# Animal Dicer and plant Dicer-like proteins

ALEKSANDER TWORAK [1], ANNA URBANOWICZ [1], JAN PODKOWINSKI [1], MAREK FIGLEROWICZ [1, 2] *

[1] Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland

[2] Institute of Computing Science, Poznan University of Technology, Poznań, Poland

* Corresponding author: marekf@ibch.poznan.pl

In eukaryotic organisms several types of 21/24-nucleotide-long RNA molecules (called small regulatory RNA; srRNA) are involved in the regulation of numerous biological processes, including developmental and stress response pathways or pathogen defense. This widespread mechanism of RNA-mediated regulation is facilitated by a range of proteins which participate in the processing of srRNA precursors and enable the functioning of mature products. The excision of 21/24 nt long RNA duplexes from longer double stranded RNA (dsRNA) precursors is a crucial step, common for most srRNA biogenesis pathways. A group of enzymes belonging to the RNase III family have been identified as the key players involved in this process (Bernstein et al., 2001). In animals these enzymes are called Dicer proteins, while in plants they are known as Dicer-like (DCL) proteins.

RNA that can function as the substrates of Dicer or DCL proteins have diverse origins. Accordingly, they may be divided into two main families. The first large family comprises single-stranded RNA (ssRNA) precursors which adopt imperfect stem-loop structures and are generated from endogenous *loci.* These precursors give rise to small non-coding microRNAs (miRNA). The second large family includes long endogenous or exogenous, perfect or nearly perfect dsRNA, being the precursors for various small interfering RNAs (siRNA). The Dicer has been shown to process dsRNA originating from transposable elements, distinct viruses as well as exogenously introduced dsRNA. High-throughput sequencing has allowed the identification of a wide range of less abundant small RNAs existing in cells (Wittmann and Jack, 2010). Some of them have also been proven to be Dicer-dependent, e.g. small RNA derived from human vault RNAs (Persson et al., 2009), small nucleolar RNAs (Taft et al., 2009), or transfer RNAs (Pederson, 2010).

In vertebrates and nematodes, a single Dicer enzyme generates a complex spectrum of Dicer-dependent srRNAs, while in insects there are two such enzymes. Their functional specialization has been demonstrated in *Drosophila melanogaster*: Dicer-1 processes miRNA precursors, while Dicer-2 is responsible for siRNA generation from exogenous and endogenous dsRNA (Lee et al., 2004). Similar to insects, the basal metazoans typically possess two and sometimes more Dicer genes. Even more Dicer homologs, showing higher structural and functional diversity, are found in plants. *Arabidopsis thaliana* possesses four DCL proteins (DCL1-4), which represent four distinct Dicer groups typically found in plants. All four groups are estimated to originate very early in plant evolution (Mukherjee et al., 2013). In many plants, duplications of genes encoding particular Dicer-like proteins have been identified. Each DCL protein is specialized in the production of a different class of small RNAs, with the DCL1 enzyme being solely involved in miRNA biogenesis (Liu et al., 2005; Park et al., 2002). Consecutive studies have demonstrated that DCL2, 3 and 4 are involved in other srRNA biogenesis pathways. The three latter enzymes mainly contribute to plant defense against a diverse range of viruses by producing virus-derived siRNAs. It has been shown that loss or suppression of the activity of one enzyme may be compensated for by the activity of another (Chapman and Carrington, 2007; Gasciolli et al., 2005). Vertebrates use several systems which protect them against pathogens and reduce the role of the Dicer protein. In contrast, the multiplicity of DCLs in plants might in part reflect the importance and complexity of antiviral strategies based on srRNAs (Blevins et al., 2006; de Jong et al., 2009).

Dicer and DCL enzymes are modular proteins composed of several highly specialized domains specifically ordered from the amino- to carboxy-terminus: DEAD
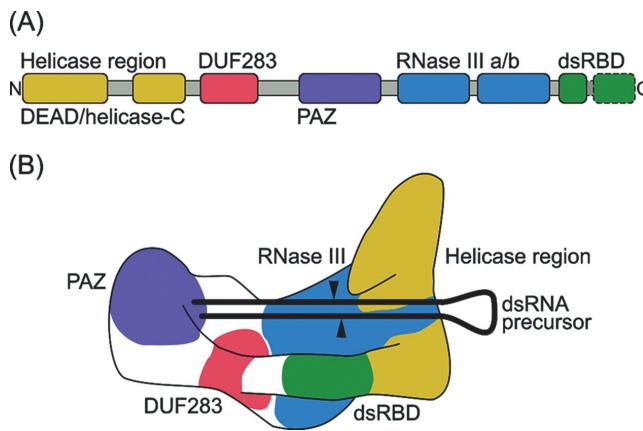
## (A)



## (B)



**Fig. 1.** Architecture of Dicer and Dicer-like (DCL) proteins.
A) Domain composition in the typical Dicer and DCL proteins
(not to scale). Presence of the second, C-terminal dsRBD domain is unique for plant DCL1, DCL3 and DCL4 enzymes.
B) Model of the human Dicer enzyme, indicating the spatial
arrangement of functional domains
(Lau et al., 2012a; Lau et al., 2012b)

box, helicase-C, DUF283, PIWI-Argonaute-Zwille (PAZ), RNase III, and dsRNA binding domain (dsRBD), as shown in Fig. 1. The full set of domains is typically found in both Dicer and DCL enzymes. There are, however, Dicer proteins that lack one or more domains. For example, the Dicer from the protozoan parasite *Giardia intestinalis* contains only the PAZ and RNase III domains. In the catalytic core of all Dicer enzymes twin ribonuclease III (RNase III) domains form an intramolecular dimer (Zhang et al., 2004). Due to the specific arrangement of the dimer a characteristic pattern is observed of the cleavage of RNA duplexes by the Dicer. The products generated carry a phosphate group at the 5′ end and the 2-nucleotide overhang at the 3′ end. The 3′ overhangs can be specifically bound by the Dicer's PAZ domain (Park et al., 2011). Structural studies of human and *Giardia intestinalis* Dicer proteins indicate that the PAZ domain and the RNases dimer are widely separated spatially. The distance between them may determine the length of the generated RNA products (Lau et al., 2012a; Macrae et al., 2006). Apparently, these two types of domains are sufficient to compose a minimal active Dicer enzyme, similar to the one found in *Giardia*.

Typical Dicer or DCL proteins contain N-terminal DEAD box and helicase-C domains. They enclose conserved motifs that are involved in the ATP hydrolysis associated with RNA or DNA unwinding and enzyme translocation (Pyle, 2008). In a human Dicer this helicase

region has been shown to form a clamp-like structure, positioned adjacent to the catalytic core domain (Lau et al., 2012a). This offers significant potential to influence the nuclease activity of the enzyme; yet the functions of both domains appear highly diverse. The helicase domain has been shown to have an inhibitory effect on dsRNA processing by human Dicer (Ma et al., 2008). This domain is also crucial for the generation of certain types of small RNAs in *Caenorhabditis elegans* (Welker et al., 2010), and can also interact specifically with the stem-loop structure of microRNA precursors in the case of the *Drosophila melanogaster* Dicer-1 enzyme (Tsutsumi et al., 2011). Recently, it has been suggested that the helicase domain facilitates the processivity of Dicer on long RNA substrates (Welker et al., 2011). The functions of the two other domains found in typical Dicers or DCLs are less clear. A domain of unknown function DUF283 connects the helicase and PAZ domains (Lau et al., 2012b). In *Arabidopsis thaliana* it has been shown to adopt a non-canonical dsRNA-binding fold and to selectively interact with the dsRNA-binding domains of some DCL cofactors (Qin et al., 2010). At least one dsRNA-binding domain (dsRBD) is also present within the carboxy-terminus of most Dicers. In human Dicer its deletion significantly decreases the efficiency of substrate cleavage, or even abolishes it completely (Ma et al., 2008; Ma et al., 2012). Both DUF283 and dsRBD are localized within the same region of the human Dicer, i.e. in the proximity of the RNase III domains, suggesting their possible interaction and role in substrate binding or digestion (Lau et al., 2012b). Although the domain architecture of all known plant DCL proteins corresponds to the general scheme, the variable number of carboxy-terminal dsRBD domains clearly distinguishes particular enzymes from this group. For example, duplicated dsRBD has been identified in the DCL1, DCL3 and DCL4 proteins found in *Arabidopsis thaliana*, *Oryza sativa*, or *Populus trichocarpa*. In contrast, the DCL2 enzymes found in those species, similar to most animal Dicers, typically possess only one dsRBD (Margis et al., 2006). An *Arabidopsis* mutant equipped with DCL1 lacking both dsRBDs is embryo-lethal, while a plant mutant producing DCL1 without even only the second dsRBD shows infertility and severe defects in the biogenesis of most miRNAs (Schauer et al., 2002). This may suggest an important role of dsRNA-binding domains, but the functional implication of dsRBD duplication in DCL proteins is still unclear (Burdisso et al., 2012).

**Table 1.** Main functions of plant DCL proteins (abbreviations explained in the main text)

| Protein | Typical product length | Main srRNA product | Role in antiviral defense |
|---------|------------------------|--------------------|---------------------------|
| DCL1 | 21 nt | miRNA | minor |
| DCL2 | 22 nt | nat-siRNA | secondary |
| DCL3 | 24 nt | hc-siRNA | minor |
| DCL4 | 21 nt | ta-siRNA | primary |

DCL1 mainly produces 21 nt long miRNAs from their single-stranded stem-loop precursors, but it is also capable of processing double-stranded precursors of siRNA (Bouche et al., 2006). The products of DCL2, DCL3, and DCL4 activities are various siRNAs of 22, 24, and 21 nucleotides in length, respectively (see Tab. 1). In *Arabidopsis* DCL4 and DCL2 are mainly associated with viral resistance and exhibit specific, hierarchical activities (Deleris et al., 2006). DCL4 is the primary viral RNA processor, while the DCL2 antiviral role arises when DCL4 is genetically compromised or suppressed, e.g. as a result of a specific virus infection. DCL4 is also associated with production of *trans*-acting siRNAs (ta-siRNA), whereas DCL2 generates siRNAs from natural *cis*-acting antisense transcripts (nat-siRNA). DCL3 is dispensable for the production of most virus-derived siRNAs, but targets viral dsRNA efficiently in the absence of DCL4 and DCL2. Under normal conditions, DCL3 generates heterochromatic siRNAs (hc-siRNA), also referred to as repeat-associated siRNAs (ra-siRNA), the most abundant small RNAs in plants (Kasschau et al., 2007; Xie et al., 2004). Interestingly, some families of miRNA precursors have been identified to be independently processed by DCL1 and DCL3 enzymes, resulting in the coexistence of both 21 nt and 23-25 nt long products (Vazquez et al., 2008). Analysis of DCL4 PAZ coding sequences reveals a relatively high variability within their RNA-binding pocket, compared to the PAZ sequences of other DCLs or Dicer proteins. This observation may reflect the deep functional adaptation of the domain, which is required to efficiently bind highly variable viral RNA substrates (Mukherjee et al., 2013). DCL4 is also one of the primary targets for multiple viral suppressors of srRNA-directed defense mechanisms (Zvereva and Pooggin, 2012).

Despite the general similarities found in Dicer and DCL1 proteins, the mechanism of srRNA biogenesis differs between animals and plants (Axtell et al., 2011).

In the canonical miRNA biosynthesis pathway identified in animals the first cleavage step occurs in the nucleus. The primary substrate for miRNA generation (pri-miRNA) is cleaved by a Drosha protein, another type of RNase III enzyme. The Drosha works within a complex known as a microprocessor, together with at least one cofactor: DGCR8, a dsRNA-binding protein (Landthaler et al., 2004). The cleavage releases a miRNA precursor (pre-mRNA) with a stem-loop structure and both a 2 nt 3′ overhang and a phosphate at the 5′-end. Next, pre-miRNA is recognized by the nuclear export protein Exportin-5 and transported to the cytoplasm, where the second cleavage occurs (Yi et al., 2003). The cytoplasmic Dicer protein (Dicer-1 in the case of *Drosophila*) generates a short RNA duplex composed of two approximately 21 nt long RNA strands. Most Dicer enzymes have been shown to interact with other proteins, including dsRNA-binding proteins which directly assist in the cleavage process. *Drosophila* Dicer-1 associates with the Loquacious (Loqs) dsRNA-binding protein (Saito et al., 2005), whereas Dicer-2 with R2D2, a paralog of Loqs (Liu et al., 2003). Loqs has been proven to be essential for efficient substrate processing and for increasing Dicer-1 specificity. In the case of the human Dicer, dsRNA-binding cofactors include TRBP and PACT, both important for the processing of certain classes of srRNAs (Kok et al., 2007).

In the canonical plant miRNA biogenesis pathway both cleavage steps occur in the nucleus and a single family of RNase III proteins, DCL, is engaged in this process. DCL1 cleaves the pri-miRNA substrate and the released product (pre-miRNA) is once again subjected to digestion. As a result, a short RNA duplex is generated. DCL1 usually starts the cleavage at the base of the stem-loop structure formed by the primary miRNA substrate, although processing in the opposite direction has also been observed (Bologna et al., 2009). Similar to their animal counterparts, plant DCL enzymes have also been shown to be assisted by specific partner proteins.

In *Arabidopsis*, HYL1 (a dsRNA-binding protein, also known as DRB1) and SERRATE (SE, a zinc-finger-domain protein) contribute to the pri-miRNA and pre-miRNA cleavage (Dong et al., 2008). It has been shown that HYL1 forms a complex with DCL1 (Kurihara et al., 2006), while SE can interact with HYL1 (Yang et al., 2006). Moreover, DCL1 and HYL1 proteins colocalize within the *Arabidopsis* nucleus in small nuclear bodies, predominantly excluded from nucleoli (Fang and Spector, 2007; Song et al., 2007). These Dicing bodies have been shown to contain pri-miRNAs and a small fraction of the nuclear SE proteins. It has also been reported that DCL1 lacking the terminal dsRBD is distributed diffusely in the nucleoplasm, which might imply that this domain is necessary for localization in the Dicing bodies (Wu et al., 2007). *Arabidopsis* possesses at least four additional HYL1 homologs, dsRNA-binding proteins 2-5 (DRB2-5). While DCL1 has been shown to preferentially interact with HYL1, DCL4 interacts exclusively with DRB4. This may suggest that each DCL protein requires the assistance of a specific DRB cofactor for its proper functioning (Hiraguri et al., 2005).

Recent advances in high-throughput sequencing technology have enabled deep surveying of small RNA fractions. Some of the newly identified srRNAs have already been proven to be Dicer-dependent. However, in many cases the biogenesis pathway is still not clear. Deeper structural and functional studies on Dicer and Dicer-like proteins may help to discover novel roles and properties of small RNAs.

## References

Axtell M.J., Westholm J.O., Lai E.C. (2011) Genome Biol. 12: 221.

Bernstein E., Caudy A.A., Hammond S.M., Hannon G.J. (2001) Nature 409: 363-366.

Blevins T., Rajeswaran R., Shivaprasad P.V., Beknazariants D., Si-Ammour A., Park H.S., Vazquez F., Robertson D., Meins F. Jr., Hohn T. et al. (2006) Nucl. Acids Res. 34: 6233-6246.

Bologna N.G., Mateos J.L., Bresso E.G., Palatnik J.F. (2009) EMBO J. 28: 3646-3656.

Bouche N., Lauressergues D., Gasciolli V., Vaucheret H. (2006) EMBO J. 25: 3347-3356.

Burdisso P., Suarez I.P., Bologna N.G., Palatnik J.F., Bersch B., Rasia R.M. (2012) Biochemistry 51: 10159-10166.

Chapman E.J., Carrington J.C. (2007) Nature Rev. Genet. 8: 884-896.

de Jong D., Eitel M., Jakob W., Osigus H.J., Hadrys H., Desalle R., Schierwater B. (2009) Mol. Biol. Evol. 26: 1333-1340.

Deleris A., Gallego-Bartolome J., Bao J., Kasschau K.D., Carrington J.C., Voinnet O. (2006) Science 313: 68-71.

Dong Z., Han M.H., Fedoroff N. (2008) Proc. Nat. Acad. Sci. USA 105: 9970-9975.

Fang Y., Spector D.L. (2007) Curr. Biol. CB17: 818-823.

Gasciolli V., Mallory A.C., Bartel D.P., Vaucheret H. (2005) Curr. Biol. CB15: 1494-1500.

Hiraguri A., Itoh R., Kondo N., Nomura Y., Aizawa D., Murai Y., Koiwa H., Seki M., Shinozaki K., Fukuhara T. (2005) Plant Mol. Biol. 57: 173-188.

Kasschau K.D., Fahlgren N., Chapman E.J., Sullivan C.M., Cumbie J.S., Givan S.A., Carrington J.C. (2007) PLoS Biol. 5: e57.

Kok K.H., Ng M.H., Ching Y.P., Jin D.Y. (2007) J. Biol. Chem. 282: 17649-17657.

Kurihara Y., Takashi Y., Watanabe Y. (2006) RNA 12: 206-212.

Landthaler M., Yalcin A., Tuschl T. (2004) Curr. Biol. CB14: 2162-2167.

Lau P.W., Guiley K.Z., De N., Potter C.S., Carragher B., MacRae I.J. (2012a) Nature Struct. Mol. Biol. 19: 436-440.

Lau P.W., Potter C.S., Carragher B., MacRae I.J. (2012b) Structure 20: 1995-2002.

Lee Y.S., Nakahara K., Pham J.W., Kim K., He Z., Sontheimer E.J., Carthew R.W. (2004) Cell 117: 69-81.

Liu B., Li P., Li X., Liu C., Cao S., Chu C., Cao X. (2005) Plant Physiol. 139: 296-305.

Liu Q., Rand T.A., Kalidas S., Du F., Kim H.E., Smith D.P., Wang X. (2003) Science 301: 1921-1925.

Ma E., MacRae I.J., Kirsch J.F., Doudna J.A. (2008) J. Mol. Bio. 380: 237-243.

Ma E., Zhou K., Kidwell M.A., Doudna J.A. (2012) J. Mol. Biol. 422: 466-476.

Macrae I.J., Zhou K., Li F., Repic A., Brooks A.N., Cande W.Z., Adams P.D., Doudna J.A. (2006) Science 311: 195-198.

Margis R., Fusaro A.F., Smith N.A., Curtin S.J., Watson J.M., Finnegan E.J., Waterhouse P.M. (2006) FEBS Lett. 580: 2442-2450.

Mukherjee K., Campos H., Kolaczkowski B. (2013) Mol. Biol. Evol. 30: 627-641.

Park J.E., Heo I., Tian Y., Simanshu D.K., Chang H., Jee D., Patel D.J., Kim V.N. (2011) Nature 475: 201-205.

Park W., Li J., Song R., Messing J., Chen X. (2002) Curr. Biol. CB12: 1484-1495.

Pederson T. (2010) RNA 16: 1865-1869.

Persson H., Kvist A., Vallon-Christersson J., Medstrand P., Borg A., Rovira C. (2009) Nature Cell Biol. 11: 1268-1271.

Pyle A.M. (2008) Ann. Rev. Biophys. 37: 317-336.

Qin H., Chen F., Huan X., Machida S., Song J., Yuan Y.A. (2010) RNA 16: 474-481.

Saito K., Ishizuka A., Siomi H., Siomi M.C. (2005) PLoS Biol. 3: e235.

Schauer S.E., Jacobsen S.E., Meinke D.W., Ray A. (2002) Trends Plant Sci. 7: 487-491.

Song L., Han M.H., Lesicka J., Fedoroff N. (2007) Proc. Nat. Acad. Sci. USA 104: 5437-5442.

Taft R.J., Glazov E.A., Lassmann T., Hayashizaki Y., Carninci P., Mattick J.S. (2009) RNA 15: 1233-1240.

Tsutsumi A., Kawamata T., Izumi N., Seitz H., Tomari Y. (2011)Nature Struct. Mol. Biol. 18: 1153-1158.

Vazquez F., Blevins T., Ailhas J., Boller T., Meins F. Jr. (2008) Nucl. Acids Res. 36: 6429-6438.

Welker N.C., Maity T.S., Ye X., Aruscavage P.J., Krauchuk A.A., Liu Q., Bass B.L. (2011) Mol. Cell 41: 589-599.

Welker N.C., Pavelec D.M., Nix D.A., Duchaine T.F., Kennedy S., Bass B.L. (2010) RNA 16: 893-903.

Wittmann J., Jack H.M. (2010) Sci. World J. 10: 1239-1243.

Wu F., Yu L., Cao W., Mao Y., Liu Z., He Y. (2007) Plant Cell 19: 914-925.

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: E104.

Yang L., Liu Z., Lu F., Dong A., Huang H. (2006) Plant J. Cell Mol. Biol. 47: 841-850.

Yi R., Qin Y., Macara I.G., Cullen B.R. (2003) Genes Develop. 17: 3011-3016.

Zhang H., Kolb F.A., Jaskiewicz L., Westhof E., Filipowicz W. (2004) Cell 118: 57-68.

Zvereva A.S., Pooggin M.M. (2012) Viruses 4: 2578-2597.