

UNIVERSIDADE FEDERAL DE OURO PRETO INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS DEPARTAMENTO DE CIÊNCIAS BIOLÓGICAS PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

Caracterização inicial dos constituintes da maquinaria de silenciamento gênico mediada por microRNAs em *Schistosoma mansoni*

AUTOR: MATHEUS DE SOUZA GOMES ORIENTADORA: PROF^a. Dr^a. RENATA GUERRA DE SÁ CO-ORIENTADOR: PROF. Dr. ELIO HIDEO BABÁ

Dissertação submetida ao programa de Pós-Graduação do Departamento de Ciências Biológicas da Universidade Federal de Ouro Preto, como parte integrante dos requisitos para obtenção do título de mestre em Ciências Biológicas, área de concentração Biologia Molecular.

Ouro Preto - MG, janeiro de 2008

G633c Gomes, Matheus de Souza. Caracterização inicial dos constituintes da maquinaria de silenciamento gênico mediada por microRNAs em Schistosoma mansoni [manuscrito] / Matheus de Souza Gomes. - 2008. 112f.: il., color; graf.; tabs.; mapas. Orientadora: Profa. Dra. Renata Guerra de Sá Co-orientador: Prof. Dr. Elio Hideo Babá. Dissertação (Mestrado) - Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Biológicas. Núcleo de Pesquisas em Ciências Biológicas. Área de concentração: Biologia molecular. 1. Schistosoma mansoni - Teses. 2. Código genético - Teses. 3. Reação em cadeia de polimerase - Teses. I. Universidade Federal de Ouro Preto. II. Título. CDU: 616.995.122

Catalogação: sisbin@sisbin.ufop.br

Este trabalho foi desenvolvido no Laboratório de Bioquímica e Biologia Molecular ICEB/NUPEB/UFOP, Centro de Pesquisa René-Rachou, Fiocruz-MG, e Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, com apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Dedico esta dissertação aos meus pais, Luiz e Rosânia, aos meus irmãos, Marcos, Júlia e Daniela, aos meus avós, e familiares e a minha querida Carol. Agradeço a todos pelo apoio e compreensão, durante todo meu percurso, e espero ainda mais, pois muito caminho tenho ainda a percorrer.

"A felicidade não depende do que nos falta, mas do bom uso que fazemos do que temos."

- Thomas Hardy

Agradecimentos

A Deus;

Aos meus pais, Luiz e Rosânia, por serem meus exemplos de vida, humildade e persistência, sempre indicando o caminho certo a trilhar, fazendo o bem;

À minhas queridas irmãs, Daniela e Júlia, pelo apoio e amor incondicionais, sempre mostrando um ombro amigo para me acolher.

Ao meu irmão Marcos pela grande amizade, amor, companheirismo e, sobretudo, por me ajudar e incentivar nos momentos mais difíceis.

Aos demais familiares, pelo carinho, preocupação e por me apoiar sempre;

Ao meu amor Carol, minha futura esposa, pela dedicação, compreensão, auxílio e, sobretudo, por ter paciência e sabedoria, contribuindo imensuravelmente para realização deste trabalho;

À família da Carol, seus pais e sua irmã Clarissa pelo incentivo e em especial Dona Lili pela paciência e acolhimento, sendo importantes para finalização deste trabalho;

À Prof.^a Dr.^a Renata Guerra de Sá, pelos ensinamentos práticos e teóricos e pelo exemplo de dedicação e profissionalismo. Pela minha formação como pesquisador e como aprendiz de professor. E principalmente, por mostrar que, para conseguir, primeiro, temos que sonhar.

Ao professor Dr. Elio Hideo Babá, pelo exemplo de pesquisador e professor. Pelos ensinamentos na pesquisa e na vida. Por sempre se mostrar disposto a ajudar. E, sobretudo, por ter aberto as portas do laboratório para mim, me acolhendo como parte integrante desta família LBBM.

Aos professores Lucinha e Elísio pelo convívio agradável e aprendizado científico transmitido.

Aos amigos da equipe do Laboratório de Bioquímica e Biologia Molecular (LBBM) da UFOP, Daniel, Eneida, Gustavo, Leonardo, Leandro, Leandro Gonzaga, Michelini. Míriam, Naiara, Natália, Nayara, Roberta, Robs, Thiago, Vanessa, pelo auxílio e momentos compartilhados;

Ao Roenick, Letícia, Cássio e Nilza, pelas trocas de conselhos, pela amizade, disponibilidade e, sobretudo, por toda ajuda nos serviços laboratoriais e extras laboratoriais.

Ao Ezequiel pelo auxílio no trabalho e pelas boas risadas;

À Helaine pelas horas de trabalho compartilhadas, pela ajuda nos problemas encontrados e pelos ensinamentos na realização deste trabalho;

À Claudinha pelo exemplo de profissionalismo, persistência, e organização, auxilando imensamente em minha formação como pesquisador.

Aos professores Liana Janotti e Omar Carvalho, por gentilmente fornecer junto ao moluscário os parasitos para a realização deste trabalho.

Ao moluscário do CPqRR em especial a Delza, Dilce e Sueleny pela disponibilização dos parasitos, e paciência em responder as perguntas.

Ao professor Vanderlei Rodrigues e sua equipe da Faculdade de Medicina de Ribeirão Preto (USP) pela oportunidade e grande colaboração nos experimentos ,em especial, Fernanda, pelo auxílio na bancada, pelo acolhimento e pela amizade.

Aos colegas de pós-graduação, em especial, Carlito, Fabiana, Mônica, Ramon e Roberta pelo companheirismo e pelos momentos de troca de conhecimento e risadas.

A todos do Laboratório de Imunoparasitologia da UFOP pelo grande suporte e auxílio para meu crescimento científico.

A todos os professores do NUPEB, pela receptividade em seus Laboratórios;

A todos os amigos dos laboratórios do NUPEB pela saudável convivência contribuindo para a realização deste trabalho;

À Cida, secretária prestativa, atenciosa e alegre sempre pronta para ajudar;

A República Tira Mágoa, moradores e ex-moradores, pela convivência, apoio e, sobretudo, pelos momentos de descontração, aprendizado de vida, amizade e cumplicidade.

Às seguintes instituições que colaboraram com recursos financeiros para a realização deste trabalho: CNPq, CAPES e FAPEMIG.

E por todos aqueles que participaram direta ou indiretamente na realização deste trabalho.

Resumo	1
Abstract	2
Lista de abreviaturas	3
Lista de Figuras e tabelas	5
1. Introdução	15
1.1 Aspectos gerais da Esquistossomose	16
1.2 O S. mansoni	16
1.3 Genoma do parasito	18
1.4 Regulação da expressão gênica em S. mansoni	20
1.5 Silenciamento gênico mediado por microRNA	22
2. Objetivos	.29
2.1 Objetivos Gerais	.30
2.1 Objetivos Específicos	30
3. Materiais e Métodos	31
3 1 Parasitos	32
3.1.1 Esquistossômulos transformados mecanicamente	32
3.2 Anotação da via de miRNA	32
3.3. Alinhamentos entre següências e Análise filogenética	34
3.4 Extração do DNA genômico	36
3.5 Extração de DNA total para PT DCP	30
2.6 Extração do DNA total para DCP quantitativo om tompo real	37
2.7 Varificação de qualidade de DNA total extraído	30
2.9 Quantificação da DNA a DNA	20
2.0 Oligonuelogidades iniciadores (Primers)	39
2.10 DCD (Decesion de Delimereses em Cadeia)	40
2.10 1 DCD conversional	40
2.10.2 PT DCD	40
3.10.2 KI-PCK	41
3.10.3 PCK quantitaviva em Tempo Real	41
3.11 Analise Estatistica.	42
3.12 Verificação da qualidade do DNA ou cDNA	42
3.13 Purificação do produto de PCR	42
3.13.1 FREEZE SQUEZE	42
3.14 Clonagem	43
3.14.1 Reação de Ligação	43
3.14.2 Preparo de células competentes	43
3.14.3 Transformação Bacteriana	44
3.14.4 PCR de Colônia	44
3.14.5 Minipreparação	44
3.14.4 PCR de Colônia	44
3.15 Sequenciamento	46
3.16 Análise computacional das seqüências	46
3.16.1 Seqüência consenso	46
3.16.2 <i>Splicing</i> alternativo	46
3.17 V Predição de miRNA	46
4 Resultados	48
4.1 Reconstituição in silico da via de miRNA	49
4.1.1 Análise in silico dos constituintes da via de miRNA	49
4.1.2 Análise das ESTs	61
4.1.3 Análise Filogenética das RNAse III e Argonautas	67
4.1.4 Análise dos domínios conservados PIWI e endoND	70

4.1.5 Análise de SmAgo4	71
4.1.6 miRNAs preditos	
4.2 Clonagem e sequenciamento parcial da SmDicer1	
4.3 Clonagem e seqüenciamento parcial de SmAgo5	
4.4 Splicing Alternativo	
4.5 Análise da expressão dos genes que codificam SmDicer1 e SmAgo2/3	/4 utilizando
qRT-PCR	
5. Discussão	
5.1 Família das RNAse III	87
5.2 Família das Argonautas	88
5.3 Outras proteínas envolvidas na via de miRNA	
5.4 MiRNAs preditos	
6. Conclusão	
7. Bibliografia	
8. Anexo	108
8.1 Anexo 1	
8.2 Anexo 2	110
8.3 Anexo 3	111
8.4 Anexo 4	

O silenciamento de RNA refere-se a uma série de processos nucleares e citoplasmáticos envolvidos na regulação da expressão gênica pós-transcricional ou silenciamento pós-transcricional (PTGS), degradando ou silenciando seqüências específicas de mRNA. O mecanismo é conservado em animais incluindo alguns pontos distintos e vários pontos coincidentes, envolvidos tanto no silenciamento de pequenos RNAs por via endógena quanto por via exógena. O pequeno RNA melhor caracterizado é o microRNA (miRNAs), que predominantemente silencia genes através de uma regulação pós-transcricional. Neste trabalho, utilizamos bioinformática para identificar possíveis genes envolvidos na via de silenciamento gênico pós-transcricional mediado por miRNA em *S. mansoni*. Nós pesquisamos os bancos de dados do genoma e do transcriptoma de *S. mansoni*, com as seqüências de aminoácidos proteínas ortólogas relacionadas à via de miRNA. Identificamos o total de 13 prováveis proteínas envolvidas na via miRNA no parasito.

Em seguida, o níveis de *SmDicer1* e *SmAgo 2, 3 e 4* foram identificados por qRT-PCR utilizando cercárias, vermes adultos, ovos e esquistossômulos com os seguintes períodos de cultivo in vitro 3,5; 8,5; 18,5; 24; 48 e 72 horas. Este resultado demonstrou que, os dois genes são diferencialmente expressos através do período de diferenciação cercária-esquistossômulo, tendo níveis similares em ovos e vermes adultos. Em cercárias *SmDicer1* tem baixa expressão. Estes resultados sugerem que esta via é um importante mecanismo de regulação da expressão gênica neste parasita. Futuros experimentos são necessários para provar esta hipótese.

Abstract

RNA silencing refers to a series of nuclear and cytoplasmatic processes involved in the post-transcriptional regulation of gene expression or post-transcriptional gene silencing (PTGS), either by sequence-specific mRNA degradation or by translational arrest. The mechanism is conserved in animals and includes at least distinct pathways with several overlapping points, involved on silencing of endogenous or exogenous sequences. The best characterized small RNA is microRNAs (miRNAs), which predominantly makes genes silencing through post-transcriptional mechanisms. In this work we used bioinformatics approach to identity gene products in parasitic trematode *S. mansoni* with sharing similarities with enzymes involved in the post-transcriptional gene silencing mediated by miRNA pathway. We searched the *S. mansoni* genome and transcriptome databases, with the amino acid sequences of othologs proteins related to the miRNA pathway. We identified the total of 13 putative proteins involved in miRNA pathway in the parasite.

Next, the levels of *SmDicer1* and *SmAgo 2, 3* and *4* were identified by qRT-PCR using cercariae, adult worms, eggs and schistosomula with the following times of in vitro cultivation 3,5; 8,5; 18,5; 24; 48 and 72 hours. This results shown that, the two genes are differentially expressed through cercariae-schistosomula differenciation, and have similar levels in eggs, and adult worms. In cercariae *SmDicer1* is down-regulated. These results suggest that this pathway is an important mechanism for the regulation of gene expression in this parasite. Future experiments are needed to prove this hypothesis.

Lista de Abreviaturas

BSA	 Albumina sérica bovina
cDNA	 DNA complementar
DNA	 Ácido desorribonucleico
RNA	 Ácido ribonucleico
DEPC	 Dietilpirocarbonato
dNTP	 Deoxinucleotídeo trifosfato (N= A, C, G
	ou T)
MOPS	 [Ácido 3-(N-Mofolino)Propanosulfônico]
mRNA	 RNA mensageiro
BLAST	 Basic Local Alignment and Search Tool
CDD	 Conserved domain database
CTAB	 Brometo de cetil-trimetil amônio
Tris	 Tris-hidroximetilaminometano
EST	 Etiquetas de seqüências expressas
IPTG	 Isopropylthio-β-galactoside
LB	 Meio de cultura Luria Bertani
PCR	 Reação em Cadeia da Polimerase
PB	 Pares de Bases
Open	 Fase Aberta de Leitura
Reading	
Frame	
RT-PCR	 Transcriptase Reversa – Reação
	Polimerásica em Cadeia
Q.S.P	 Quantidade suficiente para
NaCl	 Cloreto de Sódio
MgCl ₂	 Cloreto de Magnésio
NaOAc	 Acetato de sódio
MgSO ₄	 Sulfato de magnésio
EDTA	 Ácido etilenodiaminotetracético
CaCl ₂	 Cloreto de cálcio
SDS	 Sodio Dodecil Sulfato

NaOH	 Hidróxido de sódio
pН	 Potencial hidrogeniônico
miRNA	 MicroRNA
siRNA	 Small interfering RNA
WHO	 World Health Organization –
	Organização Mundial de Saúde
CNPq	 Conselho Nacional de Desenvolvimento
	Científico e Tecnológico
FAPEMIG	 Fundação de Amparo a Pesquisa de
	Minas Gerais
NCBI	 National Center for Biotechnology
	Information

Figura 1: Ciclo de vida do *S. mansoni*. Dentro do seu hospedeiro vertebrado, as formas larvais (esquistossômulo, B) dão origem a parasitos sexualmente maduros (verme adulto, A), os quais se acasalam e produzem ovos (C) que são liberados para o meio aquático. Os ovos eclodem (D) e liberam os miracídios (E), que penetram nos caramujos, dando origem a numerosos esporocistos (F). Os esporocistos geram cercárias (G) que saem do caramujo e são liberadas na água, as quais são as formas infectantes para o hospedeiro vertebrado (Adaptado de Wilson, 1979).

Figura 2: Expressão gênica estágio-específicas em *S. mansoni*. Genes altamente expressos estágio-específicos baseado nos projetos transcriptoma do parasito. Adaptado de Hu e cols., 2004.

Figura 3: Prováveis pré-miRNAs correspondentes a lin-4 and let-7 de *C. elegans*. As seqüências maduras dos miRNAs estão em vermelho próximo da extremidade 5' do pré-miRNA (Adaptado de Lee 1993).

Figure 4: Biogênese de miRNAs e siRNAs. A biogênese de miRNA ocorre em dois passos distintos: um nuclear e outro citoplasmático, sendo finalizada quando uma das fitas do duplex encontra RISC e seu respectivo mRNA alvo. Por outro lado, a via de siRNA, ocorre apenas no citoplasma e o silenciamento do mRNA alvo, mediado por ambas as fitas do duplex (Adaptado de Bartel, 2004).

Figura 5: Preparação de esquistossômulos de 24 horas. Os esquistossômulos foram cultivados por 24 horas em estufa de CO_2 5% a 37°C em meio 169 analisados em lupa de aumento.

Figura 6: DNA genômico de verme adulto. Foram aplicados no gel aproximadamente 2 μ g de DNA e analisados em gel de agarose 0,6% e corado com brometo de etídio.

Figura 7: Análise em gel de agarose/formaldeído a 1%. RNA total para RT-PCR obtido a partir de vermes adultos (1), cercárias (2), esquistossômulos 3,0 horas (3); esquistossômulos 8,5 horas (4); esquistossômulos 18,5 horas (5); esquistossômulos 24 horas (6); esquistossômulos 48 horas (7); esquistossômulos 72 horas (8); ovos (9)

Figura 8: Análise em gel de agarose/formaldeído a 1%. RNA total para PCR em tempo real obtido a partir de vermes adultos (1), cercárias (2), esquistossômulos 3,0

horas (3); esquistossômulos 8,5 horas (4); esquistossômulos 18,5 horas (5); esquistossômulos 24 horas (6); esquistossômulos 48 horas (7); esquistossômulos 72 horas (8); ovos (9)

Figura 9: Análise da minipreparação dos clones em gel de agarose a 0,8%. Cerca de 5 µL da minipreparação foram analisados em gel de agarose e corado com brometo de etídeo.

Figura 10: Esquema dos domínios conservados, da região codificadora da proteína SmDicer1 (Smp_169750).

Figura 11: Esquema dos domínios conservados, da região codificadora da enzima SmDrosha1 (Smp_142510.1).

Figura 12: Esquema dos domínios conservados, da região codificadora da enzima SmDrosha2 (Smp_142510.2).

Figura 13: Esquema dos domínios conservados, da região codificadora da enzima SmAgo1 (Smp_140010).

Figura 14: Esquema dos domínios conservados, da região codificadora da enzima SmAgo2 (Smp_179320).

Figura 15: Esquema dos domínios conservados, da região codificadora da enzima SmAgo3 (Smp_102690.2).

Figura 16: Esquema dos domínios conservados, da região codificadora da enzima SmAgo4 (Smp_102690.3).

Figura 17: Esquema dos domínios conservados, da região codificadora da enzima SmFmr1 (Smp_099630).

Figura 18: Esquema dos domínios conservados, da região codificadora da enzima SmTudor-SN (Sm01663).

Figura 19: Esquema dos domínios conservados, da região codificadora da enzima SmPasha (Smp_087220).

Figura 20: Esquema dos domínios conservados, da região codificadora da enzima SmExportina5.1 (Smp_152800.1).

Figura 21: Esquema dos domínios conservados, da região codificadora da enzima SmExportina5.2 (Smp_152800.2).

Figura 22: Esquema dos domínios conservados, da região codificadora da enzima SmLoquacious (Smp_023670).

Figura 23 – Árvore filogenética consenso baseada na sequências de aminoácidos do domínio endoND2. A construção da árvore e a análise do "bootstrap" foram realizadas nos programas ClustalX 2,0 e MEGA 4.0. Para árvore consenso e confiabilidade dos ramos formados foi utilizado o teste de filogenia bootstrap, utilizando 1000 réplicas para cada sequência, sendo 60% o mínimo para se considerar o ramo confiável. O primeiro colchete representa as proteínas da subfamília Dicer1, o segundo da subfamília Dicer2 e a terceira da subfamília Drosha. A barra representa o número de aminoácidos que foram substituídos de uma sequencia para outra. Os códigos de acesso referentes às seqüencias estão mostrados na seção Materiais e Métodos.

Figura 24 – Árvore filogenética consenso baseada na seqüência de aminoácidos do domínio PIWI. A construção da árvore e a análise do "bootstrap" foram realizadas nos programas ClustalX 2.0 e MEGA 4.0. Para árvore consenso e confiabilidade dos ramos formados foi utilizado o teste de filogenia bootstrap, utilizando 1000 réplicas para cada sequência, sendo 60% o mínimo para se considerar o ramo confiável. O primeiro colchete representa as proteínas da subfamília Ago, o segundo proteínas órfãs sem definição de posicionamento na árvore e a terceira da subfamília PIWI. A barra representa o número de aminoácidos que foram substituídos de uma seqüência para outra. Os códigos de acesso referentes às seqüencias estão mostrados na seção Materiais e Métodos.

Figure 25: Alinhamento do domínio PIWI e apresentação dos resíduos catalíticos responsáveis pela função Slicer. Alinhamento entre as seqüências de aminoácidos do domínio conservado PIWI presente nas proteínas Argonautas das espécies *C. elegans, D. melanogaster, Mus musculus, H. sapiens, A. thaliana* e *S. mansoni*. Está marcado com um bloco e indicado por uma seta os 3 resíduos catalíticos conservados, dois ácidos

aspárticos e uma histidina (D/D/H), determinantes da função intrínseca Slicer (RNAseH) das proteínas da família Argonauta. Os códigos de acesso referentes às seqüencias estão mostrados na seção Materiais e Métodos.

Figura 26: Alinhamento dos dominios endoND da familia RNAse III e apresentação dos resíduos catalíticos. Alinhamento entre as seqüências de aminoácidos dos domínios conservados endoND1 e endoND2 das espécies *H. sapiens*, *D. melanosgaster*, *C. elegans*, e *S. mansoni*. Está em letra maior e mais escuro os 4 resíduos catalíticos conservados, E/D e D/E, determinantes da função nucleásica das proteínas da família RNAse III. Os códigos de acesso referentes às seqüencias estão mostrados na seção Materiais e Métodos.

Figura 27: Esquema do DNA genômico, cDNA e proteína de SmAgo4. A Figura mostra o DNA genômico contendo 13381 pares de base, divididos em 17 íntrons em azul e 19 éxons em verde. O seu cDNA contendo 3102 pares de bases e a proteína 924 aminoácidos com os domínios conservados PIWI em azul e PAZ em vermelho. A seta azul mostra o splicing na região 5' não traduzida e a seta vermelha o sinal consenso de cauda poli-A na região 3' não traduzida.

Figura 28: Estrutura do pré-miRNA seus prováveis miRNAs processados. Está predito pré-miRNA foi processada no programa disponível web Mfold 3.0 gerando uma estrutura secundária com um delta G igual a -30,2 kcal/mol. Em vermelho e azul estão os preditos miRNAs 1A e 1A*, respectivamente.

Figura 29: Amplificação por RT-PCR utilizando oligonucleotídeos específicos para SmDicer1. cDNA parcial do domínio conservado endoND1 de SmDicer de tamanho 469pb utilizando como molde cDNA de vermes adultos. Na primeira canaleta padrão de peso molecular de 100 pb (Invitrogen). *Gel de agarose 1,2% e corado com brometo de etídio.

Figura 30: Amplificação por RT-PCR e PCR convencional (molde DNA genômico) utilizando oligonucleotídeos específicos para *SmAgo5*. cDNA parcial do domínio conservado PIWI de *SmAgo5* de tamanho 711pb utilizando como molde cDNA de vermes adultos. DNA genômico parcial de SmAgo5 de tamanho 1283 pares de base. Na primeira canaleta padrão de peso molecular de 100 pb (Invitrogen). *5µL do produto de PCR foram analisados em gel de agarose.

Figura 31: Alinhamento entre as seqüências de DNA genômico do geneDB e cDNA consenso (*SmAgo5*) e cDNA C604197.1 da FAPESP. As seqüências alinhadas possuem o domínio PIWI característico das proteínas Argonautas. As setas com os blocos delimitam os iniciadores F e R do fragmento amplificado e o bloco com linhas tracejadas representa o íntron/éxon inserido na seqüência consenso de cDNA.

Figura 32: Alinhamento entre as seqüências de cDNA consenso(*SmAgo5*), *SmAgo3* e *SmAgo4*. As setas com os blocos delimitam os *splicing* alternativos de *SmAgo3*, 4 e cDNA consenso parcialmente seqüenciado, *SmAgo5*, sendo o primeiro de 80 nt diferenciando *SmAgo5*, das outras duas e o segundo de 93 nt diferenciando SmAgo3 das outras duas.

Figura 33: Splicing alternativo diferenciando *SmAgo3* e *SmAgo4* mostrado no banco da FAPESP. A seta com o bloco verde aponta e delimita a seqüência consenso do contig da FAPESP C604197.1. A seta com o bloco vermelho aponta e delimita o singlet MS1-0053 da FAPESP. O alinhamento mostra as diferenças entre as seqüências e o local do splicing alternativo sendo o clone MS1-0053 similar a sequência *SmAgo3*.

Figura 34: *Splicing* alternativo diferenciando *SmAgo3* e *SmAgo4* do consenso de cDNA obtido neste trabalho (*SmAgo5*) mostrado no banco da FAPESP. A seta com o bloco verde aponta e delimita a seqüência consenso do contig da FAPESP C604197.1. A seta com o bloco vermelho aponta e delimita o singlet ML1-0048T da FAPESP. O alinhamento mostra as diferenças entre as seqüências e o local do splicing alternativo sendo o clone ML1-0048T similar ao consenso de cDNA, *SmAgo5*.

Figura 35: Expressão do gene *SmDicer1* no desenvolvimento de *S. mansoni*. A expressão de SmDicer1 foi avaliada nos estágios de cercárias, esquistossômulos mecanicamente transformados (EMT-3,5, EMT-8,5, EMT-18,5, EMT-24, EMT-48, EMT-72 horas), vermes adultos e ovos. Como normalizador (gene constitutivo) foi utilizado o gene da α-tubulina e a expressão gênica relativa foi determinada pelo método do $2^{-\Delta\Delta Ct}$, utilizando como calibrador o estágio EMT-24 horas.

Figura 36: Expressão do gene *SmAgo2/3/4* no desenvolvimento de *S. mansoni*. A expressão de *SmAgo2/3/4* foi avaliada nos estágios de cercárias, esquistossômulos mecanicamente transformados (EMT-3,5, EMT-8,5, EMT-18,5, EMT-24, EMT-48, EMT-72 horas), vermes adultos e ovos. Como normalizador (gene constitutivo) foi

utilizado o gene da α -tubulina e a expressão gênica relativa foi determinada pelo método do 2^{- $\Delta\Delta$ Ct}, utilizando como calibrador o estágio EMT-24 horas.

Figura 37: Elucidação da via de miRNA em *S. mansoni*. Prováveis proteínas envolvidas na via de silenciamento gênico mediada por miRNA em *S. mansoni*.

Tabela 1- Meio 169 com seus componentes e concentração.

 Tabela 2: Tabela de Primers. Primers utilizados em diversas reações de PCR com

 diferentes temperaturas de anelamento.

Tabela 3: Prováveis seqüências da via de miRNA em *S. mansoni*. As seqüências de *S. mansoni* foram recuperadas do banco de dados geneDB. As seqüências de seus ortólogos *D. melanogaster*, *C. elegans* e humano, foram recuperadas dos bancos Flybase, Wormbase, e Genebank, respectivamente. As seqüências dos prováveis ortólogos presentes em *S. japonicum* também foram recuperadas no Genebank.

Tabela 4A: Regiões de cobertura dos domínios conservados da enzima SmDicer1e seus respectivos alinhamentos característicos com os domínios consenso dos bancosde dados PFAM e CDD.

 Tabela 4B: Comparação entre SmDicer1 e suas ortólogas relacionando tamanho e similaridade.

 Tabela 5A: Regiões de cobertura dos domínios conservados da enzima

 SmDrosha1 e seus respectivos alinhamentos característicos com os domínios consenso

 dos bancos de dados CDD.

 Tabela 5B: Comparação entre SmDrosha1 e suas ortólogas relacionando tamanho

 e similaridade.

Tabela 6A: Regiões de cobertura dos domínios conservados da enzimaSmDrosha1 e seus respectivos alinhamentos característicos com os domínios consensodos bancos de dados CDD.

Tabela 6B: Comparação entre SmDrosha2 e suas ortólogas relacionando tamanho e similaridade.

Tabela 7A: Regiões de cobertura dos domínios conservados da enzima SmAgo1e seus respectivos alinhamentos característicos com os domínios consenso dos bancosde dados PFAM e CDD.

 Tabela 7B: Comparação entre SmAgo1 e suas ortólogas relacionando tamanho e similaridade.

Tabela 8A: Regiões de cobertura dos domínios conservados da enzima SmAgo2 e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados PFAM e CDD.

 Tabela 8B: Comparação entre SmAgo2 e suas ortólogas relacionando tamanho e similaridade.

Tabela 9A: Regiões de cobertura dos domínios conservados da enzima SmAgo3 e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados PFAM e CDD.

 Tabela 9B: Comparação entre SmAgo3 e suas ortólogas relacionando tamanho e similaridade.

Tabela 10A: Regiões de cobertura dos domínios conservados da enzima SmAgo4 e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados PFAM e CDD.

 Tabela 10B: Comparação entre SmAgo4 e suas ortólogas relacionando tamanho e similaridade.

 Tabela 11A: Regiões de cobertura dos domínios conservados da enzima SmFmr1

 e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados.

 Tabela 11B: Comparação entre SmFmr1 e suas ortólogas relacionando tamanho e similaridade.

Tabela 12A: Regiões de cobertura dos domínios conservados da enzimaSmTudor-SN e seus respectivos alinhamentos característicos com os domínios consensodos bancos de dados PFAM e CDD.

 Tabela 12B: Comparação entre SmTudor-SN e suas ortólogas relacionando tamanho e similaridade.

Tabela 13A: Regiões de cobertura dos domínios conservados da enzima SmPashae seus respectivos alinhamentos característicos com os domínios consenso dos bancosde dados PFAM e CDD.

 Tabela 13B: Comparação entre SmPasha e suas ortólogas relacionando tamanho e similaridade.

Tabela 14A: Regiões de cobertura dos domínios conservados da enzimaSmExportina5.1 e seus respectivos alinhamentos característicos com os domíniosconsenso dos bancos de dados PFAM e COG.

 Tabela 14B: Comparação entre SmExportina5.1 e suas ortólogas relacionando tamanho e similaridade.

Tabela 15A: Regiões de cobertura dos domínios conservados da enzimaSmExportina5.1 e seus respectivos alinhamentos característicos com os domíniosconsenso dos bancos de dados PFAM e COG.

 Tabela 15B: Comparação entre SmExportina5.2 e suas ortólogas relacionando tamanho e similaridade.

Tabela 16A: Regiões de cobertura dos domínios conservados da enzimaSmLoquacious e seus respectivos alinhamentos característicos com os domíniosconsenso dos bancos de dados CDD.

 Tabela 16B: Comparação entre SmLoquacious e suas ortólogas relacionando tamanho e similaridade.

Tabela 17: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmDicer1 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo

macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Tabela 18: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmDrosha1 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Tabela 19: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmDrosha2 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Tabela 20: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmAgo1 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Tabela 21: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmAgo2 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Tabela 22: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmAgo3 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Tabela 23: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmAgo4 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo

macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Tabela 24: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmFmr1 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Tabela 25: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmTudor-SN relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Tabela 26: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmPasha relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Tabela 27: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmExportina5.1 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Tabela 28: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmExportina5.1 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

1-INTRODUÇÃO

Introdução

1.1 Aspectos gerais da Esquistossomose

A esquistossomose é uma doença crônico-debilitante que afeta 200 milhões de indivíduos distribuídos em 76 países nos continentes africano, asiático e americanos e estima-se que ocorram entre 250 a 500 mil mortes por ano (Hu e cols., 2004). Dentre os infectados 10% apresentam a forma severa da doença e aproximadamente 50% exibem manifestações clínicas da doença, constituindo um sério problema de saúde pública. Além disso, estima-se que 500 a 600 milhões de pessoas estão em risco de infecção pela esquistossomose (Chitsulo e cols., 2000).

A esquistossomose humana é causada por 6 espécies do gênero *Schistosoma*, sendo: *S. mansoni, S. haematobium, S. intercalatum, S. japonicum, S. mekongi e S. malayensis.* Somente a espécie *S. mansoni* existe no Brasil. Estima-se que no Brasil 12 milhões de pessoas estejam infectadas e destas, aproximadamente 1.000.000 sofram com a forma severa da doença (Kloetzel, 1989). No Brasil, acredita-se que a parasitose tenha sido introduzida através do tráfico de escravos africanos, tendo como porta de entrada o estado da Bahia.

A espécie foi descrita primeiramente por Bilharz em 1852 e posteriormente, Weinland em 1858 denominou o gênero deste helminto de *Schistosoma*, uma vez que o macho apresenta o corpo fendido (*schisto*=fenda; soma=corpo), mas a denominação da espécie *S. mansoni* foi dada por Sambon em 1907. Os estudos de Pirajá da Silva na Bahia, realizados na mesma época contribuíram muito para a confirmação que o *S. mansoni* produzia ovos, habitava as veias mesentéricas e tratava-se de uma espécie distinta.

1.2 O S. mansoni

O parasito *S. mansoni* pertence ao Filo Platelminto, a classe Trematódea e a família *Schistosomatidae* que apresenta sexos separados e parasita vasos sanguíneos de mamíferos e aves. Essa família é dividida em duas subfamílias: *Bilharzielinae* e *Schistosomatinae*. Na primeira, encontramos os vermes sem dimorfismo sexual, que parasitam aves e alguns mamíferos domésticos ou silvestres, portanto sem interesse médico direto. Na segunda, estão incluídos os que apresentam um nítido dimorfismo sexual e que parasitam o homem e animais (Neves, 2005). Além de apresentarem

dimorfismo sexual na fase adulta estes possuem o corpo achatado dorso-ventralmente. O macho mede cerca de 1,0 cm, com a superfície coberta por tegumentos e espinhas e um canal ginecóforo para albergar a fêmea e fecundá-la. A fêmea mede cerca de 1,5 cm e possui o tegumento liso.

O *S. mansoni* é um trematodo digenético que apresenta um ciclo de vida complexo que inclui dois hospedeiros. Os vermes adultos parasitam praticamente todos os sistemas dos vertebrados. A transmissão está geralmente a cargo de dois estágios larvais e, na sua maioria, os ciclos vitais dos digenéticos dependem da água para a sua realização (Wilson, 1979).

O ciclo de vida de *S. mansoni* é complexo, compreendendo duas fases distintas: uma fase sexuada que ocorre no hospedeiro definitivo, mamífero, e uma fase assexuada, no hospedeiro intermediário que compreendem os caramujos do gênero *Biomphalaria* das espécies *B. glabrata*, *B. straminae* e *B. tenagophila* (Figura 1). O ciclo de vida deste parasito foi descrito por Leiper em 1915 e inicia-se quando as fezes de indivíduos infectados, contendo ovos, são eliminadas e entram em contato com a água. Os ovos eclodem liberando a forma infectante do hospedeiro invertebrado, os miracídios. As larvas, que têm sexo definido, penetram ativamente na hemocele do caramujo e se desenvolvem transformando-se em esporocisto primário ou mãe. Cada esporocisto primário origina de 150 a 200 esporocistos filhos que migram para as glândulas digestivas e hepatopâncreas do caramujo onde se diferenciam por expansão clonal dando origem à terceira geração, as larvas agora denominadas cercárias (Wilson, 1979).

Sob o estímulo da luz e calor, as cercárias deixam o caramujo após a quarta semana de infecção, permanecendo no meio aquático onde buscam o hospedeiro definitivo. Esta o infecta por penetração ativa na pele perdendo sua cauda bifurcada e transformando-se em esquistossômulo. A passagem pela epiderme e derme do hospedeiro pode durar de 24 a 72 horas aproximadamente. Por volta do quarto dia, os esquistossômulos chegam no sistema sanguíneo ou linfático iniciando a migração para os pulmões a dia atingindo o pico máximo entre o sétimo e nono dia após a infecção. Até a fase pulmonar os esquistossômulos sofrem apenas um elongamento, ou seja, um processo de diferenciação, não havendo sinais de divisão celular ou crescimento, mantendo a razão massa/esquistossômulo constante (Clegg, 1965). A partir do oitavo dia passa por diversos órgãos e já podem ser encontrados na veia porta-hepática onde alimentam-se do sangue do hospedeiro definitivo e inicia a divisão celular (Clegg, 1965).

Após 45 dias de infecção o parasito atinge a maturidade vivendo por vários anos no sistema porta hepático. O macho e a fêmea se acasalam e as fêmeas realizam a postura de seus ovos completando assim o ciclo evolutivo. Parte dos ovos ficam presos nos tecidos do hospedeiro gerando reações imunológicas conhecidas como granuloma (Cunha, 1970).



Figura 2: Ciclo de vida do *S. mansoni*. Dentro do seu hospedeiro vertebrado, as formas larvais (esquistossômulo, B) dão origem a parasitos sexualmente maduros (verme adulto, A), os quais se acasalam e produzem ovos (C) que são liberados para o meio aquático. Os ovos eclodem (D) e liberam os miracídios (E), que penetram nos caramujos, dando origem a numerosos esporocistos (F). Os esporocistos geram cercárias (G) que saem do caramujo e são liberadas na água, as quais são as formas infectantes para o hospedeiro vertebrado (Adaptado de Wilson, 1979).

1.3 Genoma do parasito

O S. mansoni é um organismo diplóide (2n=16), possuindo oito pares de cromossomos, sendo sete pares autossômicos e um par sexual. O sexo heterogamético no parasito é a fêmea (ZW), enquanto o macho é homogamético (ZZ) (Short e cols., 1960, Short e cols., 1979). O tamanho do genoma haplóide compreende

aproximadamente 270 Mega bases, com um conteúdo de GC de aproximadamente 34% (Simpson e cols., 1982)

O *Projeto Genoma do Schistosoma* foi criado no ano de 1992 contando com o apoio do The Institute for Genomic Research (TIGR), FAPEMIG (Fundação de Amparo à Pesquisa de Minas Gerais), Fundação Oswaldo Cruz. Em 1994, e apoio internacional da Organização Mundial da Saúde (WHO/UNDP/World Bank Special Program for Research), do Centro de Treinamento em Doenças Tropicais (Training in Tropical Diseases – TDR) e de laboratórios de pesquisas de diversos países. O alvo destas pesquisas foram as espécies *S. mansoni* e *S. japonicum* tendo como objetivo promover o desenvolvimento de novas drogas e quimioterápicos.

As primeiras ESTs (Etiquetas de seqüências transcritas) foram depositadas por Franco e colaboradores, em 1995, sendo seqüenciadas 607 de uma biblioteca de cDNA de vermes adultos, dando início a era de descoberta de genes do *S. mansoni*. Em outubro de 2003, o lançamento e conclusão de dois projetos transcriptomas para as espécies *S. mansoni* e *S. japonicum* revolucionaram e ampliaram de maneira significativa as perspectivas para a pesquisa científica neste parasito. No projeto transcriptoma do *S. mansoni* foram seqüenciados 163.586 ESTs, sendo que 151.684 foram originadas de bibliotecas do tipo ORESTES (Open reading frames ESTs), ESTs de janelas de leitura abertas, ou seja, a região central do gene, e os 11.902 seqüências restantes foram originadas de bibliotecas normalizadas de vermes adultos. Desta forma o total predito de genes foi aproximadamente 14.000 genes, cobrindo aproximadamente 92% do transcriptoma do parasito (Verjovski e cols, 2003).

A análise do transcriptoma através da classificação funcional dos transcritos pelo *Gene Ontology* (GO) confirma a presença de genes associados a eucariotos e processos específicos dos metazoários, como diferenciação eixo anterior-posterior, dorso-ventral, epitélio, processo neural, motilidade, diferenciação sexual, maturação, longevidade, parasitismo, evasão imune e resposta a stress. Além disso, genes envolvidos com o controle da expressão gênica e a diferenciação ficaram evidenciados sugerindo sua implicação no nível de silenciamento da cromatina (Verjovski e cols, 2003).

1.4 Regulação da expressão gênica em S. mansoni

O *S. mansoni* apresenta um ciclo biológico bastante peculiar, no qual diversas alterações bioquímicas e morfológicas ocorrem no parasito. Estas mudanças são adaptações evolutivas encontradas pelo parasito para sobreviver nos distintos ambientes, como na água e interior dos hospedeiros, onde se acredita que um conjunto de genes deve ser expresso coordenadamente em cada estágio para permitir a adaptação do parasito aos seus hospedeiros.

Análises do transcriptoma e proteoma têm sido usadas para investigar padrões da expressão real e da complexidade protéica em todo o parasito (Wilson e cols., 2007). Mesmo assim, existem vários problemas iminentes para busca do entendimento e do arranjo correto da estrutura e da organização molecular do parasito. Um grande problema para uma anotação correta de proteínas é a ausência de informação sobre homólogos a impedindo correlações que levariam a informações pertinentes como estrutura e função (Verjovski-Almeida e cols., 2003).

Uma das formas viáveis e de mais fácil alcance para cobrir estes problemas é o estudo do transcriptoma. Este tem propiciado a descoberta de novos genes, elaboração de modelos gênicos, desenho de sondas de microarranjos, clonagem de genes de interesse e outras informações pertinentes para o entendimento da biologia da célula (Brindley and Pearce, 2007; Wilson e cols., 2007). Plataformas de microarranjos vêm sendo utilizadas extensivamente para *S. mansoni* para análise em larga escala do seu transcriptoma enfatizando padrões de expressão entre estágios, cepas , espécies e sexos (Fitzpatrick e cols., 2005, 2006; Hoffmann, 2002; DeMarco, 2006; Dillon, 2006; Vermeire, 2006).

Outra técnica utilizada para uma análise em larga escala do transcriptoma do *S. mansoni* é a técnica de SAGE. Ojopi e colaboradores em 2007, utilizando SAGE, analisaram 6263 etiquetas de seqüências do estágio de verme adulto. O mRNA mais abundante estava relacionado com o desenvolvimento de ovos, proteína da casca do ovo, seguido de mRNA que codificam actina e miosina, proteínas do choque térmico, proteínas ribossomais e outras. O estudo destes transcriptomas, separados espacialmente, provindos de um mesmo genoma, é de extrema importância para melhor entendimento da biologia molecular do parasito e elaboração de vacinas e novas drogas específicas (Figura 2) (Hu, 2004). Além destas metodologias, outros experimentos visando desvendar a biologia celular deste parasito vêm utilizando a manipulação do genoma silenciando genes utilizando a técnica de RNA interferente (RNA i), o estudo

do proteoma através da separação de proteínas por eletroforese bidimensional e a PCR quantitativa em tempo real.

De maneira particular o período de transformação de cercária em esquistossômulo deve ser ressaltado devido a grande mudança no perfil de expressão protéica representado nesta fase. Durante esse período ocorrem várias mudanças morfológicas e bioquímicas como perda da cauda, secreção do conteúdo da glândula acetabular, duplicação da bicamada lipídica da superfície, perda do glicocálix, mudança do núcleo heterocromático para eucromático, intolerância à água e às proteínas da via do complemento do sistema imune e redução no catabolismo do piruvato. (Ramalho-Pinto, 1974; Blanton e cols., 1987). Sabe-se que a diferenciação na composição de proteínas específicas entre esses dois estágios está relacionada a alterações na regulação de genes.

O seqüenciamento do genoma e do transcriptoma do *S. mansoni* tem aberto várias perspectivas para o estudo da biologia do parasito, no sentido de manipulação de genes essenciais no desenvolvimento e manutenção do parasitismo, no estabelecimento de novas drogas contra a parasitose e também no estabelecimento de uma vacina para prevenção e terapêutica.

Estudos relacionados à manipulação do genoma utilizando o RNAi, como descrito por Skelly e cols 2003, mostrou que o gene que codifica para catepsina B no estágio de desenvolvimento de esquistossômulo a verme adulto, pode perfeitamente ser silenciado. Outros estudos demonstraram que todo ciclo de vida deste parasito, esporocistos, esquistossômulos, vermes adultos (machos e fêmeas) são susceptíveis a entrada de RNAi e sua manipulação no meio intracelular (Boyle e cols., 2003; Dinguirard, 2006; Osman e cols., 2006).



Figura 2: Expressão gênica estágio-específicas em *S. mansoni*. Genes altamente expressos estágio-específicos baseado nos projetos transcriptoma do parasito. Adaptado de Hu e cols., 2004.

O RNAi é extremamente promissor uma vez que suas aplicações abrangem, entre outros, a genética reversa, genômica funcional, terapia gênica, desenvolvimento de animais modelo, estudos do comportamento, validação de alvos para novas drogas e o combate a patógenos (Baum, 2004).

A iniciativa de descoberta de proteínas relacionadas com a via de pequenos RNAs interferentes exógenos ou endógenos tem sido extensivamente estudada em outros organismos (Meister e Tuschl, 2004). Desta forma, o estudo sistemático nos bancos de dados deste parasito das proteínas da via, podem levar tanto ao avanço da biologia celular quanto ao possível controle da parasitose.

1.5 Silenciamento gênico mediado por microRNA

O termo "silenciamento gênico" se refere a diversos mecanismos nos quais a expressão de um ou mais genes é regulada. Este mecanismo se iniciou como um sistema de defesa atuando na identificação e degradação de moléculas de RNA estranhas ao

transcriptoma de uma célula sadia como acontece com plantas resistentes a vírus (Lindbo e Dougherty, 1992). Uma dessas moléculas é o dsRNA, naturalmente gerada na replicação viral ou mobilização de elementos transponíveis, sendo reconhecidas como sinal de perigo ou alerta. Esses dsRNAs levam à destruição de RNAs fita simples (ssRNAs) que sejam similares em seqüência. A interpretação sobre esta via vai mais além, sendo como um "sistema imune" do transcriptoma e os RNAs exógenos, aplicados a fim de estudos de genética reversa, como uma doença auto-imune induzida. Atualmente este mecanismo, eficiente de regulação gênica inata, atua principalmente no controle de genes envolvidos no desenvolvimento do organismo e na manutenção da integridade do genoma (Denli e Hannon, 2003).

Estudos demonstram a existência de três vias de silenciamento de RNA com sobreposição em alguns pontos (He e Hannon, 2004). Apesar de algumas particularidades nos mecanismos, as vias são conservadas entre diferentes organismos sugerindo a existência de um mecanismo ancestral comum (Zamore, 2002). A clonagem e caracterização de diversos genes que codificam componentes destas vias em *Arabidopsis thaliana*, *C. elegans*, *Drosophila melanogaster*, *Neurospora crassa*, *Schizosaccharomyces pombe*, *Tetrahymena thermophila*, camundongos e humanos, apóiam essa hipótese (Denli e Hannon, 2003, Meister e Tuschl, 2004; Gregory, 2006).

A primeira via de silenciamento de RNA é a mediada por siRNAs. Esta está envolvida na degradação de RNA viral, transposon e transgene. A 2^a via é a de silenciamento de mRNAs endógenos, mediada por miRNAs. Os miRNAs regulam a expressão gênica negativamente por meio do pareamento de bases específicas a um mRNA alvo, resultando na clivagem do mRNA ou na repressão de sua tradução. A 3^a via é nuclear e está associada à metilação de DNA e à formação de heterocromatina. (Bartel, 2004).

Em 1993, os microRNAs ou miRNAs, foram pela primeira vez relatados. Lee e colaboradores demonstraram que o gene Lin-4, conhecido por controlar o tempo de desenvolvimento do estágio larval de *C. elegans*, não codificava nenhuma proteína, mas sim, um par de pequenos RNAs, um com 22 nt e o outro com 16 nt (Figura 3). Estes pequenos RNAs eram complementares a várias regiões da região 3' não-traduzida (3'UTR) do gene Lin-14, e mediavam sua repressão por pareamento na região UTR, inibindo sua tradução. Posteriormente, descobriram Let-7, que codificava um pequeno RNA de 22 nt, que também estava envolvido no desenvolvimento de *C. elegans* (Reinhart e cols., 2000) (Figura 3). Devido ao papel de ambos miRNAs no controle do

desenvolvimento, foram também denominados de pequenos RNAs temporais, ou small temporal RNAs (stRNAs).

Estes miRNAs, hoje, somam uma classe de milhares de exemplares. Estes pequenos RNAs de 21-24 nt são originados a partir de um precursor transcrito endógeno com capacidade de adquirir estrutura secundária em forma de grampo e ser processado por uma série de proteínas intranucleares e citoplasmáticas (Hutvagner e cols., 2001).



Figura 3: Prováveis pré-miRNAs correspondentes a lin-4 and let-7 de *C. elegans*. As seqüências maduras dos miRNAs estão em vermelho próximo da extremidade 5' do pré-miRNA (Adaptado de Lee 1993).

O modelo atual para a biogênese de miRNAs em animais propõe que a maturação dos miRNAs ocorre em duas etapas: clivagem intranuclear e clivagem citoplasmática. Antes destas duas clivagens, os miRNAs são transcritos por uma polimerase em um produto primário (pri-miRNA) complexo, com várias características peculiares como formação de grampos e voltas (loops), os tornando reconhecíveis pela maquinaria no meio intranuclear (Meister e Tuschl, 2004).

Diversas evidências indicam que os miRNAs não são transcritos pela RNApol III, pois eles são um pouco mais longos do que os demais RNAs usualmente transcritos por esta enzima. Além disso, por possuirem seqüência internas com mais de quatro uracilas seguidas, o que usualmente sinaliza o término da transcrição para a RNApol III (Lee e cols., 2002), um grande número de miRNAs e transcritos adjacentes têm sido encontrados em ESTs, incluindo exemplos com a presença de uma cauda poli-A, sugerindo que foram transcritos pela RNA polimerase II (Smalheiser, 2003).

A primeira etapa do processamento dos miRNAs consiste na clivagem nuclear do pri-miRNA, que libera um intermediário com estrutura secundária em forma de grampo com aproximadamente 60-70 nt, denominado pré-miRNA (precursor miRNA) (Lee e cols., 2002). Esta clivagem é realizada pela por um microcomplexo liderado pela enzima RNase III Drosha, que cliva ambas as fitas do pri-miRNA em sítios próximos à base da estrutura em forma de grampo, gerando um pré-miRNA com um fosfato em sua extremidade 5' e 2 nucleotídeos despareados na extremidade 3' (Lee e cols., 2003). O pré-miRNA é transportado de forma ativa, com gasto de ATP, do núcleo para o citoplasma pelo complexo Ran-GTP/Exportina-5 (Lund e cols., 2004). A clivagem nuclear define a extremidade 5' do miRNA, que é a mesma do pré-miRNA. A extremidade 3' é processada no citoplasma pela enzima Dicer, juntamente com uma proteína ligante de RNA fita dupla, que em *D. melanogaster* é chamada Loquacious e R2D2, e em *C. elegans*, R3D1. (Tabara e cols., 2002; Saito e cols., 2005; Hutvagner e Zamore, 2002).

Na segunda etapa, Dicer atua de forma idêntica à clivagem de dsRNA no processo de silenciamento de RNA. Inicialmente ocorre o reconhecimento da região de fita dupla do pré-miRNA, provavelmente por afinidade da enzima com o fosfato 5' e os 2 nt despareados na extremidade 3'. Em seguida, ocorre a clivagem, liberando um miRNA de fita dupla com aproximadamente 21-25 pb, contendo fosfatos nas extremidades 5' e dois nucleotídeos despareados nas extremidades 3' (Hutvagner e Zamore, 2002). De acordo com este modelo, a especificidade da primeira clivagem determina a posição correta da segunda clivagem do pré-miRNA, definindo assim ambas as extremidades do miRNA (Lee e cols., 2003).

Após esta clivagem, o miRNA maduro será apresentado para o complexo de silenciamento mediado por RNA, denominado RISC. Inicialmente os miRNAs foram encontrados em associação com um complexo ribonucleoprotéico denominado miRNP (miRNA ribonuclein complex), que em humanos inclui a proteína Argonauta eIF2C2, a helicase GEMIN3 e GEMIN4 (Mourelatos e cols., 2002) e em *D. melanogaster* inclue a Argonauta Ago1, Fmr1, Tudor-SN e VIG (Meister e Tuschl, 2004). Desta forma, foi sugerido que o complexo miRNP corresponde ao RISC, que direciona a clivagem de mRNAs no mecanismo de silenciamento de RNA (Hutvagner e Zamore, 2002). Somente uma das fitas do miRNA de fita dupla é incorporada ao RISC. Quando isso

ocorre, a outra fita é degradada. Estudos para determinar qual das fitas é incorporada no RISC mostraram que a fita incorporada é aquela que possui o terminal 5' com pareamento de bases mais instável (Schwarz e cols., 2002). Uma vez incorporado ao RISC, o miRNA vai direcionar a clivagem específica de mRNAs complementares ou reprimir sua tradução (Doench e cols., 2003; Meister e Tuschl, 2004) (Figura 4).

A identificação dos mRNAs alvos demonstrou que os miRNAs estão predominantemente envolvidos na regulação de genes relacionados ao desenvolvimento de órgãos e tecidos, incluindo diversos fatores de transcrição (Dugas e Bartel, 2004). Além disso, a expressão de genes relacionados com a própria maquinaria do silenciamento de RNA, é regulada por miRNAs (miRbase Target, http://microrna.sanger.ac.uk/).



Figura 4: Biogênese de miRNAs e siRNAs. A biogênese de miRNA ocorre em dois passos distintos: um nuclear e outro citoplasmático, sendo finalizada quando uma das fitas do duplex encontra RISC e seu respectivo mRNA alvo. Por outro lado, a via de siRNA, ocorre apenas no citoplasma e o silenciamento do mRNA alvo, mediado por ambas as fitas do duplex (Adaptado de Bartel, 2004).
Os miRNAs atuam de forma semelhante aos siRNAs, regulando a expressão de mRNAs alvos, sendo seus alvos de origens e naturezas distintas. Os siRNAs são derivados do próprio mRNA alvo, sendo este transcrito um transgene, vírus ou transposon. Os miRNAs são processados a partir de transcritos endógenos que não codificam proteínas e possuem como alvos mRNAs endógenos. O número de genes que codificam miRNAs está estimado em 0,5 - 1% do número total de genes do genoma em questão tornando os a maior classe de moléculas regulatórias em animais. C. elegans e D. melanogaster possuem aproximadamente 100 a 140 genes que codificam miRNAs, enquanto seres humanos possuem de 200 a 255. Esta representatividade relativa é comparável à de outras famílias gênicas envolvidas na regulação gênica como, por exemplo, a de fatores de transcrição que se ligam a DNA (Nakahara e Carthew, 2004). A maioria dos genes codificando miRNAs são conservados entre espécies relacionadas e aproximadamente 30% destes, altamente conservados, com ortólogos em vertebrados e invertebrados, sugerindo uma conservação evolutiva com base em sua função biológica. A maioria dos genes que codifica miRNAs se localiza em regiões intergênicas, sugerindo que são transcritos a partir de unidades transcricionais independentes (Lagos-Quintana e cols., 2001; Lee e Ambros, 2001).

A clonagem e o seqüenciamento direto destes pequenos RNAs tem permitido a identificação de muitos miRNAs. Estes pequenos RNAs já foram encontrados em diversos organismos como *D. melanogaster*, *C. elegans*, *A. thaliana* e *H. sapiens* (Lagos- Quintana e cols., 2001; Lee e Ambros, 2001; Llave e cols., 2002). Entretanto, algumas impossibilidades vêm dificultando a clonagem e identificação destes RNAs como: padrões de expressão estágio específicos do miRNA, presença de produtos de degradação de mRNAs e outros pequenos RNAs não codificantes. (Lai e cols., 2003). Sendo assim, a descoberta de novas formas de busca de prováveis miRNAs tem aumentado expressivamente nos últimos anos.

No início de 2003, não mais que 33 microRNA tinham sido validados em seres humanos, sem perspectiva para mais avanços utilizando técnicas não computacionais (Lim, 2003a). No mesmo ano, foi predito, utilizando técnicas computacionais, 300 miRNAs, sendo deste total, 16 validados (Berezikov, 2005). Estudos recentes estimam que 600 miRNAs esperam para sua validação em humanos (Bentwich, 2005). Sendo assim, a predição de miRNAs e seus alvos está em fase de crescimento, tendo um longo e brilhante caminho para alcançar, sendo estas ferramentas promissoras para um futuro próximo.

27

O *S. mansoni* é um parasito que passa por vários estágios evolutivos em seu ciclo de vida, apresentando um conjunto de genes com expressão estágio-específica supostamente envolvidos com a adaptação do parasito aos distintos hospedeiros e habitats. Apesar dos progressos decorrentes do seqüenciamento do genoma e do transcriptoma deste parasito, ainda permanece por ser entendido o repertório de mecanismos utilizados pelo parasito na regulação diferencial da expressão gênica. Neste sentido, o mecanismo de silenciamento de mRNAs mediado por modificações póstranscricionais foi sugerido por Blanton e Licate, em 1992, como parte da hipótese de crescimento desacoplado do processo de diferenciação no parasito. Como revisado por Nakahara e Carthew em 2004, os miRNAs e sua maquinaria possuem papel fundamental na regulação gênica pós-transcricional, com papel em processos como a proliferação celular, apoptose, sinalização e principalmente diferenciação, esta via pode ter um papel significativo na regulação da expressão gênica diferencial observado entre os diferentes estágios evolutivos deste parasito.

Objetivos

2-Objetivos

Objetivos

2.1 Objetivo Geral

O objetivo deste trabalho foi analisar *in silico* todas as seqüências relacionadas com a via de silenciamento gênico mediada por miRNA nos bancos de dados disponíveis do parasito, além de analisar por PCR em tempo real a expressão relativa das duas proteínas principais da via, Dicer e Argonauta.

2.2 Objetivos específicos

- Análisar *in silico* a via de silenciamento gênico mediada por miRNA nos bancos de dados de genoma, geneDB (Sanger Institute) e transcriptoma, FAPESP;

 Analisar a conservação dos domínios conservados das famílias de proteínas, RNAse III e Argonauta;

- Clonar e seqüenciar parcialmente o DNA genômico e cDNA da família Argonauta;

- Predizer os miRNAs do parasito;

Avaliar o perfil de expressão gênica de SmDicer1 e Argonauta2/3/4 nos estágios de: verme adulto, cercária, esquistossômulo mecanicamente transformado (EMT) de 3,0;
8,5; 18,5; 24; 48; e 72 horas de cultivo *in vitro*;

3-MATERIAIS E MÉTODOS

Materiais e métodos

3.1 Parasitos

Neste trabalho foram utilizados parasitos da espécie *S.mansoni* cepa LE, em diferentes etapas do ciclo. Vermes adultos e cercárias, foram cedidos pelo Moluscário do CPqRR/FIOCRUZ (Centro de Pesquisa René Rachou, Fundação Oswaldo Cruz) em Belo Horizonte/MG e ovos pelo Laboratório de Biologia Molecular de Parasitos, do Departamento de Bioquímica e Imunologia da Faculdade de Medicina de Ribeirão Preto - USP.

Os vermes adultos foram obtidos a partir da perfusão do sistema porta-hepático de camundongos da linhagem Swiss Webster infectados com aproximadamente 100 cercárias por via subcutânea após 50 dias, conforme descrito por SMITHERS e TERRY, 1965. Após a coleta, os vermes adultos foram armazenados em tubo de1,5 ml a -80°C para posterior utilização.

Para obtenção dos ovos inicialmete o fígado dos camundongos infectados foram triturados com auxílio de um liquidificador em solução tampão (0,9% p/v Na2HPO4 e 0,005% p/v KH2PO4) e em seguida, filtrados em peneiras de 0,180mm/80 Tyler e 0,30mm/48 Tyler sucessivamente. Os ovos foram lavados e decantados por 5 minutos. O sobrenadante foi separado e repetiu-se o passo acima por 3 vezes, ou até não se retirar mais ovos do material. Após a coleta, os ovos foram armazenados em tubo de1,5 ml a - 80°C para posterior utilização.

O hospedeiro intermediário *Biomphalaria glabrata* após 27 a 31 dias de infecção com miracídeos foi induzido à eliminação de cercárias por um processo artificial. Em água declorada aquecida a 28° C foram colocados os caramujos, e estes expostos à luz branca fria em uma estufa a 25° C. Depois de duas horas de exposição a luz as cercárias eliminadas na água foram recolhidas. Para armazenamento desta forma do parasito, o material coletado é deixado por duas horas em gelo para decantação. A forma larval, cercária, é separada do sobrenadante e das impurezas contidas no meio e armazenada em tubos de 1,5 ml a -80°C para posterior utilização.

3.1.1 Esquistossômulos transformados mecanicamente

As cercárias, após banho de gelo (2 horas) e separação das impurezas, foram transferidas para tubos falcons de 15 mL, com aproximadamente 200.000 cercárias por

tubo, contendo 5 mL de RPMI 1640 (INVITROGEN). Os tubos foram homogeneizados com auxílio de um vortex em velocidade máxima por 1 minuto repetido por 3 vezes para efetuar a separação mecânica entre a cauda e o corpo cercariano. O volume do tubo foi transferido para uma garrafa estéril de cultura de 50 mL. Em seguida, foi acrescentado no frasco 30 mL de meio de cultura RPMI suplementado com 10% de penicilina / estreptomicina100 mg/mL, e incubado em estufa de CO2 5% a 37°C por 3 horas e meia. (Harrop e Wilson, 1993)

Após este intervalo foi iniciada a retirada das caudas presentes no meio. Os esquistossômulos foram submetidos a lavagens sucessivas (5 a 8) com meio RPMI em intervalos de 4 minutos para sedimentação e remoção das caudas suspensas no sobrenadante. A detecção da existência de caudas juntamente com os esquistossômulos foi feita com auxílio de uma lupa. O precipitado de esquistossômulos sem a presença de caudas foi considerado esquistossômulos de 3,5 horas. Estes foram separados e utilizados para cultivo ou armazenados em tubos de 1,5 ml a -80°C para posterior uso.

O cultivo dos esquistossômulos foi realizado em placas de 6 poços contendo 8 mL de meio 169 (Tabela 1) suplementado com 10% de soro fetal bovino (Gibco) e 1% penicilina / estreptomicina100 mg/mL por poço. Os tempos de cultivo foram 8,5; 18,5; 24; 48 e 72 horas.

Componentes	Concentração
Hidrolisado de lactoalbumina	0,1%
Glicose	0,1%
Hipoxantina	5 X 10 ⁻⁷ M
Triiodotironina (T3)	2 X10 ⁻⁷ M
Serotonina	1 X 10 ⁻⁶ M
Hidrocortisona	1 X 10 ⁻⁶ M
Meio Mínimo Vitamina	0,5%
Meio Schneider	5%
HEPES	20 mM
RPMI	q.s.p. 500 mL

Tabela 1- Meio 169 com seus componentes e concentração.



Figura 5: Preparação de esquistossômulos de 24 horas. Os esquistossômulos foram cultivados por 24 horas em estufa de CO_2 5% a 37°C em meio 169 analisados em lupa de aumento.

3.2 Anotação da via de miRNA

Para a identificação e comparação das seqüências envolvidas na via de miRNA em S. mansoni, primeiramente foi necessário a recuperação de següências em banco de dados de organismos onde a via está bem estudada como D. melanogaster (FlyBase), C. elegans (WormBase) e H. sapiens (NCBI). Em seguida, estas seqüências foram alinhadas via BLASTp, com E-value máximo de e-5, nos bancos transcriptoma (http://verjo18.iq.usp.br/schisto/) do S. e genoma mansoni (http://www.sanger.ac.uk/Projects/S_mansoni/). Todas as sequências que apresentaram similaridade com genes relacionados á via foram extraídas dos bancos de dados e analisadas. As seqüências prováveis de participação na via presentes em S. japonicum foram recuperadas (GenBank-NCBI). As análises para a identificação dos domínios e motivos protéicos conservados foram realizadas utilizando os programas Pfam e CDD (Conserved domain database, NCBI).

3.3 Alinhamentos entre seqüências e Análise filogenética

A árvore filogenética das seqüências foi inferida usando o método Neighbor-Joining (Saitou and Nei, 1987). A árvore filogenética consenso foi montada utilizando a análise de bootstrap de 1000 réplicas. Os braços das árvores correspondendo a partes que repetiram em menos de 50% das réplicas de bootstrap foram colapsados. A distância entre os ramos horizontais, separando duas seqüências foram proporcionais a divergência entre as sequências, sendo a quantidade de identidade de aminoácidos a medida utilizada. O alinhamento entre as següências foram feitos utilizando ClustalX 2.0 e a árvore filogenética foi conduzida utilizando o programa MEGA4 (Tamura e cols., 2007). O número de acesso das sequências utilizadas neste trabalho foi: Dm.Ago1 (NP_725342), Dm.Ago2 (ABB54719.1), Dm.Ago3 (ABO26294), Dm.Aubergine (CAA64320), Dm.PIWI (AAD08705.1), Ce.ALG1 (NP_510322.2), Ce.ALG2 (NP_493837.1), Ce.RDE1 (NP_7411.1), Ce.ERGO1(NP_503362.2), Ce.SAGO2 (NP_871859.1), Ce.SAGO1 (NP_5040.1), Ce.PPW1 (NP_740835.1), Ce.PPW2 (NP_491535.1), Ce.PRG1(NP_492.1), At.AGO1(gbAAB91987.1), At.AGO4 (NP 565633.1), Ce.TAG76 (NP 499192.1), CE.PRG2(NP 500994.1), Sp.AG01 (CAA19275.1), At.AGO7 (NP_177103.1), At.Zwille (CAA11429.1), At.AGO9 (CAD66636.1), Pt.PTIWI10 (CAI39070.1), Pt.PTIWI11 (CAI39069.1), Pt.PTIWI06 (CAI39075.1), Pt.PTIWI03 (CAI39076.1), Pt.PTIWI08 (CAI39073.1), Pt.PTIWI15 (CAI39065.1), Pt.PTIWI13 (CAI39067.1), Pt.PTIWI14 (CAI39066.1), Pt.PTIWI09 (CAI39072.1), Hs.HIWI (AAK92281.1), Pt.PTIWI02 (CAI44470.1), Pt.PTIWI12 (CAI39068.1), Pt.PTIWI05 (CAI44468.1), Hs.PIWIL1 (NP_004755.2), Hs.PIWIL4 (NP_689644.1), Hs.PIWIL3 (BAC81343.1), Hs.HILI (BAC81342.1), Hs.AGO1 (Q9UL18), Hs.AGO2 (Q9UKV8), Hs.AGO3 (Q9H9G7), Hs.AGO4 (Q9HCK5), Mm.AGO2 (Q8CJG0), Mm.AGO1 (Q8CJG1), Mm.AGO3 (Q8CJF9), Mm.AGO4 (Q8CJF8), Mm.MIWI2 (AAN75583.1), Mm.MIWI (EDL19532.1), Mm.MILI (AAK31965.1), Sme.SMEDWI1 (Q2Q5Y9), Sme.SMEDWI2 (Q2Q5Y8), Sj.3 (AAW25407.1), Sj.2 (AAX25645.2), Sj.1 (AAW26476.1), Dj.PIWIL (Q2PC95), Dm.Dicer1 (NP 524453.1), Hs.Dicer1 (NP 085124.2), Mm.Dicer1 (NP 683750.2), Aa.Dicer1 (XP_001659747.1), At.DCL1 (NP_171612.1), Am.Dicer1 (XP_624510.2), Ce.DCR1 (NP_498761.1), Cb.Dicer1 (XP_0016666666.1), Dm.Dicer2 (ABB54751.1), Ds.Dicer2 (ABB54760.1), Hs.Drosha (Q9NRR4), Dp.GA21287 (XP_001361471.1), Dy.Dicer2 (ABB54764.1), Aa.Drosha (XP_001653338.1), Dm.Drosha (NP_477436.1), Am.Drosha (XP_394444.2), Aa.Dicer2 (AAW48725.1), At.Dicer2 (NP_001078101.1), Dp.Dicer2 (XP 001360634.1), Mm.Drosha (XP 001473806.1), Ce.Drsh1 (NP 492599.1).

3.4 Extração do DNA genômico

Para a extração do DNA genômico de *S. mansoni* foram utilizados vermes adultos retirados do -80°C. O protocolo adotado para extração foi descrito por Costa (Tese de Doutorado, 1997).

Para cada 200 mg de vermes adultos, foram adicionados 500 μ L do tampão de extração (Tris 0,05 M, EDTA 1 mM, pH 7,5 e 1% de N-Lauril sarcosina) e 1000 μ g/mL de proteinase K. A amostra foi mantida a 37°C durante 6 horas. A seguir, foram acrescentados 100 μ L de NaCl 5M e a mistura incubada por 10 minutos a 65°C. Posteriormente foram adicionados 50 μ L de uma solução de CTAB/NaCl a 10%, seguido por uma incubação de 20 minutos a 65°C. Após os 20 minutos foi adicionado o mesmo volume de clorofórmio. As fases aquosa e orgânica foram separadas por centrifugação a 12.000 x g por 15 minutos sendo em seguida a fase aquosa transferida para outro tubo tipo eppendorf. O DNA presente na fase aquosa foi então precipitado com igual volume de isopropanol, seguido de incubação por 30 minutos à -20°C. Posteriormente, o DNA foi centrifugado a 12000 x g por 15 minutos, lavado com etanol 70%, seco na capela por 15 minutos e ressuspenso em 100 μ L de água mili-Q estéril. O DNA foi incubado durante por 16 horas a 37°C para homogeneização. Uma alíquota de 5 μ L foi analisada em gel de agarose 0,5% em eletroforese a 90 volts. (Figura 6)



Figura 6: DNA genômico de verme adulto. Foram aplicados no gel aproximadamente 2 μ g de DNA e analisados em gel de agarose 0,6% e corado com brometo de etídio.

3.5 Extração de RNA total para RT-PCR

Cerca de 100mg de vermes adultos, cercárias e ovos foram homogenizados em 1mL de Trizol (INVITROGEN) com auxílio de um vortex, na rotação máxima, em 4 intervalos de 14 minutos de incubação e 1 de homogeneização em temperatura ambiente. Após a homogeneização, a mistura foi passada 5 vezes em uma seringa com agulha (10 mL/25x8mm) a fim de fragmentar todo DNA genômico presente. Em seguida, foram adicionados à mistura 200 µL de clorofórmio. As amostras foram agitadas lentamente em vortex por 1 minuto e incubadas a temperatura ambiente durante 20 minutos, sendo posteriormente centrifugadas a 10000 x g por 10 minutos. Após esta etapa de centrifugação, 3 fases se destacaram. Uma fase inferior de cor vermelha fenolclorofórmio, uma fase mediana incolor mas imiscível com a aquosa e uma fase superior incolor aquosa. O DNA fica na interfase, as proteínas na fase orgânica, e o RNA permanece exclusivamente na fase aquosa. A fase aquosa foi transferida para um novo tubo estéril, tipo eppendorf, e em seguida adicionados 500 µL de isopropanol para a precipitação do RNA total. O tubo foi invertido por três vezes cuidadosamente, e incubada por 30 minutos à temperatura ambiente. Após a incubação o RNA foi então recuperado por centrifugação a 10000 x g por 10 minutos. O sobrenadante foi desprezado e o precipitado lavado, invertendo 3 vezes vagarosamente, com 1,0 mL de etanol 70% em água tratada com dietilpirocarbonato (DEPC). Em seguida, o RNA foi centrifugado a 10000 x g por 5 minutos e o precipitado final foi seco em fluxo laminar por 15 minutos. Por fim RNA total foi ressuspenso em 40 µL de água livre de RNAses e incubado por 10 minutos a 56°C para completa homogeneização.

3.6 Extração do RNA total para PCR quantitativo em tempo real

Cerca de 100 mg de vermes adultos, cercárias, ovos, esquistossômulos de 3,0 horas, 8,5 horas, 18,5 horas, 24 horas, 48 horas e 72 horas foram utilizados para extração de RNA total utilizando o kit "Purelink Micro-to-Midi Total RNA purification system" (Invitrogen).

A cada amostra foi adicionado 1 mL de Trizol e com auxílio de um politron, por aproximadamente 1 minuto todo material foi completamente solubilizado. Em seguida,

a mistura foi incubada por 5 minutos a temperatura ambiente. Depois da extração do RNA, fases de separação com clorofórmio, e precipitação com isopropanol e etanol 70%, como descrito pelo método acima, procedemos à lavagem e eluição do ácido nucléico, conforme o boletim técnico do fabricante, mostrado resumidamente abaixo.

Para o procedimento de lavagem, foram transferidos aproximadamente 700 μ L da amostra de RNA com etanol para a coluna com resina, e foi centrifugada a 12.000xg por 1 minuto. Após centrifugação o volume eluido da coluna foi descartado. Esse processo foi repetido por 2 vezes sendo posteriormente, foram adicionados 700 μ L do tampão de lavagem (Wash buffer I) e centrifugado por 1 minuto a 12.000xg. A seguir foram adicionados 500 μ L do tampão de lavagem II (Wash buffer II), e centrifugado a 12.000xg por 1 minuto. Todo material eluido foi descartado e a coluna transferida para o tubo coletor onde o RNA foi eluido em 30 μ L de água livre de RNAse e incubado por 1 minuto. Para separar o RNA total da amostra, este foi centrifugado por 12.000xg por 2 minutos, e mantido a -70°C até o uso.

3.7 Verificação da qualidade do RNA total extraído

A qualidade do RNA total foi avaliada em gel de agarose-formaldeído 1%. Para a preparação do gel, inicialmente 600 mg de agarose foi dissolvido em 40,0 mL de água DEPEC autoclavada, livre de RNAses. Em outro recipiente, foram adicionados: 12,8 mL de água livre de RNAses, 6,0 mL de MOPS 10x (MOPS, acetato de sódio diidratado, EDTA tetrassódico, água livre de RNAses, para um volume final de 500 mL) e 1,2 mL de de formaldeído e 0,5 μ L de brometo de etídio. A solução foi misturada à agarose e deixada em repouso até completa solidificação. As amostras foram previamente desnaturadas em tampão (formamida, formaldeído, MOPS, azul de bromofenol, água DEPEC e brometo de etídio), incubadas por 15 minutos a 65°C, deixadas em banho de gelo durante 3 minutos e aplicadas no gel. As condições de realização da eletroforese foram: 80 volts em tampão de corrida (MOPS 1x) por aproximadamente 75 minutos (Figura 7 e 8).



Figura 7: Análise em gel de agarose/formaldeído a 1%. RNA total para RT-PCR obtido a partir de vermes adultos (1), cercárias (2), esquistossômulos 3,0 horas (3); esquistossômulos 8,5 horas (4); esquistossômulos 18,5 horas (5); esquistossômulos 24 horas (6); esquistossômulos 48 horas (7); esquistossômulos 72 horas (8); ovos (9).



Figura 8: Análise em gel de agarose/formaldeído a 1%. RNA total para PCR em tempo real obtido a partir de vermes adultos (1), cercárias (2), esquistossômulos 3,0 horas (3); esquistossômulos 8,5 horas (4); esquistossômulos 18,5 horas (5); esquistossômulos 24 horas (6); esquistossômulos 48 horas (7); esquistossômulos 72 horas (8); ovos (9).

3.8 Quantificação de DNA e RNA

As concentrações de DNA e RNA foram estimadas a partir da medida de absorbância a 260nm. Uma unidade de absorbância a 260nm corresponde aproximadamente a 50μ g/mL de DNA de fita dupla e 40μ g/mL para RNA e DNA de fita simples. O grau de pureza da preparação foi estimado através da relação entre as leituras a 260 e 280nm. As preparações foram consideradas boas, quando o valor da razão A=260/280 variou entre 1,6-2,3.

3.9 Oligonucleotídeos iniciadores (Primers)

Os oligonucleotídeos iniciadores específicos foram idealizados utilizando o programa Gene Runner e VectorNTI® (Invitrogen) para amplificação de DNA genômico, cDNA para RT-PCR e para PCR em Tempo Real. Estes foram baseados nas sequências de cDNA depositadas no banco ONSA da FAPESP (Fundação de Apoio à Pesquisa do Estado de São Paulo), e GenBank para os genes Dicer (C604739), Argonauta (C604197) e alfa tubulina (M80214). Todos primers amplificam parte dos domínios conservados e funcionais da respectiva proteína codificada pela sequência referente (Tabela 2).

Tabela 2: Tabela de Primers. Primers utilizados em diversas reações de PCR com diferentes temperaturas de anelamento.

Proteína	DNA	PCR	Foward (5' 3')	Reverse (5' 3')	Temper atura de a nelamento
Argonauta	cDNA	RT-PCR	TCCTGGTTCCACATTCCG	CGATGATTTGGCCT ACGG	59°C
	cDNA	PCR em Tempo Real e RT-PCR	TCACGCACGGTAGATCAGG	AGGTCCGCGTCAATTTGG	60°C
	DNA genômico	PCR convencio nal	TCCTGGTTCCACATTCCG	CGATGATTTGGCCT ACGG	59°C
Dicer	cDNA	RT-PCR	TTIGTTCCTTCTGGTTCC	ATGACGACGAAGAAGATG	57°C
	cDNA	PCR em Tempo Real e RT-PCR	TCTTCCGTC CACCATTCG	TGC CAAC AAC ACA ATT CC	60°C
Alfa tubulina	cDNA	PCR em Tempo Real	CGTATTCGCAAG TTGGCTGACC A	CCATCGAAGCGCAGTGATGC A	60°C

3.10 PCR (Reação da Polimerase em Cadeia)

3.10.1 PCR convencional

Para validação das temperaturas de anelamento dos primers, sequenciamento e síntese de DNA genômico foi utilzado a técnica de PCR convencional.

Foi utilizado para cada reação de amplificação 350 ng de DNA genômico, 10 mM de cada dNTP (dATP, dCTP, dTTP, e dGTP - Promega), 50 mM de MgCl₂, 10

picomoles de cada oligonucleotídeo (direto e reverso), 2,5 unidades de Platinum Taq DNA polimerase (Invitrogen), 1X do tampão PCR 10X concentrado sem Mg⁺² e com água milli-Q para completar o volume. O aparelho utilizado foi Thermo Hybaid Px2 e o ciclo das reações de amplificação foi iniciado por desnaturação de 5 minutos a 95°C, 35 ciclos pré-estabelecidos, com desnaturação a 95°C por 1 minuto, anelamento a temperatura específica de cada primer, extensão por 90 segundos a 72°C e finalizado com uma extensão de 6 minutos a 72°C.

3.10.2 RT-PCR

A primeira fita do cDNA foi sintetizada utilizando o Kit Thermoscript RT-PCR System (Invitrogen), seguindo as recomendações dadas pelo fabricante.

Foram utilizados 5 μ g de RNA total, 50 pmol de oligodT, 10 mM de dNTPs e água livre de RNAse para um volume final de 10 μ L. A desnaturação do RNA foi realizada através da incubação da mistura a 65°C em termociclador (Thermo Hybaid Px2) durante 5 minutos, seguido de banho de gelo por 1 minuto. Em seguida, foram adicionados o tampão da enzima, 0,1 M de DTT (ditiltreitol) e 15 unidades da enzima transcriptase reversa. Posteriormente, essa mistura foi incubada novamente em termociclador a 42°C durante 60 minutos e 85°C por 5 minutos. Por fim foi acrescentado 1,0 μ L de RNAseH e a amostra incubada por 20 minutos a 37°C. A amostra foi estocada a -20°C até o momento do uso.

No intuito de obter o cDNA fita dupla, amplificações foram realizadas utilizando 2µL do cDNA como molde, combinados com os mesmos reagentes e etapas de reação descritos no item de DNA genômico, com exceção dos primers por sua especificidade e temperatura de anelamento.

3.10.3 PCR quantitaviva em Tempo Real

Para análise da expressão dos genes supracitados foi utilizada a técnica de PCR em tempo real. As reações foram realizadas utilizando o kit Platinum SYBR green qPCR SuperMix-UDG ROX (INVITROGEN), conforme o manual do fabricante. A reação de PCR quantitativo em Tempo Real foi conduzida conforme programação contida no aparelho ABI 7500 Applied Biosystems. As análises foram feitas utilizando o método do $\Delta\Delta$ Ct, uilizando a fórmula 2- $\Delta\Delta$ Ct, para calcular a expressão gênica relativa dos genes em estudo, conforme boletim técnico do fabricante do aparelho de qPCR (Applied Biosystems).

3.11 Análise Estatística

A expressão relativa nos estágios do parasito foi comparada utilizando a análise da variância por ONE-WAY (Teste de Tukey). Foi considerado estatisticamente significante P<0.001. A análise estatística foi feita utilizando o programa PRISM.

3.12 Verificação da qualidade do DNA ou cDNA

Os produtos das PCR's supracitadas foram aplicados em gel de agarose 1,2% (90 volts – 75 minutos), a temperatura ambiente, juntamente com o padrão de peso molecular 100pb DNA Ladder (Invitrogen). O gel foi corado com brometo de etídio, observado ao transiluminador (Vilber Lourmat) e fotografado.

3.13 Purificação do produto de PCR

O produto obtido na reação de PCR foi transferido para um novo tubo tipo eppendorf e adicionado 1/10 do volume inicial de acetato de sódio 3M pH 7,0 (10 μ L) e 2,5 do volume (250 μ L) de etanol absoluto. Posteriormente esta mistura foi incubada a - 20°C por 30 minutos e em seguida, centrifugada a 10000 x g por 5 minutos. O precipitado foi lavado com etanol 70% e seco. O DNA foi suspenso em 50 μ L de água milli-Q estéril.

3.13.1FREEZE SQUEZE

De acordo com a metodologia descrita por Tautz e Renz (1983) as banda de interesse, contendo os fragmentos de DNA, foram cortadas do gel de agarose, com auxílio de um estilete. Em seguida, foram transferidas para um becker contendo 10 mL de uma solução composta de NaOAc 300 mM/EDTA 1 mM pH 7,0 e incubada no

escuro, sob agitação lenta por 30 min. Posteriormente, as bandas foram transferidas para um tubo contendo um lã de vidro e mantidas a -80°C por 30 minutos. O DNA livre da agarose foi obtido por centrifugação a 12000 x g por 15 minutos e recuperado a -20°C pela precipitação com 1/10 de acetato de sódio 3M pH 7,0 e 2,5 volume de etanol absoluto. Após a lavagem com etanol 70% e seco, o DNA foi suspenso em 50 µL de água milli-Q estéril.

3.14 Clonagem

3.14.1 Reação de Ligação

As amostras purificadas foram ligadas ao vetor plasmidial de sequenciamento pGEM-T easy (Promega) utilizando 5 μ L de tampão, 3 μ L do produto de PCR, 1 μ L de vetor pGEM-T easy e 1 μ L de DNA ligase. A reação foi incubada no termociclador a 4°C por 16 horas para efetuar a ligação.

3.14.2 Preparo de células competentes

Uma colônia de células de *Escherichia coli DH5a* foram inoculadas inicialmente em 5 mL de meio LB (bacto-triptona 10g/L; extrato de levedura 5g/L; NaCl 5g/L; pH=7,5) em um tubo tipo Falcon de 15 mL e incubadas a 37°C por 16 horas sob agitação de 200 rpm. Após a incubação, 250µL do pré-inóculo foram transferidos para um tubo tipo Falcon de 50 mL e o volume completado para 25 mL com meio SOB (bacto-triptona 2g/L; extrato de levedura 5g/L; NaCl 0,58; KCl 0,2g/L; pH=7,5) suplementado com 250µL de MgSO₄ 1M e 250µL de MgCl₂ 1M. A cultura foi incubada por duas horas a 37°C sob agitação (200 rpm) até que a concentração de células atingisse uma densidade ótica (600nm) entre 0,4 e 0,6. A seguir, as células foram incubadas em banho de gelo por 30 minutos e recuperadas por centrifugação a 3000 x g durante 10 minutos a 4°C. O sobrenadante foi descartado e o precipitado ressuspenso com auxílio de um pipetador automático em 10 mL de tampão Pipes (60 mM de CaCl₂, 10mM de Pipes e 15% de glicerol P.A) para a lavagem das células. As bactérias foram novamente centrifugadas a 2000 x g durante 10 minutos a 4°C. A lavagem foi repetida duas vezes nas mesmas condições. Após as lavagens, as células foram incubadas por 30 minutos em banho de gelo e em seguida centrifugadas a 2000 x g por 10 minutos a 4°C. Depois da centrifugação, o sobrenadante foi descartado e o precipitado ressuspenso em 1,2 mL de tampão Pipes. As células competentes foram então alíquotadas e armazenadas a -80°C até o momento do uso.

3.14.3 Transformação Bacteriana

Foram adicionados 5μ L do produto da reação de ligação, conforme obtido em uma alíquota de célula competente (100µL). Essa mistura foi incubada em banho de gelo durante 30 minutos, submetida a choque térmico à 42°C por 90 segundos e retornada ao banho de gelo em seguida por 2 minutos. Esta foi transferida para um tubo tipo Eppendorf contendo 950µL de meio SOB e incubada durante 90 minutos a uma temperatura de 37°C sob agitação constante (200 rpm). Após esse período, a suspensão bacteriana foi centrifugada a 3000 x g por 5 minutos e o precipitado transferido para placas de Petri com meio LB-ágar contendo ampicilina (100µg/mL), X-Gal (20µg/mL) e IPTG. A presença de X-Gal permitiu a seleção dos clones recombinates. As placas contendo as células transformadas foram incubadas durante 16 horas a 37°C em estufa.

3.14.4 PCR de Colônia

As colônias recombinantes foram transferidas para 1,5 mL de caldo LB/ampicilina durante 14 horas e submetidas à amplificação por PCR para confirmação da presença do inserto de interesse na colônia.

3.14.5 Minipreparação

Uma alíquota de 20μ L da cultura das células transformadas foram inoculadas em 5 mL de meio LB/ampicilina (100μ g/mL) e incubadas durante 16 horas a 37°C sob agitação constante (200 rpm). O método de extração utilizado foi o da lise alcalina descrito por Maniatis e cols., 1989. Um volume de 1,5 mL de cultura bacteriana foi centrifugado a 14000 x g durante 3 minutos. O sobrenadante foi removido e o procedimento realizado novamente. Em seguida, o precipitado foi suspenso em 100µL de solução TGE (Glicose 50mM, EDTA 10mM e Tris-HCl 25mM pH 8,0) suplementada com 5 μ L RNAse 20mg/mL. Após a lise alcalina das bactérias com solução composta de NaOH 0,2 N e SDS a 1% e precipitação das proteínas com acetato de potássio 3M pH 4,0, 500 μ L de isopropanol foi adicionado para a precipitação do DNA plasmidial. A amostra contendo DNA plasmidial foi incubada a temperatura ambiente por 20 minutos e centrifugada novamente a 14000 x g por 15 minutos. Após o descarte do sobrenadante o pellet foi lavado com etanol 70% e centrifugado durante 5 minutos a 14000 x g. Novamente o sobrenadante foi descartado, sendo em seguida o precipitado seco a temperatura ambiente e ressuspendido em 50 μ L de água milli-Q. Uma alíquota de 5 μ L foi aplicada em gel de agarose 0,8% para confirmação da preparação (Figura 9)



Figura 9: Análise da minipreparação dos clones em gel de agarose a 0,8%. Cerca de 5 μ L da minipreparação foram analisados em gel de agarose e corado com brometo de etídeo.

3.15 Sequenciamento

As reações de sequenciamento foram realizadas utilizando o kit Big-Dye Terminator (Applied Biosystems), de acordo com as instruções do fabricante e as reações, analisadas no seqüenciador automático de DNA, ABI 3100 Genetic Analyzer (Applied Biosystems). Os plasmídeos foram seqüenciados, nas direções "direta" e "reversa".

3.16 Análise computacional das seqüências

3.16.1 Seqüência consenso

As seqüências obtidas no seqüenciamento foram submetidas um alinhamento entre elas para obtenção da seqüência consenso. Posteriormente foram utilizadas seqüências do banco de dados de transcriptoma (<u>http://verjo18.iq.usp.br/schisto/</u>) e genoma do *S. mansoni* (<u>http://www.sanger.ac.uk/Projects/S_mansoni/</u>) para busca de similaridade. Para o alinhamento entre as seqüências foram utilizados os programas BLAST2 (NCBI) e ClustalX 2.0

3.16.2 Splicing alternativo

As seqüências consenso e aquelas presentes nos bancos de dados da FAPESP (ONSA) e geneDB representantes da família Argonauta no parasito foram alinhadas utilizando os programas BLAST2 (NCBI) e ClustalX 2.0.

3.17 Predição de miRNA

Candidatos a miRNA em *S. mansoni* foram escolhidos de 2 formas: utilizando regiões de miRNAs não maduros (pré-miRNA) de *D. melanogaster e C. elegans* como molde para busca e utilizando sequências intergênicas recuperadas do GeneDB. As seqüências que obtiveram similaridade satisfatória foram pré-selecionadas. Todas as seqüências pré-selecionadas foram rodadas no programa WEB Mfold 3.2 (www.mfold.com) para a obtenção da sua respectiva estrutura secundária. A partir deste momento as seguintes características foram analisadas: números de estruturas geradas

pela mesma seqüência, números de grampos por estrutura, tamanho do grampo ("loop"), simetria e energia livre (ΔG kcal/mol) da estrutura final.

4-RESULTADOS

4.1 Reconstituição in silico da via de miRNA

As seqüências relacionadas ao mRNA, DNA genômico e proteína da via de silenciamento mediada por miRNAs, foram recuperadas dos bancos de dados do geneDB (Sanger Institute) e FAPESP (ONSA). Estas foram selecionadas por possuírem alta similaridade com ortólogos onde a via está quase toda descrita como *D. melanogaster, C. elegans e H. sapiens*, e também com prováveis ortólogos da via como *S. japonicum*. Os ortólogos encontrados em *S. mansoni* foram nomeados utilizando as primeiras letras do nome da espécie, Sm, seguido pelo nome da proteína encontrada nos organismos relacionados e um número quando ocorrer a existência de parálogos.

Além das proteínas da via e suas correlações, também identificamos um provável microRNA com características intrínsecas expressivas de um miRNA.

4.1.1 Análise in silico dos constituintes da via de miRNA

A Tabela 3 mostra os genes identificados nos bancos de dados com uma suposta função de envolvimento na biogênese de miRNA em *S. mansoni* representando o estágio nuclear, mediado pela proteína SmDrosha1/2 e SmPasha, passando pela carioteca, com SmExportina-5.1/2, até chegar no estágio citoplasmático, representado pela SmDicer1, SmLoquacious e pelas proteínas componentes de RISC, SmArgonautas1/2/3/4, SmFmr1 e SmTudor-SN.

As Figuras 10 a 22 mostram a representação esquemática da estrutura protéica primária das proteínas identificadas. As Tabelas 4 a 16 mostram as regiões de cobertura destes domínios e sua similaridade com banco de dados de domínios conservados, e a comparação por tamanho e similaridade com seus ortólogos.

A Figura 10 mostra o contig Smp_169750 depositado no banco de dados geneDB que representa a proteína SmDicer1 com uma ORF predita de 2174 aminoácidos. A Tabela 4A mostra os domínios conservados presentes, HEL, DUF283, PAZ e RIBOc (endoND1 e endoND2), bem como a similaridade com o consenso dos bancos de dados CDD e PFAM. A Tabela 4B mostra os ortólogos de SmDicer1 e uma comparação entre os tamanhos e similaridade variando de 3e-124 em *D. melanogaster* a 1e-84 em *H. sapiens*.

Tabela 3: Prováveis seqüências da via de miRNA em *S. mansoni*. As seqüências de *S. mansoni* foram recuperadas do banco de dados geneDB. As seqüências de seus ortólogos *D. melanogaster*, *C. elegans* e humano, foram recuperadas dos bancos Flybase, Wormbase, e Genebank, respectivamente. As seqüências dos prováveis ortólogos presentes em *S. japonicum* também foram recuperadas no Genebank.

Via de miRNA	Prováveis ortólogos de S. mansoni	Função	Domínios	Ortólogos de humanos (GeneBank)	Ortólogos de D. melanogaster (FlyBase)	Ortólogos de <i>C. elegans</i> (WormBase)	Prováveis ortólogos de <i>S. japonicu</i> m (GeneBank)
RNAse III							
Dicer	Smp_169750	Processamento de pré-miRNA	RNase III, PAZ, dsRBD	NP_085124	FBgn0039016	WBGene00000939	SJCHGC08817
Drosha	Smp_142510.2 Smp_142510.1	Processamento de pri-miRNA	RNase III, dsRBD	AAF80558	FBgn0026722	WBGene00009163	SJCHGC08309
Componente de RISC							
Argonaute	Smp_140010 Smp_179320 Smp_102690.2 Smp_102690.3	Ligante de RNA e nuclease	PAZ, PIWI	Q9UL18 NP_079128	FBgn0026611 FBgn0046812	WBGene00000105 WBGene00000106	SJCHGC07755 SJCHGC07884 SJCHGC01111 SJCHGC05069 SJCHGC07014
Tudor-SN	Sm01663	Nuclease	Tudor, SNc	Q7KZF4	FBgn0035121	WBGene00006626	SJCHGC02502 SJCHGC09149 SJCHGC04700
Fmr1	Smp_099630	Ligante de RNA	KH, PRK11824	AAC50155	FBgn0028734	-	SJCHGC05245 SJCHGC09283
Out ros fato res							
Pasha	Smp_087220	Ligante de fita dupla de RNA	dsRBD	BAB83032	FBgn0039861	WBGene00011908	SJCHGC09172
Loquacious	Smp_023670	Ligante de fita dupla de RNA	dsRBD	NP_004169	FBgn0032515	-	-
Exportin-5	Smp_152800.1 Smp_152800.2	Exportadora nuclear de pré- miRNA	CRM1, Xpo1	NP_065801	FBgn0031051	-	SJCHGC09381 SJCHGC05562

	SmDicer1 (Smp_169750)		_						_	
			HELS	DUF283		PAZ		RIBOc >	RIBOc	
1		500		_	1000		1500			2174

Figura 10: Esquema dos domínios conservados, da região codificadora da proteína SmDicer1 (Smp_169750).

Tabela 4A: Regiões de cobertura dos domínios conservados da enzima SmDicer1 e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados PFAM e CDD.

Domínio	Região de Código de acesso		E-value
	cobertura (aa)	do domínio	(Blastp)
HEL	628-684	pfam00271	0,006
DUF283	802-918	pfam03368	1e-12
PAZ	1236-1337	cd02843	7e-26
RIBOc (endoND1)	1637-1719	cd00593	4e-16
RIBOc (endoND2)	1906-2086	cd00593	9e-28

Espécie	Ortóloga	Código de acesso	Tamanho	E-value
			(aa)	(Blastp)
S. mansoni	SmDicer1	Smp_169750	2174	-
D. melanogaster	Dicer-1	FBgn0039016	2249	3e-124
C. elegans	Dcr-1	WBGene00000939	1845	1e-98
H. sapiens	Dicer1	NP_085124	1922	1e-84

Tabela 4B: Comparação entre SmDicer1 e suas ortólogas relacionando tamanho e similaridade.

As Figura 11 e 12 mostram os contig Smp_142510.1 e Smp_142510.1 depositados no banco de dados geneDB que representam as proteínas SmDrosha1 e SmDrosha2 com uma ORF predita de 1531 e 1577 aminoácidos, respectivamente. As duas representam seqüências de cDNA provindas de um mesmo DNA genômico representando um splicing alternativo.

As Tabelas 5A e 6A mostram os domínios conservados presentes, RIBOc (endoND1 e endoND2), e DSRM e as similaridades com o consenso do banco de dados CDD. As Figuras 5B e 6B mostram os ortólogos de SmDrosha1 e SmDrosha2 e uma comparação entre os tamanhos e similaridade entre eles. Os valores de similaridade variaram de 0 a 1e-130.



Figura 11: Esquema dos domínios conservados, da região codificadora da enzima SmDrosha1 (Smp_142510.1).

Tabela 5A: Regiões de cobertura dos domínios conservados da enzima SmDrosha1 e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados CDD.

Domínio	Região de	Código de acesso	E-value
	cobertura (aa)	do domínio	(Blastp)
RIBOc (endoND1)	951-1064	cd00593	1e-24
RIBOc (endoND2)	1123-1258	cd00593	4e-24
DSRM	1264-1334	cd00048	1e-08

Espécie	Ortóloga	Código de acesso	Tamanho	E-value
			(aa)	(Blastp)
S. mansoni	SmDrosha1	Smp_1425.1	1531	-
D. melanogaster	Drosha	FBgn0026722	1327	0
C. elegans	Drsh-1	WBGene00009163	1086	1e-130
H. sapiens	Drosha	AAF80558	1374	0

Tabela 5B: Comparação entre SmDrosha1 e suas ortólogas relacionando tamanho e similaridade.

SmDrosha2 (Smp_142510.2)

		RIBOc	RIBOC DS	RM
1	500	1000		1577

Figura 12: Esquema dos domínios conservados, da região codificadora da enzima SmDrosha2 (Smp_142510.2).

Tabela 6A: Regiões de cobertura dos domínios conservados da enzima SmDrosha1 e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados CDD.

Domínio	Região de	Código de acesso	E-value
	cobertura (aa)	do domínio	(Blastp)
Ribonuclease III	905-1018	cd00593	1e-24
Ribonuclease III	1077-1212	cd00593	4e-24
DSRM	1218-1288	cd00048	1e-08

Tabela 6B: Comparação entre SmDrosha2 e suas ortólogas relacionando tamanho e similaridade.

Espécie	Ortóloga	Código de acesso	Tamanho	E-value
			(aa)	(Blastp)
S. mansoni	SmDrosha2	Smp_142510.2	1577	-
D. melanogaster	Drosha	FBgn0026722	1327	0
C. elegans	Drsh-1	WBGene00009163	1086	3e-105
H. sapiens	Drosha	AAF80558	1374	0

A família de Agonautas é representada em *S. mansoni* por 4 proteínas, SmAgo1, SmAgo2, SmAgo3 e SmAgo4 com seqüências completas e uma quinta provável com seqüência parcial apresentada mais abaixo nesta mesma seção Resultados. As Figuras 13, 14, 15 e 16 representam as seqüências SmAgo1, SmAgo2, SmAgo3 e SmAgo4, respectivamente. Os contigs depositado no banco de dados geneDB de cada uma estão mostrados nas Figuras. As ORFs preditas das 4 proteínas variam de 876 a 955 aminoácidos. As Tabelas 7A, 8A, 9A e 10A mostram os domínios conservados presentes, PIWI e PAZ característicos de todas as proteínas da família e suas similaridades com os consensos dos bancos de dados CDD e PFAM. Além disso, a SmAgo1 apresenta um domínio DUF1785, de função desconhecida, presentes nas Argonautas do tipo 1 como Ago1 (*D. melanogaster*) e Alg-1 (*C. elegans*). As Tabelas 7B, 8B, 9B e 10B mostram os ortólogos de SmAgos e uma comparação entre os tamanhos e sua similaridade. SmAgo1 atingiu o maior valor de similaridade (zero) em relação a suas ortólogas. SmAgo2, SmAgo3 e SmAgo4 apresentaram similaridades que variaram de 5e-99 a 3e-108, 4e-52 a 8e-62 e 1e-98 a 1e-101, respectivamente.

SmAgo1 (Smp_140010)

	DUF1785 PAZ	PIWI	
1	300	600	876

Figura 13: Esquema dos domínios conservados, da região codificadora da enzima SmAgo1 (Smp_140010).

Tabela 7A: Regiões de cobertura dos domínios conservados da enzima SmAgo1 e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados PFAM e CDD.

Domínio	Região de	Código de acesso	E-value
	cobertura (aa)	do domínio	(Blastp)
DUF1785	234-286	pfam08699	2e-16
PAZ	286-406	cd02846	2e-30
PIWI	451-834	cd04657	9e-140

Tabela	7B:	Comparação	entre	SmAgo1	e	suas	ortólogas	relacionando	tamanho	e
similario	dade.									

Espécie	Ortóloga	Código de acesso	Tamanho	E-value
			(pb)	(Blastp)
S. mansoni	SmAgo1	Smp_140010	876	-
D. melanogaster	Ago-1	FBgn0026611	984	0
C. elegans	Alg-1	WBGene00000105	1002	0
H. sapiens	EIF2-C.2	NP_036286	859	0

SmAgo2 (Smp_179320)



Figura 14: Esquema dos domínios conservados, da região codificadora da enzima SmAgo2 (Smp_179320).

Tabela 8A: Regiões de cobertura dos domínios conservados da enzima SmAgo2 e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados PFAM e CDD.

Domínio	Região de	Código de acesso	E-value
	cobertura (aa)	do domínio	(Blastp)
PAZ	363/471	pfam02170	1e-10
PIWI	510/927	cd04657	1e-120

Tabela	8B:	Comparação	entre	SmAgo2	e	suas	ortólogas	relacionando	tamanho	e
similario	dade.									

Espécie	Ortóloga	Código de acesso	Tamanho	E-value
			(pb)	(Blastp)
S. mansoni	SmAgo2	Smp_179320	955	-
D. melanogaster	Ago-1	FBgn0026611	984	7e-104
C. elegans	Alg-1	WBGene00000105	1002	5e-99
H. sapiens	EIF2-C.2	NP_036286	859	3e-108

SmAgo3 (Smp_102690.2)



Figura 15: Esquema dos domínios conservados, da região codificadora da enzima SmAgo3 (Smp_102690.2).

Tabela 9A: Regiões de cobertura dos domínios conservados da enzima SmAgo3 e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados PFAM e CDD.

Domínio	Região de	Código de acesso	E-value
	cobertura (aa)	do domínio	(Blastp)
PAZ	329/437	pfam02170	5e-09
PIWI	481/809	cd04657	2e-64

Espécie	Ortóloga	Código de acesso	Tamanho	E-value
			(pb)	(Blastp)
S. mansoni	SmAgo3	Smp_102690.2	854	-
D. melanogaster	Ago-1	FBgn0026611	984	8e-62
C. elegans	Alg-1	WBGene00000105	1002	4e-52
H. sapiens	EIF2-C.3	NP_079128	860	2e-58

Tabela 9B: Comparação entre SmAgo3 e suas ortólogas relacionando tamanho e similaridade.

SmAgo4 (Smp_102690.3)

		PAZ	PIW	/	
1	• •	300	600		924

Figura 16: Esquema dos domínios conservados, da região codificadora da enzima SmAgo4 (Smp_102690.3).

Tabela 10A: Regiões de cobertura dos domínios conservados da enzima SmAgo4 e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados PFAM e CDD.

Domínio	Região de cobertura	Código de acesso	E-value
	(aa)	do domínio	(Blastp)
PAZ	329/437	pfam02170	4e-09
PIWI	481/895	cd04657	3e-113

Tabela 10B: Comparação entre SmAgo4 e suas ortólogas relacionando tamanho e similaridade.

Espécie	Ortóloga	Código de acesso	Tamanho	E-value
			(pb)	(Blastp)
S. mansoni	SmAgo4	Smp_102690.3	924	-
D. melanogaster	Ago-1	FBgn0026611	984	1e-101
C. elegans	Alg-1	WBGene00000105	1002	8e-101
H. sapiens	EIF2-C.3	NP_079128	860	1e-98

A Figura 17 mostra o contig Smp_099630 depositado no banco de dados geneDB que representa a proteína SmFmr1 com uma ORF predita de 598 aminoácidos. A Tabela 11A mostra os domínio conservado presente, PRK11824 e sua similaridade com o consenso do banco de dados. A Tabela 11B mostra os ortólogos de SmFmr1,

uma comparação entre seus tamanhos e similaridade entre eles variando entre 5e-37 e 4e-43.

SmFmr1 (Smp_099630)



Figura 17: Esquema dos domínios conservados, da região codificadora da enzima SmFmr1 (Smp_099630).

Tabela 11A: Regiões de cobertura dos domínios conservados da enzima SmFmr1 e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados.

Domínio	Região de	Código de acesso	E-value
	cobertura (aa)	do domínio	(Blastp)
PRK11824	272-350	PRK11824	0.006

Tabela 11B: Comparação entre SmFmr1 e suas ortólogas relacionando tamanho e similaridade.

Espécie	Ortóloga	Código de	Tamanho	E-value
		acesso	(pb)	(Blastp)
S. mansoni	SmFmr1	Smp_099630	598	-
D. melanogaster	Fmr1	FBgn0028734	643	5e-37
C. elegans	-	-	-	-
H. sapiens	FXR1	AAC50155	621	4e-43

A Figura 18 mostra o contig Sm01663 representando a proteína SmTudor-SN sendo a única proteína da via oriunda de seqüências do transcriptoma (geneDB) com uma ORF de 1023 aminoácidos. A Tabela 12A mostra os domínios conservados presentes, SNc e Tudor e suas similaridades com o consenso dos bancos de dados CDD e PFAM. A Tabela 12B mostra os ortólogos de SmTudor-SN, seus tamanhos e uma similaridade que varia de 5e-159 a 0.

SmTudor-SN (Sm01663)



Figura 18: Esquema dos domínios conservados, da região codificadora da enzima SmTudor-SN (Sm01663).

do	s PFAM e CDD.			
	Domínio	Região de	Código de acesso	E-value
		cobertura (aa)	do domínio	(Blastp)
	SNc	24-164	cd00175	3e-22
	SNc	199-334	cd00175	4e-15
	SNc	364-507	cd00175	5e-16
	SNc	544-704	cd00175	3e-19
	TUDOR	788-893	pfam00567	4e-10

Tabela 12A: Regiões de cobertura dos domínios conservados da enzima SmTudor-SN e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados PFAM e CDD.

Tabela 12B: Comparação entre SmTudor-SN e suas ortólogas relacionando tamanho e similaridade.

Espécie	Ortóloga	Código de acesso	Tamanho	E-value
			(pb)	(Blastp)
S. mansoni	SmTudor-SN	Sm_01663	1023	-
D. melanogaster	Tudor-SN	FBgn0035121	926	5e-159
C. elegans	Tsn-1	WBGene00006626	914	3e-168
H. sapiens	SND-1	Q7KZF4	910	0

A Figura 19 mostra o contig Smp_087220 depositado no banco de dados geneDB representando a proteína SmPasha com uma ORF predita de 732 aminoácidos. A Tabela 13A mostra o domínio conservado presente, DSRM, e sua similaridade com o consenso do banco de dados CDD. A Figura 13B mostra os ortólogos de SmPasha, uma comparação entre seus tamanhos e a similaridade entre eles, que varia de 4e-13 a 4e-47



Figura 19: Esquema dos domínios conservados, da região codificadora da enzima SmPasha (Smp_087220).

Tabela 13A: Regiões de cobertura dos domínios conservados da enzima SmPasha e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados PFAM e CDD.

Domínio	Região de	Código de acesso	E-value
	cobertura (aa)	do domínio	(Blastp)
DSRM	443-507	cd00048	1e-08

Tabela 13B: Comparação entre SmPasha e suas ortólogas relacionando tamanho e similaridade.

Espécie	Ortóloga	Código de acesso	Tamanho	E-value
			(pb)	(Blastp)
S. mansoni	SmPasha	Smp_087220	732	-
D. melanogaster	PASHA	FBgn0039861	642	4e-47
C. elegans	Pash-1	WBGene00011908	790	4e-13
H. sapiens	DGCR8	BAB83032	773	3e-38

As Figuras 20 e 21 mostram os contig Smp_152800.1 e Smp_152800.1 depositados no banco de dados geneDB que representam as proteínas SmExportina5.1 e SmExportina5.2 com ORFs preditas de 1286 e 1026 aminoácidos, respectivamente. provindas de um mesmo DNA genômico como acontece com SmDrosha1 e 2 e SmAgo3 e 4. As Tabelas 14A e 15A mostram os domínios conservados presentes, CRM1 e Xpo1, e as similaridades com o consenso dos bancos de dadoa PFAM e COG. As Tabelas 14B e 15B mostram os ortólogos de SmExportina5.1 e SmExportina5.2 e uma comparação entre seus tamanhos e similaridade variando de 6e-35 a 1e-38 e 2e-28 a 7e-29, respectivamente.



Figura 20: Esquema dos domínios conservados, da região codificadora da enzima SmExportina5.1 (Smp_152800.1).

Tabela 14A: Regiões de cobertura dos domínios conservados da enzima SmExportina5.1 e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados PFAM e COG.

Domínio	Região de	Código de acesso	E-value
	cobertura (aa)	do domínio	(Blastp)
CRM1	19-186	COG5101	9e-06
Xpo1	110-231	pfam08389	8e-08

Tabela 14B: Comparação entre SmExportina5.1 e suas ortólogas relacionando tamanho e similaridade.

Espécie	Ortóloga	Código de	Tamanho	E-value
		acesso	(pb)	(Blastp)
S. mansoni	SmExportina5.1	Smp_152800.1	1286	-
D. melanogaster	RanBP21	FBgn0031051	1241	6e-35
C. elegans	-	-	-	-
H. sapiens	Exportin 5	NP_065801	1204	1e-38

SmExportina5.2 (Smp_152800.2)



Figura 21: Esquema dos domínios conservados, da região codificadora da enzima SmExportina5.2 (Smp_152800.2).

Tabela 15A: Regiões de cobertura dos domínios conservados da enzima SmExportina5.1 e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados PFAM e COG.

Domínio	Região de	Código de acesso	E-value
	cobertura (aa)	do domínio	(Blastp)
CRM1	19-186	COG5101	8e-06
Xpo1	110-231	pfam08389	1e-07

Espécie	Ortóloga	Código de	Tamanho	E-value
		acesso	(pb)	(Blastp)
S. mansoni	SmExportin5.2	Smp_152800.2	1021	-
D. melanogaster	RanBP21	FBgn0031051	1241	2e-28
C. elegans	-	-	-	-
H. sapiens	Exportin 5	NP_065801	1204	7e-29

Tabela 15B: Comparação entre SmExportina5.2 e suas ortólogas relacionando tamanho e similaridade.

A Figura 22 mostra o contig Smp_023670 depositado no banco de dados geneDB representando a proteína SmLoquacious com uma ORF de 356 aminoácidos. A Tabela 16A mostra os domínios conservados presentes, DSRMs, e suas similaridades com o consenso do banco de dados CDD. A Tabela 16B mostra os ortólogos de SmLoquacious e uma comparação entre os tamanhos e suas similaridade, variando de 6e-12 a 3e-22

SmLoquacious (Smp_023670)



Figura 22: Esquema dos domínios conservados, da região codificadora da enzima SmLoquacious (Smp_023670).

Tabela 16A: Regiões de cobertura dos domínios conservados da enzima SmLoquacious e seus respectivos alinhamentos característicos com os domínios consenso dos bancos de dados CDD.

Domínio	Região de	Código de acesso	E-value
	cobertura (aa)	do domínio	(Blastp)
DSRM	9-74	cd00048	6e-10
DSRM	126-193	cd00048	1e-08

Tabela 16B: Comparação entre SmLoquacious e suas ortólogas relacionando tamanho e similaridade.

Espécie	Ortóloga	Código de	Tamanho	E-value
		acesso	(pb)	(Blastp)
S. mansoni	SmLoquacious	Smp_023670	356	-
D. melanogaster	Loquacious	FBgn0032515	465	3e-22
C. elegans	-	-	-	-
H. sapiens	TARBP2	NP_004169	345	6e-12

4.1.2 Análise das ESTs

Abaixo, as Tabela de 17 a 28 mostram os alinhamentos, utilizando o algoritmo BLASTn, das prováveis seqüências de cDNA das proteínas da via de miRNA do parasito com as ESTs do banco de dados do transcriptoma da FAPESP (ONSA) e o tamanho da seqüência de EST, sendo representada por um singlets ou por um contig. Além disso, foi também representado em cada Tabela o estágio evolutivo do parasito que o clone da seqüência EST foi seqüenciado. *SmLoquacious* não apresentou sequências similares no banco de dados do transcriptoma.

A porcentagem relativa das ESTs representadas no banco de dados da FAPESP (ONSA) referentes às proteínas da via de miRNA e os estágios que elas pertencem, está mostrada nos anexos 4 e 5.

Tabela 17: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmDicer1 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Código de acesso do	Tamanho (pb) do	Estágios	Região de	E-value
contig ou singlet	contig ou singlet	evolutivos	cobertura	(Blastn)
C713374.1	468	1(L)	5436-5902	0
C604739.1	728	2(L), 1(S)	6048-6351	1e-171
C705307.1	242	1(G)	1168-1409	1e-134
C716281.1	281	1(A)	121-318	1e-100

Tabela 18: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmDrosha1 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Código de acesso do	Tamanho (pb) do	Estágios	Região de	E-value
contig ou singlet	contig ou singlet	evolutivos	cobertura	(Blastn)
C602470.1	875	4(L), 1(G)	1-564	0
C606975.1	755	2(M), 2(L), 1(E)	1905-2454	0
C700852.1	370	1(S)	3674-4043	0
C710110.1	145	1(A)	1531-1677	8e-66
C717779.1	150	1(A)	390-533	6e-48

Tabela 19: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmDrosha2 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Código de acesso do	Tamanho (pb) do	Estágios	Região de	E-value
contig ou singlet	contig ou singlet	evolutivos	cobertura	(Blastn)
C606975.1	755	2(M), 2(L),	1862-2592	0
		1(E)		
C602470.1	875	4(L), 1(G)	1- 564	0
C700852.1	370	1(S)	3812- 4181	0
C710110.1	145	1(A)	1531- 1677	8e-66
C717779.1	150	1(A)	390- 533	7e-48
Tabela 20: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmAgo1 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Código de acesso do	Tamanho (pb) do	Estágios	Região de	E-value
contig ou singlet	contig ou singlet	evolutivos	cobertura	(Blastn)
C719260.1	495	1(S)	696-1038	0
C719260.1	495	1(S)	574-695	2e-56
C705711.1	460	1(L)	1324-1580	1e-143
C705058.1	463	1(L)	1175-1325	2e-80
C716122.1	381	1(S)	1680-1818	3e-73
C716122.1	381	1(S)	1829-1938	6e-56

Tabela 21: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmAgo2 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Código de acesso do	Tamanho (pb) do	Estágios	Região de	E-value
contig ou singlet	contig ou singlet	evolutivos	cobertura	(Blastn)
C603705.1	1294	30(A), 15(F), 12(L), 2(E), 1(G)	572-1841	0
C608655.1	979	7(A),2(L)	1954-2868	0
C604197.1	3071	10(E), 9(L), 5(F), 4(A), 2(S), 1(G)	925-1484	0
C601842.1	510	3(E)	371-683	1e-174
C601842.1	510	3(E)	905-1108	1e-112
C612177.1	515	2(L), 1(F)	258-578	1e-167
C612177.1	515	2(L), 1(F)	1-101	4e-48
C702511.1	329	1(S)	1013-1151	2e-44

Tabela 22: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmAgo3 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Código de acesso do	Tamanho (pb) do	Estágios	Região de	E-value
contig ou singlet	contig ou singlet	evolutivos	cobertura	(Blastn)
C604197.1	3071	10(E), 9(L), 5(F), 4(A), 2(S), 1(G)	1-2413	0
C604197.1	3071	10(E), 9(L), 5(F), 4(A), 2(S), 1(G)	2409-2565	6e-84
C603705.1	1294	30(A), 15(F), 12(L), 2(E), 1(G)	823-1382	0
C702511.1	329	1(S)	755-1049	1e-66
C601842.1	510	3(E)	823-1006	3e-45

Tabela 23: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmAgo4 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Código de acesso do	Tamanho (pb) do	Estágios	Região de	E-value
contig ou singlet	contig ou singlet	evolutivos	cobertura	(Blastn)
C604197.1	3071	10(E), 9(L), 5(F), 4(A), 2(S), 1(G)	1-2775	0
C603705.1	1294	30(A), 15(F), 12(L), 2(E), 1(G)	823-1382	0
C702511.1	329	1(S)	755-1049	1e-66
C608655.1	979	7(A),2(L)	2332-2332	1e-66
C601842.1	510	3(E)	823-1006	3e-45

Tabela 24: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmFmr1 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Código de acesso do	Tamanho (pb) do	Estágios	Região de	E-value
contig ou singlet	contig ou singlet	evolutivos	cobertura	(Blastn)
C611449.1	1981	14(A), 9(L), 8(F), 8(E), 3(S), 1(G)	227-1797	0
C608108.1	566	2(E), 1(F), 1(S)	185-569	0
C608108.1	566	2(E), 1(F), 1(S)	570-678	2e-55

Tabela 25: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmTudor-SN relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Código de acesso do	Tamanho (pb) do	Estágios	Região de	E-value
contig ou singlet	contig ou singlet	evolutivos	cobertura	(Blastn)
C605675.1	2944	13(E), 10(A), 4(L), 4(S), 3(F), 1(G), 1(C)	1944-3452	0
C605675.1	2944	13(E), 10(A), 4(L), 4(S), 3(F), 1(G), 1(C)	730-1523	0
C605675.1	2944	13(E), 10(A), 4(L), 4(S), 3(F), 1(G), 1(C)	542-731	1e-101
C706365.1	597	1(E)	209-802	0

Tabela 26: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmPasha relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Código de acesso do	Tamanho (pb) do	Estágios	Região de	E-value
contig ou singlet	contig ou singlet	evolutivos	cobertura	(Blastn)
C704682.1	478	1(L)	1253-1730	0
C604888.1	364	2(C)	724-1090	0

Tabela 27: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmExportina5.1 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Código de acesso do	Tamanho (pb) do	Estágios	Região de	E-value
contig ou singlet	contig ou singlet	evolutivos	cobertura	(Blastn)
C606377.1	1002	5(L), 3(E), 2(F), 1(G), 1(S)	2527-3125	0
C606377.1	1002	5(L), 3(E), 2(F), 1(G), 1(S)	3148-3544	0
C711114.1	187	1 (S)	1659-1845	1e-101
C706220.1	581	1 (F)	1180-1307	2e-66

Tabela 28: Contig ou singlet do banco de dados da FAPESP de *S. mansoni* similares à proteína SmExportina5.1 relacionando tamanho de cada um, posição e os estágios evolutivos que foram obtidos. A, estágio evolutivo vermes adultos, M, estágio evolutivo macho adulto, F, estágio evolutivo fêmea adulta, E, estágio evolutivo ovos, L, estágio evolutivo, miracídios, G, esporocistos, C, cercária e S, esquistossômulos.

Código de acesso do	Tamanho (pb) do	Estágios	Região de	E-value
contig ou singlet	contig ou singlet	evolutivos	cobertura	(Blastn)
C606377.1	1002	5(L), 3(E), 2(F), 1(G), 1(S)	2527-2997	0
C711114.1	187	1 (S)	1659-1845	1e-102
C706220.1	581	1 (F)	1180-1307	1e-66

4.1.3 Análise Filogenética das RNAse III e Argonautas

Para construção da árvore da família RNAse III foi selecionado o domínio conservado endoND2 (RIBOc II) das proteínas SmDicer1, SmDrosha1, SmDrosha2 e suas homólogas Dicer1-símile, Dicer2-símile e Drosha-símile em outros organismos (Figura 23).



Figura 23 – Árvore filogenética consenso baseada na sequências de aminoácidos do domínio endoND2. A construção da árvore e a análise do "bootstrap" foram realizadas nos programas ClustalX 2,0 e MEGA 4.0. Para árvore consenso e confiabilidade dos ramos formados foi utilizado o teste de filogenia bootstrap, utilizando 1000 réplicas para cada sequência, sendo 60% o mínimo para se considerar o ramo confiável. O primeiro colchete representa as proteínas da subfamília Dicer1, o segundo da subfamília Dicer2 e a terceira da subfamília Drosha. A barra representa o número de aminoácidos que foram substituídos de uma sequencia para outra. Os códigos de acesso referentes às seqüencias estão mostrados na seção Materiais e Métodos.

Para família das Argonautas foi selecionado o domínio conservado PIWI das proteínas SmAgo1, SmAgo2, SmAgo3, SmAgo4 e suas homólogas Ago-símile e PIWIsímile de outros organismos. A correlação evolutiva está sendo enfatizada, pela presença da classificação Bilateria, Eubilateria e Coelomata proposta por Hausdorf em 2000.



Figura 24 – Árvore filogenética consenso baseada na seqüência de aminoácidos do domínio PIWI. A construção da árvore e a análise do "bootstrap" foram realizadas nos programas ClustalX 2.0 e MEGA 4.0. Para árvore consenso e confiabilidade dos ramos formados foi utilizado o teste de filogenia bootstrap, utilizando 1000 réplicas para cada sequência, sendo 60% o mínimo para se considerar o ramo confiável. O primeiro colchete representa as proteínas da subfamília Ago, o segundo proteínas órfãs sem definição de posicionamento na árvore e a terceira da subfamília PIWI. A barra representa o número de aminoácidos que foram substituídos de uma seqüência para outra. Os códigos de acesso referentes às seqüencias estão mostrados na seção Materiais e Métodos.

4.1.4 Análise dos domínios conservados PIWI e endoND

O domínio conservado completo PIWI predito das seqüências SmAgo1, 2 e 4 foi alinhado com sua seqüência de aminoácidos com seus ortólogos de *C. elegans, D. melanogaster, Mus musculus, H. sapiens* e *A. thaliana* utilizando o programa de alinhamento ClustalX 2.0. Os membros de Argonauta no parasito, representados por SmAgo1, 2 e 4, exibiram um domínio PIWI bastante conservado, contendo os 3 resíduos chaves (D/D/H), coordenados pelo Mg²⁺ com função intrínseca de Slicer (RNAse H) (Figura 25).



Figure 25: Alinhamento do domínio PIWI e apresentação dos resíduos catalíticos responsáveis pela função Slicer. Alinhamento entre as seqüências de aminoácidos do domínio conservado PIWI presente nas proteínas Argonautas das espécies *C. elegans, D. melanogaster, Mus musculus, H. sapiens, A. thaliana* e *S. mansoni*. Está marcado com um bloco e indicado por uma seta os 3 resíduos catalíticos conservados, dois ácidos aspárticos e uma histidina (D/D/H), determinantes da função intrínseca Slicer (RNAseH) das proteínas da família Argonauta. Os códigos de acesso referentes às seqüencias estão mostrados na seção Materiais e Métodos.

Os domínios endoND são estritamente conservados em todos membros da família RNAse III. Em *S. mansoni* esta família está representada por SmDicer1, SmDrosha1 e SmDrosha2. Os domínios endoND preditos destas seqüências foram alinhados com seus ortólogos de *C. elegans, D. melanogaster, Mus musculus e H. sapiens* utilizando o programa de alinhamento ClustalX 2.0. A Figura 26 mostra claramente a conservação entre os domínios dos ortólogos a presença dos resíduos catalíticos conservados E/D e D/E responsáveis pela função ribonucleásica das

RNAse III	endoND1	endoND2
Dicer1		
Hs.Dicer1	erl E mlg D D cv E	QRL E FLG D D IF E
Dm.Dicer1	erl E tig D D cv E	QRL E FLG D D VF E
Ce.DCR1	erf E tig D D av E	QRL E FLG DD IF E
Sm.Dicer1	erm E tig D D cv E	QRL E FLG DD IF E
Drosha		
Hs.Drosha	erl E flg D	QRM E FLG D D LL E
Dm.Drosha	erl E flg D	QRL E FLG D D LL E
Ce.Drsh1	erl E ylg D	QRL E WLG D D LV E
Sm.Drosha1/2	erl E flg D	QRL E FLG D D LL E

proteínas. Nossos resultados confirmam a ausência de 2 resíduos no domínio endoND1 de SmDrosha1 e 2 como em seus ortólogos.

Figura 26: Alinhamento dos dominios endoND da familia RNAse III e apresentação dos resíduos catalíticos. Alinhamento entre as seqüências de aminoácidos dos domínios conservados endoND1 e endoND2 das espécies *H. sapiens*, *D. melanosgaster*, *C. elegans*, e *S. mansoni*. Está em letra maior e mais escuro os 4 resíduos catalíticos conservados, E/D e D/E, determinantes da função nucleásica das proteínas da família RNAse III. Os códigos de acesso referentes às seqüencias estão mostrados na seção Materiais e Métodos.

4.1.5 Análise de SmAgo4

A proteína SmAgo4 está melhor representada no banco de dados de ESTs (FAPESP) pelo contig C604197.1 e genômico no banco de dados geneDB pela sequência Smp_102690.3.

A Figura 27 representa um esquema do DNA genômico possuindo 13381 pares de base, cDNA 3102 pares de base e uma ORF predita de 924 aminoácidos com os domínios conservados PIWI e PAZ determinates para sua função. O cDNA mostrado nos bancos de dados geneDB e FAPESP representa apenas a região traduzida ficando de fora a região não menos importante 5' e 3' não traduzida.

A Figura 27 mostra também um *splicing* na região 5' não traduzida, seguindo a regra canônica de emenda de éxons e retirada de íntrons GT....AG. Os íntrons representados se encontram em tamanhos que variam de 30 até 3375 pares de bases e os éxons de 27 a 457 pares de base. Além disso observamos o sítio de poliadenilação na região 3' não traduzida com o sinal consenso AATAAA, para cauda de Poli-A, localizado a 171 pares de base do códon TGA.



Figura 27: Esquema do DNA genômico, cDNA e proteína de SmAgo4. A Figura mostra o DNA genômico contendo 13381 pares de base, divididos em 17 íntrons em azul e 19 éxons em verde. O seu cDNA contendo 3102 pares de bases e a proteína 924 aminoácidos com os domínios conservados PIWI em azul e PAZ em vermelho. A seta azul mostra o splicing na região 5' não traduzida e a seta vermelha o sinal consenso de cauda poli-A na região 3' não traduzida.

4.1.6 miRNAs preditos

A Figura 28 mostra a estrutura do pré-miRNA contendo aproximadamente 80 nucleotídeos e as duas preditas estruturas dos miRNAs 1A em vermelho e 1A* em azul, com 20 nucleotídeos cada. A estrutura predita apresenta um dG (energia livre) teórico de -30,2 kcal/mol.



Figura 28: Estrutura do pré-miRNA seus prováveis miRNAs processados. Está predito pré-miRNA foi processada no programa disponível web Mfold 3.0 gerando uma estrutura secundária com um delta G igual a -30,2 kcal/mol. Em vermelho e azul estão os preditos miRNAs 1A e 1A*, respectivamente.

4.2 Clonagem e seqüenciamento parcial da SmDicer1

Para a validação parcial dos resultados *in silico* dos genes de *SmDicer1*, foram desenhados oligonucleotídeos compreendendo a região codificadora dos domínios

conservados endoND. A Figura 29 mostra a análise de um RT-PCR a a partir de vermes adultos apresentando um produto amplificado de 469pb. Este produto foi clonado e 3 colônias independentes foram seqüenciadas.

A seqüência consenso foi obtida utilizando o programa de alinhamento ClustalX 2.0. A comparação da seqüência obtida com a depositada no banco dados da FAPESP (Anexo 1) mostra uma troca de T por C no consenso.



Figura 29: Amplificação por RT-PCR utilizando oligonucleotídeos específicos para SmDicer1. cDNA parcial do domínio conservado endoND1 de SmDicer de tamanho 469pb utilizando como molde cDNA de vermes adultos. Na primeira canaleta padrão de peso molecular de 100 pb (Invitrogen). *Gel de agarose 1,2% e corado com brometo de etídio.

4.3 Clonagem e seqüenciamento parcial de SmAgo5

Após RT-PCR e PCR convencional (molde DNA genômico) para o gene *SmAgo5* utilizando um par de iniciadores (primers) que amplificava tanto no cDNA quanto no DNA genômico, ou seja, estando presentes em éxons, obtivemos produtos amplificados de 711 e 1283 pares de base, respectivamente (Figura 30). Estes produtos foram clonados em várias réplicas e seqüenciados.



Figura 30: Amplificação por RT-PCR e PCR convencional (molde DNA genômico) utilizando oligonucleotídeos específicos para *SmAgo5*. cDNA parcial do domínio conservado PIWI de *SmAgo5* de tamanho 711pb utilizando como molde cDNA de vermes adultos. DNA genômico parcial de SmAgo5 de tamanho 1283 pares de base. Na primeira canaleta padrão de peso molecular de 100 pb (Invitrogen). $*5\mu$ L do produto de PCR foram analisados em gel de agarose.

Para geração da seqüência consenso tanto de cDNA quanto de DNA genômico foi utilizado o programa de alinhamento ClustalX 2.0. As seqüências consenso de cDNA e DNA genômico formadas a partir das seqüências geradas no seqüenciamento estão mostradas no anexo 2 e 3.

O produto seqüenciado consenso cDNA não alinhou em 100% dos nucleotídeos com as seqüências de cDNA depositadas nos bancos de dados da FAPESP e geneDB. Na seqüência consenso do cDNA está inserido uma seqüência íntrônica de 80 pb, funcionando neste caso como éxon, que segue a regra canônica de splicing, GT...AG, como mostra a Figura 31. Para as seqüências *SmAgo3* (Smp_102690.2) e *SmAgo4* (Smp_102690.3), depositadas no banco de dados geneDB, e para seqüência C604197.1, depositada na FAPESP, este fragmento inserido na seqüência consenso representa um íntron, por isso, o não aparecimento deste nos mRNAs já processados.

PRIMER I	5	Inserção no cDNA consenso
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	*********************** CGATGATTTGGCCT CGATGATTTGGCCT	ACCENACEANCENCENCE ACTEACECATETICAAAATGGGGATCCCCACACAGTGTGTGCTGGTAGCAC ACCENACAATCAAACGTCTAAGTGATTTGAAAATGGGGATCCGCACACAGTGTGTTCCTGGTAGCAC ACCENACAATCAAACGTCTAAGTGATTTGAAAATGGGGATCCGCACACAGTGTGTGCGGAAGAA
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	TCTACCCAAACCTA TCTACCCAAACCTA TCTACCCAAACCTA TCTACCCAAACCTA	***** Ar green cet agoi <mark>arct treatte att atte at atte gree atter ta tree attert ta aaaceaategach</mark> ar green te ar green tega a <mark>go atet te atte atte attega attec attert attec aattet ta aaaceaatega t</mark> eg <mark>a c</mark> h
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	AATTTAGTCTAGTT	T <mark>CTUTTACCCAACCTATTATTCAAACTCAATGGGAACTTGCCGCTGUTAATTGGATTCCCAGAT CCAACCTATTATTCAAACTCAATGGGAAACTTCCTCCTCTTAATTGGATTCTCCCAGAT TCTTTTACCCAACCTATTATTGAAACTCAATGGGAAACTTGCTGATTAATTGGATTGTCCCAGAT</mark>
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	**************************************	CAA GCATTT AAT AAT GCTTTTT CGCTGET GATGTT ACTCA CCCA CCACACACACACACACACACACACAC
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	**************************************	T CROCHT CHT COATET CH CHCETCA GTE AAT COCAT AC GOGGOT AT A CHT A CACAACAA GOGAC T GCCGT T T T GGAT CT CACHT CA GTE AAT COCAT AC GOGGOT AT A CHT A CACAACAA GOGAC T CROCHT CHT COATETCA CHCETCA CHT AAT COCAT AC GOGGOT AT A CHT A CACAACAA GOGAC
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	**************************************	ACAAAACAACGA GA GAAATTATTGATAACTTGCATCTGTCACHTGCACGACAACTGCTCAG Acaaagaa ga
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	CCTTTAAATTCAAT	get at ct actor of tact act act ct caact att attent of at cotto att attents to caact
CONSENSO cDNA_FAPESP* DNAg_GeneDB**		
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	**************************************	ctototo acacedo a cette ca anatototte ce con cetta not coca teto acacedo a seconda ca a cotte coa Ctoto a cacego da cette ca anatototte ce coa cate a toto coa teta acacede a Ctoto a acecedo a cette ca anatototte ce ca cate a toto coa teta acacedo a
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	**************************************	GOR GAAGAA OCH GOR ANN AG ANNI ANG GIN GIGGA GAAGGGAGAGAGAN ATGOGANT AAAAGGAGAA GGIGAAGAACCH GON ANN AGANNY ANG GIN GIGGA GAAGGGAGAGAGAN ATGOGANT AAAAGGAGAA GGIGAAGAACCH GIN ANN AGANNY ANG GIN GIGGA GAAGGGAGAGAGAN ATGOGANT AAAAGGAGAA
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	GATCCCAG GATCCCAG GATCCCAGGTAATT	acad caaan the annoise a cracer and i cru an gebeech a course ad cours anno coancer a
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	GCTATTCGGATAAC	CCCACCET ACCACT ROTCACCCACET CACACACET CACT ATCACACACET
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	TACTACAAGGTTCA	COLIGE I TT AT GT AC ACAAAAAA AC ACCEPTT INT ACTANTE AAA CA <mark>CCETT</mark> CA CAT TER A T ACTA CO ACT AT GT
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	T <mark>AAdAGCCGATCA</mark> G	G <mark>accht Git Antit a</mark> ct act <mark>an antit act acta chita acta anticett catta conteccut chita acta anticett catta conteccut chita</mark>
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	TATGAGCCTCTCTA	GA CITIT ATT CI AA GOOGACOT AAC ACT AT GAT AAAAT CICIT A GACATOT GAT GOACHTTITT ATTITT
CONSENSO cDNA_FAPESP* DNAg_GeneDB**	CACACATACTACAT	**************************************
*C60 ** Sm	14197.1 p_102690.3	

Figura 31: Alinhamento entre as seqüências de DNA genômico do geneDB e cDNA consenso (*SmAgo5*) e cDNA C604197.1 da FAPESP. As seqüências alinhadas possuem o domínio PIWI característico das proteínas Argonautas. As setas com os blocos delimitam os iniciadores F e R do fragmento amplificado e o bloco com linhas tracejadas representa o íntron/éxon inserido na seqüência consenso de cDNA.

4.4 Splicing Alternativo

O splicing alternativo pôde ser observado duas vezes nas seqüências *SmAgos* representados pela mesma seqüência de DNA genômico. Um utilizamos seqüências

completas depositadas no banco de dados de *S. mansoni* para defini-lo, *SmAgo3* e *SmAgo4*. Outro, utilizamos cDNA consenso parcialmente seqüenciado, provável 5^a Argonauta da família no parasito (*SmAgo5*);

A seqüência de DNA genômico Smp_102690 possui duas seqüências de cDNA no banco de dados do geneDB, Smp_102690.2 e Smp_102690.3, e uma parcialmente seqüenciada neste trabalho, *SmAgo5*. O alinhamento completo entre estas seqüências está mostrado em anexo. Estas três tem propriedade de splicing alternativo na região 3', sendo que 2 íntrons ou parte deles funcionam ora funciona como íntron, e ora como éxon como mostra a Figura 32.



Figura 32: Alinhamento entre as seqüências de cDNA consenso(*SmAgo5*), *SmAgo3* e *SmAgo4*. As setas com os blocos delimitam os *splicing* alternativos de *SmAgo3*, 4 e cDNA consenso parcialmente seqüenciado, *SmAgo5*, sendo o primeiro de 80 nt diferenciando *SmAgo5*, das outras duas e o segundo de 93 nt diferenciando SmAgo3 das outras duas.

A representação no banco de dados a partir de DNA genômico pode ser confirmada com dados do transcriptoma da FAPESP. A Figura 33 mostra leituras de diferentes clones no contig C604197.1 do banco da FAPESP sendo um deles marcado por quadros pretos representando um mau alinhamento. Este clone mau alinhado, denominado MS1-0053 representa o *splicing* alternativo de 93 aminoácidos que separa *SmAgo3* e *SmAgo4*.



Figura 33: Splicing alternativo diferenciando *SmAgo3* e *SmAgo4* mostrado no banco da FAPESP. A seta com o bloco verde aponta e delimita a seqüência consenso do contig da FAPESP C604197.1. A seta com o bloco vermelho aponta e delimita o singlet MS1-0053 da FAPESP. O alinhamento mostra as diferenças entre as seqüências e o local do splicing alternativo sendo o clone MS1-0053 similar a sequência *SmAgo3*.

A Figura 34 também mostra leituras de diferentes clones no contig C604197.1 do banco da FAPESP sendo um deles marcado por quadros pretos representando um mau alinhamento. Este clone mau alinhado, denominado ML1-0048T representa um splicing alternativo de 80 aminoácidos que separa cDNA consenso parcialmente seqüenciado, *SmAgo5*, de *SmAgo3* e *SmAgo4*.



Figura 34: *Splicing* alternativo diferenciando *SmAgo3* e *SmAgo4* do consenso de cDNA obtido neste trabalho (*SmAgo5*) mostrado no banco da FAPESP. A seta com o bloco verde aponta e delimita a seqüência consenso do contig da FAPESP C604197.1. A seta com o bloco vermelho aponta e delimita o singlet ML1-0048T da FAPESP. O alinhamento mostra as diferenças entre as seqüências e o local do splicing alternativo sendo o clone ML1-0048T similar ao consenso de cDNA, *SmAgo5*.

4.5 Análise da expressão dos genes que codificam SmDicer1 e SmAgo2/3/4 utilizando qRT-PCR

Foi analisada a expressão relativa dos mRNAs de *SmAgo2/3/4* e *SmDicer1* por PCR quantitativa em tempo real como descrito em Materiais e Métodos. Foram desenhados oligonucleotídeos específicos e próprios para análise desta PCR, compreendendo a região codificadora dos domínios conservados endoND1 para *SmDicer1*, gerando um amplificado de 289 pares de bases e PIWI para *SmAgo2/3/4*, gerando um amplificado de 321 pares de bases. Previamente a qRT-PCR, os fragmentos foram amplificados, clonados e seqüenciados para validação experimental. A análise da

expressão destes genes foi executada nos seguintes estágios do parasito: verme adulto, cercária, esquistossômulo mecanicamente transformado de 3,5 horas (EMT-3,5), 8,5 horas (EMT-8,5), 18,5 horas (EMT-18,5), 24 horas (EMT-24), 48 horas (EMT-48), 72 horas (EMT-72) e ovos (Gráfico 1). Os dados foram normalizados utilizando o gene α -tubulina. Como padrão endógeno relativo foi escolhido esquistossômulo de 24 horas (EMT-24) que possue o menor índice de transcrição real. Foi notado tanto para *SmDicer1* quanto para *SmAgo2/3/4* expressão em todos os estágios sendo seus valores significativamente diferentes com P>0,001.

A Figura 35 mostra um aumento significativo nos níveis de mRNA de SmDicer1 depois da transformação de cercária em esquistossômulo. Em ovos detectamos a maior expressão deste gene, sendo estes 30 vezes mais abundantes quando comparados com cercárias e 3 vezes comparados com vermes adultos.



Figura 35: Expressão do gene *SmDicer1* no desenvolvimento de *S. mansoni*. A expressão de SmDicer1 foi avaliada nos estágios de cercárias, esquistossômulos mecanicamente transformados (EMT-3,5, EMT-8,5, EMT-18,5, EMT-24, EMT-48, EMT-72 horas), vermes adultos e ovos. Como normalizador (gene constitutivo) foi utilizado o gene da α -tubulina e a expressão gênica relativa foi determinada pelo método do 2^{- $\Delta\Delta$ Ct}, utilizando como calibrador o estágio EMT-24 horas.

A Figura 36 mostra o padrão de expressão de *SmAgo2/3/4* nos mesmos estágios descritos anteriormente. Observamos um padrão de expressão gênica diferencial com uma queda expressiva nos níveis de mRNA entre cercária e EMT-8,5, aproximadamente 15 vezes. Em EMT-18,5 horas detectamos um aumento real de transcrição (3,5 vezes) e 9 vezes após 48 horas. Após este período, foi observada uma queda em EMT-72 horas

(5 vezes). Além disso, não detectamos diferença estatisticamente significativa entre os níveis de *SmAgos2/3/4* em vermes adultos comparados com cercária. Corroborando com os resultados descritos acima de *SmDicer1*, em ovos houve a maior expressão destes genes, comparados com os outros estágios estudados.

Em conjunto estes resultados evidenciaram dois momentos em que a síntese predominou (EMT-18,5 horas e EMT-48,0 horas) sobre a degradação. E três momentos onde a degradação predominou sobre a síntese EMT-3,5, EMT-8,5 e EMT-72 horas.



Figura 36: Expressão do gene *SmAgo2/3/4* no desenvolvimento de *S. mansoni*. A expressão de *SmAgo2/3/4* foi avaliada nos estágios de cercárias, esquistossômulos mecanicamente transformados (EMT-3,5, EMT-8,5, EMT-18,5, EMT-24, EMT-48, EMT-72 horas), vermes adultos e ovos. Como normalizador (gene constitutivo) foi utilizado o gene da α -tubulina e a expressão gênica relativa foi determinada pelo método do 2^{- $\Delta\Delta$ Ct}, utilizando como calibrador o estágio EMT-24 horas.

Discussão

5-Discussão

A descoberta do mecanismo de silenciamento da expressão gênica através de pequenos RNAs e a elucidação e caracterização de proteínas envolvidas nesta via, em diferentes organismos, constituem, nos últimos anos, uma das maiores descobertas científicas visto que repercutiram por diversas áreas distintas como biologia, medicina, genômica funcional, computação, matemática, química, dentre outras (Meister e Tuschl, 2004). A fita dupla de RNA, geradora destes pequenos RNAs e unificadora dos processos do silenciamento gênico, induz uma resposta celular levando à degradação ou ao silenciamento dos mRNAs homólogos. Este processo é chamado de silenciamento gênico pós-transcricional ou PTGS (*posttranscriptional gene silencing*) e atualmente categorizado em três vias: de silenciamento mediada por miRNA, de silenciamento mediada por siRNA e nuclear associada a metilação de DNA (Bartel, 2004).

Apesar desta diversidade de vias e funções intermediárias totalmente diferentes, o objetivo final é o mesmo, o silenciamento da expressão gênica. Sendo assim, existem alguns pontos comuns e conservados entre as 3 vias e entre diferentes organismos sugerindo então a existência de um mecanismo ancestral comum (Zamore, 2002).

Neste trabalho buscamos enfatizar a via de miRNA, envolvida no silenciamento de mRNAs endógenos atuando principalmente no controle de genes envolvidos no desenvolvimento do organismo sendo um alvo de estudo promissor tanto para o avanço da biologia celular do organismo quanto para o possível controle da parasitose. Para o processamento destes miRNAs existe um grupo complexo de proteínas presentes em dois compartimentos: o núcleo e o citoplasma (Bartel, 2004).

O primeiro objetivo específico deste trabalho foi elucidar e caracterizar *in silico* a via completa no parasito. Para isto as seqüências de aminoácidos das proteínas ortólogas, cuja via está bem descrita, como os organismos modelos *D. melanogaster, C. elegans* e *H. sapiens*, foram alinhadas, utilizando o algoritmo BLASTp, com seqüências previamente anotadas de *S. mansoni*. Esta análise permitiu a identificação dos principais constituintes da maquinaria de miRNA depositadas no banco de dados geneDB do parasito (Tabela 3).

Durante estas análises também foi possível verificar algumas particularidades espécie-específica. Por exemplo, *C.elegans*, não apresentou os ortólogos para Fmr1, Loquacious ou Exportina-5. Considerando os dados experimentais que ressaltam a importância da via de miRNA em *C. elegans* a ausência destas proteínas pode ser compensada pela existência de outras que exerçam a mesma função. Além desta espécie, também comparamos os resultados obtidos em *S. mansoni* com *S. japonicum*

(Tabela 3). Estas análises mostraram várias seqüências com similaridades significativas, sugerindo a presença desta via, também neste organismo.

Recentemente, vários grupos independentes mostraram que vários estágios do parasito podem incorporar RNA exógeno e processá-los com RNA anti-senso. Esta maquinaria de processamento compartilha alguns componentes da via de síntese endógena de miRNAs. Analisados em conjunto, os experimentos com esporocistos, esquistossômulos, e vermes adultos e os obtidos neste trabalho, sugerimos que o silenciamento gênico mediado por pequenos RNA tem um papel importante na regulação da expressão gênica do parasito (Boyle e cols.., 2003; Skelly e cols., 2003; Dinguirard, 2006; Osman e cols., 2006). Entretanto, apesar de documentada a importância desta via no parasito, várias questões ainda permanecem por ser esclarecidas como, por exemplo: o padrão de expressão de todos componentes deste sistema, o repertório de genes de miRNA, o silenciamento estágio-específico mediado por esta via bem como seus alvos naturais.

Desta forma, uma das abordagens deste trabalho foi à busca por seqüências preditas envolvidas na via de miRNA depositadas nos bancos de dados de domínio público. Esta análise nos permitiu verificar que estas proteínas não eram conservadas apenas no tamanho, mas também mantinham a disposição dos domínios conservados em sua estrutura primária. Os valores obtidos, quando utilizamos a ferramenta BLASTp, comparando os domínios conservados consenso, CDD (Conserved Domain Database), PFAM e COG com as seqüências recuperadas de *S. mansoni*, ficou evidenciada sua similaridade (Figuras 10 a 22). Com relação às seqüências primárias de aminoácidos, observamos uma variação de no máximo 30% no tamanho das proteínas.

Além das características protéicas também foi nosso objetivo verificar quais estágios contribuíram com seqüências para a montagem das ORFs preditas. Para isso, utilizamos os dados do transcriptoma. Esta busca de ESTs é importante, pois é um dado físico real utilizado para a validação de seqüências de mRNA, que foram geradas a partir do seqüenciamento genômico. Considerando que o banco geneDB utiliza os programas Glimmer e Twinscan para mostrar sequências de exóns e íntrons, a presença de ESTs cobrindo estas regiões corrobora com os dados computacionais (Banco de dados da FAPESP, ONSA).

Neste trabalho foram utilizadas seqüências do banco de dados da FAPESP, composto por 125 mil seqüências, sendo 27% de vermes adultos, 15% de ovos, 15% de miracídios, 13% de esporocistos, 8% de cercária e 22% de esquistossômulos de 7 dias

correspondendo a 92% do transcriptoma do organismo para validar as seqüências de cDNA das prováveis proteínas encontradas no banco de dados do geneDB (Tabelas 17 a 28) (Verjovski-Almeida e cols., 2003).

A manutenção das características estruturais das proteínas da via são importantes para demonstrar a sua conservação. A literatura mostra que as proteínas da família RNAse III e Argonauta são elementos chaves na via de silenciamento mediada por miRNA. Desta forma, nosso próximo objetivo foi à utilização de ferramentas computacionais para localização dos resíduos conservados envolvidos com as funções catalíticas, nas proteínas destas famílias (Figuras 26 e 27).

Além disso, é possível estabelecer uma boa correlação utilizando árvores filogenéticas como parâmetro. As árvores consenso (Figura 23 e 24) foram construídas utilizando o método Neighbor-Joining a partir das seqüências de aminoácidos dos domínios conservados RIBOc (endoND2) e PIWI das proteínas RNAse III e Argonautas, respectivamente, presentes em diversos organismos.

A proximidade filogenética dos Bilateria, principalmente àqueles semelhantes evolutivamente, é ainda um pouco confusa. A relação filogenética entre os grupos Bilateria, Platelmintos (*S. mansoni*), Nematodos (*C. elegans*), Artrópodes (*D. melanogaster*) e Cordados (*H. sapiens*) diferem tanto em árvores baseadas em dados morfológicos, quanto em árvores baseadas nas seqüências do RNA ribossomal 18S (Nielsen, 1995; Aguinaldo, 1997.). Hausdorfs em 2000, utilizando seqüências de 10 genes nucleares chegou a uma árvore filogenética consenso que relaciona estes grupos Bilateria, o qual corrobora com os resultados da família Argonauta obtidos neste trabalho (Figura 24).

Também vale ressaltar que a família das proteínas Argonautas é bastante diversa possuindo variações nos números de parálogos de uma espécie para outra. Para exemplificar esta característica, temos um representante apenas desta família em *Schizosaccharomyces pombe*, aproximadamente vinte e sete em *C. elegans* (Carmell e cols., 2002), dez em *A. thaliana* (Hunter e cols., 2003), cinco em *D. melanogaster* (Williams e cols., 2002) e oito em humanos (Sasaki e cols., 2003). Atualmente, acredita-se que a diversidade gênica observada seja devido à duplicação gênica e splicing alternativo (Hutvagner, 2008). Em *S. mansoni*, até o momento, encontramos cinco prováveis representantes desta família sendo quatro com seqüências completas no banco de dados geneDB, SmAgo1, SmAgo2, SmAgo3 e SmAgo4, e uma, SmAgo5, seqüenciada parcialmente neste trabalho. Os resultados obtidos sugerem a presença de 3

genes e 5 produtos protéicos sugerindo que a diversidade protéica é mediada por *splicing* alternativo. A ocorrência deste processo nesta via, pode ser evidenciado pela presença das proteínas Ago1 e Alg-2 de *D. melanogaster* e *C. elegans*, respectivamente, que possuem 3 seqüências parálogas provindas de uma mesma seqüência genômica (FlyBase - <u>http://flybase.bio.indiana.edu/</u> and WormBase - <u>http://www.wormbase.org/</u>). Em *S. mansoni* existem vários exemplos deste processo, como exemplo podemos citar o fator de crescimento epidermal (Shoemaker, 1992), a proteína ligante de ácidos graxos Sm14 (Ramos, 2003) e o receptor de serina/treonina quinase do tipo II SmRK2 (Forrester, 2004) que a partir de uma mesma sequência genômica gera 3 transcritos. A comprovação do *splicing* alternativo nos membros da família Argonauta de proteínas neste parasito está sendo realizado no nosso laboratório.

Estas evidências nos motivaram a analisar o perfil de expressão das duas principais proteínas da via de miRNA, SmDicer1 e SmAgos, no parasito. Para esta investigação foram utilizados os seguintes estágios de vida: vermes adultos, ovos, cercárias e esquistossômulos cultivados in vitro por 3,5; 8,5; 18,5; 24; 48 e 72 horas. O tempo de cultivo in vitro foi baseado em dados anteriores de nosso grupo que sugerem que durante a transição cercária-esquistossômulo não há uma taxa real de transcrição, nas primeiras 24 horas de cultivo in vitro (Guerra-Sá, 2004). Além disso, trabalhos de proteômica desenvolvidos também pelo nosso grupo sugerem que 72 horas após a transformação cercária em esquistossômulo todas as proteínas oriundas no estágio de cercária foram degradadas (Castro-Borges, 2005). Estes resultados corroboram com os dados experimentais que mostraram a cinética de síntese de proteínas em cercária, medida pela taxa de incorporação de S³⁵ é praticamente nula, sugerindo uma inibição da tradução. Experimentos equivalentes realizados com esquistossômulos mantidos in vitro mostraram que a taxa de síntese protéica em esquistossômulo com 8 dias de cultivo é 11 vezes maior quando comparado a esquistossômulos com 24 horas (Harrop Wilson, 1993; Castro-Borges, 2007)

Para estudar o papel da via de miRNA no silenciamento gênico em *S. mansoni* nós analisamos por qRT-PCR duas proteínas chaves da via de miRNA, SmDicer1 e SmAgo2/3/4, sugerindo a atuação desta via no bloqueio dos mRNAs em esquistossômulos jovens. Essas três proteínas Argonautas possuem uma região consenso presente no domínio PAZ (Figura 36) e SmDicer1 uma seqüência cobrindo parte do domínio endoND1 (Figura 35) escolhidas para a análise de qRT-PCR. Os níveis destes mRNAs foram normalizados com um gene constitutivo endógeno, alfa-

tubulina, e em seguida comparados a um padrão endógeno relativo, EMT-24 horas. A expressão de *SmDicer1* e *SmAgos* foi verificada em todos estágios estudados, sendo todos valores encontrados estatisticamente significantes.

5.1 Família das RNAse III

A família RNAse III representa uma família de proteínas com função de clivagem de RNAs de fita dupla. Estas estão presentes em eucariotos e procariotos representadas por 4 subfamílias como mostrado por Blaszczyk e colaboradores em 2004. Nosso estudo abordou as subfamílias encontradas em eucariotos Dicer e Drosha. sendo que a atuação de cada uma acontece em compartimentos diferentes, Drosha no núcleo, e Dicer no citoplasma. Em *S. mansoni* estas proteínas são representadas por SmDicer1, SmDrosha1 e SmDrosha2.

A proteína SmDicer1, representante de Dicer 1 no parasito, mostrou alta similaridade e tamanho bem semelhante às proteínas ortólogas dos organismos metazoários (Tabela 4B). Além disso, SmDicer1 contém todos os domínios característicos de Dicer 1 como indicado na Figura 10, corroborando os resultados demonstrados pelo grupo de Krautz-Peterson, 2007. A proteína Dicer estudada por este grupo difere apenas no N-terminal da seqüência identificada no banco geneDB e apresentada neste trabalho. Estas diferenças observadas foram, provavelmente, devido a sequência do geneDB estar incompleta.

As proteínas SmDrosha, mostraram alta similaridade e semelhança em tamanho com as proteínas ortólogas dos organismos metazoários (Tabela 5B e 6B). Os domínios endoND1 e endoND2, responsáveis pela função da proteína, foram bem similares ao consenso do banco de dados de domínios conservados.

A clivagem da fita dupla de RNA é dependente dos domínios endoND1 e endoND2 e os 4 resíduos E/D e D/E conservados essencial para atividade endonucleásica. As proteínas Dicer1-símile e Drosha-símile de *D. melanogaster, C. elegans* e *H.sapiens e S. mansoni*, foram alinhadas apresentando grande conservação, tanto nos domínios endoND1 e endoND2 quanto nos 4 resíduos específicos E/D e D/E, como visto na Figura 26. Além disso os nosso resultados sugerem que SmDrosha1 e SmDrosha2 possuem os resíduos conservados E/D e D/E, nas mesmas posições que seus ortólogos, e os resíduos vizinhos, compartilham 100% de identidade entre *S. mansoni* e *D. melanogaster*.

Além da similaridade, dados filogenéticos reforçam o agrupamento das proteínas RNA III de *S. mansoni* em suas respectivas subfamílias. Podemos observar na Figura 23 que na raiz da árvore um ancestral comum divergiu nas subfamílias Drosha-Símile e Dicer-Símile. Posteriormente as subfamílias provavelmente por mecanismo de duplicação gênica divergiram em Dicer1-Símile e Dicer2-Símile. A proteína SmDicer1 agrupou dentro do ramo de Dicer1-Símile corroborando com resultados demostrados por Krautz-Peterson em 2007. O mesmo aconteceu com SmDrosha1 e SmDrosha2 agrupando-se no ramo característico de proteínas Drosha-Símile. Novos experimentos estão sendo realizados em nosso laboratório para comprovar a hipótese que SmDrosha 1 e 2 são produtos de splicing alternativo.

A conservação, durante a evolução das espécies, da seqüência, dos resíduos funcionais dos sítios ativos e da própria função e disposição do domínio conservado demonstra que as proteínas desta família são importantes para manutenção da via de miRNA ativa nos processos biológicos do organismo.

5.2 Família das Argonautas

A característica marcante na família de proteínas Argonautas é a presença dos domínios PAZ e PIWI em sua estrutura (Hutvagner, 2008). O domínio PAZ (**P**iwi/**A**rgonauta/**Z**wile) identificado nas proteínas Argonautas e Dicer, consiste de 130 aminoácidos. A função deste domínio é mediar o complexo de formação entre as proteínas Argonautas e Dicer formando um hetero-dímero (Yan, 2003; Song, 2004). PIWI tem a função de direcionar dsRNA para hidrólise. A determinação da estrutura em cristal das Argonautas revelou que PIWI se aproxima funcionalmente das RNAse H. A diferença é que as RNAse H atuam utilizando o quarteto catalítico Asp-Asp-Glu/Asp coordenando 2 íons magnésio para sua atividade nucleásica. Já as proteínas Argonautas, representadas pelo domínio PIWI, utilizam uma tríade Asp-Asp-His coordenando somente um íon magnésio (Ma, 2004; Tolia e cols, 2007). A Figura 25 enfatiza a presença desta função no parasito representada pelos resíduos Asp-Asp-His no domínio PIWI das proteínas SmAgo1, SmAgo2 e SmAgo4. A seqüência de SmAgo3 apresenta

apenas os resíduos Asp-Asp, pois seu carboxi-terminal foi reduzido provavelmente por *splicing* alternativo.

Gostaríamos de ressaltar a notável similaridade da *SmAgo1* com seus ortólogos (Tabela 7 B) o mesmo não sendo observados para seus parálogos. Entre as seqüências *SmAgo2*, *SmAgo3* e *SmAgo4* a similaridade com os ortólogos e a similaridade com o consenso do banco de dados de domínios conservados são bem semelhantes. Isto é devida a existência de regiões idênticas nas seqüências, supondo a presença de um gene ancestral comum.

Transcritos de SmAgo3, SmAgo4 e SmAgo5 provêm da mesma seqüência genômica, evidenciando que a diversidade dos membros desta família de proteínas esta relacionado ao mecanismo de splicing alternativo. Os dois splicing alternativos acima citados são confirmados com os dados do transcriptoma da FAPESP. O splicing alternativo de SmAgo3 e SmAgo4 é mostrado nas Figuras 32 e 33, representando 93 pares de base de diferença no splicing e 210 no total devido a antecipação do códon de terminação TAG devido a mudança de leitura (frame) para frente. Na Figura 33 fica evidenciado este splicing alternativo pela presença da seqüência do clone no transcriptoma da FAPESP MS1-0053. Este clone representando SmAgo3 não alinhou com o contig consenso da FAPESP C604197.1, devido a ausência de um fragmento que no caso em SmAgo3 funciona como íntron e em SmAgo4 e C604197.1 funciona como éxon. O splicing alternativo separando a provável quinta Argonauta da família no parasito, SmAgo5, de SmAgo3 e SmAgo4 é mostrado nas Figuras 32 e 34. Este splicing alternativo ocorre também na porção carboxi-terminal, sendo um íntron de SmAgo3 e SmAgo4 de 80 nucleotídeos funcionando como éxon em SmAgo5. Na Figura 34 este splicing é confirmado pela presença do clone no transcriptoma da FAPESP ML1-0048T. A sequência do clone da fase de miracídio não alinha com o contig da FAPESP C604197.1, pois este está representando esta parte como um íntron já retirado antes da montagem do contig.

As Argonautas de *S mansoni* foram separadas em 2 subfamílias: PIWI-Símile e AGO-Símile. Para confirmação desta separação foi comparada a separação filogenética dos Bilateria. A relação filogenética entre os grupos Bilateria, Platelmintos (*S. mansoni*), Nematodos (*C. elegans*), Artrópodes (*D. melanogaster*) e Cordados (*H. sapiens*) diferem tanto em árvores baseadas em dados morfológicos, quanto em árvores baseadas nas seqüências do RNA ribossomal 18S (Nielsen, 1995, Aguinaldo e cols., 1997). Hausdorfs em 2000 utilizando seqüências de 10 genes nucleares chegou a uma

árvore filogenética consenso que relaciona estes grupos Bilateria. A árvore filogenética (Figura 24) consenso construída a partir da seqüência de aminoácidos do domínio conservado PIWI das proteínas Argonautas de S. mansoni, C. elegans, D. melanogaster, H. sapiens e outras 6 espécies corroboraram a árvore consenso de Hausdorfs separando Bilateria, Eubilateria e Coelomata. A separação entre PIWI-Símile e AGO-Símile ocorre em resposta a diferenças no domínio PIWI e na proteína fazendo com que esta interaja com diferentes pequenos RNAs (Yigit, 2006; Carmell, 2002, Seto, 2007). Existem algumas proteínas que possuem uma distância evolutiva semelhante não se enquadrando em nenhuma das duas subfamílias, sendo estas proteínas órfãs. As proteínas SmAgo1, SmAgo2, SmAgo3 e SmAgo4 se enquadraram na sub-família AGO-Símile. Proteínas da subfamília AGO interagem principalmente com miRNAs e típicos siRNAs, diferentemente das proteínas da subfamília PIWI que interagem com piRNAs e rasiRNAs (siRNAs não típicos). As proteínas SmAgo2, SmAgo3 e SmAgo4 estão mais próximas evolutivamente entre si que a proteína SmAgo1. A diferença entre estas proteínas não está só na estrutura primária do domínio PIWI, mas também em um domínio denominado DUF1785, presente em SmAgo1, que possue função desconhecida, e se encontra também presente em ortólogos como Ago1 de D. melanogaster e Alg-1 de C. elegans.

Uma proteína Argonauta bem estudada pelo nosso grupo foi a proteína SmAgo4. Ela está representada nos bancos de dados da FAPESP pelo contig C604197.1 e pela seqüência Smp_102690.3 no banco de dados geneDB. A Figura 27 mostra a sequência de DNA genômico que codifica esta proteina com 13381 pares de base, 17 íntrons e 19 éxons. O *splicing* na extremidade 5' não traduzida seguiu a regra canônica GT / AG confirmada pelos dados de transcriptoma. Na extensão da região 3' não traduzida do cDNA de 3102 pares de bases encontramos o sinal consenso de poli adenilação, posicionado 171 pares de base do sinal de terminação TAG. A confirmação das seqüências completas das outras Argonautas é alvo de posteriores estudos do nosso grupo.

Tanto as proteínas RNAse III quanto as proteínas Argonautas estão bem representada no banco de dados do transcriptoma sendo expressas em todos estágios do parasito (Tabela 17 a 23). Isto demonstra a importância destas famílias nos processos biológicos do parasito independente do estágio em que se encontra.

Mesmo sendo presente em todos os estágios do parasito, confirmado pelos dados do transcriptoma, o estudo destas sequencias nas fases de transição, por exemplo

cercaria-esquistossômulos, é importante para evidenciar a contribuição desta via nos mecanismos de adaptação do parasito aos diferentes ambientes. A fase de transição cercária-esquistossômulo até 72 horas enfatiza uma característica importante: a diferenciação prevalece sobre a morfogênese no parasito. Esquistossômulos jovens são caracterizados pelo estado de semi-quiescência sem síntese real de proteína (Lawson e Wilson) e com um arraste no ciclo celular (Clegg, 1965) sugerindo um remodelamento da estrutura da cromatina. Lawson e Wilson, 1980, mostraram que esta fase representava apenas o elongamento do organismo sem aumento da massa corporal. Esta situação mostra o desacoplamento entre diferenciação e morfogênese refletindo a mudança brusca de ambiente e da forma corporal específica de esquistossômulo e posteriormente verme adulto. A baixa taxa de síntese de proteína foi também verificada por Nagai e cols. em 1977 nos esquistossômulos e relacionada ao uso de proteínas présintetizadas na fase de cercária (Hockley, 1973). Em termos de proteoma, os esquistossômulos adquirem aparato protéico após o terceiro dia de cultivo in vitro quando todas as proteínas de cercárias são degradadas como mostrado por Castro-Borges em 2007. Se não ocorre síntese real de proteínas de alguma forma os mRNAs estão sendo bloqueados. Blanton e Licate, em 1992, sugeriram que esta baixa síntese de proteínas nos estágios jovens de esquistossômulos não era devido a uma queda nas quantidades de mRNAs, mas sim devido a um bloqueio pós-transcricional da tradução.

A expressão de *SmDicer1* em verme adulto, cercária e ovos confirmaram os resultados apresentados por Krautz-Peterson e cols. em 2007. Entretanto, foi mostrado que em esquistossômulos jovens ocorre a presença de *SmAgos* e *SmDicer1* aumenta até 48 horas.

Estes dados sugerem a atuação da via de miRNA na inibição da tradução de mRNAs nesta fase de transição cercaria-esquistossomulo. Além disto, SmAgos e SmDicer1 possuíram alta expressão de seus mRNAs na fase de ovos. Sugerimos que esta alta expressão em ovos seja devido a ação desta via no bloqueio de mRNAs maternos reprimindo sua tradução ou promovendo sua degradação (Schier, 2007). Nosso grupo esta trabalhando para verificar esta hipótese. Os resultados mostram que as 2 proteinas são diferencialmente expressas durante o ciclo. Gostaríamos de ressaltar que este é o primeiro trabalho que analisa o perfil de expressão gênica temporal durante a diferenciação de cercária em esquistossômulo.

A mimetização do estado *in vivo* do parasito na forma de esquistossômulos cultivados *in vitro* possui duas linhas: uma que aceita o processo como uma

91

mimetização real da biologia do organismo levando em consideração os aspectos bioquímicos e fisiológicos e outra que defende a idéia que fatores diversos no hospedeiro, impedem esta mimetização. Tentamos neste trabalho mimetizar ao máximo este ambiente do hospedeiro, com temperatura, concentração de gases e meio de cultura contendo hormônios e outras substâncias importantes para sua sobrevivência. O RNA destas preparações foram utilizados também para analisar o padrão de expressão de outras vias metabólicas, como por exemplo SUMO E3ligase, COP9 signalossomo, proteassoma 20S mostrando uma correlação direta com a atividade bioquímica das proteínas codificadas por estes genes (Cabral, 2008).

5.3 Outras proteínas envolvidas na via de miRNA

Além das proteínas da família RNAse III e Argonautas estudamos mais 5 proteínas envolvidas na via de miRNA. Os alinhamentos com seus respectivos homólogos metazoários foram bem expressivos com valores, para maioria dos alinhamentos, aproximando-se no máximo de um e^{-5} (Tabelas 4A a 16A). Valores bem próximos de zero, e-values menores que e^{-100} , também foram encontrados. Tudo isto mostra mais uma vez que a via é conservada em *S. mansoni*.

As proteínas envolvidas em RISC são proteínas auxiliares da proteína principal Argonauta para execução do complexo. Experimentos mostram diversas composições de RISC em diversos organismos sendo que em *D. melanogaster* as principais proteínas que o compõe são VIG dFMR1 (proteína relacionada ao X frágil), Tudor-SN (nuclease), helicases, proteínas diversas Argonautas (Caudy, 2002; Caudy, 2003). Em humanos, além dessas proteínas associadas à RISC, duas helicases importantes se ligam ao complexo, Gemin3 e Gemin4 (Meister e Tuschl, 2004).

A proteína Tudor-SN (*tudor staphylococcal nuclease*) obteve e-values bastante expressivos quando comparada a seus ortólogos, chegando a 3e-168, como mostra a Tabela 12B. Tudor-SN (*tudor staphylococcal nuclease*) é uma proteína que contém 4 ou 5 domínios SNc determinantes da sua atividade nucleásica, e um domínio Tudor (Figura 18). Estes domínios mostraram-se bastante similares aos domínios consensos do banco de dados como mostra a Tabela 12A. Tudor-SN é um componente de RISC onde primeiramente foi reconhecido um domínio com função nucleásica contribuindo para ação de RISC na degradação dos mRNAs alvos.

SmFmr1 mostrou-se bastante similar a suas ortólogas de *D. melanogaster* e *H. sapiens*, apresentando valores de e-values de 5e-37 a 4e-43, respectivamente (Tabela 11B). Quando observamos o alinhamento entre os domínios consenso dos bancos de dados e o domínio conservado da proteína, observamos um valor alto para e-value, devido a pequena cobertura correlacionada entre os domínio.

Duas prováveis proteínas ligantes de fita dupla de RNA, auxiliares das proteínas RNAse III, SmDrosha e SmDicer1, foram encontradas em S. mansoni. Estas desempenham papéis importantes na via, sendo que a ausência destas provoca o acúmulo de pré-miRNA ou pri-miRNA na célula (Saito, 2005). A auxiliar de Drosha, proteína ligante de fita dupla de RNA, conhecida em D. melanogaster e C. elegans como Pasha, tem sua correspondente em S. mansoni, SmPasha. (Denli, 2004; Gregory, 2004; Han, 2005). A auxiliar de Dicer-1 em D. melanogaster, Loquacious, tem uma paráloga, R2D2, que se liga a Dicer-2. Em S. mansoni, até o momento, identificamos representante de Loquacious, SmLoquacious, mesmo porque, também identificamos apenas ortólogas de Dicer-1, SmDicer1 (Saito, 2005). Tanto SmLoquacious quanto SmPasha apresentaram o domínio DSRM, que tem função de ligação a fita dupla de RNA, na sua estrutura primária, bem conservado (Figuras 19 e 22; Tabelas 13A e 16A). Além disso, a conservação entre os ortólogos foi bastante expressiva mostrando valores de alta similaridade como entre os ortólogos de D. melanogaster Pasha, 4e-47, e Loquacious, 3e-22. Em C. elegans não foi encontrado ortólogo para SmLoquacious com similaridade significativa. Mas a literatura mostra que em C. elegans quem desempenha esta função é a proteína Rde-4, não sendo encontrado em S. mansoni ortólogos desta proteína (Grishok, 2005)

Expotina-5 é uma proteína nuclear conservada em diversos organismos com diversas funções sendo a principal exportar pequenos RNAs não codificantes assim como precursores de miRNA (Shibata, 2006). Em *S. mansoni* foi identificada duas proteínas SmExportina5.1 e SmExportina5.2 provindas de uma mesma seqüência de DNA genômico, mas com diferentes mRNAs, processadas diferentemente pela propriedade de splicing alternativo. Nosso grupo está trabalhando para validação deste splicing. Estas proteínas apresentaram sinilaridade alta com ortólogas de *D. melanogaster e H. sapiens* além de significativos valores de similaridade entre os domínios conservados da proteína e os domínios conservados consenso, mostrando grande conservação desta proteína no parasito.

Todas proteínas, exceto SmLoquacious, apresentaram ESTs correlacionadas em diversos estágios do parasito no banco de dados do transcriptoma da FAPESP, confirmando a presença destas em *S. mansoni*.

Baseado nos resultados apresentados neste trabalho a via de miRNA em S. mansoni é composta por 13 proteínas, como resumido na Figura 37.



Repressão da tradução

Figura 37: Elucidação da via de miRNA em *S. mansoni*. Prováveis proteínas envolvidas na via de silenciamento gênico mediada por miRNA em *S. mansoni*.

5.4 MiRNAs preditos

A alta similaridade entre as proteínas ortólogas envolvidas na via de processamento de miRNAs em *S. mansoni*, *D. melanogaster*, *C. elegans* e *H.sapiens* ficou evidenciada. O mesmo não aconteceu com as seqüências de miRNAs homólogas

em diversos organismos mantidas na evolução. É o caso da família de miRNA let-7 que é encontrada em diversos organismos e não foi encontrada em S. mansoni até o momento como D. melanogaster, C. elegans e H. sapiens e a família lin-4 que é encontrada em C. elegans e H. sapiens e também não foi encontrada no parasito (Lim, 2003b). Na busca por miRNAs homólogos, como let-7 e lin-4, no genoma do S. mansoni, não obtivemos resultados satisfatórios até o momento. A única seqüência provável que possui características para preditos miRNAs encontrada até o momento esta mostrada na Figura 28, não possuindo homologia com miRNA de outra espécie. Para as buscas destes miRNAs preditos levamos em conta algumas considerações, tais como: tamanho do grampo (hairpin) formado, a necessidade de no mínimo 50 nucleotídeos integrados entre si na estrutura secundária; distância da haste do grampo em relação à volta (loop) (com no mínino 21 nucleotídeos em cada extremidade); e a característica mais expressiva: estabilidades termodinâmicas do grampo, utilizando um ·G de energia livre de formação de -25 kcal/mol. Esta estabilidade termodinâmica máxima adotada foi baseada na termodinâmica das estruturas secundárias geradas por pré-miRNAs das espécies D. melanogaster (-25 kcal/mol), e C. elegans (-23 kcal/mol) que possuem uma maquinaria de miRNA próxima em termos evolutivos com o S. mansoni (Lim, 2003; Lai, 2003).

Na Figura 28 podemos observar que a estrutura do miRNA predito e seu precursor no parasito, se enquadram nas características abordadas como grampo com uma haste longa bem alinhada com as extremidades 5' e 3'terminal, pouco estáveis e um loop para fechamento do grampo. Além disso, a estrutura secundária do RNA de fita dupla com uma energia livre de formação espontânea de -30,2 kcal/mol, tornando a estrutura bastante estável. Os miRNAs preditos encontrados no parasito foram denominados miRNA1A e miRNA1A*. Novos experimentos estão sendo realizados para conhecer os níveis de expressão deste miRNA durante a transição cercária-esquistossômulo.

Conclusão

6- Conclusão

Os principais resultados obtidos neste trabalho foram:

- 1- Identificação *in silico* de 14 prováveis proteínas envolvidas na via de silenciamento mediada por miRNA, sendo uma sem ORF completa;
- 2- Identificação de prováveis *splicing* alternativos, ocorrendo com um membro nuclear, SmDrosha, um membro de transporte núcleo-citoplasma, SmExportina5, e um membro citoplasmático, SmAgo.
- 3- Distância evolutiva da família Argonauta e RNAse III no parasito em relação aos ortólogos;
- 4- Expressão diferencial de *SmDicer1* e *SmAgo2/3/4* são na fase de transição cercária-esquistossômulo e com maior abundância em ovos.

Diante destes resultados obtidos podemos concluir que a via de processamento de miRNA em *S. mansoni* é bastante conservada, mas, requer mais estudos para seu entendimento completo, podendo assim, auxiliar futuramente, na elucidação dos mecanismos envolvidos na regulação gênica pós-transcricional, proliferação celular, apoptose, sinalização e diferenciação no parasito.

7-BIBLIOGRAFIA
- Adams, M.D.; Kelley, J.M.; Gocayne, J.D.; Dubnick, M.; Polymeropoulos, M.H.; Xiao, H.; Merril, C.R.; Wu, A.; Olde, B.; Moreno, R.F. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252, 1651–1656.
- Aguinaldo, A. M. A.; Turbeville, J. M.; Linford, L. S.; Rivera, M. C.; Garey, J. R.; Raff, R. A.; Lake, J. A. (1997). Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387:489–493.
- Bartel, D.P. (2004). MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell, 116, 281–297.
- Baum, B.; Craig, G. (2004). RNAi in a postmodern, postgenomic era. Oncogene 23, 8336–8339
- 5. Bentwich, I. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. Nat. Genet. 37, 766–770.
- Berezikov, E.; Guryev, V.; van de, B.J.; Wienholds, E.; Plasterk, R.H., Cuppen, E. (2005). Phylogenetic shadowing and computational identification of human microRNA genes. Cell 120, 21–24.
- Blanton, R.; Loula, E.C.; Parker, J. (1987). Two heat-induced proteins are associated with transformation of *Schistosoma mansoni* cercariae to schistosomula. *Proc. Nat. Acad. Sci. USA* 84(24), 9011-9014.
- 8. Blanton, R.E., Licate, L.S., (1992). Developmental regulation of protein synthesis in schistosomes. Mol. Biochem. Parasitol. 51, 201-208.
- Blaszczyk, J.; Gan, J.; Tropea, J.E.; Court, D.L.; Waugh, D.S.; Ji, X. (2004). Noncatalytic Assembly of Ribonuclease III with Double-Stranded RNA. Structure 12, 457-466.
- Boyle, J.P.; Wu, X.J.; Shoemaker, C.B.; Yoshino, T.P. (2003). Using RNA interference to manipulate endogenous gene expression in Schistosoma mansoni sporocysts. Molecular and Biochemical Parasitology 128, 205–215.
- 11. Brindley, P.J. (2005). The molecular biology of schistosomes. *Trends Parasitol.*;21(11), 533-536.
- Brindley, P.J.; Pearce, E.J. (2007). Genetic manipulation of schistosomes. International Journal for Parasitology 37, 465–473.
- Carmell, M.A.; Xuan, Z.; Zhang M.Q.; Hannon, G.J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes e Dev*. 16, 2733-2742

- 14. Castro-Borges, W. (2005). Diversidade de proteassoma 20S e perfil de ubiquitinação durante o desenvolvimento do parasita Schistosoma mansoni: uma abordagem proteômica. Tese defendida no Programa de Pós-Graduação do Departamento de Bioquímica da Universidade de São Paulo (FMRP), como parte integrante dos requisitos para obtenção do título de Doutor em Ciências – Área de Concentração: Bioquímica.
- Castro-Borges, W.; Cartwright, J.; Ashton, P.D.; Braschi, S.; Guerra-sa, R., Rodrigues, V.; Wilson, R.A.; Curwen, R.S. (2007). The 20S proteasome of Schistosoma mansoni: A proteomic analysis. Proteomics. (7):1065-1075..
- 16. Caudy, A. A.; Ketting, R.F.; Hammond, S.M.; Denli, A.M.; Bathoorn, A.M.P; Tops, B.B.J.; Silva, J.M.; Myers, M.M.; Hannon, G.J.; Plasterk, R.H.A. (2003). A micrococcal nuclease homologue in RNAi effector complexes. *Nature* 425, 411–414.
- Caudy, A. A.; Myers, M.; Hannon, G. J.; Hammond, S. M. (2002). Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev.* 16, 2491–2496.
- 18. Chitsulo, L.; Engels, D.; Montresor, A.; Savioli, L. (2000). The global status of schistosomiasis and its control. *Acta Trop.* **77**, 41-51.
- 19. Clegg, J.A. (1965). In vitro cultivation of *schistosoma mansoni*. *Exp Parasitol.*,16:133-147.
- 20. Coelho, M.V. (1970). O parasito *Schistosoma mansoni*. In: CUNHA, A.S. *Esquistossomose mansoni* 1:1-12 São Paulo: Editora Savier.
- 21. Cunha, A. S. (1970). Esquistossomose mansoni. USP, 1970, 13-39.
- 22. DeMarco, R.; Oliveira, K.C.; Venancio, T.M.; Verjovski-Almeida, S: (2006). Gender biased differential alternative splicing patterns of the transcriptional cofactor CA150 gene in Schistosoma mansoni. Mol Biochem Parasitol, 150, 123-131.
- Denli, A. M., Tops, B. B., Plasterk, R. H., Ketting, R. F. and Hannon, G. J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature 432(7014): 231-235.
- Denli, A.M.; Hannon, G.J. (2003). RNAi: An ever-growing puzzle. Trends in Biochemical Sciences 28, 196-201.
- Dillon, G.P.; Feltwell, T.; Skelton, J.P.; Ashton, P.D.; Coulson, P.S.; Quail, M.A.; Nikolaidou-Katsaridou, N.; Wilson, R.A.; Ivens, A.C. (2006). Microarray

analysis identifies genes preferentially expressed in the lung schistosomulum of Schistosoma mansoni. Int J Parasitol, 36, 1-8.

- 26. Dinguirard, N.; Yoshino, T.P.; (2006). Potential role of a CD36-like class B scavenger receptor in the binding of modified low-density lipoprotein (acLDL) to the tegumental surface of Schistosoma mansoni sporocysts. Molecular and Biochemical Parasitology 146, 219–230.
- 27. Doench, J.G.; Petersen, C.P.; Sharp, P.A. (2003). siRNAs can function as miRNAs. Genes e Development 17, 438-442.
- Dugas, D.V; Bartel, B. (2004). MicroRNA regulation of gene expression in plants. Plant Biology, 7 (5), 512-520.
- 29. El-Sayed, N.M; Bartholomeu D.; Ivens, A.,; Johnston, D.A.; LoVerde, P.T. (2004). Advances in schistosome genomics. *Trends Parasitol*. 20(4), 154-157.
- 30. Fire, A.; Xu, S.; Montgomery, M.K.; Kostas, S.A.; Driver, S.E. e Mello, C.C. (1998). Potent and specific genetic interference by doublestranded RNA in *Caenorhabditis elegans*. Nature 391, 806-811.
- Fitzpatrick, J.M.; Hoffmann , K.F. (2006). Dioecious Schistosoma mansoni express divergent gene repertoires regulated by pairing. Int. J Parasitol, 36, 1081-1089.
- 32. Fitzpatrick. J.M.; Johnston, D.A.; Williams, G.W.; Williams, D.J.; Freeman, T.C.; Dunne, D.W.; Hoffmann, K.F. (2005). An oligonucleotide microarray for transcriptome analysis of Schistosoma mansoni and its application/use to investigate gender-associated gene expression. Mol Biochem Parasitol, 141, 1-13.
- 33. Forrester, S.G; Warfel, P.W.; Pearce, E.J. (2004). Tegumental expression of a novel type II receptor serine/threonine kinase (SmRK2) in Schistosoma mansoni. Molecular and Biochemical Parasitology 136, 149-156.
- 34. Franco, G.R.; Adams, M.D.; Soares, M.B.; Simpson, A.G.J.; Venter, J.C.; Pena, S.D. (1995). Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. Gene;152, 141–147.
- 35. Gregory, R. I., Yan, K. P., Amuthan, G., Chendrimada, T. and Doratotaj, B. (2004) The Microprocessor complex mediates the genesis of microRNAs. Nature 432: 235-240.

- 36. Gregory, R.I.; Chendrimada, T.P.; Shiekhattar, R. (2006). MicroRNA biogenesis: isolation and characterization of the microprocessor complex. Methods Mol. Biol. 342, 33-47.
- Grishok, A. (2005). RNAi mechanisms in Caenorhabditis elegans. FEBS Letters 579, 5932–5939
- Guerra-Sá, R. (2004). Genômica funcional do Schistosoma mansoni. Projeto de Pós-Doutorado, USP-FMRP.
- 39. Han, J., Lee, Y., Yeom, K. H., Kim, Y. K. and Jin, H. (2005). The Drosh-DGCR8 complex in primary microRNA processing. Genes Dev 18: 3016-3027.
- 40. Harrop R.; Wilson, R.A. (1993). Protein synthesis and release by cultured schistosomula of Schistosoma mansoni. Parasitology.107 (3), 265-274.
- 41. Hausdorf, B. (2000). Early Evolution of the Bilateria. Systematic Biology, 49, 130-142.
- 42. He, L.; Hannon, G.J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. Nature Reviews Genetics 5, 522-531.
- Hockley, D.J., McLaren, D.J. (1973). Schistosoma mansoni: changes in the outer membrane of the tegument during development from cercaria to adult worm. Int J Parasitol. 3, 13-25.
- 44. Hoffmann, K.F.; Johnston, D.A.; Dunne, D.W. (2002). Identification of Schistosoma mansoni gender-associated gene transcripts by cDNA microarray profiling. Genome Biol, 3, RESEARCH0041.
- 45. Hu, W; Brindley, P.J.; McManus, D.P.; Feng, Z.; Han, Z. (2004). Schistosome transcriptomes: new insights into the parasite and schistosomiasis. Trends in Molecular Medicine 10 (5), 217-225.
- 46. Hunter, C.; Sun, H.; Poethig, R.S.(2003). The Arabidopsis heterochronic gene ZIPPY is an ARGONAUTE family member. Curr Biol. 13, 1734–1739.
- 47. Hutvagner, G.; Mclachlan, J.; Pasquinelli, A.E.; Balint, E.; Tuschl, T. e Zamore,
 P.D. (2001). A cellular function for the RNAinterference enzyme Dicer in the maturation of the *let-7* small temporal RNA. Science 293, 837-838.
- Hutvagner, G.; Simard, M.J. (2008). Argonautes proteins: key players in RNAsilencing. Nature Review Molecular Cell Biology 9, 22-32.
- 49. Hutvagner, G.; Zamore, P.D. (2002). A microRNA in a multipleturnover RNAi enzyme complex. Science 297, 2056-2060.

- 50. Ketting, R.F.; Fischer, S.E.; Bernstein, E.; Sijen, T.; Hannon, G.J.; Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. Genes e Development 15, 2654-2659.
- 51. Kloetzel K. (1989) Schistosomiasis in Brazil: Does social development suffice? Parasitol.Today 5, 388-391.
- 52. Krautz-Peterson, G.; Skelly, P.J.; (2007). Schistosoma mansoni: The dicer gene and its expression. Exp. Parasitol. ARTICLE IN PRESS.
- 53. Lagos-Quintana, M.; Rauhut, R.; Lendeckel, W.; Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science 294, 853-858.
- 54. Lai, E.C.; Tomancak, P.; Williams, R.W.; Rubin,G.M.; (2003). Computational identification of *Drosophila* microRNA genes. *Genome Biology*, 4, R42
- 55. Lawson, J.R.; Wilson, R.A. (1980). Metabolic changes associated with the migration of the schistosomulum of Schistosoma mansoni in the mammal host. Parasitol. 81, 325-336.
- 56. Lee, R.C.; Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. Science 294, 862-864.
- 57. Lee, R.C.; Feinbaum, R.L.; Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843-854.
- 58. Lee, Y.; Ahn, C.; Han, J.; Choi, H.; Kim, J.; Yim, J.; Lee, J.; Provost, P.; Radmark, O.; Kim, S. e Kim, V.N. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature 425, 415- 419.
- 59. Lim, L.P.; Glasner, M.E.; Yekta, S.; Burge, C.B.; Bartel, D.P. (2003a). Vertebrate microRNA genes. Science 299, 1540.
- 60. Lim, L.P; Lau, N.C; Weinstein, E.G.; Abdelhakim, A.; Yekta, S.; Rhoades, M.W.; Burge, C.B.; Bartel, D.P. (2003b). The microRNAs of *Caenorhabditis elegans. Genes e Dev.* 17, 991-1008.
- 61. Lindbo, J.A.; Dougherty, W.G. (1992). Untranslatable transcripts of the tobacco etch virus coat protein gene sequence can interfere with tobacco etch virus replication in transgenic plants and protoplasts. Virology 189, 725-733.
- 62. Llave, C.; Kasschau, K.D.; Rector, M.A.; Carrington, J.C. (2002). Endogenous and silencing-associated small RNAs in plants. Plant Cell 14, 1605-1619.

- 63. Lund, E.; Guttinger, S.; Calado, A.; Dahlberg, J.E.; Kutay, U. (2004). Nuclear export of microRNA precursors. Science 303, 95-98.
- 64. Ma, J. B.; Ye, K.; Patel, D. J. (2004). Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* 429, 318–322.
- 65. Meister, G.; Tuschl, T. (2004). Mechanisms of gene silencing by doublestranded RNA. Nature 431, 343-349.
- 66. Mourelatos, Z.; Dostie, J.; Paushkin, S.; Sharma, A.; Charroux, B.; Abel, L.; Rappsilber, J.; Mann, M. e Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes e Development 16, 720-728.
- 67. Nagai, Y., Gazzinelli, G., de Moraes, G.W., Pellegrino, J., 1977. Protein synthesis during cercaria-schistosomulum transformation and early development of the Schistosoma mansoni larvae. Comp Biochem Physiol B. 57, 27-30.
- 68. Nakahara, K.; Carthew, R.W. (2004). Expanding roles for miRNAs and siRNAs in cell regulation. Current Opinion in Cell Biology 16:127-33.
- 69. Neves, D.P.; Melo, A.L.; Genaro, O. ; Linardi, P.M. (2005). Parasitologia Humana, 11º edição, Atheneu Editora.
- Nielsen, C. (1995). Animal evolution. Interrelationships of the living phyla. Oxford Univ. Press, Oxford, UK.
- 71. Ojopi, E.P.B.; Oliveira, P.S.L.; Nunes, D.N.; Paquola, A.; DeMarco, R.; Gregório, S.P.; Aires, K.A.; Menck, C.F.M; Leite, L.C.C; Verjovski-Almeida, S.; Dias-Neto, E. (2007). A quantitative view of the transcriptome of *Schistosoma mansoni* adult-worms using SAGE. *BMC Genomics*, 8, 186.
- 72. Osman, A.; Niles, E.G.; Verjovski-Almeida, S.; LoVerde, P.T. (2006). Schistosoma mansoni TGF-beta receptor II: role in host ligand-induced regulation of a schistosome target gene. PLoS Pathogens 2, e54.
- 73. Ramalho-Pinto, F.J.; Gazzinelli, G.; Hopoços, R.E.; Mota-Santos, T.A.; Figueiredo, E.A.; Pellegrino, J. (1974). *Schistosoma mansoni*: defined system for stepwise transformation of cercaria to schistosomule in vitro *Experimental Parasitology* 36, 360-372
- 74. Ramos, C.R. e cols. (2003). Gene Structure and M20T Polymorphism of the Schistosoma mansoni Sm14 Fatty Acid-Binding Protein: Molecular, Functional and Immunoprotection Analysis. J. Biol. Chem. 278, 12745-12751.

- 75. Reinhart, B.J.; Slack, F.J.; Basson, M.; Pasquinelli, A.E.; Bettinger, J.C.; Rougvie, A.E.; Horvitz, H.R. e Ruvkun, G. 2000. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. Nature 403, 901-906.
- 76. Saito, K.; Ishizuka, A.; Siomi, H.; Siomi, M.C. (2005). Processing of premicroRNAs by the Dicer-1–loquacious complex in Drosophila cells. PLoS Biol 3(7), e235.
- 77. Sasaki, T.; Shiohama, A.; Minoshima, S.; Shimizu, N. (2003). Identification of eight members of the Argonaute family in the human genome small star, filled. Genomics 82, 323-330.
- 78. Schier, A. F. (2007). The Maternal-Zygotic Transition: Death and Birth of RNAs. Science 316, 406-407.
- 79. Schwarz, D.S.; Hutvagner, G.; Haley, B. e Zamore, P.D. 2002. Evidence that siRNAs function as guides, not primers, in the *Drosophila* and human RNAi pathways. Molecular Cell 10, 537-548.
- 80. Seto, A.G.; Kingston, R.E, Lau, N.C. (2007). The Coming of Age for Piwi Proteins. Molecular Cell, 26.
- 81. Shibata, S.; Sasaki, M.; Miki, T.; Shimamoto, A.; Furuichi, Y.; Katahira, J., Yoneda,Y. (2006). Exportin-5 orthologues are functionally divergent among species. Nucleic Acids Research, 0(0), 1–11.
- 82. Shoemaker, C. B.; Ramachandran, H.; Landa, A.; Dos Reis, M. G.; Stein, L. D. (1992). Alternative splicing of the Schistosoma mansoni gene encoding a homologue of epidermal growth factor receptor. Molecular and biochemical parasitology, 53, 17-32.
- 83. Short, R.B.; Menzel, M.Y. (1960). Chromosomes of nine species of schistosomes 1. J.Parasitol. 46, 273-287.
- Short, R.B.; Menzel, M.Y.; Pathak, S. (1979). Somatic chromosomes of Schistosoma mansoni 1. *J.Parasitol.* 65, 471-473.
- 85. Simpson, A.J.; Sher A.; Mccutchan, T.F. (1982). The genome of Schistosoma mansoni: isolation of DNA, its size, bases and repetitive sequences 1. *Mol.Biochem.Parasitol.* 6, 125-137.
- 86. Skelly, P.J.; Da0dara, A.; Harn, D.A. (2003). Suppression of cathepsin B expression in Schistosoma mansoni by RNA interference. International Journal for Parasitology 33, 363–369.

- 87. Smalheiser, N.R. (2003). EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues. Genome Biology 4:403.
- 88. Song, J. J., Smith, S. K., Hannon, G. J. and Joshua-Tor, L. (2004) Crystal Structure of Argonaute and its implications for RISC slicer activity. Science 305: 1434-1437.
- 89. Tabara, H.; Yigit, E.; Siomi, H.; Mello, C.C. (2002). The dsRNA Binding Protein RDE-4 Interacts with RDE-1, DCR-1, and a DExH-Box Helicase to Direct RNAi in *C. elegans*. Cell, 109, 861–871.
- Tolia, N.H.; Joshua-Tor, L. (2007). Slicer and the Argonautes Nature Chemical Biology 3, 36-43.
- 91. Verjovski-Almeida, S.; Demarco, R.; Martins, E.A.; Guimaraes, P.E.; Ojopi, E.P.; Paquola, A.C.; Piazza, J.P.; Nishiyama, M.Y.; Jr.; Kitajima, J.P.; Adamson, R.E.; Ashton, P.D.; Bonaldo, M.F.; Coulson, P.S.; Dillon, G.P.; Farias, L.P.; Gregorio, S.P.; Ho, P.L.; Leite, R.A.; Malaquias, L.C.; Marques, R.C.; Miyasato, P.A.; Nascimento, A.L.; Ohlweiler, F.P.; Reis, E.M.; Ribeiro, M.A.; Sa, R.G.; Stukart, G.C.; Soares, M.B.; Gargioni, C.; Kawano, T.; Rodrigues, V.; Madeira, A.M.; Wilson, R.A.; Menck, C.F.; Setubal, J.C.; Leite, L.C.; As-Neto, E. (2003) Transcriptome analysis of the acoelomate human parasite Schistosoma mansoni. *Nat.Genet.* 35, 148-157.
- 92. Vermeire, J.J.; Taft. A.S.; Hoffmann, K.F.; Fitzpatrick, J.M.; Yoshino, T.P.: (2006) Schistosoma mansoni: DNA microarray gene expression profiling during the miracidium-to-mother sporocyst transformation. Mol Biochem Parasitol, 147:39-47.
- 93. Williams, R.W.; Rubin, G.M. (2002). ARGONAUTE1 is required for efficient RNA interference in Drosophila embryos. Proc. Natl. Acad. Sci. 99, 6889-6894.
- 94. Wilson, R.A. (1979). Introdução à Parasitologia. Ed. EDUSP.
- 95. Wilson, R.A.; Ashton, P.D.; Braschi, S.; Dillon, G.P.; Berriman, M.; Ivens, A. (2007). Oming in on schistosomes: prospects and limitations for post-genomics. Trends in Parasitology 23, 14–20.
- 96. Yan, K. S., e cols. (2003). Structure and conserved RNA binding of the PAZ domain. Nature 426(6965): 468-74.
- 97. Yigit, E., Batista, P.J.; Bei, Y.; Pang, K.M.; Chen, C.G.; Tolia, N.J.; Joshua-Tor, L.; Mitani, S.; Simard, M.J.; Mello, C.C. (2006). Analysis of the *C. elegans*

Argonaute Family Reveals that Distinct Argonautes Act Sequentially during RNAi. Cell 127, 747–757.

98. Zamore, P.D. (2002). Ancient pathways programmed by small RNAs. Science 296, 1265-9.

Anexo

8-ANEXO

8.1 Anexo 1

Alinhamento entre os fragmentos de cDNA representantes do domínio endoND1 seqüenciados e a recuperados do banco de dados da FAPESP de S. mansoni.



*C604739.1

8.2 Anexo 2

Alinhamento entre as seqüências de DNA genômico representantes do domínio PIWI seqüenciadas e a recuperada do banco de dados geneDB de *S. mansoni.*



As setas com os blocos delimitam os iniciadores F e R do fragmento amplificado.

8.3 Anexo 3

Alinhamento entre os fragmentos seqüenciados de cDNA representantes do domínio PIWI.



As setas com os blocos delimitam os iniciadores F e R do fragmento amplificado e o bloco com linhas tracejadas representa o íntron/éxon inserido na seqüência consenso de cDNA.

8.4 Anexo 4



Porcentagem relativa das ESTs referentes às proteínas da via de miRNA.

Foi levado em consideração o número de ESTs seqüenciadas por estágio no projeto transcriptoma da FAPÈSP para o cálculo de porcentagem relativa. Em amarelo o estágio de ovos, em vermelho, esporocistos, em verde miracídios, em roxo cercária, em azul esquistossômulos e em alaranjado vermes adultos.

9.5 Anexo 5

Porcentagem das ESTs presentes nas proteínas depositadas no banco da FAPESP SmDicer1, SmDrosha1 e 2, SmAgo1, 2, 3 e 4, SmFmr1, SmTudor-SN, SmPasha, SmExportina5.1, 5.2.

