International Letters of Natural Sciences
ISSN: 2300-9675, Vol. 61, pp 85-94
doi:10.56431/p-p7a5s1

Submitted: 2016-10-30
Revised: 2016-12-18
Accepted: 2016-12-19
Online: 2017-01-10

DCL and Associated Proteins of *Arabidopsis thaliana* - An Interaction Study

Paushali Roy[1*] and Abhijit Datta[2]

[1]Department of Botany, Vidyasagar College, Kolkata, West Bengal, India

[2]Department of Botany, Jhargram Raj College, Jhargram, West Bengal, India

[*]E-mail address: paushali.06@gmail.com

Keywords: RNA interference, Dicer-like protein, DCL- associated proteins, domain-domain interactions.

Abstract. During RNA interference in plants, Dicer-like/DCL proteins process longer double-stranded RNA (dsRNA) precursors into small RNA molecules. In *Arabidopsis thaliana* there are four DCLs (DCL1, DCL2, DCL3, and DCL4) that interact with various associated proteins to carry out this processing. The lack of complete structural-functional information and characterization of DCLs and their associated proteins leads to this study where we have generated the structures by modelling, analysed the structures and studied the interactions of *Arabidopsis thaliana* DCLs with their associated proteins with the homology-derived models to screen the interacting residues. Structural analyses indicate existence of significant conserved domains that may play imperative roles during protein-protein interactions. The interaction study shows some key domain-domain (including multi-domains and inter-residue interactions) interfaces and specific residue biases (like arginine and leucine) that may help in augmenting the protein expression level during stress responses. Results point towards plausible stable associations to carry out RNA processing in a synchronised pattern by elucidating the structural properties and protein-protein interactions of DCLs that may hold significance for RNAi researchers.

1. Introduction

RNA interference regulates gene expression with the help of small non-coding RNA molecules such as siRNA, miRNA, trans-acting siRNAs (ta-siRNAs), natural antisense transcript short interfering RNAs (natsiRNAs) using the cell's natural machinery [1]. In plants, Dicer-like/DCL proteins, an RNaseIII-like enzyme processes longer double-stranded RNA (dsRNA) precursors into small RNA molecules. *Arabidopsis thaliana* contains four DCL proteins- DCL1, DCL2, DCL3 and DCL4 [2]. DCL expression varies according to several factors such as the developmental stage of the organism and the environmental stress [3]. DCL1 produces microRNAs, DCL2 generates both stress-related natural antisense transcript short interfering RNAs and siRNAs, DCL3 produces siRNAs and DCL4 generates both trans-acting siRNAs and siRNAs [4]. Dicer-like/DCL proteins interact with a number of associated proteins for functioning. The associated interacting proteins identified through databases and literature sources include Hyponastic leaves1 (HYL1or DRB1), DRB protein family (Double stranded RNA binding protein family including DRB2, DRB3, DRB4, DRB5), Serrate, Dawdle, HUA enhancer1 (HEN1), RNA dependent RNA polymerase 2 (RDR2). DCL1 interacts with HYL1, Serrate, Dawdle, HEN1 and DRB proteins, DCL2 with HYL1 and DRB5, DCL3 with HYL1, DRB proteins, Dawdle and RDR2, DCL4 with DRB4, HYL1 and Dawdle. Hyponastic leaves 1 (HYL1/DRB1) combines with DCL to produce pre-small RNA (precursor small RNA) from pri-small RNA (primary small RNA) molecules [5, 6, 7, 8, 9, 11, 12, 14]. Serrate associates with DCL1/HYL1 complex for splicing of primary miRNAs [9, 10]. These small RNAs are protected by 2'O methylation by Hua enhancer 1 (HEN1) [7, 12, 16]. Dawdle (DDL) associates with DCL and helps in accumulation of sRNA molecules [9]. The DRB family proteins help in locating the small RNA target sequences [11, 13, 14, 16, 17, 18]. RDR2 associates with DCL3 to stabilize DCL3 protein complex [15, 19]. In this study, we have tried to characterize and provide insights on the complete structural properties of all four DCLs and nine

associated proteins of *Arabidopsis thaliana* along with deeper insights into the protein-protein interactions of DCLs.

2. Materials and Method

2.1. *Sequence retrieval*: The protein sequences were retrieved from NCBI GenPept and cross verified against UniprotKB and Phytozome databases.

2.2. *Homology Modelling*: For template identification, the retrieved sequences were subjected to BLASTp search against the Protein Data Bank. We employed the GenThreader and Dompred tools from PSIPRED for identifying homologues for sequences with poor BLAST results. Model quality was checked with the Q-MEAN scores [20, 21] and the structures were validated by analysing the Ramachandran plot.

2.3. *Identification of conserved domains*: The conserved domains of *Arabidopsis thaliana* DCL and associated proteins were identified by searching against the NCBI Conserved Domain Database (CDD) and were mapped onto the constructed models.

2.4. *DCL-Associated protein docking*: The four *Arabidopsis thaliana* DCLs were docked to their respective associated proteins using the FFT-based rigid body docking algorithm of ClusPro v 2.0, followed by clustering and refinement. The lowest scoring complex of the largest cluster was selected in each case. Complexes were scored using the "balanced" Cluspro score, which considers van der waals attraction and repulsion terms, together with electrostatic and desolvation contributions [23, 24, 25, 26].

2.5. *DCL-Associated protein interactions*: Hydrogen bonds and hydrophobic interactions between the DCLs and their associated proteins were determined with Ligplot+. The interacting residues were mapped onto the conserved domains to identify any domain-domain interactions.

3. Results and Discussion

3.1. *Sequence retrieval*: The accession numbers of the selected sequences after cross checking the retrieved sequences from NCBI Genpept against UniprotKB and Phytozome databases are shown in Table 1.

Table 1. Accession Numbers of Protein Sequences

Protein	*Accession Number*	*Protein*	*Accession Number*
DCL1	332189100	DAWDLE	15232296
DCL2	332640405	DRB2	18401724
DCL3	332644322	DRB3	75273549
DCL4	322510021	DRB4	22331912
HYL1	18391056	DRB5	30693732
SERRATE	14486602	RDR2	75318859
HEN1	75333381		

Table 2. Homology Model Validation

PROTEIN	*RAMACHANDRAN PLOT ANALYSES (% of Amino Acids in Allowed Regions)*	*MODEL QUALITY SCORES*	
		Q-MEAN score	*Z-score*
DCL1	95.3%	0.111	-7.21
DCL2	94.5%	0.086	-7.49
DCL3	94.7%	0.031	-8.10
DCL4	95.9%	0.018	-8.23
HYL1	95.2%	0.103	-7.84
DRB2	93.8%	0.290	-5.67
DRB3	96.9%	0.364	-4.78
DRB4	94.9%	0.101	-7.81
DRB5	95.7%	0.254	-5.97
HEN1	96.9%	0.563	-2.21
RDR2	96.9%	0.094	-7.40
SERRATE	96.7%	0.023	-8.18

3.2. *Identification of conserved domains*: The domain compositions of all the four DCLs are similar, only the residue intervals vary (ex. in DCL3 where DSRM domain is absent). Rnc and MPH1 are multi domains (Rnc domain consists of RIBOc and DSRM domains within it in DCL1, DCL2 and DCL4 and only RIBOc in DCL3 and MPH1 domain consists of DEXDc and HELICc domains within it) [Refer Table 3A and 3B]. The modelled structures of the proteins with the mapped conserved domains are shown in Fig.1 and 2 respectively.

Table 3A. Domains of Dicer-like proteins (DCLs)

Dicer-like protein (DCL)			
DCL1		*DCL2*	
Domain Name$^{\Phi}$	Residue Composition	Domain Name$^{\Phi}$	Residue Composition
PAZ	1177-1321	PAZ	807-937
RIBOc	1576-1731	RIBOc	1164-1318
RIBOc	1358-1537	RIBOc	980-1123
DEXDc	263-415	DEXDc	38-192
DSRM	1837-1905	HELICc	376-489
HELICc	648-764	Dicer_dimer	559-638
Dicer_dimer	840-935	DSRM	1323-1383
DSRM	1735-1785	Rnc	1151-1384
Rnc	1558-1785	*DCL4*	
DEXDc	250-456	Domain Name$^{\Phi}$	Residue Composition
DCL3		RIBOc	1308-1452
Domain Name$^{\Phi}$	Residue Composition	RIBOc	1104-1261
PAZ	824-962	DEXDc	138-289
RIBOc	1214-1361	HELICc	467-600
RIBOc	1006-1175	DSRM	1621-1681
DEXDc	58-203	Dicer_dimer	656-747
HELICc	392-506	PAZ	937-1080
Rnc	1200-1435	DSRM	1463-1527
		MPH1	124-640
		Rnc	1285-1531

Table 3B. Domains of Associated proteins

Associated Proteins			
SERRATE		*DAWDLE*	
Domain Name[Φ]	Residue Composition	Domain Name[Φ]	Residue Composition
DUF3546	197-280	FHA	195-302
HYL1		COG1716	208-304
Domain Name[Φ]	Residue Composition	DRB2	
DSRM	16-82	Domain Name[Φ]	Residue Composition
DSRM	103-168	DSRM	2-68
DRB3		DSRM	88-153
Domain Name[Φ]	Residue Composition	DRB4	
DSRM	2-68	Domain Name[Φ]	Residue Composition
DSRM	88-150	DSRM	83-148
DRB5		DSRM	5-71
Domain Name[Φ]	Residue Composition	HEN1	
DSRM	2-68	Domain Name[Φ]	Residue Composition
DSRM	88-153	Methyltransf_12	718-806
RDR2		DSRM	11-85
Domain Name[Φ]	Residue Composition		
RdRP	389-965		

[Φ] PAZ- Piwi, Argonaut, and Zwille domain for nucleic-acid binding; RIBOc- Ribonuclease III C terminal domain-a dsRNA-specific endonuclease; DEXDc- DEAD-like helicases superfamily; DSRM- Double-stranded RNA binding motif; HELICc- Helicase superfamily c-terminal domain; DEXDc-, DEAD-, and DEAH-box proteins; Dicer_dimer- Dicer dimerization domain; Rnc- dsRNA-specific Ribonuclease; MPH1- ERCC4-like helicases; DUF3546- Domain of unknown function; FHA- Forkhead associated domain (FHA); COG1716- FOG: FHA domain; DSRM- Double-stranded RNA binding motif; Methyltransf_12- Methyltransferase domain; RdRP- RNA dependent RNA polymerase. (*Source- NCBI Conserved domain database*)

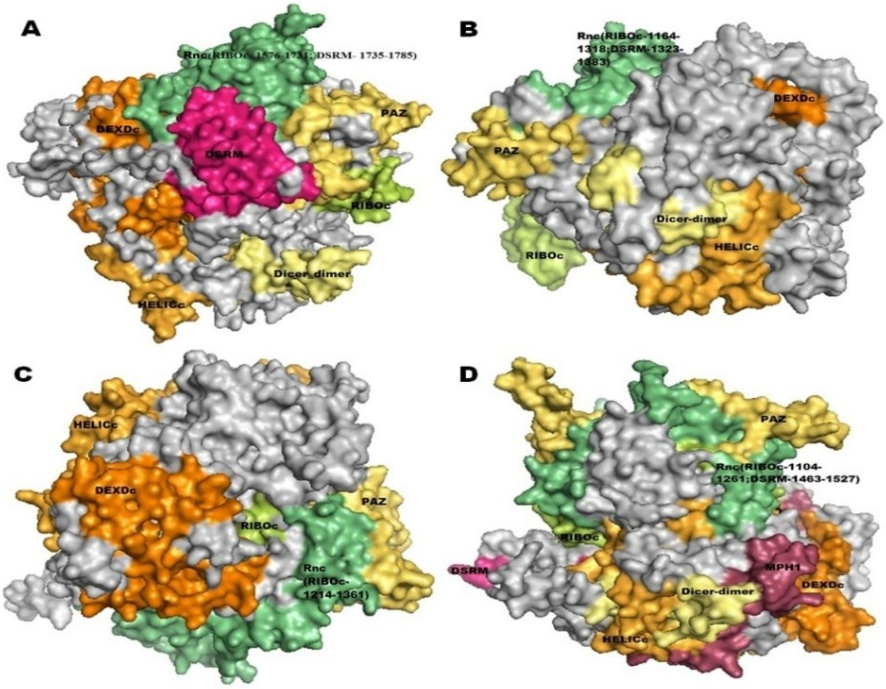


Figure 1 (A-D). Domain composition of DCL proteins: (A) Structure of DCL1. (B) Structure of DCL2. (C) Structure of DCL3. (D) Structure of DCL4. *Colour Code*- DEXDc- orange; HELICc- brightorange; Dicer_dimer- pale yellow; PAZ- yellow orange; RIBOc-lemon; RnC-limegreen; DSRM- hotpink; MPH1- raspberry

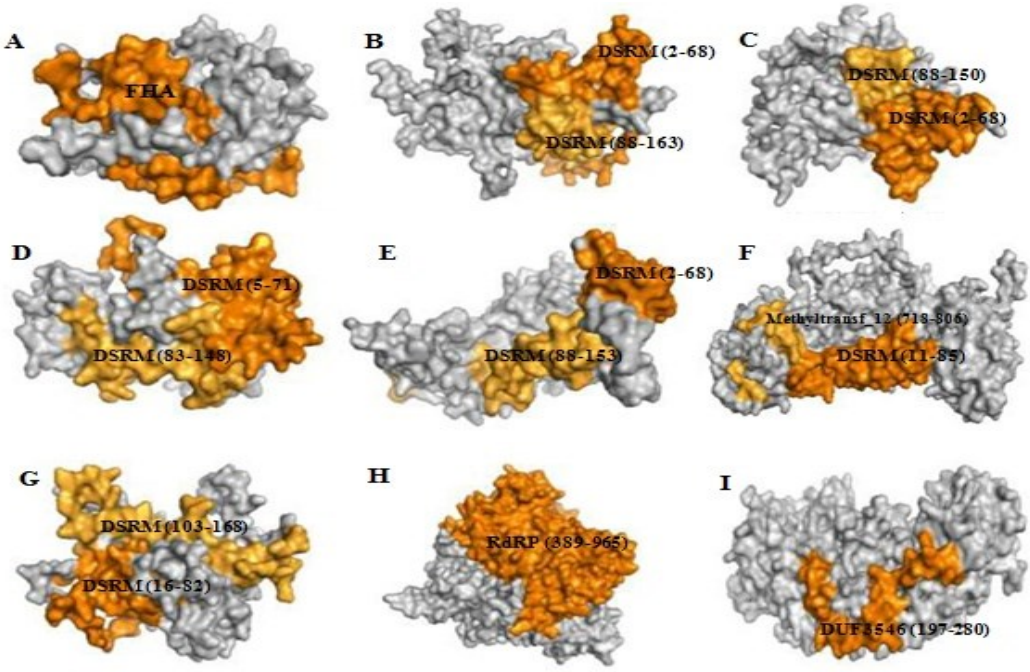


Figure 2. Domain composition of Associated Proteins: (A) DAWDLE. (B) DRB2. (C) DRB3. (D) DRB4. (E) DRB5. (F) HEN1. (G) HYL1. (H) RDR2. (I) SERRATE

3.3. *DCL-Associated protein docking*: The lowest cluster energy score for each DCL-associated protein docking is represented graphically in Fig.3. DCL1-DRB4 (DCL1-associated proteins interactions), DCL2-HYL1 (DCL2-associated proteins interactions), DCL3-DRB5 (DCL3-associated proteins interactions) and DCL4-HYL1 (DCL4-associated proteins interactions) show the lowest cluster energy scores and thus, maximum stability. A propensity of stronger DCL-DRB/HYL1 associations has been observed suggesting a probable necessity of firm protein-protein interactions for selection of primary small RNA targets to construct precursor small RNAs. The

most strong interaction of DCL3-DRB5 complex among all the associations may be indicative of the fact that the DRB proteins (primarily DRB5) may help DCL3 in locating the target RNA sequences binding as the DSRM domain is absent in *Arabidopsis thaliana* DCL3. The docked complexes of all the four DCLs with their respective associated protein partners are shown in Fig.4 (A-D).

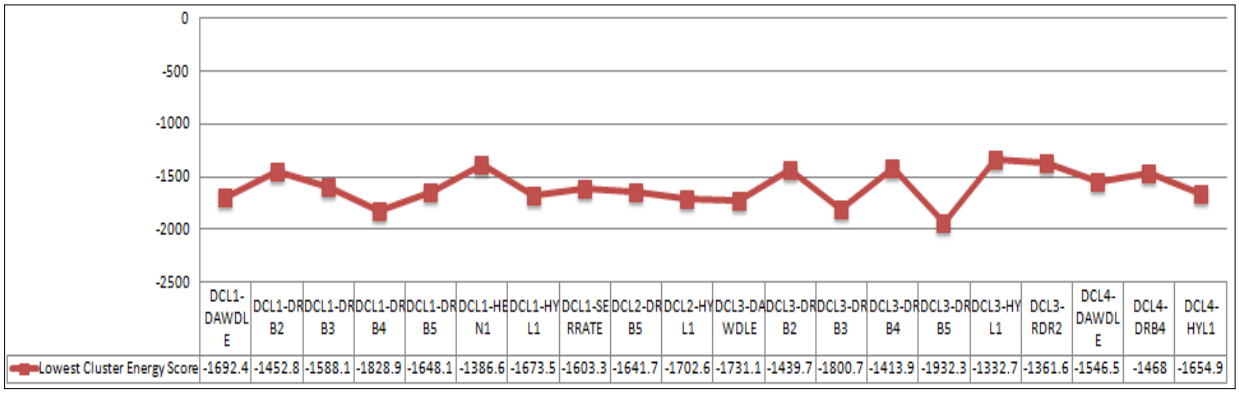
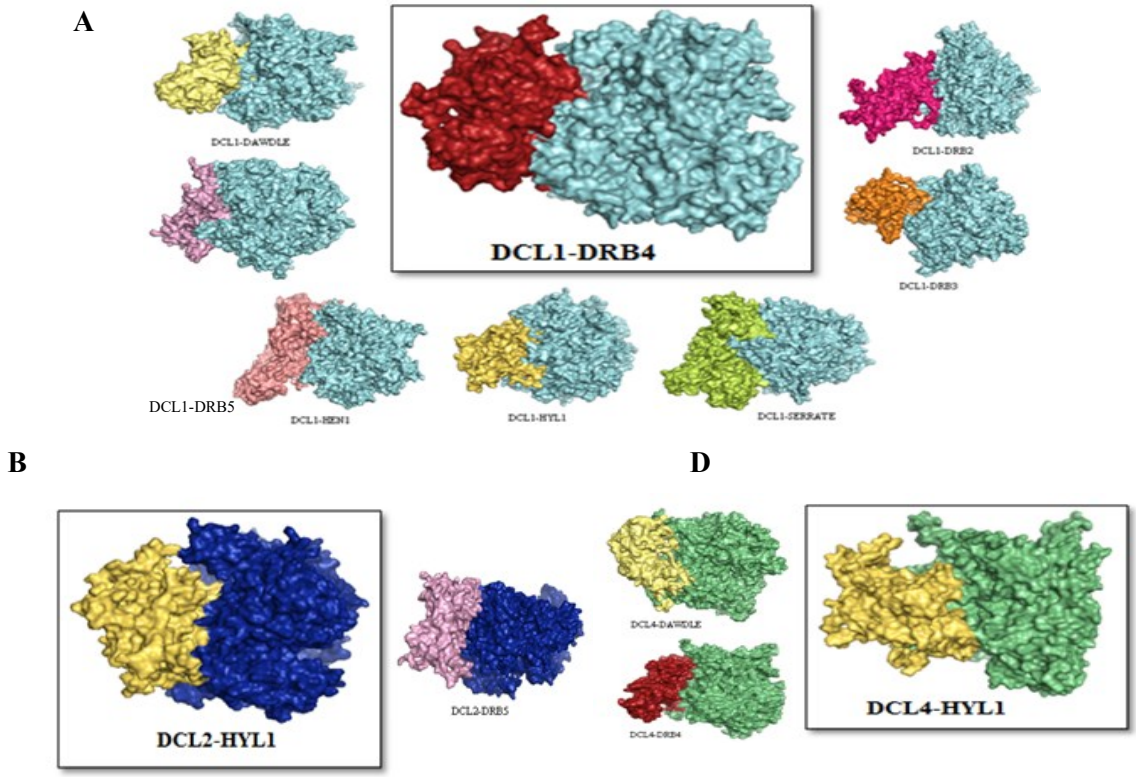


Figure 3. Graphical representation of DCL-associated proteins interactions

3.4. *DCL-Associated protein interactions*: There is predominance of Rnc, RIBOc, HELICc and PAZ domains of DCLs in forming domain-domain interactions. The participation of multidomains may possibly assist in maintaining the stability of the docked complexes [Table 4]. There is a bias for arginine usage in hydrogen bonds of DCLs and associated proteins interactions. The occurrence of arginine in hydrogen bonds formation as well as in hydrophobic interactions may be due its amphipathic nature. Hydrophobic interactions of DCLs-associated proteins show leucine bias and associated proteins-DCLs show arginine bias. Interestingly, arginine is found to suppress protein interactions that differ from our results [Table 5A and 5B]. The role of leucine in functional stability is not apparent till date and requires further investigation. We have also observed inter-residue interactions in the interfaces. Arginine is also found to play important role in plant stress tolerance and perhaps this arginine bias help the DCL proteins to regulate their mechanism during different environmental stresses. All these findings though require further experimental validation.



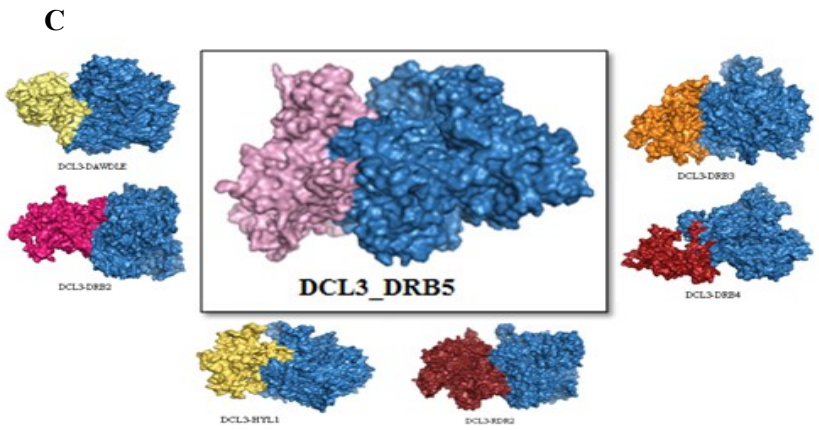


Figure 4 (A-D). *Anticlockwise*- (A) DCL1-associated protein interactions (B) DCL2-associated protein interactions (C) DCL3- associated protein interactions (D) DCL4-associated protein interactions. (Colour code: DCL1- *Cyan*; DCL2- *Blue*; DCL3- *Skyblue*; DCL4- *Limegreen*; DAWDLE- *Pale yellow*; DRB2-*Hot pink*;DRB3- *TV_Orange*;DRB4- *Firebrick*;DRB5-*Pink*; HEN1- *Salmon*;HYL1- *Yelloworange*; RDR2- *Ruby*; SERRATE- *Lemon*) **[**docked complex with lowest energy score are emphasized for each interactions**]**

Table 4. Domain-Domain Interactions

DCL Protein	DCL Protein Domain	Associated Protein	Associated Protein Domain
DCL1	*HELICc*	DRB2	*DSRM*
DCL1	*PAZ*; *RIBOc*	DRB5	*DSRM*
DCL1	*RIBOc*	SERRATE	*DUF3546*
DCL3	*RIBOc, RnC*	DAWDLE	*FHA, COG1716*
DCL3	*RIBOc*, *Rnc*, *HELICc*	DRB2	*DSRM*
DCL3	*HELICc*	DRB3	*DSRM*
DCL3	*RIBOc*, *Rnc*	DRB5	*DSRM*
DCL3	*HELICc*	HYL1	*DSRM*
DCL3	*RIBOc*, *Rnc*	RDR2	*RdRP*
DCL4	*PAZ*	DAWDLE	*FHA*, *COG1716*
DCL4	*RIBOc*, *Rnc*, *PAZ*	DRB4	*DSRM*
DCL4	*RIBOc*, *Rnc*, *PAZ*	HYL1	*DSRM*

Table 5A. Hydrogen bonds and Hydrophobic Interactions of DCLs with Associated Proteins[Δ]

Hydrogen Bonds					Hydrophobic Interactions				
DCL Protein	Associated Protein	Preferred Residue/protein	Residue Preferred (in total)	Most Preferred	DCL Protein	Associated Protein	Residue Preferred Residue/protein	Residue Preferred(in total)	Most Preferred Residue
DCL1	DAWDLE	Asp 6, Glu 6	Arg 31	Arg 80	DCL1	DAWDLE	Leu 9	Glu 22	Leu 67
	DRB2	Ser 4				DRB2	Ala 12		
	DRB3	Glu 11				DRB3	Glu 10		
	DRB4	Glu 9				DRB4	Glu 6		
	DRB5	Arg 6, Asp 6				DRB5	Leu 8		
	HEN1	Arg 15				HEN1	Arg 8		
	HYL1	Arg 10				HYL1	Arg 6, Glu 6		
	SERRATE	Arg 14				SERRATE	Lys 5		
DCL2	DRB5	Glu 4	Asn 6		DCL2	DRB5	Leu 7	Leu 17	
	HYL1	Asn 6				HYL1	Leu 10		
DCL3	DAWDLE	Arg 6	Arg 46		DCL3	DAWDLE	Leu 8	Leu 28	
	DRB2	Tyr 2, Glu 2				DRB2	Leu 5		
	DRB3	Arg 4				DRB3	Leu 9		
	DRB4	Arg 8				DRB4	Phe 6		
	DRB5	Arg 15				DRB5	Ser 7		
	HYL1	Arg 3				HYL1	Leu 6		
	RDR2	Arg 10				RDR2	Arg 8		
DCL4	DAWDLE	Tyr 5	Lys 7		DCL4	DAWDLE	Phe 7, Tyr 7	Phe 14	
	DRB4	Arg 3				DRB4	Leu 5		
	HYL1	Lys 7				HYL1	Phe 7		

Table 5B. Hydrogen bonds and Hydrophobic Interactions of Associated Proteins with DCLs [Δ]

Hydrogen Bonds					Hydrophobic Interactions				
Associated Proteins	DCL Proteins	Preferred Residue/protein	Residue Preferred(in total)	Most Preferred Residue	Associated Proteins	DCL Proteins	Preferred Residue/protein	Residue Preferred(in total)	Most Preferred Residue
DAWDLE	DCL1	Arg 13	Arg 41	Arg 89	DAWDLE	DCL1	Arg 23	Arg 75	Arg 159
	DCL3	Arg 18				DCL3	Arg 32		
	DCL4	Arg 10				DCL4	Arg 20		
DRB2	DCL1	Arg 4	Arg 5		DRB2	DCL1	Tyr 10	Tyr 10	
	DCL3	1 residue per interaction				DCL3	Arg 6, Pro 6		
DRB3	DCL1	Arg 15	Arg 23		DRB3	DCL1	Arg 25	Arg 44	
	DCL3	Arg 8				DCL3	Arg 19		
DRB4	DCL1	Asn 4, Gln 4	Gln 7		DRB4	DCL1	Ser 17	Ser 17	
	DCL3	Arg 4				DCL3	Leu 12		
	DCL4	Gln 3, Thr 3				DCL4	Gln 9		
DRB5	DCL1	Arg 7	Arg 16		DRB5	DCL1	Arg 12	Arg 20	
	DCL2	Arg 5				DCL2	Arg 8, Leu 8		
	DCL3	Arg 4				DCL3	Leu 11		
HEN1	DCL1	Asp 13	Asp 13		HEN1	DCL1	Asp 18	Asp 18	
HYL1	DCL1	Gln 6, Lys 6	Gln 9 Lys 9		HYL1	DCL1	Lys 13, Thr 13	Thr 39	
	DCL2	Asn 4, Thr 4				DCL2	Thr 12		
	DCL3	Gln 3				DCL3	Glu 9		
	DCL4	Ile 3, Leu 3, Lys 3, Thr 3				DCL4	Ile 14, Thr 14		
RDR2	DCL3	Asp 4, Glu 4	Asp 4 & Glu 4		RDR2	DCL3	Asp 10	Asp 10	
SERRATE	DCL1	Lys 5	Lys 5		SERRATE	DCL1	Arg 14	Arg 14	

[Δ]Arg-Arginine, Lys- Lysine-, Asp- Aspartic acid, Glu- Glutamic acid, Gln- Glutamine, Asn- Asparagine, Ser- Serine, Thr- Threonine, Tyr-Tyrosine, Ile- Isoleucine, Leu- Leucine, Phe- Phenylalanine.

4. Conclusion

The analyses reveal that the DCL proteins composed of a complex set of conserved domains that are involved in some key interactions to function specifically in processing the small RNAs. The strong DCL-HYL1/DRB interactions suggest a processive screening of RNA targets. The predominance of multi-domain interactions may help in stabilising the protein complexes. Arginine bias in hydrogen bond formation and hydrophobic interactions differ from the common perception of restraint of protein-protein interaction by arginine, though it requires further analyses. The predominance of charged as well as polar residues in forming hydrophobic interactions and inter-residue interactions may further helps in better understanding of protein stability. Specific residue biases may also play imperative role during abiotic stress responses and subsequent expressions of DCL proteins. All these interactions along with the key interacting residues possibly suggest that RNA processing by DCL is a synchronized process. Furthermore, the interaction data may possibly help towards better understanding of the role of Dicer-like/DCL proteins during miRNA and siRNA biogenesis pathways.

References

[1] B.L. Davidson, P.B. McCray Jr., Current prospects for RNA interference-based therapies, Nature Reviews Genetics. 12 (2011) 329-340.

[2] E.J. Finnegan, M.A. Matzke, The small RNA world, Journal of Cell Science. 116(23) (2003) 4689-4693.

[3] Q. Liu, Y. Feng, Z. Zhu, Dicer-like (DCL) proteins in plants, Funct. Integr. Genomics. 9(3) (2009) 277-286.

[4] A.F. Fusaro et al., RNA interference-inducing hairpin RNAs in plants act through the viral defence pathway, EMBO reports. 7(11) (2006) 1168–1175.

[5] X. Chen, microRNA biogenesis and function in plants, FEBS Letters. 579(26) (2005) 5923-5931.

[6] N.S. Mishra, S.K. Mukherjee, A Peep into the Plant miRNA World, The Open Plant Science Journal. 12 (2007) 1-9.

[7] A.L. Eamens et al., DRB2 Is Required for MicroRNA Biogenesis in *Arabidopsis thaliana*, PLoS ONE. 7(4) (2012) 1-15.

[8] Y. Kurihara, Y. Takashi, Y. Watanabe, The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis, RNA. 12 (2006) 206-212.

[9] B. Yu et al., The FHA domain proteins DAWDLE in *Arabidopsis* and SNIP1 in human's act in small RNA biogenesis, Proceedings of the National Academy of Sciences. 105(29) (2008) 10073–10078.

[10] Y. Fang, D.L. Spector, Identification of Nuclear Dicing Bodies Containing Proteins for MicroRNA Biogenesis in Living *Arabidopsis* Plants, Curr. Biol. 17(9) (2007) 818–823.

[11] A. Hiraguri et al., Specific interactions between Dicer-like proteins and HYL1/DRB-family dsRNA-binding proteins in *Arabidopsis thaliana*, Plant Mol. Biol. 57(2) (2005) 173-188.

[12] A. Eamens et al., RNA Silencing in Plants: Yesterday, Today, and Tomorrow, Plant Physiology. 147 (2008) 456–468.

[13] H. Qin et al., Structure of the Arabidopsis thaliana DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein–protein interaction, RNA. 16(3) (2010) 474–481.

[14] F. Vazquez et al., Arabidopsis endogenous small RNAs: highways and byways, Trends in Plant Science. 11(9) (2006) 460-468.

[15] Z. Xie et al., Genetic and functional diversification of small RNA pathways in plants, PLoS Biol. 2(5) (2004) 642-652.

[16] F. Qu et al., Arabidopsis DRB4, AGO1, AGO7, and RDR6 participate in a DCL4-initiated antiviral RNA silencing pathway negatively regulated by DCL1, Proceedings of the National Academy of Sciences. 105(38) (2008) 14732-14737.

[17] Y. Nakazawa et al., The dsRNA-binding protein DRB4 interacts with the Dicer-like protein DCL4 in vivo and functions in the trans-acting siRNA pathway, Plant Mol. Biol. 63(6) (2007) 777-785.

[18] A. Fukudome et al., Specific requirement of DRB4, a dsRNA-binding protein, for the in vitro dsRNA-cleaving activity of Arabidopsis Dicer-like 4, RNA. 17(4) (2011) 750-760.

[19] A. Mallory, H. Vaucheret, Form, Function, and Regulation of ARGONAUTE Proteins, The Plant Cell. 22 (2010) 3879–3889.

[20] P. Benkert et al., QMEAN: A comprehensive scoring function for model quality assessment, Proteins: Structure, Function, and Bioinformatics. 71(1) (2008) 261-277.

[21] P. Benkert, et al., QMEAN Server for Protein Model Quality Estimation, Nucleic Acids Res. 37(1) (2009) 1-5.

[22] P. Benkert et.al., Toward the estimation of the absolute quality of individual protein structure models, Bioinformatics. 27(3) (2011) 343-350.

[23] S.R. Comeau et al., ClusPro: an automated docking and discrimination method for the prediction of protein complexes, Bioinformatics. 20(1) (2004) 45-50.

[24] S.R. Comeau et al., ClusPro: a fully automated algorithm for protein-protein docking, Nucleic Acids Research. 32(2) (2004) 96-99.

[25] D. Kozakov et al., PIPER: An FFT-based protein docking program with pairwise potentials, Proteins. 65(2) (2006) 392-406.

[26] D. Kozakov et al., How good is automated protein docking? Proteins. 81(12) (2013) 2159-2166.