL
CD39_MOUSE	NDTGVVQQL	ECQVK.....	..GPG.....	..ISKYAQKTD	E.IGAYLAEC
CD39_RAT	NDTGVVQQL	ECQVK.....	..GPG.....	..ISKYAQKTD	E.IAAYLAEC
CD39_HUMAN	NDTGVVHQVE	ECRVK.....	..GPG.....	..ISKVFQKVN	E.IGIYLTDC
CD39_BOVIN	NDTGVVTQIE	ESNVK.....	..GPG.....	..ISGFAKKVN	E.INVYLTAC
CD39_PIG	NDTGVVSQVE	ECKLK.....	..GPG.....	..ISEFAKKLG	E.IDIYLEAC
CD39_CHICK	NETGIVSQVE	SCTVN.....	..GSG.....	..ISSYADDPK	G.AGASLKPC
etdiphosphohydrolase3_Homo_sapiens	NNTGVVSQVE	KCSVK.....	..GSG.....	..ISSYGNNPT	D.VPRAFEEC
apyschistofull	RSNGKVDVTF	NEKL.....	..SPG.....	..ISDYINDTI	K.AYDTLKP
Yeast_Nucleoside_Diphosphatase	H....IHQEK	DW.....TFK	LNPG.....	..LSSFEKKPQ	DAYKSHIKPL
apyrase_lyosomal_Homo_sapiens	HDLLDIRQMR	DKNRKPVVMK	IKPG.....	..ISEFATSP.	EKVS DYISPL
	201				250
NTP1_TOXGO	EEDTVAILDS	QLNEKQKVQV	KALGIPVMLC	STAGVROFHE	WYRDALFVLL
NTP2_TOXGO	EEDTVAILDS	QLNEEQKVQV	KALGIPVMLC	STAGVROFHE	WYRDALFVLL
P_sativum_NTPase	LEQAEDVVPD	DLQPKT....	PVRLG	ATAGLRLL..	..NGDASEKIL
apyrase_GS52_Glycine_soja	LEEAESVVE	DLYPTT....	PVKLG	ATAGLRLL..	..EGDASERIL
apyrase_Arabidopsis	LDKAEASVPR	ELRPKT....	HVRVG	ATAGLRLL..	..GHDASENIL
apyrase_potato	LDGAEGVVPQ	ELQSET....	PLELG	ATAGLRLL..	..KGDAAEKIL
etdiphosp5_homo_sapiens	LEVAKDSIPR	SHWKKT....	PVVLK	ATAGLRLL..	..PEHKAKALL
nucleoside_triphosphatase_Mus_musculus	LEVAKDSIPR	SHWERT....	PVVLK	ATAGLRLL..	..PEQKAQALL
etdiphosphohydrolase_6_homo_sapiens	LDVAKQDIPF	DFWKAT....	PLVLK	ATAGLRLL..	..PGEKAQKLL
CD39_MOUSE	MELSTELIPT	SKHHQT....	PVYLG	ATAGMRLLRM	ESEQSADEVL
CD39_RAT	MKMSTERIPA	SKQHQT....	PVYLG	ATAGMRLLRM	ESQSADEVL
CD39_HUMAN	MERAREVIPR	SQHQT....	PVYLG	ATAGMRLLRM	ESEELADRVL
CD39_BOVIN	MERAQKVIPS	IQHMET....	PVYLG	ATAGMRLLRM	ENKQMA DKIL
CD39_PIG	MERARTVVPK	SQHAET....	PVYLG	ATAGMRLLRM	KNENLASKIL
CD39_CHICK	LDKAMAVIPV	EQOWQT....	PTYLG	ATAGMRLLRE	QNSTKAEQVF
etdiphosphohydrolase3_Homo_sapiens	MQKVKGQVPS	HLHGST....	PIHLG	ATAGMRLLRL	QNETAANEVL
apyschistofull	LLKLTNSLTF	EQKQHT....	PIYLA	ATAGMRLLKI	EDPLGSLDLF
Yeast_Nucleoside_Diphosphatase	LDFAKNIIPE	SHWSSC....	PVFIQ	ATAGMRLL..	..PQDIQSSIL
apyrase_lyosomal_Homo_sapiens	LNFAAEHVPR	AKHKET....	PLYIL	CTAGMRLL..	..PESQQA KAIL
				ACR2	
	251				300
NTP1_TOXGO	RHLINNPSPA	H.GYKFFTNP	FWTRPIIGAE	EGLFAFITIN	HLSRRLGEDP
NTP2_TOXGO	RHLINNPSPA	H.GYKFFTNP	FWTRPIIGAE	EGLFAFITIN	HLSRRLGEDP
P_sativum_NTPase	QSVRDMLSNR	...STFNVPQ	DAVSIINGTQ	EGSYLWVTYN	YALGNLGKKY
apyrase_GS52_Glycine_soja	QAVRDLLKNR	...STLNAQP	DAVAIINGTQ	EGSYLWVTIN	YLLGKLGKRF

apyrase_Arabidopsis	QAVRELLRDR	...SMLKTEA	NAVTVLIGTQ	EGSYQWVTIN	YLLRNLGKPY
apyrase_potato	QAVRNLVKNQ	...STFHSKD	QWVTILIGTQ	EGSYMMAAIN	YLLGNLQKDY
etdiphosp5_homo_sapiens	FEVKEIFRK.	...SPFLVPK	GSVSIMIGSD	EGILAWVTIN	FLTGQLHGHR
nucleoside_triphosphatase_Mus_musculus	LEVEEIFKN.	...SPFLVPD	GSVSIMIGSY	EGILAWVTIN	FLTGQLHGHR
etdiphosphohydrolase_6_homo_sapiens	QKVKEVFKA.	...SPFLVGD	DCVSIMIGTD	EGVSAWITIN	FLTGSCLKTPG
CD39_MOUSE	AAVSTSLKSY	PFDF....QG	A..KIITGQE	EGAYGWITIN	YLLGRF....
CD39_RAT	AAVSRSLKSY	PFDF....QG	A..KIITGQE	EGAYGWITIN	YLLGRF....
CD39_HUMAN	DVVERSLSNY	PFDF....QG	A..RIITGQE	EGAYGWITIN	YLLGKF....
CD39_BOVIN	AAVASSISEY	PFDF....QG	A..RIISGQE	EGAYGWITIN	YLLGKF....
CD39_PIG	STVAESITRY	PFDF....QG	A..RIITGQE	EGAYGWITIN	YLLDKF....
CD39_CHICK	AEVSKAIREF	PVDF....RG	A..QILTGNE	EGSFGWITIN	YLLETL....
etdiphosphohydrolase3_Homo_sapiens	ESIQSYFKSQ	PFDF....RG	A..QIISGQE	EGVYGWITIN	YLMGNF....
apyschistofull	SVIRQYLKQS	GFQIETPNR	I..RLLYGSE	EGLYGWVSVN	YILGII....
Yeast_Nucleoside_Diphosphatase	DGLCQGLK.H	PAEFLVEDCS	AQIQVILGET	EGLYGWLGIN	YLYG...HFN
apyrase_lyosomal_Homo_sapiens	EDLLTDIP.V	HFDLFLSDSH	A..EVISGKQ	EGVYAWIGIN	FVLGRFEHIE
ACR3					
		301			
NTP1_TOXGO	ARC.....	.MIDEYGVKQ	CRNDLAGVVE	VGGSAQIVF	PLQEGTVLPS
NTP2_TOXGO	ARC.....	.MIDEYGVKH	CRNDLAGVVE	VGGSAQIVF	PLQEGTVLPS
P_sativum_NTPase			..TKTVGVID	IGGSSVQIMAY	AVS.KKTAKN
apyrase_GS52_Glycine_soja			..SKTVGVVD	IGGSSVQIMAY	AIS.RNTAKN
apyrase_Arabidopsis			..SDTVGVVD	IGGSSVQIMAY	AIS.EEDAAS
apyrase_potato			..KSTTATID	IGGSSVQIMAY	AIS.NEQFAK
etdiphosp5_homo_sapiens			..QETVGTLD	IGGASTQITF	LPQFEKTLQ
nucleoside_triphosphatase_Mus_musculus			..QETVGTLD	IGGASTQITF	LPQFEKTLQ
etdiphosphohydrolase_6_homo_sapiens			..GSSVGMLD	IGGASTQIAF	LPRVEGTLQA
CD39_MOUSE	TQEQ	SWLSLI.SDS	QKQETFGALD	IGGASTQITF	.VPQNSTIES
CD39_RAT	TQEQ	SWLNFI.SDS	QKQATFGALD	IGGASTQITF	.VPLNQTLQA
CD39_HUMAN	SQKT	RWFESIVPYET	NNQETFGALD	IGGASTQITF	.VPQNQITIES
CD39_BOVIN	TQKL	SWFNLPKPSK	DTQETYGALD	IGGASTQITF	.VPQNQETTES
CD39_PIG	IQKS	GWENLKPCKG	DTQETYGALD	IGGASTQITF	.VPQNQVLES
CD39_CHICK	IKFS	FAGKWEH..P	QNTQVGLGALD	IGGASTQITF	.QP.GVTIED
etdiphosphohydrolase3_Homo_sapiens	LEKN	LWHMNVH..P	HGVETTGALD	IGGASTQISF	.VA.GEKM.D
apyschistofull	KE..	GKQT	NPSDTVGSGLD	IGGASTQIAF	.IPKVYSNIP
Yeast_Nucleoside_Diphosphatase	DYNPEVSD..		..HETFGEMD	NGGASTQIAF	.APHDSGEIA
apyrase_lyosomal_Homo_sapiens	DDDEAVVEVN	IPGSESSEAI	VRKRTAGILD	IGGASTQIAF	EVPKTVSFAS
ACR4					
		351			
NTP1_TOXGO	SVRAVNLQRE	RLL	PERYPSADV	SVSFMQLGMA	SSAGLFLKEL
NTP2_TOXGO	SVRAVNLQRE	RLL	PERYPSADV	SVSFMQLGMA	SSAGLFLKEL
P_sativum_NTPase	AP.....KVA	DGDDPYIKKV	VLKGIPYDLY	VHSYLHFGRE	ASRAEILKLT
apyrase_GS52_Glycine_soja	AP.....KPP	DGEEPIYIKKL	VLKGKEYDLY	VHSYLHYGRE	ASRAEILKVT
apyrase_Arabidopsis	AP.....KPL	EGEDSYVREM	YLGKRYFLY	VHSYLHYGLL	AARAEILKVS
apyrase_potato	AP.....QNE	DGE.PYVQK	HLMSKDYNLY	VHSYLNHYQL	AGRAEIFKAS
etdiphosp5_homo_sapiens	TP.....R..	GYLTSF	EMFNSTYKLY	THSYLGFLGK	AARLATLGAL
nucleoside_triphosphatase_Mus_musculus	TP.....R..	GYLTSF	EMFNSTFKLY	THSYLGFLGK	AARLATLGAL
etdiphosphohydrolase_6_homo_sapiens	SP.....P..	GYLTAL	RMFNRTYKLY	SYSYLGGLIM	SARLAILGGV
CD39_MOUSE	PENSLQFR..	LYGED.....	YTVY	THSFLCYGKD	QALWQKL.AK
CD39_RAT	PETSLQFR..	LYGTD.....	YTVY	THSFLCYGKD	QALWQKL.AQ
CD39_HUMAN	PDNALQFR..	LYGKD.....	YVNY	THSFLCYGKD	QALWQKL.AK
CD39_BOVIN	PNNNLYFR..	LYGKN.....	YSVY	THSFLCYGKD	QALLQKL.AL
CD39_PIG	PENTLHFR..	LYGKN.....	YSVY	THSFLCYGKD	QALLQKL.TK
CD39_CHICK	KNTSVLFR..	LYGTN.....	YSLY	THSYLCYQOI	QASKRLM.AA
etdiphosphohydrolase3_Homo_sapiens	LNTSDIMQVS	LYGYV.....	YTLY	THSFQCYGRN	EAEKKFL.AM
apyschistofull	KEKLDIFYPLR	LYGND.....	FSVY	SHSFLCYGKS	EFERRVI.TS
Yeast_Nucleoside_Diphosphatase	RHRDDIATIF	LRSVN..GDL	QK...WDVF	VSTWLGFGAN	QARRRYL.AQ
apyrase_lyosomal_Homo_sapiens	SQQEEVAKNL	LAEFNLGCDV	HQTEHVYRVY	VATFLGFPGN	AARQRYE.DR
		401			
NTP1_TOXGO	CSNDEFLLQGG	ICSNPCLFKG	FQQSCSAGEV	EVRPDGSASV	NEDVRKNRLK
NTP2_TOXGO	CSNDEFLLQGG	ICSNPCLFKG	FQQSCSAGEV	EVRPDGSASV	NEDVRKNRLK
P_sativum_NTPase	.PRSPN....	PCLLAG	FNGIYTYSGE	EFK.....	
apyrase_GS52_Glycine_soja	.DGSAN....	PCILAG	YDGTYYTSGA	DYK.....	
apyrase_Arabidopsis	.EDSEN....	PCIVAG	YDGMKYGGK	EFK.....	
apyrase_potato	.RNESN....	PCALEG	CDGYYSYGGV	DYK.....	
etdiphosp5_homo_sapiens	.E.TEGTDGH	TFRSACLPRW	LEAEWIFGGV	KYQ.....	
nucleoside_triphosphatase_Mus_musculus	.E.AKGTGDH	TFRSACLPRW	LEAEWIFGGV	KYQ.....	
etdiphosphohydrolase_6_homo_sapiens	.EGQPAKDGK	ELVSPCLSPS	FKGEWEHAHV	TYR.....	
CD39_MOUSE	.DIQVSS.GG	VLKDPFCFNP	YKK.VVNVSE	LYG.....	
CD39_RAT	.DIQVSS.GG	ILKDPFCFYP	YKK.VVNVSE	LYG.....	
CD39_HUMAN	.DIQVAS.NE	ILRDPFCFHP	YKK.VVNVSD	LYK.....	

CD39_BOVIN_	.GLQ.GT.NG	IIHEPCFHRS	YMR.KIKMSV	LNE.....	
CD39_PIG	.DLK.NT.NG	TIHEPCFHSG	YQR.RMNVSH	LYE.....	
CD39_CHICK	.LHQDGSYVQ	NISHPCYPKG	YRR.IITIAE	IYD.....	
etdiphosphohydrolase3_Homo_sapiens	.LLQNSPTKN	HLTNPCYPRD	YSI.SFTMGH	VFD.....	
apyschistofull	.IAAASLLQS	KIPNPCFLQG	YKSDLYNAFE	WFS.....	
Yeast_Nucleoside_Diphosphatase	.LI.NTLPEN	TND.....	YENDDFST.R	NLN.....	
apyrase_lyosomal_Homo_sapiens	.IFANTIQKN	RLL.....	GKQTGLTPDM	PYL.....	
451					
NTP1_TOXGO	PLATYCSVNN	PEISFKVTNE	MQCRENSIDP	TKPLAERMKI	ENCSIIGKTG
NTP2_TOXGO_	PLATYCSVNN	PEISFKVTNE	MQCRENSIDP	TKPLAERMKI	ENCSIIEGTG
P_sativum_NTPase	A	TAYT.....			SGAN
apyrase_GS52_Glycine_soja	A	FAPL.....			SGSS
apyrase_Arabidopsis	A	PASQ.....			SGAS
apyrase_potato	V	KAPK.....			KGSS
etdiphosp5_homo_sapiens	Y	GGNQ.....			EGEV
nucleoside_triphosphatase_Mus_musculus	Y	GGNQ.....			EGEM
etdiphosphohydrolase_6_homo_sapiens	V	SGQK.....			AAAS
CD39_MOUSE	T	PC.....TKR	...FEKKLP.	.FDQFR....	IQGTG
CD39_RAT	T	PC.....TKR	...FEKKLP.	.FNQFQ....	VQGTG
CD39_HUMAN	T	PC.....TKR	...FEMTLP.	.FQQFE....	IQGIG
CD39_BOVIN_	G	FC.....TKR	HELNSSFP.	.LVDIE....	IRGAG
CD39_PIG	A	PC.....TRR	...FLTSLP.	.FPELE....	IQGTG
CD39_CHICK	S	PC.....V..	PTP.SMLSP.	.AQILT....	VTGTG
etdiphosphohydrolase3_Homo_sapiens	S	LC.....TVD	QRP.ESYNP.	.NDVIT....	FEGTG
apyschistofull	G	SCLSGTYVKK	TFAEIEFRPP	NMNSFS....	FNGTG
Yeast_Nucleoside_Diphosphatase	D	PCM.....	PRGSSTDFEF	KDTIFH....	IAGSG
apyrase_lyosomal_Homo_sapiens	D	PCL.....	PLDIKDEIQQ	NGQTIY....	LRGTG
501					
NTP1_TOXGO	NFDKCVSQVE	SILVAPKLPL	PANIEAASSG	.FESVDQVFR	FASSTAPMIV
NTP2_TOXGO_	NFDKCVSQVE	SILVAPKLPL	PANIEAASSG	.FESVDQVFR	FASSTAPMIV
P_sativum_NTPase	.FNKCKNTIR	KALKLNYP.C	PY.QNCTFGG	IWNGGGGNGQ	KNLFASSSFF
apyrase_GS52_Glycine_soja	.YDECREVLV	QALKLNES.C	PH.QNCTFGG	IWDGGRGSGQ	KILFGTSSFY
apyrase_Arabidopsis	.LDECRRTIT	NALKVNDTLC	TH.MKCTFGG	VWNGGRRGGGQ	KNMFVASSFF
apyrase_potato	.WKRCRRLTR	HALKIN.AKC	NI.EECTFNG	VWNGGGGGGQ	KNIHASSFFY
etdiphosp5_homo_sapiens	GFEPYCAEVL	RVVRGKLL..	.H.QPEEVQR	G.....	.SFYAFSYYY
nucleoside_triphosphatase_Mus_musculus	GFEPYCAEVL	RVVQKLL..	.H.QPEEVQR	S.....	.AFYAFSYYY
etdiphosphohydrolase_6_homo_sapiens	LHELCAARVS	EVLQNRV...	.H.RTEEVKH	V.....	.DFYAFSYYY
CD39_MOUSE	DYEQCHQSIL	ELF..NNSHC	PYSQ.CAFNG	VFLPPLH...	GSFGAFSAFY
CD39_RAT	DYEQCHQSIL	KFF..NNSHC	PYSQ.CAFNG	VFLPPLQ...	GSFGAFSAFY
CD39_HUMAN	NYQQCHQSIL	ELF..NTSYC	PYSQ.CAFNG	IFLPLPQ...	GDFGAFSAFY
CD39_BOVIN_	NFQRCRQSII	QLF..NTSYC	PYSS.CSFNG	VFLPPLH...	GQFGAFSAFY
CD39_PIG	DFQKCRQSIR	PLF..NTSYC	PYSR.CSFDG	VFLPLPQ...	GDFAAFSAFY
CD39_CHICK	NPAACPTAIL	KLF..NL.TC	GANRTCGRFDG	VYQPPVR...	GQFFAFAGFY
etdiphosphohydrolase3_Homo_sapiens	GFSLCKEVA	SIF..DFKAC	HDQETCSFDG	VYQPKIK...	GPVAFAGFY
apyschistofull	QPNECVNYIL	KHF..QTK.C	DHS.SCSFNN	VFPAPF...	GKFMAYSFGS
Yeast_Nucleoside_Diphosphatase	NYEQCTKSIY	PLL..LKNMP	CDDEPCLFNG	VHAPRIDFAN	DKFIGTSEYW
apyrase_lyosomal_Homo_sapiens	DFDLCRETIQ	PFM..NKT..	.NETQTSLNG	VYQPPIHFNQ	SEFYGFSEFY
551					
NTP1_TOXGO	TGGGMLAAIN	TLKDHRLRLS	DFSGDVEELA	EAAREFCSSSE	V.IIRTDGPV
NTP2_TOXGO_	TGREMLASID	TLKDHRLRLS	DFSGDVEELA	EAAREFCSSSE	V.IIRTDGPV
P_sativum_NTPase	YLPEDTGMVD	ASTPNFILR.	PVDIE	TKAKEACALN	FEDAKSTYPF
apyrase_GS52_Glycine_soja	YLPTEIGIID	LNKPNISKIH.	PVDLE	IEAKRACETK	LEDAKSTYPN
apyrase_Arabidopsis	DRAAEAGFVD	PKQPVATVR.	PMDFE	KAACKACSMK	LEEGKSTFPL
apyrase_potato	DIGAQVGIVD	TKFPSALAK.	PIQYL	NAKVACQTN	VADIKSIFPK
etdiphosp5_homo_sapiens	DRAVDTDMID	YEK.GGILK.	VEDFE	RKAREVCD.N	LENFTSGSP.
nucleoside_triphosphatase_Mus_musculus	DRAADTHLID	YEK.GGVLLK.	VEDFE	RKAREVCD.N	LGSFSSGSP.
etdiphosphohydrolase_6_homo_sapiens	DLAAGVGLID	AEK.GGSLV.	VGDFE	IAAKYVCR.T	LETQPPQSSP.
CD39_MOUSE	FV...MDFFK	KVAKNSVIS.	QEKMT	EITKNFCSKS	WEETKTSY..
CD39_RAT	FV...MDFFK	KMANDSVSS.	QEKMT	EITKNFCSKP	WEEVKASY..
CD39_HUMAN	FV...MKFLN	LTSEK..VS.	QEKVT	EMMKKFCAPQ	WEEIKTSY..
CD39_BOVIN_	YV...MEFLN	LTSEES.VS.	VEQLT	EKLEEFCAQR	WEEVQKNF..
CD39_PIG	YV...MGFLN	LTSEEG.SF.	QSKVT	STLEAFCSR	WAELOMYF..
CD39_CHICK	YT...FSFLN	LTGQQSL...	SHVN	ATVWDFCNKN	WSELVETP..
etdiphosphohydrolase3_Homo_sapiens	YT...ASALN	LSGSFSL...	DTFN	SSTWNFCSQN	WSQLPLLL..
apyschistofull	YV...MRYLF	PNKNTGFTL.	TE.VT	DAVMKFCKKP	WKDVAKIT..
Yeast_Nucleoside_Diphosphatase	YT...ANDVF	KLK..GEYN.	FDKFS	KSLREFCNSN	WTQILANSK
apyrase_lyosomal_Homo_sapiens	YC...TEDVL	RMG..GDYN.	AAKFT	KAAKDYCATK	WSILRERFDR
600					

		601		650
	NTP1_TOXGO	IQLPNARGEQ	KLNSLNFDLC	KTMALTVSLL ...RHMAAGE NQPSFIKWEK
	NTP2_TOXGO	IQLPNARGEQ	KLNSLNFDLC	KTMALTVSLL ...RHMAAGE NQPSFIKWEK
	P_sativum_NTPase	LD.....KKN	VASY....VC	MDLIYQYVLL VDGFGLDPLQ K.....ITS
	apyrase_GS52_Glycine_soja	.P.....AED	RLPY....VC	LDIAYQYALY TDGFSLDPWQ E.....ITV
	apyrase_Arabidopsis	VE.....EEN	.LPY....LC	MDLVYQYTLL IDGFGLEPSQ T.....ITL
	apyrase_potato	TQ.....DRN	.IPY....LC	MDLIYEYTL VDGFGLNPHK E.....ITV
	etdiphosp5_homo_sapiens		...F....LC	MDLSYITALL KDGFGFADST V.....LQL
	nucleoside_triphosphatase_Mus_musculus		...F....LC	MDLYITALL KDGLGFAERH P.....LTA
	etdiphosphohydrolase_6_homo_sapiens		...F....SC	MDLTYVSLLL QE.FGFPRSK V.....LKL
	CD39_MOUSE	...PSVKEK	YLSE....YC	FSGAYILS.L LQGYNF..TD SSWEQIHF..
	CD39_RAT	...PTVKEK	YLSE....YC	FSGTYILSLL LQGYNF..TG TSWDQIHF..
	CD39_HUMAN	...AGVKEK	YLSE....YC	FSGTYILSLL LQGYHF..TA DSWEHIHF..
	CD39_BOVIN	...GEVKEK	YLSE....YC	FSGTYILVLL LNGYHF..TA ESWKNIHF..
	CD39_PIG	...GDVKEK	YLSE....YC	FSGTYILTLL LSGYHF..TA ETWKNIHF..
	CD39_CHICK	...PQNKE	HLHT....YC	VVGLYILTLL VDGYKF..DE HTWSNIHF..
	etdiphosphohydrolase3_Homo_sapiens	...PKFDEV	YARS....YC	FSANYIYHLF VNGYKF..TE ETWPOIHF..
	apyschistofull	...KLSDQG	FTAK....YC	FDGLYIITLL .KMYGFT.TD ESWKTITF..
	Yeast_Nucleoside_Diphosphatase	GVYNS.IPEN	FLKD...AC	FKGNWVLNLL HEGFDMPRID VDAENVNDRP
	apyrase_lyosomal_Homo_sapiens	GLYASHADLH	RLKY....QC	FKSAMMFVEF HRGFSFP... ..VNYKS
		651		700
	NTP1_TOXGO	SIAGPDGKPL	ADLCW.VGVI	LHHVLFTEEW GRNAYEAGYS HNL~.....
	NTP2_TOXGO	SIAGPDGKPL	ADLCW.VGVI	LHHVLFTEEW GRTAYEAGYS HNL~.....
	P_sativum_NTPase	GKEIEYQDAI	VEA...W.LGNA	VEAISALPKF ERLMYFV~..
	apyrase_GS52_Glycine_soja	ANEIEYQDAL	VEA...W.LGTA	IEATSSLPKF DRIMYFI~..
	apyrase_Arabidopsis	VKKVKYGDQA	VEA...W.LGSA	IEAVSS.P~..
	apyrase_potato	IHDVQYKNYL	VGA...W.LGCA	IDLVSTTNK IRVASS~..
	etdiphosp5_homo_sapiens	TKKVVN...	IET...W.LGAT	FHLQSLGIS H~.....
	nucleoside_triphosphatase_Mus_musculus	HKESEQ...	.HRLW.LGLG	GHLSPAPVSG HHQLRPSSTS EACISEPVFS
	etdiphosphohydrolase_6_homo_sapiens	TRKIDN...	VET...W.LGAI	FHYIDSL... NRQKSPAS~..
	CD39_MOUSE	...MGKIKD	SNAG...W.LGYM	LNLTNMIPAE QPLSPPLPHS TYIGLMVLF
	CD39_RAT	...MGKIKD	SNAG...W.LGYM	LNLTNMIPAE QPLSPPLPHS TYISLMVLF
	CD39_HUMAN	...IGKIQG	SDAG...W.LGYM	LNLTNMIPAE QPLSTPLSHS TYVFLMVLF
	CD39_BOVIN	...MNKVR	TDV...W.LGYM	LNLTNKIPAE EPMSPPLPHS TYVFLMVLF
	CD39_PIG	...MGKVQS	TSV...W.LGYM	LNLTNMIPSE EPSTRLSHS TYVFLMVLF
	CD39_CHICK	...SQKAGN	ADIG...W.LGFM	LNLTNMIPTE ALE.HVKGHE PSLWAGATSF
	etdiphosphohydrolase3_Homo_sapiens	...EKEVGN	SSI...W.LGYM	LSLTNQIPAE SPL.IRLPIE PPVFGTLLAF
	apyschistofull	...DSKVNG	KSV...W.LGYM	LDQSGHLPSE SP...KVSVS TQLFVCLFIL
	Yeast_Nucleoside_Diphosphatase	LFQSVKVEE	REL...W.LGRI	LLYASGSILA GNDDFMVGIA PSERRTKLTG
	apyrase_lyosomal_Homo_sapiens	LKTALQ.VYD	KEV...W.LGAI	L.....
			ACR5	Transmembrana 750
	NTP1_TOXGO			
	NTP2_TOXGO			
	P_sativum_NTPase			
	apyrase_GS52_Glycine_soja			
	apyrase_Arabidopsis			
	apyrase_potato			
	etdiphosp5_homo_sapiens			
	nucleoside_triphosphatase_Mus_musculus	QEGVDSETFS	DLSGKAWPET	R.....
	etdiphosphohydrolase_6_homo_sapiens			
	CD39_MOUSE	LLLAV.VIT	GLFIYSKPSY	FWKEAV~..
	CD39_RAT	LVLVAM.VIT	GLFIYSKPSY	FWKEAV~..
	CD39_HUMAN	LVLFTV.AII	GLLIYHKPSY	FWKDMV~..
	CD39_BOVIN	LILLAV.IIV	GIVVFHKPSY	FWKDMV~..
	CD39_PIG	LILVIV.VII	GLFVCHRP	FWKDMV~..
	CD39_CHICK	IVLAIV.AGL	VAILLQC...	FWKSK~..
	etdiphosphohydrolase3_Homo_sapiens	FTVAAL.LCL	AFLAYLC...	SATRRKRHSE HAFDHAVDSD
	apyschistofull	LLLLIIIGSII	GLLLTRNFTS	KSKKSTSQV~
	Yeast_Nucleoside_Diphosphatase	KKFIPGKLL	SDQLRKQSSS	LSNKGFLMWF AIICCFYLI FHRSHIIRR
	apyrase_lyosomal_Homo_sapiens	.RFLPLRDIQ	QEAFRASHT.	HWR GV.....S FVYNHYL...
		751	Transmembrana	800
	NTP1_TOXGO			
	NTP2_TOXGO			
	P_sativum_NTPase			
	apyrase_GS52_Glycine_soja			
	apyrase_Arabidopsis			
	apyrase_potato			
	etdiphosp5_homo_sapiens			

nucleoside_triphosphatase_Mus_musculus	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
etdiphosphohydrolase_6_homo_sapiens	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
CD39_MOUSE	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
CD39_RAT	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
CD39_HUMAN	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
CD39_BOVIN	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
CD39_PIG	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
CD39_CHICK	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
etdiphosphohydrolase3_Homo_sapiens	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
apyschistofull	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
Yeast_Nucleoside_Diphosphatase	FSGLYNITKD	EKTGIRRRLK	FLRRSDPFS.	RLEEG	ELGTDVDGFK
apyrase_lyosomal_Homo_sapiens	FSGCFLVLL	AILLYLLRLR	RIHRRTPRSS	SAAALWMEEG	LPAQNAPGTL
	801				850
NTP1_TOXGO	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
NTP2_TOXGO	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
P_sativum_NTPase	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
apyrase_GS52_Glycine_soja	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
apyrase_Arabidopsis	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
apyrase_potato	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
etdiphosp5_homo_sapiens	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
nucleoside_triphosphatase_Mus_musculus	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
etdiphosphohydrolase_6_homo_sapiens	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
CD39_MOUSE	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
CD39_RAT	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
CD39_HUMAN	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
CD39_BOVIN	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
CD39_PIG	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
CD39_CHICK	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
etdiphosphohydrolase3_Homo_sapiens	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
apyschistofull	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
Yeast_Nucleoside_Diphosphatase	DVYRMKSSSM	FDLGKSSATM	QREHEPQRTA	SQSANLAPSN	LRPAFSMADF
apyrase_lyosomal_Homo_sapiens	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
	851				
NTP1_TOXGO	~~~~~				
NTP2_TOXGO	~~~~~				
P_sativum_NTPase	~~~~~				
apyrase_GS52_Glycine_soja	~~~~~				
apyrase_Arabidopsis	~~~~~				
apyrase_potato	~~~~~				
etdiphosp5_homo_sapiens	~~~~~				
nucleoside_triphosphatase_Mus_musculus	~~~~~				
etdiphosphohydrolase_6_homo_sapiens	~~~~~				
CD39_MOUSE	~~~~~				
CD39_RAT	~~~~~				
CD39_HUMAN	~~~~~				
CD39_BOVIN	~~~~~				
CD39_PIG	~~~~~				
CD39_CHICK	~~~~~				
etdiphosphohydrolase3_Homo_sapiens	~~~~~				
apyschistofull	~~~~~				
Yeast_Nucleoside_Diphosphatase	SKFKDSRLYD				
apyrase_lyosomal_Homo_sapiens	~~~~~				

alinhados com as outras apirases. Entretanto podemos notar que a apirase de *S. mansoni* possui motivos bem característicos nas regiões adjacentes aos domínios. Tal fato é esperado pois a maioria das apirases descritas pertencem a vertebrados que possuem uma distância filogenética muito grande em relação ao *S. mansoni*.

Através da pesquisa dos bancos públicos via o programa de alinhamento local BLASTN foi possível encontrar três ESTs de *S. mansoni* (GenBank accession numbers AW017414, BF936655 e AA269045) que possuem identidade com a região não codificante do gene da apirase (Figura 28-a barras marrons). O fato deles se localizarem na região 3' não codificante explica porque não foi possível deduzir que eles faziam parte deste gene através da busca com o programa BLASTx. Isto demonstra a utilidade de técnicas de produção de EST com o uso de RT-PCR de baixa estrigência, como a utilizada neste projeto, que tendem a amplificar o centro do gene onde se situam as regiões conservadas, permitindo a descoberta de novos genes.

A apirase de *S. mansoni* possui maior similaridade com o grupo das NTPDases 1 (CD39). Estas já foram descritas na literatura como possuidoras de uma relação de hidrólise ATP:ADP próxima de 1, e possuem inserção na membrana via N- e C- terminal (Heine et al., 2001). Tal caracterização é reforçada pelo fato de nossa proteína possuir duas regiões transmembranas preditas (sublinhadas no alinhamento da figura 30), que se alinham com as regiões transmembrânicas das CD39.

Entretanto, nossa sequência possui uma região amino-terminal significativamente maior que as CD39 se considerarmos a primeira metionina como o início da proteína. Neste aspecto ela seria semelhante a CD39L2 humana (etdiphosphohydrolase_6_homo_sapiens no alinhamento da figura 30), que possui uma sequência sinal postulada para os primeiros 78 aminoácidos (Chadwick and Frischauf, 1998) ou com a apirase lisosomal humana que tem duas regiões transmembranares preditas em sua região amino terminal (Biederbick et al., 1999).

Se considerarmos a segunda metionina como o aminoácido inicial teremos uma proteína que tem seu início praticamente no mesmo ponto que as CD39 alinhadas. Não podemos afirmar qual das predições é a correta, visto que a pequena diferença de tamanho entre as duas (30 aminoácidos) não nos permite fazer uma avaliação a partir do tamanho da proteína detectada no parasita anteriormente (Vasconcelos et al.,1996), além da falta de outras evidências experimentais.

4.2.3- Detecção do gene da apirase no genoma do parasita.

Para confirmar a existência desta sequência no DNA genômico do parasita realizou-se um experimento de southern blot utilizando o EST original como sonda. Foram feitas digestões por do DNA genômico de *S. mansoni* com as enzimas de restrição BamHI, Hpa I e uma digestão dupla com esta duas enzimas. O resultado foi o aparecimento de duas bandas em para cada reação (Figura 31). Entretanto, não existe nenhum sítio de corte para nenhuma das duas enzimas dentro da sequência utilizada como sonda e portanto seria de se esperar apenas uma banda. Isto pode indicar a existência de um intron que possuísse os dois sítios dentro desta região, ou a existência de isoformas dentro do genoma

Foi encontrada uma sequência proveniente de sequenciamento de pontas de BACs de *S. mansoni* (GenBank accession number BH209132, 470pb) que possui um trecho de 109 pb idênticos a nossa sequência, sendo que após este ponto o alinhamento é interrompido (figura não mostrada). Isto indica a existência de um íntron nesta região do gene.

Este BAC não possui a sequência completa do gene da apirase, pois uma de suas pontas contém uma região do meio do gene. Decidiu-se utilizar o DNA do gene da apirase para a realização de um varredura em uma membrana de alta densidade contendo BACs com insertos de

DNA genômico de *S. mansoni* de uma biblioteca construída no “Children’s Hospital Oakland Research Institute” (CHORI) para a obtenção de um clone contendo a sequência completa do gene da apirase. Foram obtidos quatro clones positivos que foram comprados da instituição. Estes clones foram digeridos com a enzima *DraI* e três fragmentos de tamanhos diferentes desta digestão foram reconhecidos pela sonda do gene da apirase (Figura 32).

4.2.4- Expressão heteróloga da apirase de *S. mansoni*.

Foram desenhados *primers* que permitiam amplificar toda a região codificante da apirase, com o objetivo de clonagem em um vetor de expressão. A finalidade é produzir esta proteína em um sistema heterólogo. Utilizou-se cDNA de vermes adultos gerados a partir da transcrição com poly T, como substrato para uma reação de PCR com estes *primers* e obteve-se uma banda única (Figura 33).

O vetor escolhido para a clonagem foi o pET 21-b (Novagen). O vetor permite que a metionina inicial utilizada seja a da nossa própria sequência através da ligação em um sítio *NdeI*. Na porção carboxi terminal, o vetor possui um trecho adicional contendo uma sequência de hexahistidina, permitindo purificação com o uso de uma coluna de níquel.

Os *primers* foram desenhados de modo a possuir extremidades 5’ com sítios de restrição *Nde I* e *Sac I*. Deste modo o fragmento amplificado e digerido pode ser inserido direcionalmente no vetor pET digerido com as mesmas enzimas.

Foram seqüenciados vários clones para determinação de um clone sem mutações introduzidas na reação de PCR. Estes clones foram alinhados e percebeu-se que existiam mutações pontuais que se repetiam em vários clones, sendo que alguns vinham de diferentes reações de PCR. Portanto estas mutações não podiam ser provenientes da reação de PCR mas

Figura 31- Southern blot utilizando DNA genômico de *S. mansoni* e sonda de fragmento da apirase de *S. mansoni* obtida por RT-PCR. **BamHI** refere-se à reação do DNA com a enzima de restrição BamHI (Na figura é visível apenas uma banda no poço devido a baixa resolução, no filme fotográfico é possível perceber duas bandas.), **Hpa I** refere-se à reação do DNA com a enzima de restrição Hpa I e **BamHI/Hpa I** refere-se à reação do DNA com as duas enzimas.

Figura 32- Detecção da apirase em Bacs através do uso de sonda radioativa. Revelação de autoradiografia de membrana de nylon com fragmentos da digestão de BACs com a enzima DraI. A amostra de digestão foi submetida a uma eletroforese em gel de agarose 0,8% em TAE. Após a transferência, a membrana foi incubada com uma sonda do gene da apirase marcada com fósforo radioativo. Os nomes acima da raia indicam a placa e a posição do clone na biblioteca CHORI-103.

sim ocorrentes naturalmente na população (Figura 34). Os dados da figura 34 sugerem a existência de duas populações que podem representar SNPs da apirase ou ainda duas isoformas.

As mutações eram na maior parte silenciosas (não modificavam o aminoácido), com exceção de uma que modificava uma leucina para uma isoleucina no aminoácido 57 da proteína. Como se pode ver na Figura 29 esta mudança ocorre na região da hélice transmembranar. Devido ao fato de ambos aminoácidos terem um caráter hidrofóbico tal mutação não deve gerar uma grande mudança no comportamento da proteína.

O clone M1clone4 foi então escolhido para a realização da expressão desta proteína em *E. coli* cepa BL21(DE3) pYlsS. Foi realizada a indução da expressão desta proteína com a adição de IPTG para concentração final de 1mM. O extrato solúvel obtido foi submetido a um SDS-PAGE e posterior western-blot com anticorpo anti apirase de batata, que possui imunoreatividade cruzada com a apirase de *S. mansoni* conforme descrito em Vasconcelos et al.(1996). O resultado do western-blot está na Figura 35.

Pode-se notar uma banda de aproximadamente 61kDa reativa no western blot, sugerindo a presença da apirase de *S. mansoni* que está sendo expressa. Esta afirmação baseia-se no fato de que a apirase de *S. mansoni* já havia sido detectada em western blots de extratos de tegumento com o mesmo anticorpo, e que através do programa Comput pI/Mw (disponível no site http://ca.expasy.org/tools/pi_tool.html) foi calculado um peso molecular de 63 kDa para nossa construção da apirase com uma cauda de hexahistidina.

Realizou-se um ensaio de western-blot com o lisado bacteriano contra um anticorpo monoclonal anti-hexahistidina para verificar sua presença na proteína recombinante. Conforme se vê na figura 36 ocorreu a identificação de bandas de baixo peso molecular na fração insolúvel, o que pode indicar que está havendo uma degradação das extremidades da proteína retirando a

Figura 33- Produto da amplificação do cDNA de *S. mansoni* com os *primers* Apy MIS (GGAATTCCATATGGTTTTTGG) e Apypet(TCCGAGCTCGATACCTGACTAGTTG) sob o seguinte programa de PCR: 95°C, 1 minuto;+ 25 ciclos de 95°C, 30 segundos; 55°C, 30 segundos; 68°C, 3 minutos + 68°C, 3 minutos. O produto amplificado esta dentro do tamanho esperado para a região codificante da apirase de *S. mansoni*.

	451				500
ApymII_clone3	TATATAAATG	GATCGATGAA	CCATTTTCGAT	CGAATGGAAA	AGTTGATGAA
ApymII_clone4	TATATAAATG	GATCGATGAA	CCATTTTCGAT	CGAATGGAAA	AGTTGATGAA
ApyMI_clone3	TATATAAATG	GATCGATGAA	CCATTTTCGAT	CGAATGGAAA	AGTTGATGAA
ApymI_clone4	TATATAAATG	GATCGATGAA	CCATTTTCGAT	CAAATGGAAA	AGTTGATGAA
apyrase_full_lenght	TATATAAATG	GATCGATGAA	CCATTTTCGAT	CAAATGGAAA	AGTTGATGAA
					
	701				750
ApymII_clone3	ATTCTCAGTA	ATTCGTCAAT	ATTTAAAACA	GTCTGGTTTT	CAAATTGAAA
ApymII_clone4	ATTCTCAGTA	ATTCGTCAAT	ATTTAAAACA	GTCTGGTTTT	CAAATTGAAA
ApyMI_clone3	ATTCTCAGTA	ATTCGTCAAT	ATTTAAAACA	GTCTGGTTTT	CAAATTGAAA
ApymI_clone4	ATTCTCAGTA	ATTCGTCAAT	ATTTAAAACA	GTCTGGTTTT	CAAATTGAGA
apyrase_full_lenght	ATTCTCAGTA	ATTCGTCAAT	ATTTAAAACA	GTCTGGTTTT	CAAATTGAGA
					
	1451				1500
ApymII_clone3	GTTTGTGAAA	AAACCATGGA	AAGATGTGGC	TAAAATTACA	AAATTATCTG
ApymII_clone4	GTTTGTGAAA	AAACCATGGA	AAGATGTGGC	TAAAATTACA	AAATTATCTG
ApyMI_clone3	GTTTGTGAAA	AAACCATGGA	AAGATGTGGC	TAAAATTACA	AAATTATCTG
ApymI_clone4	GTTTGTGAAA	AAACCATGGA	AAGATGTGGC	TAAAATTACA	AAATTATCTG
apyrase_full_lenght	GTTTGTGAAA	AAACCATGGA	AAGATGTGGC	TAAAATTACA	AAATTATCTG

Figura 34- Mutações pontuais encontradas a partir do alinhamento dos diferentes clones da apirase de *S. mansoni*. Acima são apresentados trechos do alinhamento dos diferentes clones gerados. O numero acima das seqüências representa o numero de bases em relação ao inicio da seqüência completa (incluindo não codificante). As 9 letras em vermelho representam as bases que possuem mutação em relação à seqüência original.

cauda de hexahistidina. Nossa hipótese para explicar este fato seria a presença de regiões transmembrânicas de nossa proteína que não estavam sendo inseridas em membrana, levando a um enovelamento incorreto que resultaria na ativação de mecanismos de degradação de *E. coli* de proteínas mal enoveladas.

Devido à degradação e perda da cauda carboxi-terminal de hexahistidina, decidiu-se realizar uma nova construção retirando as duas regiões hidrofóbicas correspondentes a regiões transmembrânicas da proteína, previstas como estando localizadas na porção amino-terminal (aminoácidos 44 até 60) e carboxi-terminal (aminoácidos 510 até 527). Para isso desenhou-se novos *primers* para a região desejada (cDNA correspondentes aos aminoácidos 65 a 510) e clonou-se este novo fragmento no vetor de expressão (figura 37).

A proteína recombinante desta vez se expressou em grandes quantidades em *E. coli* e se localizou na fração insolúvel. Realizando-se um western-blot com anticorpo anti hexahistidina é possível ver que a proteína possuía a cauda de hexahistidina (Figura 38). Um peso molecular abaixo do esperado (52kDa) foi obtido no western blot. Entretanto após solubilização desta proteína dos corpos de inclusão com o uso de SDS em alta concentração (10%), fervura por aproximadamente 30 minutos e diálise para a retirada do SDS, a migração da proteína no gel passou a ser próxima da correta em relação a seu tamanho esperado (52 kDa) (ver figura 39).

A proteína não apresentou reconhecimento pelo anticorpo anti-apirase de batata (dados não mostrados), mas devido ao fato de estarmos lidando com um fragmento é possível que o epítipo reconhecido não esteja mais presente na proteína ou não esteja exposto devido ao fato da proteína apresentar-se insolúvel.

Figura 35- Expressão heteróloga da apirase de *S.mansoni* em *E. coli*. **A.** SDS-PAGE corada com Comassie Blue da fração solúvel da bactéria recombinante após 3 horas da indução com 1mM de IPTG. **B.** Western-blot com anti-soro anti apirase de batata da mesma amostra (NOTA: a imagem do western-blot está ampliada em relação ao gel corado). + Amostra da fração solúvel da bactéria induzida. – controle negativo da mesma fração em bactéria sem indução.

Figura 36- Western blot com anticorpo anti hexahistidina (1:1000) das frações solúvel e insolúvel da expressão da apirase em sistema heterologo. As bandas de reconhecimento da hexahistidina estão na fração insolúvel e com baixo peso molecular.

Figura 37- Esquematização da clonagem do fragmento hidrofílico da apirase e sua expressão. A) Esquema da amplificação do fragmento. Os números representam a posição em bases no cDNA. As setas indicam os *primers* utilizados. TM-Região transmembranica. B) Esquema da proteína expressa em sistema heterólogo.

4.2.5- Purificação em coluna de níquel e imunização para produção de soro policlonal.

Decidiu-se realizar a purificação desta proteína utilizando coluna de níquel com o objetivo de se obter uma proteína pura para imunização em coelhos. A purificação foi feita utilizando imidazol como competidor nas lavagens (20mM) e eluição (250 mM) da coluna (Figura 39).

É possível perceber que parte de nossa proteína é perdida no volume não ligado da coluna e outra fração significativa da proteína é eluída na fração de lavagem. Em nossas amostras das frações de eluição temos uma amostra purificada em relação a amostra bruta. Foi levantada a hipótese de que a perda de proteína na fração não ligada e de lavagem era causada pela presença de SDS na amostra, que interfere na ligação na proteína na coluna. Apesar disso pode-se

Figura 38- Expressão em sistema heterólogo de um fragmento de 52 kDa da apirase de *S. mansoni*. **A.** SDS-PAGE das frações insolúveis resultante da lise de bactérias expressando o fragmento da apirase (+) e um controle com plasmídeo sem o inserto (-). **B.** Western blot com anticorpo anti-hexahistidina (1:1000) das mesmas amostras.

Figura 39- Ensaio de purificação do fragmento da apirase de *S. mansoni* a partir de proteínas isoladas dos corpos de inclusão de *E. coli* e aplicadas em coluna de níquel. SDS-PAGE das frações corado com comassie-blue: **FT.** Fração não ligada (Flow-through). **W.** Fração de lavagem (20mM imidazol). **E.** fração de eluição (250 mM imidazol).

considerar esta perda como aceitável visto que ainda existia uma quantidade suficiente para realizar a imunização em coelhos.

Este fragmento purificado da apirase foi utilizado para a produção de um soro policlonal em coelhos. Primeiramente aproximadamente 50 ug de proteína purificada com igual volume de adjuvante de Freund completo foram injetados por via subcutanea em um coelho. Após 15 dias outra dose de 50 ug foi aplicada desta vez com igual volume de adjuvante de Freund incompleto.

Após 15 dias foi retirado sangue do coelho e extraiu-se o soro. Foi testada então a reatividade de anticorpos em um experimento de western blot usando-se no gel a proteína recombinante e revelando-se as membranas tanto com o soro imune como com o soro pre-imune (Figura 40). Podemos ver que há uma banda do tamanho esperado na membrana revelada com o soro imune e que não é possível detectar esta banda na membrana revelada com o soro pré imune. Este dado permite-nos afirmar que conseguimos produzir anticorpos reativos para nossa proteína.

Foi possível reconhecer na fração de tegumento solúvel em SDS 0,1 % um fragmento de aproximadamente 63 kDa via western blot utilizando o soro policlonal contra a proteína recombinante gerado em coelho (Figura 41). Há outra banda com uma massa acima de 110 kDa reconhecidas pelo soro policlonal. A abundancia relativa de nossa proteína não deve ser baixa visto que para conseguirmos um bom sinal na detecção precisamos de uma amostra relativamente bem concentrada (Ver materiais e métodos item 3.19).

Figura 40- Western blot da proteína recombinante revelado com soro imune de coelho anti-apirase recombinante e pre-imune (1:1000). A figura mostra que foi possível produzir um anticorpo com reatividade para a proteína.

figura 41-Anticorpo policlonal anti-apirase- recombinante reconhece apirase no tegumento. **A-** SDS-PAGE do tegumento de *S. mansoni* solubilizado em SDS 0,1% corado com comassie blue. **B.**Western blot revelado com anticorpo anti apirase recombinante (1:1.000).

5-Conclusões e perspectivas.

Este trabalho conseguiu atingir uma serie de objetivos que permitiram a obtenção de maiores conhecimento referentes ao parasita *S. mansoni* em relação a sua diversidade gênica e permitiu o estudo mais aprofundado de uma de suas proteínas.

Além disso foi validada uma técnica de construção de bibliotecas de ESTs que pode ser aplicada a qualquer eucarioto para a descoberta de novos genes.

5.1-Construção de bibliotecas de cDNA.

Foi possível realizar um aperfeiçoamento da técnica de construção de minibibliotecas de cDNA anteriormente descrito (Dias Neto et al., 1997 e 2000) agora com a utilização de primers híbridos consenso degenerados. Em nossos experimentos conseguimos obter bibliotecas com cerca de 400 seqüências diferentes com um nível de redundância aceitável. Isto representa um grande avanço em relação ao numero médio de 20 a 40 seqüências descritos anteriormente na literatura.

Para tanto foi necessário a alteração de alguns parâmetros como a introdução do *touchdown* no ciclo de PCR, a diluição do tampão de amostra e a diminuição da velocidade de rampa da transição dos passos de anelamento para extensão no ciclo de PCR. Tais modificações se mostraram criticas e contribuíram em muito para um aumento da diversidade de mensagens obtidas.

Outro fator que se revelou critico para a construção das bibliotecas de EST foi a qualidade do RNA mensageiro utilizado. Protocolos de extração que geraram um RNA mais puro permitiram a construção de uma biblioteca com um numero muito menor de

seqüências contaminantes, este fato se refletiu positivamente em todos os outros parâmetros medidos levando a um aumento significativo da eficiência da técnica.

Tais mudanças não levaram a uma mudança significativa em parâmetros desejáveis descritos anteriormente para esta técnica, como a amplificação preferencial do centro das mensagens e a normalização da abundancia relativa das mensagens.

Foi possível constatar que os primers CODEHOP não se comportaram conforme havia sido proposto por Rose et al., pois não ocorreu a estabilização prevista do anelamento do primer pela alça. Apesar disso foi possível verificar que a presença da alça era benéfica para a obtenção de diversidade nas seqüências amostradas, devido a um mecanismo ainda não compreendido.

Devido ao grande numero de seqüências amostradas foi possível realizar uma contribuição significativa ao banco existente de seqüências expressas de *S. mansoni*. Além disso foi possível conseguir novas seqüências de genes cuja função pode ser inferida devido a semelhança com proteínas já descritas de outros organismos e que podem ser futuramente estudadas para uma maior compreensão do parasita.

Esta parte do trabalho ainda gerou um artigo publicado em uma revista internacional (Fietto, DeMarco and Verjovski-Almeida., 2002) e um adendo a uma patente (Fietto et al., construction of ORESTES cDNA lnraries using Touch-down PCR, Addendum to patent application # 196,716 Instituto Ludwig de pesquisa para o câncer, 2000).

O sucesso dos experimentos aqui descritos incentivou nosso grupo a liderar um projeto, do qual o doutorando fez parte, envolvendo vários laboratórios para a produção em larga escala de seqüências ESTs. Este projeto produziu cerca de 160.000 seqüências ate

outubro de 2002 e permitirá uma aumento significativo da compreensão da biologia molecular do parasita.

Estas 160.000 ESTs produzidas permitirão que iniciemos um estudo em larga escala do nível de expressão relativa de milhares de genes ao longo do ciclo de vida do parasita com a utilização da tecnologia de microarray.

5.2-Clonagem da apirase de *S. mansoni*

Este trabalho descreveu a clonagem de um gene de apirase do parasita. Esta clonagem foi possível devido a ESTs geradas no projeto paralelo do qual o doutorando participou, isto reforça a importância da geração de ESTs para auxiliar a descoberta de novos genes. Além disso fica caracterizada a importância da construção de bibliotecas de EST por PCR de baixa stringência como método complementar as ESTs tradicionais, pois já existiam anteriormente ESTs para o gene da apirase que não podiam ser identificados por não terem regiões codificantes em sua extensão.

Uma vez obtida esta EST foi possível obter a sequência completa do gene da apirase contendo toda a sua região codificante. Não há muitas dúvidas sobre a função deste gene encontrado visto estarem presentes todos domínios conservados da família das apirases. Foram detectados motivos únicos em sua sequência indicando uma divergência deste gene em relação as apirases de vertebrados depositadas nos bancos públicos.

Foi possível detectar a presença do gene da apirase no genoma do parasita. Este dado exclui a possibilidade da sequência ser originária de outro organismo, como por exemplo do hospedeiro do parasita, uma possibilidade que sempre deve ser levada em conta quando obtemos sequência via PCR. Além disso fragmentos da sequência gerados por grupos

independentes foram encontradas nos bancos de dados públicos o que reforça ainda mais esta certeza.

A partir da sequência do RNA mensageiro obtida foi possível expressar a apirase de *S. mansoni* em *E. coli* que não apresentou atividade. Entretanto isto já era esperado visto a apirase ser uma proteína de membrana; muitas vezes procariotos não possuem a maquinaria celular adequada reconhecimento de sinais de proteínas de eucariotos para inserção em membrana. Além disso a apirase pode ser alvo de modificações pos-traducionais importantes para sua atividade, tais como glicosilação, que também não seriam reproduzidas em um sistema procarioto.

Apesar disso foi possível obter grandes quantidades de um fragmento da proteína, o que permitiu a geração de um soro policlonal em coelho. Este soro permitiu a identificação da enzima no tegumento do parasita o que reforça a hipótese que esta seria a apirase descrita anteriormente por Vasconcelos et al.(1993 e 1996)

Pretende-se ainda realizar a imunocalização desta proteína via microscopia confocal com o uso do anticorpo gerado e obter-se a sequência genômica completa para este gene. A expressão deste gene em um sistema eucarioto pode gerar uma proteína ativa que então poderá ter suas propriedades bioquímicas analisadas.

Por se localizar na superfície externa do parasita esta proteína pode ser alvo de drogas terapêuticas ou utilizado para imunização. Estudos adicionais sobre propriedade bioquímicas e antigênicas desta proteína poderão nos dizer a viabilidade desta aplicações.

6-Bibliografia

- Abath, F.G.C. and Werkhauser, R.P. (1996). The tegument of *Schistosoma mansoni*: functional and immunological features. *Parasite Immunol.* 18: 15-20.
- Adams, M.D., Kelley, J.M., Gocyene, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merrill, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. and Venter, J.C. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science.* 252: 1651-1656.
- Adams, M.D., Celniker, S.E., Holt, R.A. et al 2000. The genome sequence of *Drosophila melanogaster*. *Science* 287: 2185-2195.
- Angeli, V., Faveeuw, C., Delerive, P., Fontaine, J., Barriera, Y., Franchimont, N., Staels, B., Capron, M. and Trottein, F. (2001). *Schistosoma mansoni* induces the synthesis of IL-6 in pulmonary microvascular endothelial cells: role of IL-6 in the control of lung eosinophilia during infection. *Eur. J. Immunol.*: 31:2751-61.
- Argiro, L.L., Kohlstadt, S.S., Henri, S.S., Dessein, H.H., Matabiau, V.V., Paris, P.P., Bourgois, A.A. and Dessein, A.J. (2000) Identification of a candidate vaccine peptide on the 37 kDa *Schistosoma mansoni* GAPDH. *Vaccine* 18: 2039-2048
- Biederbick, A., Rose, S. and Elsasser, H.P. (1999). A human intracellular apyrase-like protein, LALP70, localizes to lysosomal/autophagic vacuoles. *J. Cell. Sci.* 112: 2473-2484
- Boros, D.L. (1999). T helper cell populations, cytokine dynamics, and pathology of the schistosome egg granuloma. *Microbes Infect.* 1: 511-516
- Brindley, P.J., Kalinna, B.H., Dalton, J.P., Day, S.R., Wong, J.Y.M., Smythe, M.L. and McManus, D.P. (1997). Proteolytic degradation of host hemoglobin by schistosomes. *Mol. Biochem. Parasitol.*: 89: 1-9.
- Brouwers, J.F., Skelly, P.J., Van Golde, L.M. and Tielens A.G. (1999). Studies on phospholipid turnover argue against sloughing of tegumental membranes in adult *Schistosoma mansoni*. *Parasitology.* 119: 287-294.
- Butterworth AE, Capron M, Cordingley JS, Dalton PR, Dunne DW, Kariuki HC, Kimani G, Koech D, Mugambi M, Ouma JH, et al. (1985) Immunity after treatment of human

- schistosomiasis mansonii. II. Identification of resistant individuals, and analysis of their immune responses. *Trans. R. Soc. Trop. Med. Hyg.* 79:393-408
- Capron, A., Capron, M., Dombrowicz, D. and Riveau, G. (2001). Vaccin strategies against schistosomiasis: from concepts to clinical trials. *Int. Arch. Allergy Immunol.* 24: 9-15.
- Chadwick, B.P. and Frischauf, A.M. (1998). The CD39-like gene family: identification of three new human members (CD39L2, CD39L3, and CD39L4), their murine homologues, and a member of the gene family from *Drosophila melanogaster*. *Genomics* 50(3):357-367
- Chen, L., Rao, K.V., He, Y.X. and Ramaswamy, K. (2002). Skin-stage schistosomula of *Schistosoma mansoni* produce an apoptosis-inducing factor that can cause apoptosis of T cells. *J. Biol. Chem.* 277: 34329-34335.
- Craig, S.P. 3rd, Muralidhar, M.G., McKerrow, J.H. and Wang, C.C. (1989). Evidence for a class of very small introns in the gene for hypoxanthine-guanine phosphoribosyltransferase in *Schistosoma mansoni*. *Nucleic Acids Res.* 17:1635-1647.
- Cummings, R. D. and Nyame, A.K. (1999). Schistosome glycoconjugates. *Biochim. Biophys. Acta* 1455: 363-374.
- Da'dara, A.A., Skelly, P.J., Wang, M.M. and Harn, D.A. (2001). Immunization with plasmid DNA encoding the integral membrane protein, Sm23, elicits a protective immune response against schistosome infection in mice. *Vaccine* 20: 359-369.
- Daniel, J.L., Dangelmaier, C., Jin, J., Ashby, B., Smith, J.B. and Kunapuli, S.P. (1998). Molecular basis for ADP-induced platelet activation. I. Evidence for three distinct ADP receptors on human platelets. *J. Biol. Chem.* 273: 2024-2029.
- Davies, S.J, Shoemaker, C.B. and Pearce E.J. (1998). A divergent member of the transforming growth factor β receptor family from *Schistosoma mansoni* is expressed on the parasite surface membrane. *J. Biol. Chem.* 18: 11234-11240.
- Davis, R.E., Hardwick, C., Tavernier, P., Hodgson, S. and Singh, H. (1995). RNA trans-splicing in flatworms. Analysis of trans-spliced mRNAs and genes in the human parasite, *Schistosoma mansoni*. *J. Biol. Chem.* 270: 21813-21819.

- de Souza S.J., Camargo A.A., Briones, M.R., Costa, F.F., Nagai, M.A., Verjovski-Almeida, S., Zago, M.A., Andrade, L.E., Carrer, H., El-Dorry, H.F., Espreafico, E.M., Habr-Gama, A., et al. (2000) Identification of human chromosome 22 transcribed sequences with ORF expressed sequence tags. *Proc Natl Acad Sci U S A*. Nov 7;97(23):12690-3.
- Deprès, L., Imbert-Establet, D. and Monnerot, M. (1993). Molecular characterization of mitochondrial DNA provides evidence for the recent introduction of *Schistosoma mansoni* into America. *Mol. Biochem. Parasitol.* 60: 221-230.
- Dias Neto, E., Harrop, R., Correa-Oliveira, R., Wilson, R.A., Pena, S.D.J. and Simpson, A.J.G. (1997). Minilibraries constructed from cDNA generated by arbitrarily primed RT-PCR: an alternative to normalized libraries for generation of EST from nanogram quantities of mRNA. *Gene*. 186 : 135-142.
- 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. et al. (2000) Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc. Natl. Acad. Sci. U.S.A.* 97: 3491-3496.
- Doenhoff, M.J., Kusel, J.R., Coles, G.C. and Cioli, D. (2002). Resistance of *Schistosoma mansoni* to praziquantel: is there a problem? *Trans. R. Soc. Trop. Med. Hyg.* 96:465-469.
- Ellinghaus, P., Badehorn, D., Blümer, R., Becker, K. and Seedorf, U. (1999) Increased efficiency of arbitrarily primed PCR by prolonged ramp times. *BioTechniques* 26: 626-630.
- Enjyoji, K., Sevigny, J., Lin, Y., Frenette, P.S., Christie, P.D., Esch, J.S. 2nd, Imai, M., Edelberg, J.M., Rayburn, H, Lech, M., Beeler, D.L., Csizmadia, E., Wagner, D.D., Robson, S.C. and Rosenberg, R.D.(1999). Targeted disruption of cd39/ATP diphosphohydrolase results in disordered hemostasis and thromboregulation. *Nat. Med.* 5: 1010-1017.
- Esch, J.S. 2nd, Sevigny, J., Kaczmarek, E., Siegel, J.B., Imai, M., Koziak, K., Beaudoin, A.R. and Robson, S.C. (1999). Structural elements and limited proteolysis of CD39 influence ATP diphosphohydrolase activity. *Biochemistry* 38: 2248-2258.
- Ewing, B., Hillier, L., Wendl, M.C. and Green, P. (1998) Base-calling of automated sequencer traces using Phred I: Accuracy assessment. *Genome Res.*, 8, 175-185.

- Fallon, P.G., Fookes, R.E. and Wharton, G.A. (1996). Temporal differences in praziquantel- and oxamniquine-induced tegumental damage to adult *Schistosoma mansoni*: implications for drug-antibody synergy. *Parasitology* 112 :47-58
- Fallon, P.G., Smith, P. and Dunne, D.W. (1998). Type 1 and type 2 cytokine-producing mouse CD4⁺ and CD8⁺ T cells in acute *Schistosoma mansoni* infection. *Eur. J. Immunol.* 28: 1408-1416.
- Fietto, J.L.R., DeMarco R and Verjovski-Almeida, S. (2002). Use of degenerate primers and touchdown PCR for construction of cDNA libraries. *Biotechniques* 32(6):1404-1411.
- File, S. (1995). Interaction of schistosome eggs with vascular endothelium. *J Parasitol* 81: 234-238
- Franco, G.R., Adams, M.D., Soares, M.B., Simpson, A.J., Venter, J.C. and Pena, S.D. (1995). Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. *Gene* 152: 141-147.
- Franco, G.R., Rabelo, E.M., Azevedo, V., Pena, H.B., Ortega, J.M., Santos, T.M., Meira, W.S., Rodrigues, N.A., Dias, C.M., Harrop, R., Wilson, A., Saber, M., Abdel-Hamid, H., Faria, M.S., Margutti, M.E., Parra, J.C. and Pena, S.D. (1997). Evaluation of cDNA libraries from different developmental stages of *Schistosoma mansoni* for production of expressed sequence tags (ESTs). *DNA Res* 4: 231-240.
- Fulford, A.J.C., Webster, M., Ouma, J.H., Kimani, G. and Dunne, D.W. (1998). Puberty and age-related changes in susceptibility to schistosome infection. *Parasitol. Today* 14: 23-26.
- FUNASA (2002). Situação da prevenção e controle das doenças transmissíveis no Brasil.
- Gao, X.D., Kaigorodov, V. and Jigami, Y. (1999). YND1, a homologue of GDA1, encodes membrane-bound apyrase required for Golgi N- and O-glycosylation in *Saccharomyces cerevisiae*. *J Biol Chem* 274: 21450-21456
- Gayle, R.B. 3rd, Maliszewski, C.R., Gimpel, S.D., Schoenborn, M.A., Caspary, R.G., Richards, C., Brasel, K., Price, V., Drosopoulos, J.H., Islam, N., Alyonycheva, T.N., Broekman, M.J. and Marcus, A.J. (1998). Inhibition of platelet function by recombinant soluble ecto-ADPase/CD39. *J. Clin. Invest.* 101: 1851-1859

- Gear, A.R., Suttitanamongkol, S., Viisoreanu, D., Polanowska-Grabowska, R.K., Raha, S. and Camerini, D. (2001). Adenosine diphosphate strongly potentiates the ability of the chemokines MDC, TARC, and SDF-1 to stimulate platelet function. *Blood* 97: 937-945.
- Girolomoni, G., Santantonio, M.L., Pastore, S., Bergstresser, P.R., Giannetti and A., Cruz, P.D. Jr. (1993). Epidermal Langerhans cells are resistant to the permeabilizing effects of extracellular ATP: in vitro evidence supporting a protective role of membrane ATPase. *J. Invest. Dermatol.* 100:282-287
- Goffeau, A., Aert, R., Agostini-Carbone, M.L. et al., (1997). The yeast genome directory. *Nature* 387: supl 5-6
- Grevelding, C.G., Sommer, G. and Kunz W. (1997). Female-specific gene expression in *Schistosoma mansoni* is regulated by pairing. *Parasitology* 115: 635-640.
- Grinthal, A. and Guidotti, G. (2000). Substitution of His59 converts CD39 apyrase into an ADPase in a quaternary structure dependent manner. *Biochemistry.* 39: 9-16.
- Grinthal, A. and Guidotti, G. (2002). Transmembrane domains confer different substrate specificities and adenosine diphosphate hydrolysis mechanisms on CD39, CD39L1, and chimeras. *Biochemistry* 41: 1947-1956.
- Haberl, B., Kalbe, M., Fuchs, H., Strobel, M., Schmalfuss, G. and Haas, W.(1995). *Schistosoma mansoni* and *S. haematobium*: miracidial host-finding behaviour is stimulated by macromolecules.*Int J Parasitol.* 25: 551-560.
- Hagan, P., Ndhlovu, P.D. and Dunne, D.W. (1998). Schistosome immunology: more questions than answers. *Parasitol. Today* 14: 407-412.
- Halton, D.W. (1997) Nutritional adaptations to parasitism within the platyhelminthes. *Int. J. Parasitol.* 27: 693-704.
- Handa, M. and Guidotti, G. (1996). Purification and cloning of a soluble ATP-diphosphohydrolase (apyrase) from potato tubers (*Solanum tuberosum*). *Biochem. Biophys. Res. Commun.* 218: 916-23.

- Heine, P., Braun, N., Seigny, J., Robson, S.C., Servos, J. and Zimmermann, H. (2001). The C-terminal cysteine-rich region dictates specific catalytic properties in chimeras of the ectonucleotidases NTPDase1 and NTPDase2. *Eur. J. Biochem.* 268:364-373.
- Henerariu, O., Heerema, N.A., Dlouhy, S.R., Vance, G.H. and Vogt, P.H. (1997) Multiplex PCR: Critical parameters and step-by-step protocol. *BioTechniques*, **23**, 504-511.
- Hoffmann, K.F., James, S.L., Cheever, A.W. and Wynn, T.A. (1999). Studies with double cytokine-deficient mice reveal that highly polarized Th1- and Th2-type cytokine and antibody responses contribute equally to vaccine-induced immunity to *Schistosoma mansoni*. *J. immunol.* : 163: 927-938.
- Ismail, M., Metwally, A., Farghaly, A., Bruce, J., Tao, L.F. and Bennett, J.L. (1996). Characterization of isolates of *Schistosoma mansoni* from Egyptian villagers that tolerate high doses of praziquantel. *Am. J. Trop. Med. Hyg.* 55:214-218.
- Kalbe, M., Haberl, B. and Haas, W. (1996). *Schistosoma mansoni* miracidial host-finding: Species specificity of an Egyptian strain. *Parasitol. Res.* 82 : 8-13.
- Kan, Z., Rouchka, E.C., Gish, W.R. and States, D.J. (2001). Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res* 11:889-900
- Khalife, J., Cêtre, C., Pierrot, C. and Capron M. (2000). Mechanisms of resistance to *S. mansoni* infection: the rat model. *Parasitol. Int.* 49: 339-345.
- Khalil, S.B. and Mansour, N.S. (1995). Worm development in hamsters infected with unisex and cross-mated *Schistosoma mansoni* and *Schistosoma haematobium*. *J. Parasitol.* 81: 8-11.
- Kirley, T.L., Yang, F. and Ivanenkov, V.V. (2001). Site-directed mutagenesis of human nucleoside triphosphate diphosphohydrolase 3: the importance of conserved glycine residues and the identification of additional conserved protein motifs in eNTPDases. *Arch. Biochem. Biophys.* 395: 94-102.
- Köhler, P. (2001). The biochemical basis of antihelminthic action and resistance. *Int. j. parasitol.* 31: 336-345.
- Laliberté, J.F. and Beaudoine, A.R. (1983) Kinetic effects of Ca^{2+} e Mg^{2+} on ATP hydrolysis by the purified ATP diphosphohydrolase. *J. Biol. Chem.* **257**: 3869-3874.

- Liang YS, Coles GC, Dai JR, Zhu YC, Doenhoff MJ. (2001) Biological characteristics of praziquantel-resistant and -susceptible isolates of *Schistosoma mansoni*. *Ann. Trop. Med. Parasitol.* 95:715-723.
- Macke, J., Smallwood, P. and Nathans, J. (1996) Human retina cDNA randomly primed sublibrary. GenBank accession no. W26664 – W29128.
- Markell, E.K., Voge, M. and John, D.T. (1992). Medical parasitology. W.B. Saunders company. 7th edition. Pag: 199-215.
- Marshall, I., (1995). Schistosomiasis. In *Medical Parasitology*. IRL press. 1st edition. Pag: 191-208.
- McKerrow, J.H. and Salter, J. (2002). Invasion of skin by *Schistosoma mansoni*. *Trends parasitol.* 18: 193-195.
- Miller, K.G., Emerson M.D., McManus J.R. and Rand J.B. (2000). RIC-8 (Synembryn): a novel conserved protein that is required for G(q)alpha signaling in the *C. elegans* nervous system. *Neuron* 27: 289-299.
- Miller K.G. and Rand J.B. (2000). A role for RIC-8 (Synembryn) and GOA-1 (G(o)alpha) in regulating a subset of centrosome movements during early embryogenesis in *Caenorhabditis elegans*. *Genetics*. 156: 1649-1660.
- Morales-Montero, J., Newhouse, E. Mohamed, F., Baghdadi, A. and Damian, R.T. (2001). Altered levels of hypothalamic-pituitary-adrenocortical axis hormones in baboons and mice during the course of infection with *Schistosoma mansoni*. *J. Infect. Dis.* 183: 313-320.
- Nakazawa, M., Fantappie, M.R. Freeman jr, G.L. Eloi-Santos, S., Olsen, N.J., Kovacs, W.J., Secor, W.E. and Colley, D.G. (1997). *Schistosoma mansoni*: susceptibility differences between male and female mice can be mediated by testosterone during early infection. *Exp. Parasitol.* 85: 233-240.
- Nessim, N.G., Hassan, S.I., William, S. and el-Baz H. (2000). Effect of the broad spectrum anthelmintic drug flubendazole upon *Schistosoma mansoni* experimentally infected mice. *Arzneimittelforschung* 50: 1129-1133

- Olds G.R., King C., Hewlett J., Olveda R., Wu G., Ouma J., Peters P., McGarvey S., Odhiambo O., Koech D., Liu C.Y. et al., (1999). Double-blind placebo-controlled study of concurrent administration of albendazole and praziquantel in schoolchildren with schistosomiasis and geohelminths. *J Infect Dis.* 179: 996-1003.
- Pancera, C.F., Alves, A.L., Paschoalotti, M.A. and Chieffi, P.P. (1997). Effect of wide spectrum anti-helminthic drugs upon *Schistosoma mansoni* experimentally infected mice. *Rev Inst Med. Trop. Sao Paulo.* 39: 159-163.
- Pearce, E.J., Basch, P.F. and Sher, A. (1986). Evidence that the reduced surface antigenicity of developing *Schistosoma mansoni* schistosomula is due to antigen shedding rather than host molecule acquisition. *Parasite Immunol.* 8: 79-94.
- Popiel I, Basch PF. (1984). Reproductive development of female *Schistosoma mansoni* (Digenea: Schistosomatidae) following bisexual pairing of worms and worm segments. *J. Exp. Zool.* 232: 141-150.
- Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M. and Henikoff, S. (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res.* 26: 1628-1635.
- Roux, K.H., (1994). Using mismatched primer-template pairs in touchdown PCR. *BioTechniques* 16: 812-814.
- Reynolds S.R., Dahl C.E. and Harn D.A (1994). T and B epitope determination and analysis of multiple antigenic peptides for the *Schistosoma mansoni* experimental vaccine triose-phosphate isomerase. *J. Immunol.* 152:193-200
- Salzet, M., Capron, A. and Stefano, G. B. (2000). Molecular crosstalk in host-parasite relationships: Schistosome- and Leech-host interactions. *Parasitology today* 16: 536- 540.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning. A laboratory manual. Cold spring harbor laboratory press. Second edition.
- 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. (1999). Analysis of the gene expression profile

- of *Schistosoma mansoni* cercariae using the expressed sequence tag approach. *Mol Biochem Parasitol* 103: 79-97.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1997). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74: 5463.
 - Saribas, A.S., Lustig, K.D., Zhang, X. and Weisman, G.A. (1993). Extracellular ATP reversibly increases the plasma membrane permeability of transformed mouse fibroblasts to large macromolecules. *Anal. Biochem.* 209: 45-52.
 - Schuler, G.D., Boguski, M.S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodriguez-Tome, P., Aggarwal, A., Bajorek, E. et al. (1996). A gene map of the human genome. *Science* 274:540-546
 - Sarkis, J.J.F. and Salto, C (1991). Characterization of a synaptosomal ATP diphosphohydrolase from the electric organ of *Torpedo marmorata*. *Brain. Res. Bull.* 26: 871-876.
 - Simpson, A.J.G., Sher, A. and McCutchan (1982). The genome of *Schistosoma mansoni*: isolation of DNA, its size, bases and repetitive sequences. *Mol. Biochem. Parasitol.* 6: 125-137.
 - Skelly, P.J. and Shoemaker, C.B. (1996). Rapid appearance and asymmetric distribution of glucose transporter SGLT4 at the apical surface of intramammalian-stage *Schistosoma mansoni*. *Proc. Natl. Acad. Sci. USA* 93: 3642-3646.
 - Skelly PJ, Pfeiffer R, Verrey F, Shoemaker CB. (1999). SPRM1lc, a heterodimeric amino acid permease light chain of the human parasitic platyhelminth, *Schistosoma mansoni*. *Parasitology* 119: 569-76.
 - Smith, T.M. and Kirley, T.L. (1999). Site-directed mutagenesis of a human brain ecto-ATPase: evidence that E-type ATPases are related to the actin/heat shock 70/sugar kinase superfamily. *Biochemistry* 38: 321-328.
 - Smith, T.M., Lewis Carl S.A. and Kirley T.L.(1999). Mutagenesis of two conserved tryptophan residues of the E-type ATPases: inactivation and conversion of an ecto-ATPase to an ecto-NTPase. *Biochemistry.* 38: 5849-5857.

- Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L. and Efstratiadis, A. (1994). Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci. USA* 91: 9228-9232.
- SUCEN, (2001). Esquistossomose- informações para profissionais da saúde.
- The *C.elegans* Sequencing Consortium. 1998. Genome sequence of nematode *C. elegans*: A platform for investigating biology. *Science* 282: 2012-2018.
- Van Lieshout, L., Polderman, A.M. and Deelder, A.M. (2000). Immunodiagnosis of schistosomiasis by determination of the circulating antigens CAA and CCA, in particular in individuals with recent or light infections. *Acta tropica* 77: 69-80.
- Vasconcelos E.G., Nascimento P.S., Meirelles M.N., Verjovski-Almeida S. and Ferreira S.T. (1993) Characterization and localization of an ATP-diphosphohydrolase on the external surface of the tegument of *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* 58: 205-14.
- Vasconcelos, E.G., Ferreira, S.T., Carvalho, T.M.U., Souza, W., Kettlun, A.M., Mancilla, M., Valenzuela, M.A. and Verjovski-Almeida, S. (1996). Partial purification and immunohistochemical localization of ATP diphosphohydrolase from *Schistosoma mansoni*. *J. Biol. Chem.* 271:22139-22145.
- Veronesi, R. (1972). Doenças infecciosas e parasitárias. Guanabara Koogan, 5ª edição, pag 870-892.
- Xie, H., Zhu, W., Wasserman, A., Grebinskiy, V., Olson, A. and Mintz, L. (2002). Computational Analysis of Alternative Splicing Using EST Tissue Information. *Genomics* 80: 326-330.
- WHO fact sheet # 115 (1996).
- WHO (1991). PEEM guidelines. WHO/CWS/91.2, 2nd ed. Annex 1.
- Wilson, R.A. and Coulson P.S. (1999). Strategies for a schistosome vaccine: can we manipulate the immune response effectively. *Microbes Infect.* 1: 535-543.
- Zheng Y., Wong M.L., Alberts B. and Mitchison T. (1995). Nucleation of microtubule assembly by a gamma-tubulin-containing ring complex. *Nature* 378:578-83.