Published in final edited form as:

Exp Parasitol. 2008 January ; 118(1): 122–128.

***Schistosoma mansoni* : the Dicer Gene and its Expression**

Greice Krautz-Peterson and Patrick J. Skelly

Molecular Helminthology Laboratory, Division of Infectious Diseases, Department of Biomedical Sciences, Tufts University, Cummings School of Veterinary Medicine, Grafton, MA 01536, USA

RNA interference (RNAi) is a mechanism by which gene-specific, double stranded RNA (dsRNA) triggers the degradation of homologous mRNA transcripts. This results in a “knock-down” or “silencing” of the expression of the target gene. The process first involves the recognition of dsRNA by a multi-domain nuclease called Dicer. Dicer proteins play a pivotal role in RNAi since they are responsible for cutting dsRNAs into smaller molecules called short interfering RNAs (siRNAs) approximately 21 nucleotides long. The siRNAs then join with an effector nuclease complex, known as the RNA-induced silencing complex (RISC), which recognizes and destroys homologous target mRNAs in an endonucleolytic manner (Grishok et al., 2001; Grishok and Mello, 2002). As a molecular tool for analyzing gene function, RNAi has permeated all fields of eukaryotic biology.

Proposed functions for endogenous gene-silencing phenomena like RNAi include the destruction of viral RNAs or suppressing the expression of potentially harmful segments of the genome, such as transposons (Hannon, 2002). RNAi may also participate in the regulation of nuclear microdomain organization, mRNA stability or in other pathways of gene regulation (Allshire, 2002; Hannon, 2002). Endogenous RNA substrates called microRNAs, or miRNAs, are short noncoding RNAs, generated by the action of Dicer, that can enter silencing pathways, leading to gene silencing by either mRNA destruction or the suppression of protein synthesis (Hutvagner and Zamore, 2002). The importance of miRNAs in the regulation of gene expression is becoming increasingly apparent.

RNAi has been described in organisms of diverse phylogeny including vertebrates, plants, fungi, insects, nematodes, and platyhelminths (flatworms). Within the phylum platyhelminthes, RNAi has been described in the human parasitic worm, *Schistosoma mansoni* where the addition of exogenous dsRNA has resulted in demonstrable suppression of target gene expression (Boyle et al., 2003; Skelly et al., 2003). These reports show that schistosomes possess the molecular machinery that mediates RNAi. Furthermore, analysis of the schistosome transcriptome has revealed several expressed sequence tags with homology to components of the RNAi pathway (Verjovski-Almeida et al., 2003). A particularly important gene in RNAi is Dicer. In many systems including zebrafish and mice, Dicer enzymes are essential for normal animal development and in cultured vertebrate cells Dicer has been shown to be involved in the formation of heterochromatin structure (Bernstein et al., 2003). Here we characterize the structure and expression of the Dicer gene from *S. mansoni*.

Transcriptome analysis of *S. mansoni* revealed two ESTs (designated C604739.1 and C713374.1) with sequence similarity to proteins belonging to the Dicer family (Verjovski-Almeida et al., 2003). These sequences encode two different regions of a single parasite protein and both were used to search the *S. mansoni* genome assembly (version 3) at

http://www.sanger.ac.uk/Projects/S_mansoni/ for genomic clones containing the sequences. Matches were found on contigs c0028249 and c0028250. An analysis of open reading frames (ORFs) within these contigs revealed several new sequences at various points along each contig with further similarity to Dicer family proteins. Dicer oligonucleotides listed in table 1 were synthesized based on these conserved sequences and were used, with adult parasite cDNA, to amplify overlapping fragments of the predicted *S. mansoni* Dicer (SmDicer) coding DNA. In this manner the majority of the SmDicer coding DNA was determined. The remaining 3' end of SmDicer was identified by first examining EST C604739.1, since this derives from the extreme 3' end of the Dicer ORF, and includes much of the final exon. Using genomic sequence as a guide, this predicted exon sequence could be extended to an in-frame stop codon, representing the likely end of the protein coding region. Confirmation of the 3' end of the SmDicer coding sequence was obtained by direct sequencing of a PCR product obtained from adult parasite cDNA using oligonucleotides (Dicer E and Dicer F, table 1) which span this region.

The extreme 5' end of the SmDicer cDNA is not well conserved and could not be identified by analysis of the genomic contig sequences. Instead, the 5' end was isolated by the RACE (rapid amplification of cDNA ends) method using the Smart™ RACE cDNA amplification kit from Clontech, CA, following the manufacturer's instructions. For this, the oligonucleotides Dicer L and Dicer race3 (table 1) were both used for first strand cDNA synthesis and the upstream Dicer race4 oligonucleotide was used for subsequent amplification of the 5' end of the Dicer coding sequence. All amplified DNA fragments were sequenced at the Tufts University Core Sequencing Facility. The GenBank accession number for the full SmDicer coding sequence is **EF204544**. The accession numbers of Dicer-1 coding DNAs from other species examined in this study are as follows: *Drosophila melanogaster*, **NP524453**; *Aedes aegypti*, **AAW48724**; *Caenorhabditis elegans*, **NP498761** and *Homo sapiens*, **NP803187**. Comparisons between SmDicer and related sequences were undertaken using the UPGMA best tree building method, with gaps distributed proportionally, using DS Gene software (Accelrys Inc.).

A diagrammatic representation of the SmDicer gene is given in Figure 1A. The size of the gene is at least 54,198 bp; its precise size cannot be given at present because the entire sequence of intron 24 is not known. Attempts to span this intron using PCR were unsuccessful. Intron 24 separates genomic contigs c0028249 and c0028250; our analysis shows that the contigs are juxtaposed. The SmDicer gene has 30 exons ranging in size from the smallest at 43 bp (exon 20) to the largest at 652 bp (exon 5). Mean exon size is 264 bp. All exons combine to form a 7923 bp open reading frame, potentially encoding a 2641 amino acid protein. This is the largest Dicer protein yet described. For comparison, the human Dicer1 protein (GenBank accession number NP803187) is 1922 amino acids long and the *D. melanogaster* Dicer1 protein (accession number NP524453) is 2249 amino acids in length. The predicted pI of SmDicer is 6.1 and its molecular weight is predicted to be 301,321 Da.

SmDicer contains all of the domains that are characteristic of metazoan dicers and these are indicated in figure 1B. Included is an amino terminal helicase domain, a domain of unknown function (DUF) 283, a PAZ domain (so called since the sequence is also found in proteins belonging to the Piwi, Argonaut and Zwiille families), two RNase III catalytic domains and a carboxyl terminal, double stranded RNA binding domain (dsRBD).

Exon shuffling refers to the notion that a particular exon, encoding a specific protein domain, can, through duplication and rearrangement, create novel genes with reused functional properties during evolution (Liu and Grigoriev, 2004). However, each Dicer domain identified here spans more than one exon (Figure 1B) and could not easily be shuffled as one functional unit.

A helicase domain exists at the amino terminus of the schistosome Dicer protein and such domains are proposed to possess an ATP-dependent RNA unwinding activity. Figure 1 C,i shows an alignment of Dicer helicase domains. All are related to the DEAD-box protein family of RNA helicases, so-called because of the existence of a conserved DEAD/DExH motif that is reported to be involved in ATPase activity. The schistosome protein possesses this motif (¹⁵⁸DECH¹⁶¹, motif II in figure 1C,i) as well as several other domains conserved within this family (motifs I to V, figure 1C,i). The SmDicer helicase domain is substantially larger than the equivalent domain of other dicers.

SmDicer also contains a conserved domain of unknown function (DUF) 283. An alignment of this domain with other DUF 283 motifs is shown in figure 1C,ii. As for other domains, the schistosome DUF 283 is larger than its counterparts from other organisms. DUF283 is a domain that is characteristic of members of the Dicer protein family but its role in protein function is currently not clear.

The PAZ (Piwi-Argonaut-Zwille) domain is a nucleic acid binding region conserved in some but not all Dicer proteins and in proteins of the Argonaut family that are core components of RISC. In some systems, the PAZ domain shows low-affinity nucleic acid binding and can bind the characteristic 2-nucleotide 3' overhangs of siRNAs (Lingel et al., 2004). This has led to the suggestion that the PAZ domain contributes to the specific incorporation of siRNAs into the RNAi pathway. The schistosome Dicer PAZ domain is larger than its counterpart from other orthologs (Figure 1C,iii).

The RNase III catalytic signature sequence is repeated twice in Dicer proteins and alignments of both domains with orthologs are given in figure 1C,iv and v. RNase III enzymes exhibit specificity for dsRNA. Current models of Dicer action suggest that each RNase III domain cleaves one strand of the dsRNA. The first RNase III domain is predicted to cut the RNA strand bearing a 3'-hydroxy group, approx. 21 nucleotides from the end. In conjunction with the PAZ domain, it is thought to be responsible for determining the distance from the terminus of the RNA to the cleavage site (Zhang et al., 2004).

The carboxyl terminus of Dicer proteins contains a conserved double stranded RNA binding domain (dsRBD) (Figure 1C,vi). RNA binding through this domain as well as through the PAZ domain, in conjunction with intramolecular dimerization of the two RNase III domains, is thought to generate a processing center where dsRNAs are cleaved to produce the characteristic 2-nucleotide 3' overhang (Jaronczyk et al., 2005).

There are 29 introns in the SmDicer gene and their size is variable; the smallest intron is 40 bp (intron 2); the largest intron identified in this study is 5596 bp (intron 6). Mean known intron size is 1590 bp. All of the introns possess canonical GT:AG splice donor:acceptor sites. Nine of the SmDicer gene introns contain domains with considerable sequence identity (>80%) to fragments of transposable elements. These domains vary in size from 233 bp, in intron 23, to 3,815 bp, in intron 6. The positions of these sequences are indicated as rectangles beneath the illustration of the Dicer gene in figure 1A. All of the sequences belong to the non-long terminal repeat (LTR) retrotransposon families and, within this group, with one exception, all belong to the schistosome retrotransposon 2 (SR2) family (Drew et al., 1999). The 682bp sequence in intron 16 is the exception since it shows greatest sequence identity to members of the SR3 family (white rectangle, figure 1A). Within the SR2 family, the sequences in introns 4 and 12 exhibit greatest similarity to the Perere-3 retrotransposon (accession number 67625700) (black rectangles, figure 1A) (DeMarco et al., 2005). Two elements are relatively large. The Perere-3 element in intron 4 is 3088 bp. The SR2 element in intron 6 is 3815 bp and exhibits 83 % identity with the SR2 consensus sequence (accession number AF025672). It seems unlikely that either of these large SR2 elements is active and mobile since they do not possess sizable,

intact, open reading frames. Since suppressing the expression of potentially harmful segments of the genome, such as transposons, has been proposed as one function of Dicer enzymes and the RNAi machinery (Hannon, 2002), it is ironic that the schistosome Dicer gene itself is “infected” with retrotransposon fragments.

An examination of the available *S. mansoni* genome sequence suggests that the Dicer gene described here is the only Dicer gene in the parasite genome. Like *S. mansoni*, many other organisms also possess a single Dicer gene including *C. elegans*, *S. pombe*, and several vertebrates (Grishok et al., 2001; Wienholds et al., 2003).

Developmental expression of the Dicer gene across different life cycle stages of the Puerto Rican strain of *S. mansoni* was measured by TaqMan Gene Expression Assay (Applied Biosystems, CA). Cercariae were obtained from infected *Biomphalaria glabrata* snails and isolated parasite bodies were prepared as described (Skelly et al., 2003). Schistosomula were cultured for 15 days in RPMI medium supplemented with 10mM Hepes, 2mM glutamine, 5% fetal calf serum and antibiotics (100U/ml penicillin and 100 µg/ml streptomycin) at 37°C, in an atmosphere of 5% CO₂. Adult parasites were recovered by perfusion from Balb/c mice that were infected with 125 cercariae, 7 weeks previously. Parasite eggs were isolated from infected mouse liver tissue as described (Hackett, 1993).

For gene expression analysis, total RNA was first extracted from all parasites using the Trizol method (Invitrogen, CA), following the manufacturer’s instructions. Next, 1 µg of total RNA, pre-treated with TurboDNase (Ambion, TX), was reverse transcribed to cDNA using random hexamers and Superscript reverse transcriptase (Invitrogen, CA). Quantitative real-time PCR (qRT-PCR) was performed using cDNA derived from 50 ng of total RNA and primer sets/reporter probes labeled with 6-carboxyfluorescein (FAM), custom synthesized by Applied Biosystems (Foster City, CA). For Dicer the Dcr-dcr2F, and Dcr-dcr2R oligonucleotides were used with the FAM-labeled probe, Dcr-dcr2M2 (table 1). For the endogenous control *S. mansoni* triosephosphate isomerase (SmTPI) gene the SmTPI-TPI3F and SmTPI-TPI3R primers were used with the FAM-labeled probe, SmTPI-TPI3M2 (table 1). The GenBank accession number for SmTPI is M83294. Probe positions were designed to span exon/exon boundaries to minimize detection of any contaminating genomic DNA. All samples were run in triplicate and underwent 45 amplification cycles on a 7500 ABI PRISM® Sequence Detection System Instrument. For relative quantification, the $\Delta\Delta C_t$ method was employed (Livak and Schmittgen, 2001).

Figure 2 shows the expression pattern of the Dicer gene during different life stages of *S. mansoni* as determined using qRT-PCR. Comparable levels of expression are seen in cercariae and in adult male and female parasites. The highest relative expression is detected in schistosomula which show over 3 fold the level of expression as the adult worms. Since these schistosomula had been cultured for 15 days and were actively growing, it is possible that higher Dicer gene expression at this time has been selected for to control retrotransposon activation that may be more prone to occur during this period of active larval cell division and growth. This speculation assumes that higher Dicer expression is indicative of higher activation of the entire RNA interference pathway, and while this has not been proven, it is clear that older schistosomula are more susceptible to RNAi than younger forms (Krautz-Peterson et al., 2007). However, this does not explain the higher relative level of Dicer expression in eggs which is over twice that seen in cercariae or adults. The expression of the Dicer gene in all life stages examined suggests that the protein does provide an important function for the parasite throughout its life, likely involving nucleic acid metabolism and gene regulation. The expression of Dicer throughout schistosome development further suggests that all life stages are amenable to RNA interference from exogenous dsRNA. This has already been demonstrated experimentally for sporocysts, schistosomula, and adult male and female worms

(Boyle et al., 2003;Skelly et al., 2003;Dinguirard and Yoshino, 2006;Osman et al., 2006). The widespread expression of Dicer bodes well for the use of RNAi technologies in deciphering gene function in all developmental stages of *S. mansoni*.

Acknowledgements

This work was funded by NIH-NIAID grant AI-056273. Schistosoma-infected snails were provided by the Biomedical Research Institute through NIH –NIAID Contract N01-AI-30026. We thank David Ndegwa for technical assistance and Dr. C. Shoemaker for critically reviewing the manuscript.

References

- Allshire R. RNAi and Heterochromatin--a Hushed-Up Affair. *Science* 2002;297:1818–1819. [PubMed: 12193643]
- Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, Mills AA, Elledge SJ, Anderson KV, Hannon GJ. Dicer is essential for mouse development. *Nature Genetics* 2003;35:215–217. [PubMed: 14528307]
- Boyle JP, Wu XJ, Shoemaker CB, Yoshino TP. Using RNA interference to manipulate endogenous gene expression in *Schistosoma mansoni* sporocysts. *Molecular and Biochemical Parasitology* 2003;128:205–215. [PubMed: 12742587]
- DeMarco R, Machado AA, Bisson-Filho AW, Verjovski-Almeida S. Identification of 18 new transcribed retrotransposons in *Schistosoma mansoni*. *Biochemical and Biophysical Research Communications* 2005;333:230–240. [PubMed: 15939396]
- Dinguirard N, Yoshino TP. 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* 2006;146:219–230. [PubMed: 16427708]
- Drew AC, Minchella DJ, King LT, Rollinson D, Brindley PJ. SR2 elements, non-long terminal repeat retrotransposons of the RTE-1 lineage from the human blood fluke *Schistosoma mansoni*. *Molecular Biology and Evolution* 1999;16:1256–1269. [PubMed: 10486981]
- Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* 2001;106:23–34. [PubMed: 11461699]
- Grishok A, Mello CC. RNAi (Nematodes: *Caenorhabditis elegans*). *Advances in Genetics* 2002;46:339–360. [PubMed: 11931230]
- Hackett, F. The culture of *Schistosoma mansoni* and production of life cycle stages. In: Hyde, JE., editor. *Protocols in Molecular Parasitology, Methods in Molecular Biology*. 21. Humana Press; 1993. p. 89-99.
- Hannon GJ. RNA interference. *Nature* 2002;418:244–251. [PubMed: 12110901]
- Hutvagner G, Zamore PD. RNAi: nature abhors a double-strand. *Current Opinion in Genetics and Development* 2002;12:225–232. [PubMed: 11893497]
- Jaronczyk K, Carmichael JB, Hobman TC. Exploring the functions of RNA interference pathway proteins: some functions are more RISCy than others? *Biochemical Journal* 2005;387:561–571. [PubMed: 15845026]
- Krautz-Peterson G, Radwanska M, Ndegwa D, Shoemaker CB, Skelly PJ. Optimizing gene suppression in schistosomes using RNA interference. *Molecular and Biochemical Parasitology* 2007;153:194–202. [PubMed: 17420062]
- Lingel A, Simon B, Izaurralde E, Sattler M. Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. *Nature Structural and Molecular Biology* 2004;11:576–577.
- Liu M, Grigoriev A. Protein domains correlate strongly with exons in multiple eukaryotic genomes--evidence of exon shuffling? *Trends in Genetics* 2004;20:399–403. [PubMed: 15313546]
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(- $\Delta\Delta C(T)$) Method. *Methods* 2001;25:402–408. [PubMed: 11846609]
- Osman A, Niles EG, Verjovski-Almeida S, LoVerde PT. *Schistosoma mansoni* TGF-beta receptor II: role in host ligand-induced regulation of a schistosome target gene. *PLoS Pathogens* 2006;2:e54. [PubMed: 16789838]

- Skelly PJ, Da'dara A, Harn DA. Suppression of cathepsin B expression in *Schistosoma mansoni* by RNA interference. *International Journal for Parasitology* 2003;33:363–369. [PubMed: 12705930]
- Verjovski-Almeida S, DeMarco R, Martins EA, Guimaraes PE, Ojopi EP, Paquola AC, Piazza JP, Nishiyama MY Jr, Kitajima JP, Adamson RE, Ashton PD, Bonaldo MF, Coulson PS, Dillon GP, Farias LP, Gregorio SP, Ho PL, Leite RA, Malaquias LC, Marques RC, Miyasato PA, Nascimento AL, Ohlweiler FP, Reis EM, Ribeiro MA, Sa RG, Stukart GC, Soares MB, Gargioni C, Kawano T, Rodrigues V, Madeira AM, Wilson RA, Menck CF, Setubal JC, Leite LC, Dias-Neto E. Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. *Nature Genetics* 2003;35:148–157. [PubMed: 12973350]
- Wienholds E, Koudijs MJ, van Eeden FJ, Cuppen E, Plasterk RH. The microRNA-producing enzyme Dicer1 is essential for zebrafish development. *Nature Genetics* 2003;35:217–218. [PubMed: 14528306]
- Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W. Single processing center models for human Dicer and bacterial RNase III. *Cell* 2004;118:57–68. [PubMed: 15242644]

A

B

i. Amino terminus and helicase domain (amino acids 1 - 850)

Sm MVVCRDVKYQIAKKILKNVPNDHFLPPYQLDLLDAAYERNIIICSSNEL
 Dm MAFHWCDNNLNTTVETPRDFQVELLATAYERNIIICIGHRS
 Ae MAYHWTNDNIHTALTTPRDYQVELLASAKERNIILCLAHNS
 Ce MVRLRADLQCFNPRDYQVELLDKATKKNTIVQLGTGS
 Hs PASSPMGPFFGLPWQQAIIHDNIYTPRKYQVELLEAALDHNIVCLNTGS

I

Sm NKMFFVVNLIREFRFDNPGKQIIYFTDDTKMNSIDNFLTHHTGLSCYSFK
 Dm SKEFTALKLLQELSRAR-----RHGRVSVYLSCEVGTSTPECSIYTM
 Ae SKEFTALKLIHELGSQLR-----KPAGKRKRTIYISQNDVSFNLIRD
 Ce GKTFTIAVLLKKEYGVQLFAPLDQGGKRAFFVVEKVNLEQQAIHIEVHTS
 Hs GKTFTIAVLLTKELSYQIRGDFSRNGKRTVFLVNSANQVAQQVSAVRTHSD

Sm YDEQNNIKKDFMDRLTSVVRKYDIILLSPWLIPVIVMLSSVNFFTKLIHQ
 Dm LTHLTDL-RVWQEQPDMPDHCWTDYHVSILRPEGFLYLLETRELLIS
 Ae TDLKVIN-VNDLEDADEDYDWEQIVDDYQVIITDERKCLDAICGYLDLN
 Ce FKVGQVHGQTSSGLWDSKEQCDQFMKRHHVVITAQCLLDLIRHAKLKIE
 Hs LKVGEYS-NLEVNASWTKERWNQEFKTKHQVLIMTCYVALNVLKNGYLSLS

II

Sm YFYLVVWDECHMVLDNSHPLSIVFGPNGPMLQNTEDFGPTATRQNMSDFN
 Dm SVELIVLEDCHDSAVYQIRIRPLFENHIMPAPP-----
 Ae EVNLLVIDDCHKVYGNEEISKLFIDYINVCR-----
 Ce DMCVLIFFDECHHALGSQHPYRSIMVDYKLLKKD-----
 Hs DINLLVWFDECHLAILDHPYREIMKLCENCPS-----

III

Sm VSVTGKVRILCFTSALLPREITDPEKARLRQTLKRTSCRLETASSELLT
 Dm ---ADRPRILGLAGP-LHSAGCELQQLSAMATLEQSVLCQIETASDIVT
 Ae ---EKPILGLAGP-LHNAGCIEGRISAETEQLYCLQAKAETASDIVT
 Ce ---KPVPRVLGLTAS-LIKAKVAPEKIMEQLKKLESAMDSVIETASDLVS
 Hs -----CPRILGLTASIL-NGKCDPEELEEKIQKLEKILKSNAETATDLVV

Sm MLSLGARFPQERVILCSKSSSIAAIPFHQFMLGVFREIKEFIIDIESNLPC
 Dm VLRVCSRPHYIVQCAPFEMDELSLVLADVLNTH---KSFLLDHRYPY-
 Ae VLRVCTKPKKEILLQCAPPANSNLASYLKEIVLSQ---ITFLEDHRYDPS-
 Ce LSKYGAKPYEVVIIICKDFEIGCLGIP-----NFDTVIEIFDET-
 Hs LDRVTSQPCIEIVVDCGFPTDRSG-----LYERLLMELEEA-

Sm TKSSSIGDSTSVVINSGGFRICVLDYCKRAISQCEEILNELGLWCAQIV
 Dm -----EIYGTDQFMDELKDIPDPKVDPLNVINSLVLVHEMGPWCQORAA
 Ae -----EIYEDDEFLEELKNIPDPRADPLKFLHEFLAVLEEMGPWCADRAA
 Ce -----VAFVNTTTEFHPDLDDPRRPIKDSLKTTRAVFRQLGPWAAWRTA
 Hs -----LNFINDCNISVHSKERDSTLISKQILSDCRAVLVVLGPWCADKVA

Sm RVFAKHLIALDRQRSEILKNCELVNNDKEKEKEPNSSSTCKSSYKADI
 Dm HHFYQCNEKLKVKTPH-ERHYLLYCLVS-----
 Ae LAMIVQIEKQIKTPY-ERHFLLLCLVS-----
 Ce QVWEKELGKIISQVLPDKTLRFLNMAK-----
 Hs GMMVRELQYIKHEQE-ELHRKFLFLTD-----

Sm LKDNEYSITIPKINDNEDRISYLLRLTVTQMCFLSRLFQMEFDSIITLEE

Dm -----TALIQLYSLCEHAFHRRHLGSGSDSRQT
 Ae -----TVFVQIRSHCDQIFQQYS---NEKEK
 Ce -----TSMITIKRLLPEPMKKIKS-----IEA
 Hs -----TFLRKIHALCEEHFSPAS-----L

Sm FORMISPKVNLNLIHQKLYKPSMNFRIEVAELPTAKNPPIITNNKKYRK
 Dm IERYSSPKVRRLLQTLRCFKPEE-----
 Ae IQAHSTPKVLRLLLEIFRLESPEG-----
 Ce LRPYVPQVRIRLEFELLEENPEF-----
 Hs DLKFVTPKVIKLLLEILRKYPYE-----

Sm GGSQRNSRCQTSKISNSSLSMGMFISCDVDSLSDSMSDTMSSLSDDDDDD
 Dm -----VHTQADGLR
 Ae -----TRNKEKAIS
 Ce -----QKERMKLEK
 Hs -----RQQFESVEW

Sm TRSVRSLNSTKSRLSTQSMKNLNSSSKSRKRKNRSNLSNLSFNLHDLHFV
 Dm RMRHQVDQADFNRLSHTLESKCRMVDQMDQPPTETRALVATLEQILHTE
 Ae EIFPGEPPDTCQSESQNLLEEIKLIDFKKLIADVDDVTKSIDSTITESIA
 Ce AEHLS-----
 Hs YNNRNQDNYVSWSDSEDDDEDEEIEEKEKPETNFPSPFTN-----

Sm PASTLNNGGIRPDVDPSTLVYRAVLNTDNRHNKNHNRSFGFIGQESSDMTA
 Dm DRQTNRSAARVTPPTPAHAKPKPSSGANTAQPRTRRRVYTR-----
 Ae NLKTSCLKQISEPSIAEADVATQSKKLRSPTKNGIRQLRQRRKPGMPNRPY
 Ce -----
 Hs -----

IV

Sm STNNNDVSNRLCGLLILPCQFSAYALSRLIDELCIWVDLYFIKIGHL
 Dm RHHRDHNDGSDTLCALIYCQNHTARVLFELAEISRRDEDLKFLRCQYT
 Ae RSNYYNQNDPDALCALIFCNSKFIKILYSIFYESIRSDEQLAYINVQYT
 Ce -----AIIIFVDQRYTAYSLLMMRHIKSWEEKFKFVNPDYV
 Hs -----ILCGIIFVERRYTAVVLNRIKEAGKQDEELAYISSNFI

Sm FCRQTLLEDNDLSTTTNQTNLNKTGHGTNFPYKTLVKFSNEPVMNNDSS
 Dm TDR-----VAD
 Ae VDK-----TAD
 Ce VGAS-----GRN
 Hs TGHG-----IGK

V

Sm TNNNSNCRNPFSVNOEETITNFRGAINLIVATQAAISTVTTSGTELPRC
 Dm PTTEPKAELEHRRQEEVLKFRFMHDCNVLIGT----SVLEEGILVPKC
 Ae PIKDPKEAEIEHRRQEEVLKFRFMHECNLLIGT----SVLEEGILVPKC
 Ce LASSD--SQGLHKRQTEVLRFRHNEINCLLAT----SVLEEGVLDVKC
 Hs NQPRNKQMEAEFRKQEEVLKFRHETNLLIAT----SIVEEGVLDIPKC

Sm NLVIALRLNSLAEYLSKARSRLVNYGAKVIYLMHDHSIEKINNNNNNN
 Dm NLVVRWDPTTYSYVQCKGRARAAPAYHVILVAP-----
 Ae NLVIRWNEPVSYRSYVQCKGRARASTAYHILFVTPKSDAQNCANLCEDLS
 Ce NLVTKFDRPLDMRSYVQSKGRARRAGSRYVITVEE-----
 Hs NLVVRFDLPTESYRSYVQSKGRARAPISNYIMLADT-----

ii. Domain of Unknown Function 283 (DUF 283, amino acids 938 - 1058)

```

Sm  AINIINQYCARLPSDYITNLTPKWKIFPQPGFTPSSTGPGSTCGSLK
Dm  AIALVNKYCARLPSDTFTKLTALWRCTRNERAG-----
Ae  AIQMVNKYCAKLPSDTFTKLTPLWRGAKTVRNG-----
Ce  --KDTAAYCSKLPSDITFTRLVPHNQIPIEENG-----
Hs  AIGHINRYCARLPSPFTHTLAPKCTRELDPG-----

```

```

Sm  DLQPDGTFNLYQCVLRPLPINSVSKETIVG-EPMVCKKLAKYSAAFNAIHL
Dm  -----VTLFQYTLRLPINSPLKHDIVG-IPMTQTILARRLAALQACVE
Ae  -----QELYQCTIRMPINSPLKEDIVG-IPMTEILARRMAAYTACRV
Ce  -----VTKYCAETLLPINSPIKHAIVLKNPMNKKTAQMAVALEACRQ
Hs  -----TFYSTLYLPINSPLRASIVG-FPMSCVRLAERVVVALICCEK

```

```

Sm  IYLSGEMDSKWELINRETNPM
Dm  LHRIGELDDQLQPICKEG---
Ae  LHKAGELDDSLQPICKES---
Ce  LHLEGELDDNLLPKGRES---
Hs  LHKIGELDDHLMVCKET---

```

iii. PAZ domain (amino acids 1371- 1564)

```

Sm  QKSGIIESNSVNNKPSIFEERMEEFVNAVVTGGRNLDQPOYYVVAITRN
Dm  NGNTMPRAVPDEERQA-QPFDPPQRFQDAVVMWPYRNQDQPOYYFYVAEICP
Ae  NVDKMPTFIPDENRKN-YTFDSNQEFKDAVVMWPYRNQDQPOYYFYVAEICY
Ce  NMENMPER-IPKDEVRRQYKFNADYKDAIIVMPWYRNLEQPVFFYVAEILP
Hs  SEARIGIPSTKYTKETPFVEKLEDYQDAVLIIPRYRNFDQPHRFYVADVYT

```

```

Sm  DMSPLSPFPSPDKFRNFAAYWINKYNALITNNQPLLDVDLTVLRNLNLLIP
Dm  HLSPLSCFPDGNRYTFKHXYLVKYGLTIQNTSQPLLDVDHTSARLNFLTP
Ae  HLSPKSSFPGLNYSFEEYWKYKYSIQIQNCKQPLLDVDHTSARLNFLTP
Ce  EWRPSSKFPDTHFEFNEFYFKKYKLEIYDQNSLLDVDHTSTRNLNLLQ
Hs  DLTPLSKFSPFEYETFAEYWKYKYNLDLTNLNQLLDVDHTSSRLNLLTP

```

```

Sm  RYMNIYGHNNTNANDNNGDINNSKSNVYTHKQYYQQNHDDHQGEDLTN
Dm  RYVNRKGVALPTSSEETKRAKRENLE-----Q
Ae  RYVNRKGVALPTSSEETKRAKRENLE-----Q
Ce  RIQNPFRSRTVSNSTSNIPQASASDSKESNTSVPHS-----SQ
Hs  RHLNQKQKALFLSSAEKRAKWESTIQ-----N

```

```

Sm  KQLLVPELCFRHSFEASVWRKAVCLPSILYRLEHLLLAELRHR
Dm  KQILVPELCTVHPFEASLWRKAVCLPCILYRINGLLLADDIRKQ
Ae  KQILVPELCTIHPFEASLWRAAVCLPCILYRINALLADEIRKE
Ce  RQILVPELMDIHPISATLWNVIAALPSIFYRVNQLLTDELRET
Hs  KQILVPELCATHPIASLWRKAVCLPSILYRLHCLLTAEELRAQ

```

iv. RNase III domain 1 (amino acids 1939 - 2141)

Sm RAYQEGPTTILQALTMSCSNDFINLERMETIGDSFLKFVVTVHLYLTWPE
 Dm GHPGSPSPSIILOALTMSNANDGINLERLETIGDSFLKYAITTYLYITVEN
 Ae NHPGESPSIILOALTMSNANDGINLERLETIGDSFLKYAITTYLYCTVEN
 Ce IGLGVSPCLLLTALTTSNAADGMSLERFETIGDSFLKFATTDLYHTLLD
 Hs RTLGENPGIILQALTLSNASDGFNLERLEMLGDSFLKHAITTYLFTCPD

Sm AHEGKLSHRSRIVCNLSNLYRLGKAKDLQNRMIGCKFEFPHENWIPPGYYV
 Dm VHEGKLSHRSKQVANLNLYRLGRKRRLGEYMIATKFEPHDNWLPPCYVYV
 Ae VHEGKLSHRSKQVSNLNLYRLGRKRVLGESMIATKFEPHDNWLPPCYVYV
 Ce QHEGKLSFARSKEVSNCLNLYRLGKKGIPQLIVANKFDAHDSWLPPCYIP
 Hs AHEGRLSYMRSKKVSNCNLYRLGKKKGIPSRMVVSIFDFPVNWLPPGYVYV

Sm R-----
 Dm P-----KELEKAL
 Ae P-----KELEQAL
 Ce T-----CDFKAP
 Hs NQDKSNTDKWEKDEM TKDCMLANGKLDEDEYEEDEEEESLMWRAPKEAD

Sm -----QDKRLNNEI IKKFESNRNLVIWSTD TLMDDVLRNIDFIDENK
 Dm IEAKIPTHHWKLADLLDIKNLSSVQICEMVREKADALGLEQNGGAQNG--
 Ae IDAKIPACHWNLADLPDIKQLSSEEICRLVKERALALGLLDDSDNEEGET
 Ce NTDDAEEKDNEIERILDGVIEEKPENKTGWDIGGDVSKSTTDGIIETIT-
 Hs YEDDFLEYDQEHIRFIDNMLMGSGAFVKKISLSFFSTTDSAYEWKMPKKS

Sm IKPIENFPPISEWDPNDPKVLHAQHLNN-----
 Dm -----QLDDSDNSCND FSCFIP-----
 Ae MKEIQQLETRNDEFIKNAEYFACFIP-----
 Ce -----FPKQARVGNDDISPLP-----
 Hs SLGSMFPSSDFEDFDYSSWDAMCYLDPSKAVEEDDFVVGFWNPSEENGCV

Sm -----QYLITIQQAIFDKSIADCV EALIGCYLTTTRGER
 Dm -----YNLVSCHSIEPKSIADCV EALIGAYLIECGPR
 Ae -----YNLVTCHSIEPKSVADCV EALIGAYLIECGPR
 Ce -----YNLLTQCHISDKSIADAVEALIGVHLLTLGPN
 Hs DTGKQSISYDLHTEQCIADKSIADCV EALIGCYLTSOGER

v. RNase III domain 2 (Amino acids 2246 - 2531)

Sm DCYQRLEFLGDAVLDYVITRFLYEDSKQSPGVLTDLRSALVNNTIFASL
 Dm DCYQRLEFLGDAVLDYLI TRHLYEDPRQSPGALTDLRSALVNNTIFASL
 Ae DCYQRLEFLGDAVLDYLI TRHLYEDPRQSPGALTDLRSALVNNTIFASL
 Ce GCYQRLEFLGDAVLDYMITRYLYEDSRQSPGVLTDLRSALVNNTIFASL
 Hs DCYQRLEFLGDAVLDYLI TKHLYEDPRQSPGVLTDLRSALVNNTIFASL

Sm AVRIGLHKYFRASSPQLLHTIDVFVRYKDVAKDDLDFITNEEIEIRQPE
 Dm AVRHGFGHKFFRHLSPGLNDVIDRFVRIQQENG-----
 Ae AVRHGFGHKFFRHLSPGLNDVIDRFVRIQQDNG-----
 Ce AVKFEFQKHFIAMCPGLYHMIKFKVKKCSERN-----
 Hs AVKYDYHKYBKAVSPGLFHVIDFVQFQLEKN-----

Sm LVDTFTTEETINNLTSSNLSSSNSYPVTKTSMVMNNQNTPSIISISKQRHC
 Dm -----
 Ae -----

Ce -----
Hs -----

Sm QTHVHGDVSEEEGGVKEKDNDNDEEDDLYNDEEEKHASILNHAKES
Dm --HCISEEYLLS-----EECDDAED-----
Ae --HIISEEYLLS-----EDECDAED-----
Ce --FDTNFNAEMYMTTEEIIDEQGQED-----
Hs --EMQGMDSLELRRS-----EEDEEKEED-----

Sm KWQEKDTQELTGLFVQNSSENQSSKPQHQTIEQHLKSGKSNNANNTIPT
Dm -----
Ae -----
Ce -----
Hs -----

Sm NTGHLTTNRLSDDVEIPKALSDIFESLAGAIFLDSN
Dm -----VEVPKALGDFESLAGAIFLDSN
Ae -----VEVPKALGDFESVAGAIIFLDSN
Ce -----IEVPKAMCDIFESVAGAIYLD SG
Hs -----IEVPKAMGDI FESLAGAIYMD SG

vi. Double stranded RNA binding domain (dsRBD, amino acids 2557 – 2620)

Sm PKSPVRQLLELEPEGTKEFERARRMVDG-RISVCAHVLGKGRFYGVGRNYR
Dm PKSPIRELLLELEPETAKFGKPEKLADGRRVRVTVDVFCKGTFEGIGRNYR
Ae PKSPIRELLLELEPETAKFGKPEKLADGRRVRVTVEVFGKTFEGIGRNYR
Ce PRSPIRELMFEQSKVRFKMERILESGKVRVTVEVVNNMRFTGMGRNYR
Hs PRSPVRELLEMEPETAKFSPAERTYDG-KVRVTVEVVGKGFEGVGRSYR

Sm IAKSLAAKRALRVLR
Dm IAKCTAAKCALRQLK
Ae IAKCTAAKCALRQLK
Ce IAKATAAKRALKYLH
Hs IAKSAAARRALRSLK

Figure 1.

A. The *S. mansoni* Dicer (SmDicer) gene. Exons 1 – 30 are indicated by white boxes. The gap in intron 24 indicates that this intron has not been fully characterized. “K” indicates kilobase pairs. The rectangles beneath the depiction of the Dicer gene indicate intron sequences with similarity to fragments of retrotransposons; grey rectangles are most similar to members of the SR2 retrotransposon family, black rectangles are most similar to the Perere-3 SR2 retrotransposon and the white rectangle indicates greatest similarity to the SR3 retrotransposon family.

B. SmDicer coding DNA. Exons 1 - 30 are indicated. “K” indicates kilobase pairs. The coding positions for domains conserved among members of the Dicer protein family (Helicase, DUF 283, PAZ, RNaseIII domains and dsRBD) are indicated on top by rectangles.

C. Alignment of domains conserved among members of the Dicer family. Accession numbers are given in parentheses. Sm, *Schistosoma mansoni* (EF204544); Dm, *Drosophila melanogaster* (NP 524453); Ae, *Aedes aegypti* (AAW48724); Ce, *Caenorhabditis elegans* (NP 498761) and Hs, *Homo sapiens* (NP 803187). Shading denotes amino acid conservation. The amino acids encompassed by each domain in the *S. mansoni* protein are indicated: i, Helicase (Numbers I – V refer to conserved motifs); ii, Domain of Unknown Function 283; iii,

PAZ domain; iv, RNase III domain 1; v, RNase III domain 2; vi, double stranded RNA binding domain (dsRBD).

Figure 2.

Expression of the SmDicer gene in different developmental stages. E, egg; C, cercariae, S, 15-day cultured schistosomula; M, males (7-week old) and F, females (7-week old).

TABLE 1

Oligonucleotides used to characterize SmDicer. Sequences are given 5'-3', each position recorded is relative to the first base of the coding DNA. SmTPI represents the *S. mansoni* triphosphatase isomerase endogenous control, FAM signifies 6-carboxyfluorescein.

Name	Sequence	Position
Dier B	GTCATCATCCTGAAAATGAACCG	3527–3549
Dier C	CCTACAACTATTTACAGGCTTTAAC	5830–5855
Dier D	CGTAAATGACTTAATTTGCCTTCATG	5965–5990
Dier E	CAAGGAATGTCAAAATATACACTGACG	7973–8000
Dier F	GATTGTTATCAACGTTTAGAATTTT	6736–6760
Dier G	ATCTTGTTATGACATAATCTAATAC	6772–6796
Dier H	TGAAGCAGGAAATGAATGCCG	4591–4611
Dier I	AAAATCATGATGATCATCAGGG	4523–4544
Dier J	CTTGAAAAAGAACTAGTTGTCGACTT	700–726
Dier K	TTTCGTCGTGGTGCTATCAATTACT	2314–2339
Dier L	AATTACAACGTGGCAATTCAGTACC	2380–2404
Dier race2	TGTTGTTCTGCAAAACGGCTTAACC	3483–3506
Dier race3	CTGATGGAAAGGTATTGCAGCAATTGAGGACGATTGGAACAC	792–834
Dier-der2F	GCAAGAAAAAGATACTCAAGAATTAACTGGT	7341–7371
Dier-der2R	CGTATGTTTGGTGTTGAGGTTTTGA	7405–7429
Dier-der2M2	FAM-ATTGATTTTCAGAAGAATTTT	7382–7404
SmTPI-TPI3F	CATACTTGGACATTCTGAGCGTAGA	276–300
SmTPI-TPI3R	ACCTTCAGCAAGTGCATGTTGA	342–363
SmTPI-TPI3M2	FAM-CAATAAGTTCATCAGATTCAC	311–331