International Journal of Genetics, ISSN: 0975–2862, Volume 2, Issue 1, 2010, pp-08-17

Catalytic RNA world relics in Dicer RNAs

Sayak Ganguli*, Dey S.K., Priyanka Dhar, Protip Basu, Paushali Roy and Abhijit Datta
DBT-Centre for Bioinformatics, Presidency College, Kolkata, sayakbif@yahoo.com

Abstract- RNA interference (RNAi) is a naturally occurring phenomenon of RNA-mediated gene silencing that is highly conserved among multicellular organisms. In the first step of the pathway, long double-stranded RNA molecules are chopped into shorter duplexes with 2 nucleotide overhangs at both 3' ends by an endonuclease dubbed Dicer, the structure of which has been solved only recently. This results in the formation of small 21 nucleotide long RNAs, aptly named small or short interfering RNAs (siRNAs), which are incorporated into a multimeric protein complex, the RNA-induced silencing complex (RISC). One of the two-siRNA strands guides RISC to a complementary RNA. After hybridization the endonucleolytic "slicer" activity of RISC cleaves the target RNA, thus preventing its translation. While long double-stranded RNA molecules can be employed to induce RNAi in lower eukaryotes, siRNAs being 21 nucleotides in length have to be used for gene silencing in mammalian cells in order to prevent the activation of an unspecific interferon response [1]. In contrast to siRNAs, however, miRNAs are capable of inhibiting translation of the targeted mRNA without degrading it (at least in mammalian cells)[2-4]. The need for in silico analysis of the components of the RNA interference pathway arises from the fact that very little is known about the structural and interacting properties of the components. With the above background the analysis was performed to identify putative catalytic motifs in the mRNA of the DICER enzyme.
Key Words: RNAi, Drosha, RISC, miRNA, ribozymes, motifs

Introduction

Over the past two decades it has become clear that a variety of the RNA molecules have important or essential biological functions in cells. RNA is proficient at forming complex and varied tertiary structures as revealed by high resolution structures of a handful of RNAs. The secondary and tertiary structures of RNA are key for understanding their biological activity. Motifs that stabilize RNA 3D folds are relatively small and often involve backbone functional groups, making them impossible to detect even when in large families of secondary structures. Tetraloops and their receptors, U – turns, dinucleotide platforms, ribose zippers and S – turns all consist of 4 – 11 nucleotides and occur within a variety of sequence contexts. In addition, non – canonical base pairs often create context dependent helical geometries or surfaces used in RNA – RNA and RNA – Protein recognition. One of the most essential structures of RNA is the RNA hairpin. It can guide RNA folding, determine interactions in a ribozyme, protect messenger RNA from degradation, serve as recognition motif for RNA binding proteins and act as substrates for enzymatic reactions. Eukaryotic small RNAs of approximately 21–24 nucleotides function as guide molecules in a remarkably wide range of biological processes, including developmental timing and patterning, formation of heterochromatin, genome rearrangement, and antiviral defense [6-7]. They belong to at least two general classes, miRNA and siRNA. miRNAs (approximately 21–22 nucleotides) are found in plants and animals and are often phylogenically conserved within their respective kingdoms. These miRNAs are formed from a precursor which is transcribed from genes which are non protein coding. A part of this nascent precursor adopts a fold-back structure that interacts with a multidomain RNaseIII-like enzyme termed DICER or DICER-LIKE (DCL1 in Arabidopsis), which catalyzes accurate excision of the mature miRNA[8]. The miRNAs then associate with ribonucleoprotein complexes that function to negatively regulate target genes controlling a range of developmental events, such as timing of cell fate decisions, stem cell maintenance, apoptosis, organ morphogenesis and identity, and polarity [6]. siRNAs are chemically similar to miRNAs, although in plants they typically range in size between 21 and 24 nucleotides [9-11]. They are associated with both post-transcriptional forms of RNA interference and transcriptional silencing involving chromatin modification [7]. siRNAs are processed from precursors containing extensive or exclusive double-stranded RNA (dsRNA) structure, such as transcripts containing inverted repeats or intermediates formed during RNA virus replication . siRNA precursors can also be formed by the activity of one or more cellular RNA-dependent RNA polymerases (RdRp), as was shown genetically in several screens for RNA silencing-defective mutants [13-16]. Arabidopsis plants contain at least three active RdRp genes, termed RDR1, RDR2, and RDR6 (also known as SDE1/SGS2) [14,15 and 19]. RDR6 is necessary for sense transgene mediated RNAi, but not for silencing of constructs that encode transcripts with hairpins containing extensive dsRNA structure [14,15 and 19]. In many animals, both miRNAs and siRNAs are formed by the activity of the same DICER enzyme [20-27], although in plants they are formed by distinct DCL activities [7]. Arabidopsis contains four DCL genes (DCL1 to DCL4), only one of which (DCL1) has been assigned a

definitive function in small RNA biogenesis [28-30]. Biochemical data indicate, however, that multiple DCL activities or pathways catalyze formation of siRNAs of small-sized (approximately 21 nucleotides) and large-sized (approximately 24 nucleotides) classes [11]. Endogenous siRNAs in plants arise from many types of retroelements and transposons, other highly repeated sequences, pseudogenes, intergenic regions (IGRs), and a few expressed genes [9,10 and 31]. Exogenous siRNAs can arise from both sense and hairpin transcript-forming transgenes and by viruses [32 and 33]. Both siRNAs and miRNAs function post-transcriptionally to suppress or inactivate target RNAs. siRNAs guide sequence-specific nucleolytic activity of the RNA-induced silencing complex to complementary target sequences. Among other proteins, RISCs contain ARGONAUTE (AGO) family members that likely bind siRNAs or target sequences [34]. In plants and insects, post-transcriptional RNAi serves as an adaptive antiviral defense response [35 and 36]. miRNAs are fully competent to guide nucleolytic function of RISC, provided that a target sequence with sufficient complementarity is available [37,38 and 11]. Many plant miRNAs function as negative regulators through this cleavage-type mechanism [10-11, 39-43]. In animals, the level of complementarity between target and miRNA sequences is generally low, which inhibits nucleolytic activity. Animal miRNAs suppress translation of target mRNAs [44 and 29]. Some plant miRNAs may also function as translational suppressors [45, 47]. siRNAs also guide chromatin-based events that result in transcriptional silencing. Two lines of evidence support this view. First, in Schizosaccharomyces pombe and Arabidopsis, endogenous siRNAs from repeated sequences corresponding to centromeres, transposons, and retroelements are relatively abundant [10,31,46]. RNAi-related factors (DICER, RdRp, and AGO proteins) are required to maintain S. pombe centromeric repeats and nearby sequences in a transcriptionally inactive, heterochromatic state. Mutants that lose RNAi component activities lose heterochromatic marks, such as histone H3 methylation at the K9 position (H3K9), as well as centromere function [16]. In plants, AGO4 is necessary to maintain transcriptionally silent epialleles of SUPERMAN. The ago4 mutants lose both cytosine methylation, particularly at non-CpG positions, and H3K9 methylation at SUPERMAN and other constitutive heterochromatic sites (the Arabidopsis thaliana short interspersed element 1 [AtSN1] locus) [48]. And, second, heterochromatin formation of nuclear DNA can be triggered, in a sequence-specific manner, by post-transcriptional silencing of cytoplasmic RNAs [49-51].

Materials and Methods

Since the work was performed absolutely in silico hence the materials are existing database sequences available at the public databases such as SWISS – PROT and GenBank of NCBI. The Accession numbers of the sequences used are provided below:

- >gi|50897086|dbj|AB182481.1| *Tetrahymena thermophila* DCL1 mRNA for Dicer-related RNase III protein Dcl1p.
- >gi|86565498|ref|NM_068617.4| *Caenorhabditis elegans* Dicer Related Helicase family member (drh-1) (drh-1) mRNA, complete cds.
- >gi|89114031|gb|DQ398891.1| *Drosophila melanogaster* isolate Dicer-1 (Dcr1) gene.
- >gi|78099742|gb|DQ208406.1| *Oryza sativa* (indica cultivar-group) dicer-like protein mRNA.
- >gi|30677869|ref|NM_099986.2| *Arabidopsis thaliana* DCL1 (DICER-LIKE1); ATP-dependent helicase/ ribonuclease III (DCL1) mRNA, complete cds.
- >gi|125841883|ref|XM_678382.2| PREDICTED: *Danio rerio* Dicer1.
- >gi|117168270|ref|NM_148948.2| *Mus musculus* Dicer1.
- >gi|29294650|ref|NM_177438.1| *Homo sapiens* Dicer1.

Transcriptomic Analysis

The mRNA sequences corresponding to the query sequences were derived using the BIOINFX server and were analyzed using the RNA analyzer tool at the Vienna RNA suite. The secondary structures of the snRNP motifs were modeled using the Quick Fold server.

RESULTS AND DISCUSSION

Secondary structures of the RNA sequences were obtained for all the sequences under study and the structures showed multiple bulges and stem loop structures. When these RNA secondary structures were analyzed using the Vienna RNA suite then multiple snRNP motifs were identified along with few regulatory protein binding motifs. The results of the analysis show that all dicer mRNAs form secondary structures resulting in higher stability and prevent themselves from being digested by cellular nucleases [53]. The presence of multiple sites for regulatory factor binding indicates that these areas may also serve as riboswitches or RNA aptamers which regulate when the mRNAs will be translated[54]. One of the most interesting observation is the fact that the mRNAs all posses multiple snRNP motifs which indicate that these might have catalytic properties. The *Drosophila* mRNA as indicated by the RNA analyzer tool does not possess a functional product, but posseses multiple stem loop structures in

clusters – a property common to many catalytic RNA and group I introns. This is a breakthrough result as this proves beyond doubt that the RNAi machinery is nothing but RNA world relics. Probably in the ancient period they acted to regulate the production of proteins in a more active manner but with the gradual emergence of DNA as the genetic material their functions have become restricted towards the silencing of the expression of the unwanted mRNAs in the cell.

As evident from the secondary structure that there are multiple stem / hairpin loop structures in all the mRNAs. Most RNA secondary structures exhibit the presence of an RNA hairpin. This unique structure possesses the ability to guide RNA folding, determine interactions in a ribozyme, protect messenger RNA from degradation, serve as a recognition motif for RNA binding proteins or act as a substrate for enzymatic reactions [55]. Most of the regulatory snRNP motifs showed inclination towards purine - purine and pyrimidine – pyrimidine base pairing (FIG1) As no such regulatory mechanisms have been reported until now we can conclude that probably these stem loop structures play important roles in localization of the mRNAs and thus regulate translation [56].

List of abbreviations used

RNAi : RNA interference
siRNAs : small or short interfering RNAs
RISC : RNA-induced silencing complex
miRNAs : microRNAs
dsRNA : double-stranded RNA
RdRp : RNA-dependent RNA polymerases
IGRs : intergenic regions
AGO : ARGONAUTE
mRNA : messenger RNA

Acknowledgement

The authors are acknowledging the financial support from DBT, Govt. of India.

References

[1] Wallace M., Worrall E., Pettersson S., Hupp, T. R. & Ball K. L. (2006) *Molecular Cell* 23, 251-263.
[2] Wallace M. and Ball K.L. (2004) *Mol Cell Biol* 24, 5606-6619.
[3] Ball K.L., Lain S. Fahraeus R., Smythe C. and Lane D.P. (1997) *Current Biology* 7, 71-80.
[4] Behrens S. E., Tomei L. and De Francesco R. (1996) *EMBO J.* 15: 12–22
[5] Kato J. and Niitsu Y. (2002) *Int. J. Hematol.* 76: 208–212
[6] Carrington J.C. and Ambros V.(2003) *Science* 301 (5631): 336-338.
[7] Finnegan E. J. and Matzke M. A.(2003) *Journal of Cell Science* 116, 4689-4693.
[8] Denli A.M., Hannon G.J. (2003) *Trends Biochem Sci* 28(4):196-201.
[9] Hamilton A., Voinnet O., Chappell L. and Baulcombe D. (2002) *EMBO J.* 21: 4671–4679
[10] Llave C., Kasschau K.D., Rector M.A., and Carrington J.C. (2002a) *Plant Cell* 14: 1605–1619.
[11] Tang Guiliang Reinhrt, Brenda J., Bartel David P. and Zamore Phillip D. (2003) *Genes and Development*, vol. 17, no. 1, p. 49-63
[12] Cogoni C., Macino G. (1999a) *Current opinion in microbiology*;2(6):657-62.
[13] Cogoni C., Macino G. (1999b) *Nature*;399(6732):166-9.
[14] Dalmay T., Hamilton A., Rudd S., Angell S. and Baulcombe D. (2000) *Cell* 101, 543-553
[15] Mourrain P., Béclin C., Elmayan T., Feuerbach F., Godon C., Morel J.B., Jouette D., Lacombe A.M., Nikic S., Picault N., Rémoué K., Sanial M., Vo T.A., Vaucheret H. (2000) *Cell* 101:533-42.
[16] Volpe T.A., Kidner C., Hall I.M., Teng G., Grewal S.I., Martienssen R.A. (2002) *Science* 297:1833-7.
[17] Smardon A.M., et al. (2000) *Current Biology* 10: R393-R394
[18] Yu D., Fan B., Macfarlane S. A. and Chen Z. (2003) *Mol. Plant-Microbe Interact.* 16: 206–216.
[19] Beclin C., Boutet S., Waterhouse P. and Vaucheret H. (2002) *Curr. Biol.* 12: 684–688.
[20] Grishok A., Pasquinelli A.E., Conte D., Li N., Parrish S., Ha I., Baillie D.L., Fire A., Ruvkun G., Mello C.C. (2001) *Cell.*106:23–34.
[21] Ketting R.F., Fischer S.E., Bernstein E., Sijen T., Hannon G.J., Plasterk R.H. (2001) *Genes & Dev*:15:2654–2659.
[22] Hutvagner G., McLachlan J., Pasquinelli A.E., Balint E., Tuschl T., Zamore P.D. (2001) *Science*. 293:834–838.
[23] Knight S.W., Bass B.L. (2001) *Science*:293:2269–2271.
[24] Provost P., Dishart D., Doucet J., Frendewey D., Samuelsson B. & Radmark O. (2002a) *EMBO Journal* 21 5864–5874.
[25] Provost P., Silverstein R.A., Dishart D., Walfridsson J., Djupedal I., Kniola B., Wright A., Samuelsson B., Radmark O. & Ekwall K. (2002b) *PNAS* :99 16648–16653.
[26] Zhang H., Kolb F.A., Brondani V., Billy E. & Filipowicz W. (2002) *EMBO Journal* 21 5875–5885.
[27] Myers J.W., Jones J.T., Meyer T. and Ferrell J.E. (2003) *Nat. Biotechnol.* 21: 324–328.

[28] Park W., Li J., Song R., Messing J. and Chen X. (2002) *Curr. Biol*. 12: 1484-1495.
[29] Reinhart B.J., Weinstein E.G., Rhoades M.W., Bartel B. and Bartel D.P. (2002) *Genes Dev*. 16: 1616-1626.
[30] Schauer S.E., Jacobsen S.E., Meinke D.W. and Ray A. (2002) *Trends Plant Sci*. 7: 487-491.
[31] Mette M. F., Winden J. van der, Matzke M. and Matzke A. J.M.(2002) *Plant Physiol*, Vol. 130: 6-9.
[32] Hamilton A.J., Baulcombe D.C. (1999) Science: 29;286(5441):950-2.
[33] Mette M.F., Aufsatz W., van der Winden J., Matzke M., Matzke A.J.M. (2000) *EMBO J* 19: 5194-5201
[34] Carmell M.A., Xuan Z., Zhang M.Q., Hannon G.J. (2002) *Genes Dev*. 16(21): 2733--2742.
[35] Waterhouse P.M., Wang M.B. and Lough T. (2001) *Nature* 411, 834–842.
[36] Li X., Scuderi A., Letsou A., Virshup D.M. (2002) *Mol. Cell. Biol*. 22(11): 3674--3684.
[37] Hutvágner G. and Zamore P. D. (2002) *Science* 20: Vol. 297. no. 5589: 2056 – 2060 ,DOI: 10.1126/science.1073827
[38] Doench J.G., Sharp P.A. (2004) *Genes Dev*.: 1. 18(5):504-11.
[39] Rhoades M.W., Reinhart B.J., Lim L.P., Burge C.B., Bartel B. and Bartel D.P. (2002) *Cell* 110: 513–520.
[40] Emery J.F., Floyd S.K., Alvarez J., Eshed Y., Hawker N.P., Izhaki A., Baum S.F., Bowman J.L. (2003) *Curr Biol* 13: 1768–1774
[41] Kasschau K.D., Xie Z., Allen E., Llave C., Chapman E.J., Krizan K.A., Carrington J.C. (2003) *Dev Cell* 4: 205–217
[42] Palatnik J.F., Allen E., Wu X., Schommer C., Schwab R., Carrington J.C., Weigel D. (2003) *Nature* 425: 257–263
[43] Xie Z., Johansen L.K., Gustafson A.M., Kasschau K.D., Lellis A.D., Zilberman D., Jacobsen S.E., Carrington J.C. (2004) *PLoS Biol* 2: 642–652
[44] Olsen P.H., Ambros V. (1999) *Dev Bio*l. 15;216 (2):671-80.
[45] Aukerman M. J. and Sakai H.(2003) *The Plant Cell*, Vol. 15:2730-2741.
[46] Reinhart B.J. and Bartel D. P.(2002) *Science DOI*: 10.1126/science.1077183.
[47] Chen X.(2003) *Science DOI:* 10.1126/science.1088060.
[48] Zilberman D., Cao X. and Jacobsen S. E.(2003) *Science* : Vol. 299. no. 5607, pp. 716 – 719.DOI: 10.1126/science.1079695
[49] Jones A.L., Hamilton A.J., Voinnet O., Thomas C.L., Maule A.J. (1999) *Plant Cell*. 11:2291–2230.
[50] Aufsatz W., Mette M.F., van der Winden J., Matzke A.J., Matzke M. (2002) *Proc Natl Acad Sci USA*; 99(Suppl 4):16499–16506.
[51] Schramke V., Allshire R. (2003) *Science*.301:1069–1074.
[52] Ernst J., Ghanem L., Bar-Joseph Z., McNamara M., Brown J. (2009) *PLoS ONE* 4(10): e7469. doi:10.1371/journal.pone.0007469
[53] Mandal M. and Breaker R. R.(2004) *Nature Reviews, Molecular Cell Biology*, 5: 451-463
[54] Svoboda P., Cara A. D.(2006) *Cell Mol Life Sci*, Vol. 63, No. 7-8: 901-908.
[55] Eeden F. J.M. van, Palacios I. M., Petronczki M., Weston M.J.D., and Johnston D.I St.(2001) *The Journal of Cell Biology*, 154 (3), 511–523.

WEB REFERENCES

[1] www.ncbi.nlm.nih.gov
[2] www.ebi.ac.uk.
[3] http://users.ugent.be/~mdgroeve/bioinfx/
[4] http://molbiol-tools.ca/RNA_analysis.htm
[5] http://wb2x01.biozentrum.uni-wuerzburg.de/

Table 1- DICER MOUSE (Length:9851, Origin: RNA, Promoter: none, Exon: 281- 6001, type-coding, PlyA-Signal:9832-9837)

snRNP motifs	Position	Sequence	Quality
snRNP motif:	173	gguucugg	+
snRNP motif:	766	gguguuuga	+
snRNP motif:	790	gauucuag	+
snRNP motif:	814	aauuauga	+
Put. sm site:	853	aauuuugg	++
snRNP motif:	1006	gauugugg	+
snRNP motif:	1097	gauuguaa	+
snRNP motif:	1174	aguguugg	+
snRNP motif:	1439	aguguuga	+
snRNP motif:	1933	aauguuag	+
snRNP motif:	2002	gaucuuaa	+
snRNP motif:	2575	ggucuuaa	+
snRNP motif:	2945	gauauuga	+
Put. sm site:	3062	aauuuuga	++
snRNP motif:	3215	gauguuga	+
Put. sm site:	3512	gauuuuag	++
snRNP motif:	3812	aguuauga	+
snRNP motif:	3920	aauuauga	+
snRNP motif:	4590	aguuuauag	+
Put. sm site:	4736	gauuuuga	++
snRNP motif:	4814	gguuucugg	+
Put. sm site:	5362	gauuuugg	++
snRNP motif:	5422	ggucuuga	+
snRNP motif:	5906	gguguugg	+
Put. sm site:	6171	aauuuugg	++
Put. sm site:	6270	gguuuuaa	++
Put. sm site:	6291	gguuuuugg	++
snRNP motif:	6340	gguguuag	+
Put. sm site:	6548	gauuuuga	++
Put. sm site:	6794	gauuuuuaa	++
snRNP motif:	7648	aguauuuaa	+
snRNP motif:	7668	gauuguag	+
snRNP motif:	7771	gaucuuuuga	+
snRNP motif:	7826	gauugugg	+
snRNP motif:	7869	gauuuauaa	+
snRNP motif:	8035	aauauuga	+
snRNP motif:	8761	gauuuguuga	+
snRNP motif:	8795	aauguuuugg	+
snRNP motif:	9822	aauuguag	+

Table 2- C elegans DICER (Length: 3281, Origin: RNA, Promoter: none, Exon: 7- 3066, type- coding, Poly A Signal: none)

snRNP motifs	**position**	**sequence**	**quality**
snRNP motif:	390	gguucugg	+
snRNP motif:	438	aguucuga	+
snRNP motif:	1062	gauucuga	+
snRNP motif:	1200	gauuguaa	+
snRNP motif:	1791	gguuuguaa	+
snRNP motif:	2052	gauuucuaa	+
Put. sm site:	2157	aguuuuga	++
snRNP motif:	2756	gauucugg	+
snRNP motif:	2851	aauugugg	+
snRNP motif:	3066	gguauuuga	+
snRNP motif:	3142	aguauuuga	+

Table 3- DICER HUMAN(Length: 10276, Origin: RNA, Promoter: none, Exon: 270- 6008, type-coding, Exon: 6360- 6565, type- coding, Poly A Signal 1: 6159- 6164, Poly A Signal 2: 6334- 6339)

snRNP motif	position	sequence	quality
snRNP	708	aauguuuuga	+
snRNP	755	gguguuuga	+
snRNP	803	aauuauga	+
snRNP	923	aauucuuaa	+
snRNP	962	ggucuuag	+
snRNP	995	gauugugg	+
snRNP	1005	gauugugg	+
snRNP	1086	gauuguaa	+
snRNP	1172	aguucugg	+
snRNP	1351	aauuuguaa	+
snRNP	1479	gauucuga	+
snRNP	1812	gguguuga	+
snRNP	1922	aauguuag	+
Put. Sm site	1950	aguuuuga	++
snRNP	1991	gaucuuga	+
snRNP	2315	aguuguag	+
snRNP	2503	aguguuuga	+
Put. Sm site	2564	gguuuuaa	++
snRNP	2910	gauauuga	+
snRNP	2934	gauauuga	+
Put. Sm site	3051	aauuuuga	++
snRNP	3231	aaucuuuuga	+
Put. sm site:	3501	gauuuuag	++
snRNP	3566	aauuucuaa	+
snRNP	3801	aguuauga	+
snRNP	4065	gauucuga	+
snRNP	4355	aauauuuga	+
snRNP	4591	gauuuauag	+
Put. sm site:	4728	gauuuuga	++
Put. sm site:	4737	gauuuuga	++
snRNP	4816	gguucugg	+
snRNP	4845	gguguuga	+
Put. sm site:	5241	aauuuuga	++
Put. sm site:	5369	gauuuugg	++
snRNP	5595	gauucuga	+
snRNP	5640	gauauuga	+
snRNP	5667	gauauuuuuga	+
snRNP	5913	gguguugg	+
snRNP	6062	aauuauuuaa	+
snRNP	6091	aguuguuag	+
Put. sm site:	6184	aauuuugg	++
snRNP	6196	aauuguag	+
snRNP	6315	gguuuuguuuuuuuuuuuuga	+
snRNP	6356	aguguuag	+
Put. sm site:	6582	gauuuuuaa	++
snRNP	6626	aguuuucuga	+
snRNP	6694	agucuuuaa	+
snRNP	6708	gaucuuuuugg	+
snRNP	6903	gaucuuugg	+
snRNP	7050	gauuuuauaa	+
Put. sm site:	7221	aauuuuga	++
snRNP	7326	aguuucuuuaa	+
Put. sm site:	7436	aauuuuaa	++
snRNP	7618	gauucuga	+
snRNP	7785	gguauuaa	+

snRNP	7820	aaucuuaa	+
Put. sm site:	7857	aguuuuuga	++
snRNP	7890	aauauuuaa	+
Put. sm site:	7928	gauuuuuuaa	++
snRNP	7976	aauucuuga	+
snRNP	8045	gauuuguga	+
snRNP	8193	aguuauga	+
Put. sm site:	8393	gauuuuaa	++
snRNP	8483	gguuauga	+
Put. sm site:	8507	gauuuuuaa	++
Put. sm site:	8801	aauuuuuuaa	++
snRNP	8993	gauuuaugg	+
snRNP	9016	aauuauaa	+
snRNP	9109	gauuuguuaa	+
snRNP	9280	gauucuga	+
snRNP	9454	aauauugg	+
snRNP	9562	agucuuuaa	+
snRNP	9624	gguuguuuuaa	+
Put. sm site:	9731	aguuuuaa	++
Put. sm site:	9769	gauuuuuaa	++
snRNP	9778	aguuguugg	+
snRNP	10019	aauuucugg	+
snRNP	10236	aauuguag	+

Table 4- Danio rerio DICER (Length: 8708, Origin: RNA, Promoter: none, Exon: 158 - 4440, type- coding, Exon: 4495 – 5914, type- coding, Poly A-Signal: 7585 -7590)

snRNP motif	position	sequence	quality
Put. sm site:	238	gauuuuaa	++
Put. sm site:	781	gguuuuuga	++
snRNP	838	gauuuguga	+
snRNP	1031	gauugugg	+
snRNP	1100	aauuuucuaa	+
snRNP	1505	gauucuga	+
snRNP	1627	gguuguuuuaa	+
snRNP	1774	agucuuga	+
Put. Sm site:	2057	gauuuuga	++
snRNP	2664	gauguuuugg	+
snRNP	2960	gauauuga	+
Put. sm site:	3077	aauuuuga	++
snRNP	3209	aauguuaa	+
Put. sm site:	3358	gauuuuag	++
snRNP	5259	aguuucuugg	+
snRNP	5275	aauucuag	+
snRNP	6072	aguguuaa	+
snRNP	6320	gguguugg	+
snRNP	6529	aguuguga	+
snRNP	6696	gauauuga	+
snRNP	6705	gauauuuag	+
Put. sm site:	6757	aguuuuga	++
snRNP	6962	gauauuga	+
Put. sm site:	7579	aauuuuaa	++
snRNP	8184	aguuauugg	+
Put. sm site:	8438	gguuuuga	++

Table 5- Arabidopsis DICER (Length: 6188, Origin: RNA, Promoter: none, Exon: 381- 6104, type- coding, Poly A Signal: none)

snRNP motif	Position	Sequence	Quality
SnRNP motif:	371	aauuaugg	+
snRNP motif:	758	gguuuguag	+
snRNP motif:	813	aauucuag	+
snRNP motif:	832	gguuuauga	+
snRNP motif:	939	gguuauugg	+
snRNP motif:	1017	gauguuaa	+
snRNP motif:	1103	gguugugg	+
snRNP motif:	1421	gguucuag	+
snRNP motif:	1870	aauuuaugg	+
snRNP motif:	1916	gguuuaugg	+
snRNP motif:	2586	gauauuag	+
snRNP motif:	2787	gaucuuag	+
snRNP motif:	2846	aguuuauaa	+
snRNP motif:	3299	aguuuguga	+
Put. sm site:	3354	gauuuugg	++
snRNP motif:	3404	gauucuuuuugg	+
snRNP motif:	3579	gauguuga	+
snRNP motif:	3861	gguauuag	+
snRNP motif:	4448	gauucuuga	+
snRNP motif:	4535	gguuguuag	+
snRNP motif:	5010	aauguuaa	+
snRNP motif:	5055	ggucuuga	+
snRNP motif:	5164	aauuuguugg	+
snRNP motif:	5180	ggucuugg	+
snRNP motif:	5362	aauuuguga	+
snRNP motif:	5441	aguucuugg	+
snRNP motif:	6036	aguguuaa	+
snRNP motif:	6163	gguuguag	+

Table 6- Drosophila DICER (Length: 4136, Origin: RNA, Promoter: none, Exons: none, Poly A Signal: none)

snRNP	position	sequence	quality
snRNP motif:	23	aauauuga	+
Put. sm site:	89	aauuuuuuuuuuag	++
snRNP motif:	107	aauuauuag	+
snRNP motif:	273	gguuauga	+
snRNP motif:	488	aauuuauga	+
snRNP motif:	606	gaucuuga	+
snRNP motif:	615	gauuauaa	+
snRNP motif:	698	aauuuauaa	+
Put. sm site:	718	gauuuuuuaa	++
Put. sm site:	729	gauuuuuag	++
snRNP motif:	802	aauuucuuuaa	+
Put. sm site:	879	aauuuuga	++
Put. sm site:	942	gauuuuga	++
snRNP motif:	966	aaucuuuag	+
Put. sm site:	1029	aauuuuag	++
snRNP motif:	1171	aauuauaa	+
snRNP motif:	1521	gauuuauuuuaa	+
Put. sm site:	1652	gauuuuuga	++
snRNP motif:	1867	aauuuguaa	+
snRNP motif:	1911	gauuauuag	+
snRNP motif:	1941	aauauuuaa	+
snRNP motif:	2264	gauuguuag	+
snRNP motif:	2494	aauauuuag	+
snRNP motif:	2626	aaucuuuuuag	+

snRNP motif:	2672	aguuguugg	+
Put. sm site:	2773	aguuuuag	++
snRNP motif:	2866	aauuauaa	+
snRNP motif:	2881	aauuuuuauaa	+
Put. sm site:	3052	aauuuuaa	++
snRNP motif:	3252	gauuauuag	+
Put. sm site:	3268	aauuuuugg	++
Put. sm site:	3293	gauuuuag	++
snRNP motif:	3457	aauauuag	+
snRNP motif:	3481	aauuuauag	+
snRNP motif:	3551	aguucuag	+
snRNP motif:	3561	gauuuaugg	+
snRNP motif:	3597	gauucuga	+
snRNP motif:	3623	aauuuaugg	+
snRNP motif:	3680	aauuauuuuuaa	+
snRNP motif:	3797	gauucuaa	+
snRNP motif:	4076	aguauuuag	+

Table 7- Oryza sativa DICER (Length: 5193, Origin: RNA, Promoter: none, Exon 1: 287 – 1359, type- coding, Exon 2: 1584 – 4888, type- coding, Poly A Signal: none)

snRNP motif	**Position**	**Sequence**	**Quality**
snRNP	147	gauguugg	+
snRNP	574	aauauuuga	+
snRNP	633	aauuuuauuuugg	+
snRNP	719	gauuguga	+
snRNP	820	aauuguga	+
snRNP	1025	gaucuugg	+
snRNP	1091	gauucuga	+
snRNP	1333	aauuauga	+
snRNP	1359	gguuuauga	+
snRNP	1372	aauuguuaa	+
snRNP	2445	aaucuuugg	+
snRNP	2639	aguugugg	+
Put. sm	3219	gauuuugg	++
snRNP	3693	gguuauga	+
snRNP	3914	aauuucuugg	+
snRNP	3976	gauguuga	+
snRNP	4044	aguuguaa	+
snRNP	4188	gguucuugg	+
snRNP	4224	aguauuuuuag	+
Put. sm	4240	gauuuuaa	++
snRNP	4398	aauuauag	+
snRNP	4504	gaucuuaa	+
snRNP	4766	aauuugugg	+

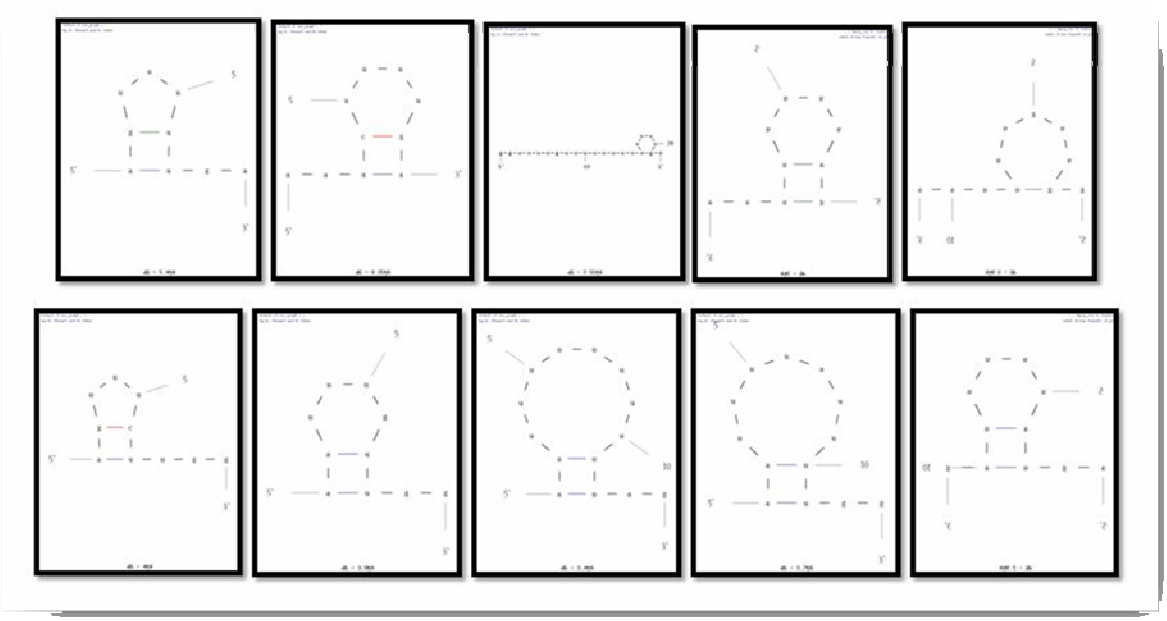

Fig. 1-Secondary Structure of important snRNP motifs in the DICER mRNAs