MOLECULAR INSIGHTS INTO MICRORNA PROCESSING BY ARABIDOPSIS THALIANA SERRATE AND DAWDLE

SATORU MACHIDA
(Bachelor of Agriculture and Life Science *(Hons.)**, Hirosaki University, Japan)***

A THESIS SUBMITTED FOR

THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2011

Acknowledgement

The majority of my experiments were performed in Temasek Life Sciences Laboratory Ltd. I would like to express my deepest gratitude to my supervisor A/Prof. Adam Yuan for his guidance in project planning, crystallographic data collection, data analysis, structural modelling and preparation of the manuscripts. I would also like to thank A/Prof. J Sivaraman, A/Prof. He Yuehui and A/Prof. Wang Shu for taking parts in the oral panel and Prof. Hew Choy Leong in QE panel. I am grateful to Dr. Yang Jing, Ms. Wang Meimei, Dr. Yang Xia, Ms. Ng Ailing, Dr. Tang Xuhua and Dr. Huang Jinshan for kind advice in biochemical analyses. I am thankful to Dr. Yang Shoong Wook, Dr .Ye Jian and Dr. Geng Yun Feng for providing some plant materials. I am indebted to Dr. Toshiro Ito for the opportunity to use the real-time PCR thermal cycler, Dr. Ke Yu for lending me a camera, and Chen Hong Ying for valuable help in protein crystallization and laboratory management. I thank Academia Sinica, Taipei, and Suranaree University of Technology, Thailand, for training me partly in crystallization and data collection. NUS Research Scholarship is gratefully acknowledged for supporting me for the first four years.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

mRNA	messenger RNA
TGS	transcriptional gene silencing
PTGS	post-transcriptional gene silencing
RNAi	RNA interference
pre-mRNA	precursor-mRNA
nt	24 nucleotides
siRNA	small interfering RNA
miRNA	microRNA
TAS	trans acting siRNA
pri-miRNA	primary-miRNA
RNaseIII	ribo-nuclease type III enzyme
RISC	RNA-induced silencing complex
DUF	domain of unknown function
dsRBD	double-stranded RNA binding domains
DCL1	Dicer-like1
SE	Serrate
HYL1	Hyponastic Leaves1
DDL	Dawdle
HEN1	HUA ENHENCER1
CBC	cap binding complex
CBP	cap binding protein

Ars2	Arsenate resistance protein 2
FHA	fork head associated
cDNA	complementary DNA
RT	reverse transcription
PCR	polymerase chain reaction
PMSF	phenylmethanesulfonylfluorideand
PBS	phosphate buffered saline
GST	glutathione-s-transferase
MAD	multiwavelength anomalous dispersion
DIG	digoxigenin
EMSA	electrophoretic mobility shift assay
IPTase	Isopentenyltransferase
Nt domain	Amino-terminal domain
Mid domain	Middle domain
core+C	SE core to carboxyl-terminus
ZF+C	SE Zinc-finger domain to carboxyl terminus
D-loop	Displacement loop
PCNA	Proliferating Cell Nuclear Antigen
helix-hairpin-helix-helix	HhH_2
tRNA	transfer-RNA
RNaseP	Ribonuclease-P
SNIP1	Smad Nuclear Interacting Protein 1

Summary

MicroRNAs are a class of small RNAs which regulate gene expression sequence-specifically as a part of a multi-protein complex containing Argonaute (Ago). The long precursor of a microRNA is typically transcribed from intergenic region and forms a stem-loop secondary structure. *Arabidopsis thaliana* uses ribonucleaseIII-type enzyme Dicer-like1 (DCL1) for step-wise processing of pri-miRNA into a single-stranded microRNA, 21-22 nucleotides in length. Serrate (SE) and Hyponastic-Leaves 1 (HYL1) are known to direcly improve the rate and accuracy of the processing by DCL1 both *in vitro* and *in vivo*. Dawdle (DDL) was recently reported to help processing of pri-miRNA *in vivo* but apart from SE and HYL1. Here we report the crystal structure of *Arabidopsis* SE core (residues 194-543) at 2.7Å and Dawdle FHA domain (residues 180-314) at 2.0Å. SE core adopts the "walking man-like" topology with amino-terminal α-helix (Nt), Middle (Mid) and caboxyl-terminal zinc-finger (ZF) domains. Pull-down assay shows that SE core provides the platform for HYL1 and DCL1 binding, whereas *in vitro* RNA binding and *in vivo* mutant rescue experiments suggest that the non-canonical zinc finger domain coupled with carboxyl-terminal tail binds miRNA precursors. We propose that SE works as a scaffold-like protein capable of binding both protein and RNA to guide the positioning of miRNA precursor toward DCL1 catalytic site within miRNA processing machinery. On the other hand, the crystal structure of DDL FHA domain reveals β-sandwich architecture with a uniquely long two strands and outwardly emanating loops. Superimposition of DDL FHA on structurally homologous proteins suggests the loop region connecting the strands forming β-sandwich of DDL FHA domain is a canonical phospho-threonine recognizing domain as it is seen among mammalian phospho-signalling molecules. Furthermore, the crystal packing in an asymmetric unit of DDL FHA has a glutamate residue inserted into the putative phospho-recognition cleft of a

symmetrically related molecule, indicating the ability of the strand-connecting loops to recognize an acidic residue that mimics a phosphate group. The sequence alignment of various FHA proteins reveals that a lysine or arginine residue following a conserved glycine, typical of phospho-recognizing FHA but not of phospho-independent FHA domain, is present in DDL. Taken together, we propose that, by the putative phospho-threonine binding cleft discovered in the crystal structure, DDL FHA is a phospho-recognition domain, and that, through its affinity to DCL, serves as a signal-inducible bridge between phospho-signalling and miRNA pathways.

Chapter One: Introduction

1. A brief account of landmarks toward post-transcriptional gene regulation

1.1. The thesis statement

The development of a plant body from an embryo to an adult plant involves construction of trillions of cells in a precise architecture in many different tissues. Arrangement of the cells under a certain condition renders each cell to a cellular environment where a cell participates in coordination of physiological processes for maintenance and stress response. Physiological processes are driven by proteins taking part in morphogenetic and metabolic pathways by elaborate interaction networks. The concatenation of molecular events in a physiological process is coordinated with timely availability of specific molecules and their accessibility to their interacting partners, substrates, or cellular structures in the compartment under a specific cellular environment. The feat to coordinate availability of specific proteins and nucleic acids is achieved by regulation of gene expression (gene regulation) as well as the stability and transportation of the molecules. This thesis concerns the molecular mechanism of two proteins from a post-transcriptional gene regulation pathway in *Arabidopsis*.

1.2. The transition of eukaryotic gene concepts

It is established that the cellular output of genetic information is manifested in the structures of proteins that eventually execute their roles three-dimensionally. On the other hand, RNAs had been considered mere carriers of information from chromosome to translation machineries in cytoplasm (1). In 1941, genes were considered nearly synonymous to protein production units, as advocated in “one gene-one protein hypothesis” proposed by George Beadle and Edward Tatum (2). Soon later, the theory accepted modifications to be “One gene one polypeptide hypothesis” as Vernon Ingram and others conducted genetic studies on

multimeric protein complexes (3). The protein-centric view of genetic information laid foundation for most processes of development and stress responses. That was quite reasonable given proteins' extraordinary versatility involving biochemical catalysis, anatomical structure, intra-and inter-cellular communications, not to mention regulatory pathways of gene expression.

1.3. Major discoveries in post-transcriptional regulation

One gene one polypeptide hypothesis had been considered essentially true for prokaryotic system where protein-coding genes are more densely packed on the genome than eukaryotic counterparts, with limited number of exceptions, namely, non-coding sequences playing marginal regulatory roles (4). The assumptions had dominated, for almost half a century, that the protein-centric view of gene applied to multi-cellular organisms as well, and that prokaryotic gene regulatory pathways were analogous to eukaryotic counterparts. In the mean time, Phillip Sharp and Richard J. Roberts described RNA splicing in 1977 whereby transcripts are modified to remove introns and join exons making a protein sequence shorter than the length predicted by translation of intact genomic sequence (5,6). Since then, the notion fell untenable that RNA carried the genetic information unscathed between genomic and protein sequences and spatially from nucleus to cytoplasm. Along the same line, it was Thomas Cech working on splicing mechanism that discovered self-splicing RNA and, together with Sidney Altman, described ribozymes (7-9). It was a seminal discovery of catalytic RNA after decades when RNAs were thought as a passive agent.

1.4. The disproportionate scaling of morphological complexity and protein-coding genes

As the advanced genome-sequencing technologies became more accessible, the question arose of why the number of protein-coding sequences in a genome appears disproportionate to morphological complexity of the organism (10).

These assumptions led logically to two subsidiary ideas:

1. The increased regulatory sophistication of more morphologically complex organism is achieved through the interactions of regulatory proteins intersecting with more complex regulatory sequences in promoters and untranslated regions of messenger RNAs (mRNA) (11).
2. The vast amounts of non-coding sequences in eukaryotic organisms are, apart from *cis*-acting regulatory sequences, are functional junkyards.

The latter idea was reinforced by the fact that many non-coding sequences are transposons and repetitive sequences (12), as well as introns to be "discarded" during mRNA maturation. The enigma was further deepened by the findings in human genomic sequence (13,14): The entire range of genomic sequence is covered by 19,438 known genes and 2,188 predicted genes. These genes have a total of 231,667 exons, with ~10.4 exons per locus and ~9.1 exons per transcript. The total length covered by the coding exons is ~34 Mb or ~1.2% of the euchromatic genome (13).

1.5. The indication of non-protein-coding regions

The genome sequences from other organisms also revealed that, while it was consistent with non-linear scaling of morphological complexity and exon numbers, there was shown to be a strong correlation between the extent of non-coding dominance in a genome and the relative complexity of higher organisms (10). Most of the non-coding sequences in genomes are in fact expressed. Indeed the majority of genomes ranging from yeasts, insects to mammals are

transcribed on both strands, even overlapping and interleaving the genes. Moreover, large distinctive populations of non-poly-adenylated transcripts were discovered both in the nuclei and cytoplasm (1,4). These observations confronted traditional protein-centric view of genetic information, resulting in a reasonable anticipation that this massive population of non-protein-coding sequences might uncover a hidden layer of systemic and tissue-specific regulatory factors. The potential functions of non-protein-coding RNAs are supported by the following facts (1):

1. All well-studied loci in mammals and insects expressed a large number of non-protein coding transcripts
2. Many of the experimentally detected non-coding RNA were differentially expressed and dynamically regulated.
3. Some non-coding RNAs have specific cellular localization.

1.6. The advantage of hybridization-based gene regulation

With these facts in hands, it follows to ask what could be the key advantages of non-coding RNAs over proteins if most non-coding RNAs were not junk but functional. RNAs can direct a precise interaction with its target poly-nucleotides by base-pairing even over a short stretches of nucleotides. This mode of interaction is far more efficient than protein-protein interactions. Hybridization-based interaction allows for additional number of regulatory controls embedded in genomes, because these regulatory controls can harbour flexibility within the genetic information without altering essential requirements for development and stress responses. RNAs intrinsically possess much more precise specificity of interaction with RNAs and DNAs than could be achieved by protein-nucleic acid interactions. This improves precision of the control and minimizes genetic noise arising from cross-talks (1). However, the concept of regulatory roles directly played by RNA remained vague until Richard A.

Jorgensen reported co-suppression in 1990 (15). It eventually led to discovery of RNA hybridization-based gene regulation known as RNA silencing. In animal system, it is more commonly called RNA interference (RNAi) and, the discovery is credited to Craig Cameron Mello and Andrew Z. Fire (16). RNA silencing can regulate gene expression at both transcriptional and post-transcriptional levels, which are termed transcriptional gene silencing (TGS) and post-transcriptional gene silencing (PTGS). These regulations are in part mediated by RNA-binding proteins crucial to processing of mRNA. In fact, all RNA transcripts are accompanied with RNA-binding proteins on the course of processing from the initiation of transcription to degradation of transcripts in the cytoplasm (17).

1.7. RNA processing

During a transcription of messenger RNA, a protein-coding DNA sequence is read by RNA polymerase II complex to produce precursor-mRNA (pre-mRNA). Several rounds of processing and modification are carried out by respective machineries to convert the precursors into mature mRNAs that qualifies for transportation and translation. The pre-mRNA acquire 7mG cap at 5' terminus, have intron removed, and are poly-adenylated at 3' terminus. These processes all affect the downstream events such as mRNA transport, stability and translational efficiency. It was reported that some 30% of *Arabidopsis* protein-coding transcripts are alternatively spliced to increase the coding capacity of the genome by expansion of proteome diversity. Many mRNAs transcribed by RNA polymerase II are differentially expressed and regulated in different cell types specific to developmental stages and environments. This proteomic diversity is further contributed to by RNA stability that is controlled by ribonucleases. PTGS conventionally features the class of ribonucleases that choose their target by hybridization with small RNAs.

2. Literature review of miRNA processing and its components

2.1. Small RNAs

Small RNAs refers to olig-ribonucleotides, 21-24 nucleotides (nt) in length derived from longer transcripts that form either partially structured or double-stranded RNA of endogenous or exogenous origins. The transcripts can arise from a non-coding sequence, introns, or double-stranded regions of viral genes, as well as mRNA. Based on their unique biogenesis pathways, endogenous small RNAs are broadly categorized into small interfering RNAs (siRNAs) and microRNAs (miRNAs). Plant miRNAs suppress coding genes in a sequence specific manner via messenger RNA cleavage or translational repression (18,19). Some miRNAs target on non-protein coding transcripts called *TRANS ACTING SIRNA (TAS)*, the cleavage of which yields trans-acting small interfering RNAs (tas-siRNAs) to initiate another round of silencing on protein-coding targets (20,21).

2.2. *MIR* locus

Mature miRNAs are commonly detected as 21-22nt small RNAs expressed endogenously in plants from defined loci called *MIR* genes. Most plant *MIR* genes exist as independent transcription units located in non-protein-coding regions. Nearly 200 *MIR* loci have been identified in *Arabidopsis* genome. The initial transcript encoded by a *MIR* gene is called primary-miRNA (pri-miRNA). It is transcribed by RNA polymerase II, followed by addition of 5’ cap and 3’ polyadenylation. The promotor elements include TATA box and transcription initiation elements found among other RNA polymerase II transcripts. The key feature of pri-miRNA includes a single stem-loop structure (22,23) containing mismatches within the fold-back of a transcript. But exceptions exist such as polycistronic *MIR* loci where a single pri-miRNA can form two or more stem-loops resulting in distinct mature

miRNA species (24). Some miRNAs are derived outside independent *MIR* loci as byproducts of host gene expression, such as spliced out introns or transposable elements (25).

2.3. Components of miRNA processing

Processing of miRNAs and siRNAs shares the unifying feature of PTGS, namely, initial cleavage of their precursors by (ribo-nuclease type III enzyme) RNaseIII dicer and subsequent incorporation into effecter complex called RNA-induced silencing complex (RISC) for target cleavage (18,26-29). In *Arabidopsis*, four dicer homologues called Dicer-likes1-4 (DCL1-4) have been reported to exist. All of them are predicted to contain amino-terminal helicase domain, domain of unknown function (DUF), PAZ domain, two RNaseIII domains and two double-stranded RNA binding domains (dsRBD). Here, small RNA pathways ramify according to the structural features of precursors allocated to different DCLs. Among them, DCL1 is known to process pri-miRNA. At least five components of miRNA processing machinery have been indentified hitherto: Dicer-like1 (DCL1), Serrate (SE), Hyponastic Leaves1 (HYL1), DAWDLE (DDL) and HUA ENHENCER1 (HEN1) (Figure1.1.). HEN1 is a methyltransferase to modify 2' hydroxyl group of 3' terminus on mature miRNAs. Terminal methylation of miRNA stabilizes the mature miRNA for RISC loading (26,30,31). SE and HYL1 are known to directly interact with DCL1 and co-localize in sub-nuclear bodies containing pri-miRNAs (19,32). SE-null mutants are destined to embryonic lethality. *Se-1*, a mutant line with short frame-shift near carboxyl-terminus, has miRNA production severely impaired (33-35) and leaf serrated. *Hyl1* mutants also show leaf serration and reduction in miRNA expression (36-40).

Xie et al (2010)

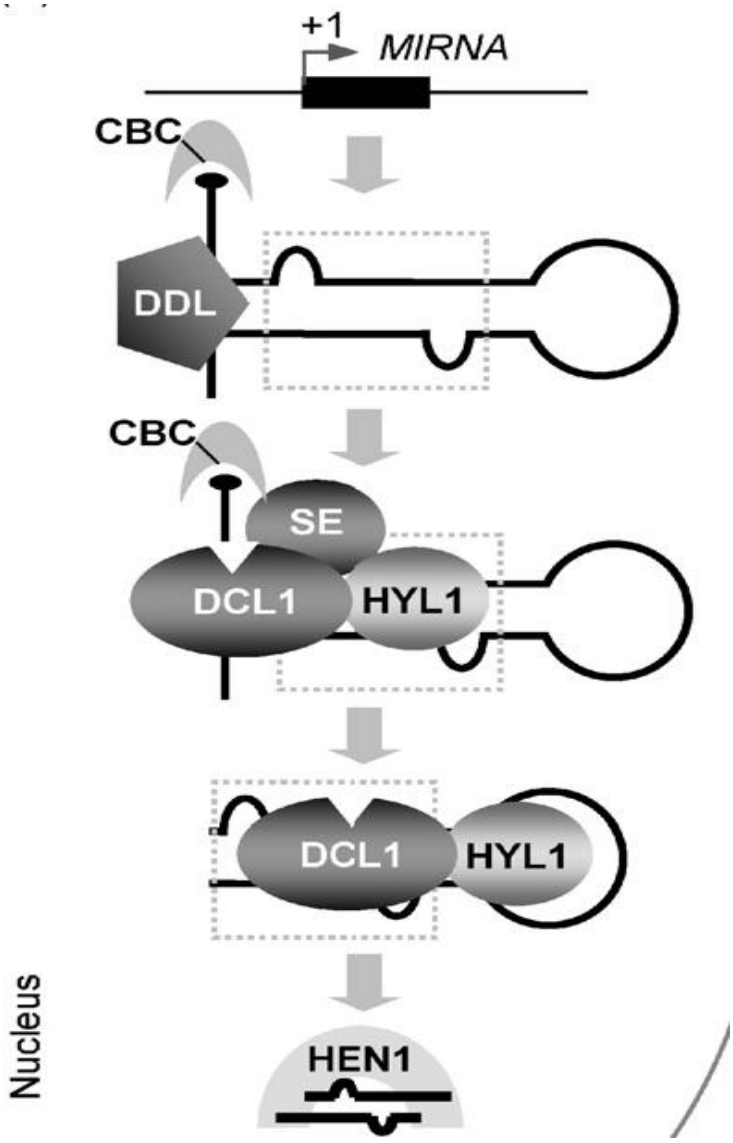


Figure 1.1. Canonical model of miRNA processing.
The stepwise processing of a pri-miRNA by the DCL1 complex ultimately gives rise to a miRNA duplex that is subsequently recognized and end-methylated by HEN1. The miRNA precursor is transcribed, go through modification processes common to mRNAs, and accompanied by cap-binding complex (CBC). SE and HYL1 recognizes the substrate and guide to DCL1,

2.4. The effect of HYL1 and SE *in vitro*

Though the effect of interaction between DCL1, HYL1 and SE had been reported by several groups (34-37,41-45), the gist of observations is found in the report by Nina Fedoroff's group (46). They purified recombinant full-length DCL1 expressed in insect cell and performed *in vitro* dicing assay to observe the dicing kinetics and accuracy in presence and absence of HYL1 and/or SE. The dicing efficiency was improved *in vitro* when the reaction system was supplemented with HYL1 or SE, and the effect was more significant in presence of both. Even though DCL1 alone was sufficient to sustain the low enzyme activity, the majority of the products were outside the correct sequence of the mature miRNAs that occur *in vivo*. The presence of both HYL1 and SE rescued the dicing accuracy up to 80% *in vitro* (Figure 1.2.). Yet, the suboptimal recovery of processing accuracy in presence of HYL1 and SE suggests existence of a missing component in the dicing system as noted by Dong et al (2008).

2.5. Effect of DDL *in vivo*

A fork-head associated protein DAWDLE (DDL) appears the most promising candidate for the missing component of the dicing complex. Loss of function mutation of DDL leads to compromised production of pri-miRNAs as well as mature miRNAs and siRNAs, and have a pleiotropic phenotype (47,48). Furthermore, DDL can interact with pri-miRNA and DCL1 *in vitro*, suggesting its role in upstream probably by facilitating the interaction between DCL1 and pri-miRNA. A notable observation was the reduced level of pri-miRNA as well as of siRNA and miRNA in *ddl* mutants (48). Reduction in activity of HYL1 and SE leads to cleavage defect and reduced miRNA and, accordingly pri-miRNA accumulates. On the other hand, loss of DDL activity decreases pri-miRNA as well (48), which suggests the role of DDL distinct from those of HYL1 and SE.

Dong et al (2008)

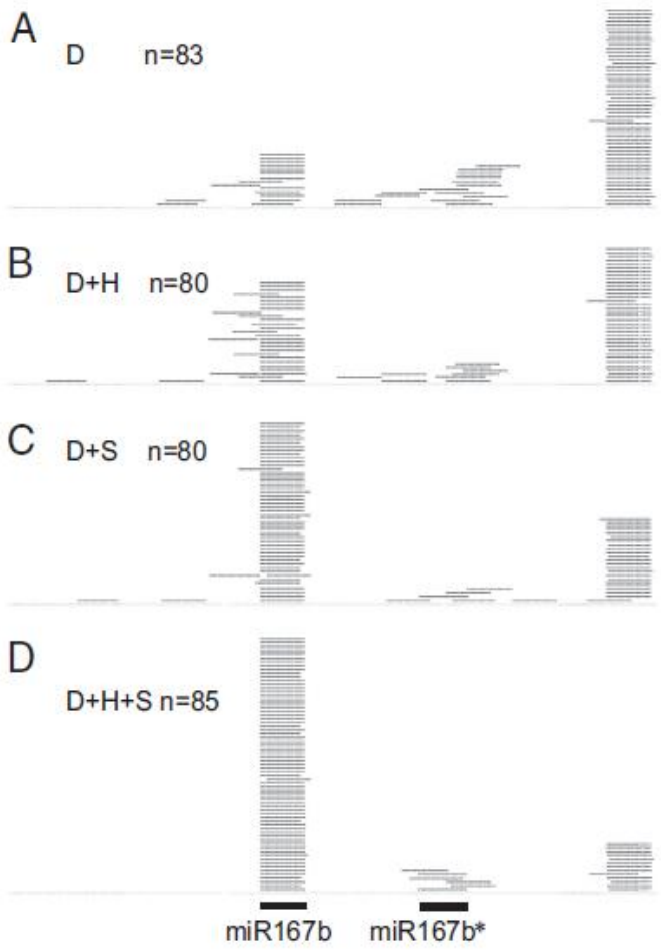


Figure 1.2. Effects of HYL1 and SE on the processing accuracy of DCL1 *in vitro*.
Distribution of sequenced small RNAs from *in vitro* processing reactions within that of the pri-miR167b substrate. Each small bar represents a single small RNA sequence. The recombinant proteins added to each reaction is indicated as:
A. Only DCL1 was added to the reaction.
B. DCL1 with HYL1
C. DCL1 with SE
D. DCL with HYL1 and SE

2.6. Involvement of other RNA processing proteins

The early events of miRNA processing involve at least some of mRNA maturation machineries responsible for capping, splicing, and poly-adenylation of protein-coding transcripts. Not surprisingly, nuclear-cap binding complex (CBC) has recently been reported to participate in pri-miRNA processing. CBC is a hetero-dimer complex made of cap-binding proteins (CBP) 20 and 80. Loss-of-function mutation in CBP80 and CBP20 resulted in accumulation of unspliced transcripts and pri-miRNAs, reduced level of mature miRNAs (49-51) and the pleiotropic phenotype resembling miRNA-defective *se-1*. Moreover, *se-1*mutant, along with *cbp80* and *cbp20,* shows elevated accumulation of unspliced transcripts (50). These observations established the dual roles of SE and CBC in pre-mRNA and pri-miRNA processing. However, these roles played by CBC in miRNA processing were only revealed by genetic studies and a direct interaction between components of mRNA- and miRNA-processing machineries remains to be detected in plant systems (49). The possibility of interaction between CBC and SE is conceivable from an analogous example in animal system. Arsenate resistance protein 2 (Ars2) from Chinese hamster is the only protein homologous to *Arabidopsis* SE and is known to interact with animal CBP20 and CBP80, and mediate miRNA processing (52-57). The knock out mutation of Ars2 causes embryonic lethality as with SE.

2.7. Unresolved mechanisms of miRNA processing

The molecular mechanism whereby DCL1 complex works on pri-miRNA and accurately produce miRNAs is yet to be elucidated. In animal system, pri-miRNA undergoes two successive cleavages of loop-distal and loop-proximal sites by nucleic Drosha and cytoplasmic Dicer respectively. In plant system, two steps of cleavage are both mediated by DCL1 complex in nucleus before exportation of 21nt-miRNA to cytoplasm (58). In addition,

the hairpin portion of the pri- and pre-miRNAs is longer and more diverse than those of animal counterparts (19). The structural determinants of the plant miRNA appear similar to animal ones, i.e. initial cleavage of loop-distal site and loop-proximal next (20,59,60). Yet, at least two *MIR* transcripts were found to be inconsistent with the canonical model of miRNA processing. MiR319 and 159 with prominently long loop go through first cleavage at loop-proximal site and subsequent three cuts before reaching mature miRNA (61,62). It implies existence of different miRNA processing pathways in plants. Further, the structural roles of HYL1 and SE may explain why lack of these proteins in dicing assay shifts the miRNA excision site to the loop-distal side of the negative strand of pri-miRNA (46). To initiate the structural efforts into the components of miRNA processing machinery in *Arabidopsis*, we report the crystal structure of *Arabidopsis* SE core (residues 194-543) at 2.7 Å, which displays a "walking man-like" topology. Our pull-down assays shows that SE core provides the platform for HYL1 and DCL1 binding, whereas *in vitro* RNA binding and *in vivo* mutant rescue experiments suggests that the conserved non-canonical zinc finger domain coupled with carboxyl-terminal tail recognizes miRNA precursors. Moreover, we present the crystal structure of fork head associated (FHA) domain of DDL (residues 180-314) at 2.0 Å which displays a β-sandwich architecture with loops harboring a putative phospho-recognition cleft. The functional implications obtained from those structures will be discussed.

Chapter Two: Materials & Methods

1. Construction of *E. coli* expression vectors

1.1. Truncation of SE and DDL

SE core (residues 194-543) was cloned from cDNA in frame into pET28b vector (Novagen) to be expressed with a carboxyl-terminal hexahistidine-tag. SE middle domain (residues 248-470) was cloned into pET28b and pGEX6p-1 vectors to bear carboxyl-terminal hexahistidine and amino-terminal glutathione-s-transferase (GST) tags, respectively. HYL1 dsRBD1 (residues 15-84), HYL1 dsRBD2 (residues 100-172) and HYL1 dsRBD1+2 (residues 15-172) were cloned from cDNA into pGEX6p-1 vector with amino-terminal GST. Cloned likewise were the fragments of DCL1, namely DCL1 DUF283 (residues 836-942), DCL1 PAZ (residues 1176-1353) and DCL1 RBD1+2 (residues 1732-1909) with amino-terminal GST. DDL FHA domain (180-314) was cloned from cDNA into pET28b to bear carboxyl-terminal hexahstidine.

1.2. Preparation of cDNA and amplification

Complementary DNA (cDNA) of DDL was synthesized from total RNA extracted by RNeasy (Qiagen). The reverse transcription (RT) was prepared based on the manufacturer's protocol, using AMV reverse transcriptase (Roche), followed by reaction 65^{o}C 5min, 42^{o}C 1h, and 95^{o}C 5min with one cycle each. The RT product was purified by ethanol precipitation. The protein-coding sequence of DDL was amplified by Polymerase Chain Reaction (PCR) using ***Taq*** 2X Master Mix (New England Biolabs). 50µl reactions were prepared with 0.1nmol forward and reverse primers, and 0.5-1.0µg first strand cDNA. The protein-coding sequence of SE was amplified from a 50-100ng plasmid template. The PCR program was 1cycle of 95^{o}C 15min, 5cycles of 95 ^{o}C 45sec, 48 ^{o}C 45sec and 72 ^{o}C 2min, 30 cycles of 95 ^{o}C 45sec, 56 ^{o}C 45sec and 72 ^{o}C 2min, and 1 cycle of 72 ^{o}C 10min.

1.3. Restriction digestion and ligation

The fragments were digested with restriction enzymes (New England Biolabs), gel-purified using QIAquick Gel Extraction Kit (Qiagen). The respective vectors were digested and purified likewise. The elution volume of the insert and the vector from the gel-extraction column was adjusted to have 40-100ng/µl DNA in the eluent. The insert and the vector were ligated with Quick DNA Ligation Kit (Roche Applied Science) in 11µl reaction containing at least 100ng insert at room temperature for 10-30min. The ligation product was immediately subjected to transformation with *Escherichia. coli* DH5α strain.

1.4. Chemical competent cell

The preparation of chemical competent cell is described briefly. The non-transformed competent cell was spread with a sterile needle on agar plate without antibiotics and incubated at 37^{o}C overnight. A colony was picked into 50ml LB liquid medium without antibiotics in an autoclaved conical flask, and shaken at 37^{o}C 230-250rpm for 6-8h till the turbidity reaches OD_{600} 0.5-1.0. The cell was collected by centrifugation and re-suspended in 0.1M chilled $CaCl_2$ and left on ice for 10-30min. The cell was again spun and re-suspended in 0.1M chilled $CaCl_2$ containing 25% glycerol and incubated likewise. 40µl aliquots were made in pre-cooled eppendorf tubes by a dropper, flash-frozen with liquid nitrogen and stored in -80^{o}C.

1.5. Bacterial transformation

The competent cell was thawed on ice while the ligatyion products were chilled on ice. 40µl competent cell was added into the eppendorf tube containing 11ul of the ligation product and incubated on ice for 40min, followed by heat-shock at 42^{o}C for 40-60sec. After cooling down on ice for 2min, the cell was recovered by addition of 0.5ml LB liquid medium and shaking at 37^{o}C 230-250rpm for 1h. The cells were collected by brief centrifugation and spread on LB agar plates containing respective antibiotics. pET28b vector is resistant to Kanamycin and pGEX-6p-1 vector to Ampicilin.

1.6. Screening of clones

Single colonies were picked from the LB agar plate into a culture tube containing 3ml LB liquid medium with antibiotics and shaken overnight at 37°C, 250rpm. The cells are spun down, re-suspended and extracted using AxyPrep column. The plasmids were screened for correct inserts by double-digestion using restriction enzymes (New England Biolabs). 10µl reaction containing 100-150ng plasmid was incubated at 37^{o}C for 10-30min and resolved on 1.0% agarose gel containing 1XTAE (0.04M Tris-acetate, 1mM EDTA, pH8.0) at 120V for 20min. The positive clones are selected by the electrophoretic mobility of the insert from the digested plasmid. The digestion releases the ligated insert, the size of which indicates successful cloning as it is electrophoresed at the same speed as the insert for ligation.

1.7. Sequencing and re-transformation

The positive clones are sequenced by dideoxynucleotide method at DNA sequencing service at Temasek LifeSciences Laboratory. Prior to submission for sequencing, the chain termination reaction was performed with BigDye (ABI) in 20µl reaction with 30 cycles of 96 ^{o}C 10sec, 50^{o}C 10sec and 60^{o}C 4min. The plasmids were subjected to re-transformation with

chemically competent *E. coli* (BL21/DE3 strain) for protein expression. The re-transformation procedure followed that of *E. coli* (DH5α) except that the cell was spread on LB agar plates containing chloramphenicol in addition to the antibiotics to which the respective vectors are resistant.

2. Protein expression and purification

2.1. Screening of protein expression

The protein expression was first tested in a small scale. A colony was picked from agar plates into culture tube containing 3ml LB liquid medium with antibiotics and shaken for 6-8h at 37^{o}C, 230-250rpm. 300ul of the incubated medium was further transferred to fresh LB liquid media containing the same combination of antibiotics and shaken likewise until the optical density at 600nm (OD_{600}) reached 0.5-0.6 measured by VivaSpec (Sartorius). The protein expression was induced with 0.4mM IPTG (isopropyl β-d-thiogalactoside). The tubes were moved to 20^{o}C shaker (230rpm) for 12-16h. Spun down at 4000rpm for 1min, cells were re-suspended in Buffer Z (25mM Tris-HCl pH7.4, 25mM Potassium Phosphate pH6.8, 500mM Sodium chloride, 10% Glycerol, 1mM DTT), sonicated, and spun down to collect the supernatant into a fresh tube. The pellet and supernatant were separated into different tubes and boiled with Laemmli loading dye (100mM Tris-HCl pH6.8, 100mM DTT, 2%SDS, 0.10%Bromophenol-blue, 10%Glycerol) for 5min at 100^{o}C and resolved by 15% SDS-PAGE gel with power supplied by Mini-PROTEAN Tetra Cell (BioRad) at 200-250V. The protein expression level and solubility were estimated from the band intensity of the protein from the pellets and supernatants run in parallel on the gel to be stained with Coomassie Brilliant Blue colloidal solution (0.1% R-250 dye, 10% acetic acid and 25% Isopropanol) at room temperature.

2.2. Large scale protein expression

Upon confirming the expression of soluble protein, the expression was scaled up to 2L for biochemical experiments and 8L for protein crystallization. The bacteria were grown in 2L conical flask with 1L LB medium in each. The cell suspension of buffer Z was supplemented with 0.5mM phenylmethanesulfonylfluorideand (PMSF) and was mechanically lyzed with High-Pressure EmulsiFlex-C3 (Avestin). To remove the cell debris, 200ml lysate was centrifuged at 40000rpm 4°C for 1h. The supernatant was transferred to a chilled bottle, ready to load into the affinity column.

2.3. Affinity chromatogrphy

The lysate was first loaded onto affinity column. The cell lysate containing hexahistidine tag was loaded onto Ni^{2+} affinity resin (GE Healthcare) and the one with GST tag was loaded onto glutathione sepharose resin (GE Healthcare). Prior to loading, the columns had been washed with washing buffers, specifically equilibirium buffer (250mM Sodium Chloride, 10mM Tris-HCl pH7.4, 10% Glycerol) for Ni^{2+} affinity resin and phosphate buffered saline (PBS) (140mM Sodium Chloride, 10mM Sodium hydrogen phosphate, 1.8mM Potassium dihydrogenphosphate , 2.7mM Potassium chloride pH7.3) for glutathione sepharose resin. The lysate was loaded using peristaltic pumps at flow rate 1.0ml/min in $4^{o}C$ until ~10 ml of the lysate remained in the feeding bottle. The loaded column was washed using ÄKTA Prime Plus purification system (GE) with the respective washing buffers until the chromatograph stabilized to the baseline. The bound protein was eluted from Ni^{2+} affinity resin with equilibrium buffer supplemented with successive concentration of imidazole 25, 250, 500mM at 4ml/min, while 5ml per fraction was collected into 14ml culture tubes. The elution from glutathione sepharose resin was performed with GST elution buffer (0.62% m/v reduced

glutathione, 10mM Tris-HCl pH8.0). 8-10 tubes of 5ml fraction were collected at 4ml/min. The concentration and purity of the target proteins were visually assessed from SDS-PAGE gel. The promising fractions were pooled together for subsequent purification steps.

2.4. Size-exclusion chromatography

Carboxyl-terminal hexahistidine-tagged SE core, SE N-terminal and Mid domains, and DDL FHA domains were purified with size exclusion chromatography following affinity purification. The fragments of DCL1 were fused with amino-terminal glutathione-s-transferase (GST) and finished with purification in the same way as hexahistidine-tagged proteins. As to RBD1 and RBD2 of HYL1, the GST tags were cleaved off by PreScission enzyme under a dialysis condition (100mM Sodium chloride, 10mM Tris-HCl pH7.4). The cleaved tag was removed by size-exclusion chromatography. Hexahistidine-tagged SE core and DDL FHA were concentrated up to ~15mg/ml using VivaSpin (Sartorius), while the sodium chloride concentration was maintained at 500mM.

2.5. Selenium-labelling of DDL FHA

Selenium-labeled DDL FHA was expressed in the same host cell and was grown in M9 minimal media (20% glucose, 1 mM $MgSO_4$ and 0.3 mM $CaCl_2$). Cells were grown at 37°C up to an OD_{600} of 0.5-0.6, followed by addition of lysine, phenylalanine and threonine at 50 mg/ml and isoleucine, leucine, valine and L-selenomethionine at 100 mg/ml into the culture and incubated for 15 minutes before 0.4 mM IPTG was added for induction at OD_{600} ~ 0.6. Chromatographic purification of the selenomethionine-labeled protein followed the same procedure as the native counterpart.

3. Crystallization, data Collection, and structure determination

3.1. SE core crystallization

Crystals of SE core were grown by hanging drop vapor diffusion at 20 ºC. Typically, a 2.0 µl hanging drop contained 1.0 µl of protein (15 mg/ml) mixed with 1.0 µl of reservoir containing 13.5 % PEG 4000, 0.7 M Potassium Formate and 100mM Tris (pH 7.4), and equilibrated over 1ml of reservoir solution. These crystals grew to a maximum size of 0.35 mm×0.2 mm×0.2 mm over the course of 2 days. For data collection, crystals were flash frozen (100K) in the above reservoir solution supplemented with 30% Glycerol. A total of 360 frames per wavelength of 1º oscillation were collected for each crystal on two wavelengths near the zinc edge (1.2814Å and 1.2818Å, respectively) and one wavelength at 1.1 Å; and subsequently processed by HKL2000 (www.hkl-xray.com). The crystals belong to space group $P2_12_12_1$, with unit cell dimensions a = 50.97 Å, b = 80.66 Å, c = 112.67 Å, and $\alpha = \beta = \gamma = 90°$ with one molecule per asymmetric unit (Figure 2.1., Table 2.1.).

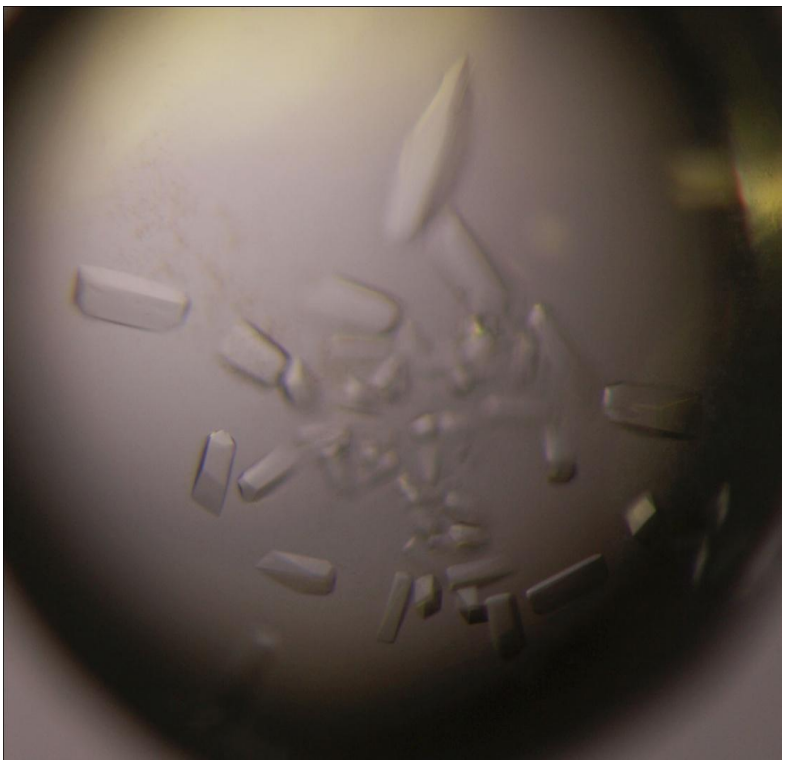


Figure 2.1. Protein crystal of SE core (aa194-543).
The crystal was obtained with hanging drop 20° over 13.5%PEG4k, 0.7M Pottasium Formate, 0.1M Tris pH7.4. The structure was solved with MAD method by zinc edge. The crystals belong to space group $P2_12_12_1$, with unit cell dimensions a = 50.97 Å, b = 80.66 Å, c = 112.67 Å, and $\alpha = \beta = \gamma = 90°$ with one molecule per asymmetric unit

Data collection			
Space group		$P2_12_12_1$	
	Peak	Inflection	Remote
Cell dimensions *a, b, c (Å)*		50.965, 80.663, 112.670	
Wavelength (Å)	1.2814	1.2818	1.1
Resolution (Å) [a]	20~2.7 (2.8~2.7)	20~2.7 (2.8~2.7)	20~2.7 (2.8~2.7)
R_{sym} (%)	7.3 (47.6)	7.6 (54.5)	7.7 (60.6)
$I/\sigma(I)$	32.1 (3.5)	29.6 (2.6)	28.5 (3.1)
Completeness (%) [a]	99.9 (99.6)	99.8 (98.2)	99.9 (99.9)
Redundancy	7.6 (6.8)	7.5 (6.0)	7.7 (7.3)
Initial Figure of Merit (Acentric/Centric)		0.348/0.268	
Refinement			
Resolution Range (Å)	20~2.7		
No. reflections	11,896		
R_{work} (R_{free}) (%)	24.3/28.6 (40.2/47.1)		
No. atoms			
Protein	2,437		
Zn	1		
Water	11		
B-factors (Å^2)			
Protein	74.1		
Zn	94.7		
Water	55.2		
R.m.s. deviations			
Bond lengths (Å)	0.011		
Bond angles (°)	1.3		
% favored (disallowed) in Ramachandran plot	92.5 (0)		

Table 2.1. Crystallography statistics of SE core.

3.2. DDL FHA crystalization

Crystals of native DDL FHA domain was grown by hanging drop vapour diffusion at 20 ºC. A 2.0 µl hanging drop contained 1.0 µl of protein (15 mg/ml) mixed with 1.0 µl of reservoir containing 30% PEG#8000 and 100mM HEPES (pH 6.8), and equilibrated over 1.0 ml of reservoir solution. The selenomethionine labeled DDL FHA domain failed to crystallize at the first instance. After supplementing the protein suspension solution with 5mM DTT, the crystals were successfully reproduced under the same condition with a comparable size to the native one. These crystals grew to a maximum size of 0.2 mm×0.1mm×0.1 mm for 1-3 days. For data collection, labeled crystals were flash frozen (100K) in the above reservoir solution supplemented with 30% Glycerol. A total of 3X360 frames per wavelength of 1º oscillation were collected on wavelengths 0.9795, 0.9798 and 0.9600 Å using the beamline X12C at Brookhaven synchrotron light source (NewYork, USA). The data were processed by HKL2000 (www.hkl-xray.com). The crystals belonged to space group $P2_1P2_1P2_1$ with unit cell dimensions a = 48.56 Å, b = 48.71 Å, c = 54.6 Å, and $\alpha = \beta = \gamma = 90°$ having one molecule per asymmetric unit with 50.0-54.6 % solvent content (Figure 2.2., Table 2.2.).

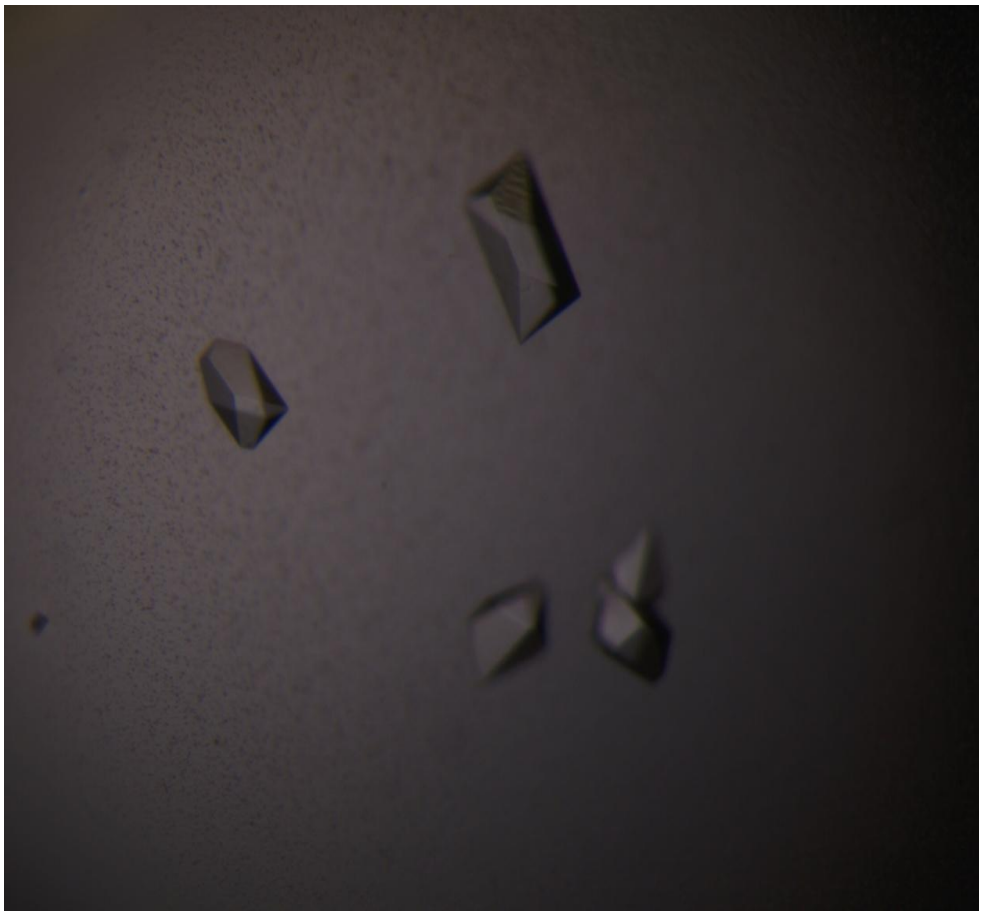


Figure 2.2. Protein crystal of DDL FHA (aa180-314).
Carboxyl-terminally hexahistidine tagged DDL FHA domain was crystallized by hanging drop method in 20º over 1-3 days. The native crystal was optimized in 30%PEG8k, 0.1M HEPES pH6.8. Seleno-methionine labeled DDL FHA was crystallized with additional 5mM DTT. The structure was solved by MAD method. The crystals belonged to space group $P2_1P2_1P2_1$ with unit cell dimensions a = 48.56 Å, b = 48.71 Å, c = 54.6 Å, and $\alpha = \beta = \gamma = 90°$ having one molecule per asymmetric unit with 50.0-54.6 % solvent content.

Data collection			
Space group		$P2_12_12_1$	
	Peak	Inflection	Remote
Cell dimensions *a, b, c (Å)*		48.564, 48.709, 54.688	
Wavelength (Å)	0.9795	0.9798	0.95
Resolution (Å) [a]	50~1.7 (1.73~1.7)	50~1.7 (1.73~1.7)	50~1.7 (1.73~1.7)
R_{sym} (%)	6.4 (52.1)	6.5 (57.8)	6.6 (59.4)
$I/\sigma(I)$	40.4 (4.4)	39.6 (3.9)	37.8 (3.6)
Completeness (%) [a]	100.0 (100.0)	100.0 (100.0)	99.9 (99.9)
Redundancy	7.6 (7.5)	7.6 (7.5)	7.7 (7.3)
Initial Figure of Merit		0.64	
Refinement			
Resolution Range (Å)	50~1.7		
No. reflections	13,269		
R_{work} (R_{free}) (%)	19.3/23.6 (40.2/47.1)		
No. atoms			
Protein	1,143		
Water	121		
B-factors (Å^2)			
Protein	22.5		
Water	30.4		
R.m.s. deviations			
Bond lengths (Å)	0.011		
Bond angles (°)	1.32		
% favored (disallowed) in Ramachandran plot	92.6 (0)		

Table 2.2. Crystallography statistics of DDL FHA.

3.3. Structure Determination

The crystal structure of SE core was determined by multiple anomalous dispersion method (MAD) based on the anomalous scattering signals from the bound zinc atom using SOLVE/RESOLVE (www.solve.lanl.gov). The initial phase was further improved by density modification assuming a solvent content of ~53% using the SHARP program (www.globalphasing.com). The model was built by using the program O (http://xray.bmc.uu.se/alwyn) and refined using REFMAC/CCP4 (www.ccp4.ac.uk). The R-free set contained 5% of the reflections chosen at random. The model comprises residues 194–551 (including carboxyl-terminal 8 extra residues from the vector). Disordered regions, including loop segments 290–324, 366-380 and 435-444, were not included in the model. The crystal structure of DDL FHA was determined likewise by MAD except that the anomalous scattering signal was observed from the methionine-incorporated selenium atom, not bound zinc. The model was constructed by the same software for residues 180-324 with extra 10 residues at the carboxyl-terminus including the tag.

4. Electrophoretic Mobility Shift Assay

4.1. DIG-labelling of RNA

Pre-miR164C (miRBase accession MI0001087) was transcribed *in vitro* containing digoxigenin(DIG)-labeled uracile (Roche), using RiboMAX Large Scale RNA Production Systems-SP6 and T7 (Promega). The reaction was formulated according to the manufacturer's instruction. The reaction was incubated at 37^{o}C on water bath for 3-5h and stopped by addition of chloroform followed by vortexing. The transcribed RNA is phased out in the upper aqueous layer after centrifugation. The aqueous phase was collected into a fresh eppendorf tube and precipitated with 5% volume of 5M Potassium acetate and 2.5 times volumes of ethanol. Chilled in -20^{o}C freezer for an hour, the pellet was collected by

centrifugation 14000rpm 10min and washed once with 70% ethanol. Air-dried pellet was re-suspended in nuclease-free water supplanted with Protector RNase Inhibitor (Roche Applied Science).

4.2. The *in vitro* binding reaction

The DIG-labelled RNA was incubated at room temperature with different concentrations of the protein in EMSA binding buffer (10mM Tris-HCl pH7.4, 50mM Potassium chloride, 1mM DTT, 10mM Magnesium chloride, 0.1% Nonedit P-40, 5% Glycerol) in 20ul reaction. Following 20min incubation, the samples were immediately loaded onto 4% native poly-acryl amide gel with non-denaturating dye, and resolved at 100V for 90min in 0.5XTBE (45mM Tris-HCl, 45mM Boric acid, 1mM EDTA, pH 8.3). The resolved RNA was electro-blotted onto Hybond-N+ (GE healthcare), and cross-linked twice with UV-linker (Stratagene) at 254nm 1200μJ. Blocking and washing of the membrane were performed according to the instruction of DIG GelShift Kit 2nd Generation (Roche).

4.3. Biotin labelling of RNA

For EMSA using biotin-labeled pre-miR164c, native pre-miR164c was transcribed *in vitro* and subsequently labelled with biotin using UV. EZ-Link Psoralen-PEO_3-Biotin (PIERCE) was dissolved in DMSO to 20mM under subdued light to make biotin stock for labelling reaction. The double-stranded RNA was denatured at 95^{o}C for 5min and immediately chilled on ice for 2min. 1% volume of biotin stock solution was added into the diluted RNA and vortex briefly. The RNA was irradiated under UV 365 nm for 30min on ice, and precipitated with 0.15M potassium acetate and 2.5X volumes of ethanol. After 1h in -80^{o}C, The RNA was recovered by centrifugation and washed once with 70% ethanol. The pellet was air-dried and re-suspended in nuclease-free water. The binding reaction, separation and blotting were

performed as above. Blocking and washing of the membrane were performed using Chemiluminescent Nucleic Acid Detection Module Kit (PIERCE). EMSA using siRNA duplex followed the same procedure, except that the chemically synthesized oligo-ribonucleotide was purchased from Thermo, not transcribed *in vitro*. The 21nt siRNA was designed to be a palindromic duplex with 5' phosphate and 3' 2-nt overhang (5'-P-AGACAGCAUUAUGCUGUCUUU-3').

5. *In vitro* GST Pull-Down Assay

5.1. *In vitro* binding and washing

Glutathione sepharose resin (GE healthcare) was used to detect interaction between recombinant proteins *in vitro*. The procedure deviates from the manufacturer's instruction, since non-specific interaction between prey proteins and glutathione sepharose resin was observed. Various salt and protein concentrations were adopted to minimize non-specific interaction. 20-30µg bait proteins were incubated with the resin overnight in GST binding buffer (25mM Tris-HCl pH7.4, 1mM EDTA, 0.01% Nonedit-40, 2M Sodium chloride). Hexahistidine-tagged prey proteins were added and the binding reaction was left for 2-12h with gentle rotation at 4^{o}C. The bound proteins were washed 10-13 times with 500µl binding buffer. The proteins were eluted with 20ul Laemmli loading dye at 100^{o}C on Accublock Ditital Drybath (LabNet) for 5min. The samples were spun down for 10min and the aqueous phase was collected to load on SDS PAGE gel.

5.2. Immuno-blot

The resolved proteins were electro-blotted onto PVDF membrane Immobilin-P (Millipore) in 1Xtransfer buffer (25mM Tris Base, 192mM Glycine, 20% Methanol) on ice at 100V for 60min. The membrane was incubated with blocking buffer (25mM Tris-HCl, 150mM

Sodium chloride, 2.5mM Potassium chloride, 0.14%Tween 20, 5% milk) for 1h and subsequently with antibody solution (25mM Tris-HCl, 150mM Sodium chloride, 2.5mM Potassium chloride, antibody variable) for 1h. The membrane was washed thrice for 10 min each before incubation with secondary antibody for 1h. The membrane was again washed with three changes of TBST and was subjected to exposure, wrapped in plastic bag with Immobilon Western HRP substrate (Millipore).

5.3. Antibody purification

The polyclonal anti-SE and anti-HYL1 antibodies were raised in rabbit and purified using Protein A agarose resin (BioRad). Prorein A agarose was washed twice with 1ml binding buffer (1M potassium phosphate pH 9.0), prior to addition of serum diluted with equal volume of the phosphate buffer, and rotated at 4^{o}C for 3-5h. The bond antibody was washed 5 times with the binding buffer and was eluted by 1M citrate pH3.0 for 5min with rotation at 4^{o}C and adjusted to pH7.0 with 1.5M Tris-HCl pH~11. The volume of 1.5M Tris-HCl to be added is empirically determined. The antibody concentration was estimated using Bradford Reagent (Sigma).

6. Generation of Transgenic Plants

6.1. SE expression constructs

To make transgenic plants, we used pBA-HA and pBA-myc vectors that harbor 3 repeats of Hemagglutinin (HA) and 6 repeats of c-Myc tags respectively. SE full length (1-720), SE core (194-543), SE core to carboxyl-terminus (core+C, 194-720) and SE Zinc-finger domain to carboxyl terminus (ZF+C, 470-720) were cloned into pBA-HA and pBA-myc. The enzyme digestion of DNA fragments and vectors, ligation and transformation with competent cell DH5α were performed as above. These clones were re-transformed with *Agrobacterium*

tumefaciens strain EHA105 by electroporation using Gene pulser-Xcell (BioRAD) and spread on LB agar plates containing 100mg/L spectinomycin and 25 mg/L rifampicin. Single colonies were picked into 3ml LB liquid medium containing the same antibiotics and shaken at 28^{o}C 250rpm for 16-20h to make glycerol stock

6.2. Floral dipping

The *se-1* is a mutant line in which carboxyl-terminal region of SE is disrupted by T-DNA insertion. The *se-1* and wild type *Arabidopsis Columbia-0* were sown directly on soil. When flower buds appeared 4-5 weeks post germination, the plants were transformed by floral dip method. The bacterial culture was scaled up to 250mL, grown at 28^{o}C overnight, and collected by centrifugation at 4000rpm 4^{o}C for 10min. The cell was re-suspended in 5% sucrose, supplemented with 0.02% Vac-In-Stuff Silwet L-77 (LEHLE) and mixed immediately. The cell suspension was transferred into 200ml beaker. The aerial part of the plant was submerged into the cell suspension so that the rosette laves touch the liquid surface for 10 seconds or longer. Excess liquid was removed with tissue paper. The plants were covered with cling wrap to keep the moisture and stored in darkness at room temperature over night. T1 seeds were selected on Murashige-Skoog medium containing 10mg/L BASTA and 100ug/L Carbenicillin. The protein expression was confirmed with immuno-blot from fluorescence lysate of T2 and T3 generations.

6.3. Total RNA extraction

The miRNA levels of SE full length (1-720), SE core (194-543), core+C (194-720), SE, and ZF+C (470-720) expression lines under 35S promoter in *se-1* background were tested with northern blot. Total RNA was prepared by Trizol (Invitrogen). Flash-frozen whole seedlings 20-23 days old were ground with mortar and pestle, and transferred to eppendorf tubes. With

1ml Trizol added, the tubes were vortexed vigorously till the mixture is homogeneous. After being for 5 min in room temperature, the tubes were spun at 14000rpm, 4^{o}C for 10min to remove tissue debris. Upper liquid phase was vortexed with 0.2 volumes of chloroform in a fresh tube and spun likewise. The upper aqueous phase was transferred to a fresh tube and vortexed with equal volume of chloroform. The upper aqueous phase was precipitated with equal volume of isopropanol in room temperature for 30-60min. The total RNA was recovered with centrifugation and washed with 1ml 70% ethanol. The air-dried pellet was re-suspended in nuclease-free water, ready for northern blot analysis.

6.4. Northern blot

12µg total RNA was mixed on ice with native PAGE loading dye saturated with urea, and applied using loading tip into 15mm well of urea polyacrylamide gel pre-run for 1h. The RNA was electrophoresed until the bromo-phenol-blue reached one third of gel height from the bottom normally after 3h. The resolved RNA was electro-blotted onto Hybond-N+ (GE healthcare) at 100V for 30min, and cross-linked. Hybridization and detection followed a standard protocol except the hybridization temperature adjusted to DNA/RNA hybridization instead of RNA/RNA. The membrane was gently shaken in 10ml ULTRAhyb (Ambion) 42^{o}C for 30min, followed by another 12-20h in the same buffer with 0.2nmol 5'-terminally biotinylated oligo-deoxyribonucleotide at 42^{o}C. Biotinylated oligo-deoxyribonucleotides were purchased from 1stBase (Singapore) as a probe designed complementary to the guide strand of a miRNA with the following sequences:

164biotin: Biotin-5'-TGCACGTGCCCTGCTTCTCCA-3'

319biotin: Biotin-5'- AGGGAGCTCCCTTCAGTCCAA -3'

The hybridized membrane was washed twice with 2XSSC (0.3M Sodium chloride, 30mM Trisodium citrate), 0.1 %SDS at 42^{o}C for 5min and twice with 0.1X SSC (0.015M Sodium

chloride, 1.5mM Trisodium citrate), 0.1% SDS at 42^{o}C for 15min. The hybridized probes were detected by Nucleic Acid Detection Module Kit (PIERCE).

Chapter Three: Results

1. Crystal structure of SE core

1.1. Overall structure

Arabidopsis SE core adopts a "walking man-like" topology with N-terminal two α-helices resembling the leading leg, the Middle α-helix dominant domain resembling the body and the carboxyl-terminal non-canonical C_2H_2 zinc finger domain featuring Helix-kink-Helix (HkH) motif resembling the lagging leg (Fig. 3.1.C.). The green-colored N-terminal domain (residues 195-240) consists of a short α-helix (α1) followed by an orthogonally oriented long α-helix (α2) that shares the fold of N-terminal half of histone H3 core (Z score 4.4, r.m.s.d. 2.0 Å, 43 Cα) (Fig. 3.2.A. right panel) (63). The N-terminal domain is connected to the Middle domain through a short α-helix (α3). The cyan-colored Middle domain (residues 241-469) displays a novel fold dissimilar to any solved structure hitherto, according to Dali (www2.embl-ebi.ac.uk/dali). The Middle domain consists of loosely packed three α-helices (α5, α6 and α7) oriented orthogonally against a central ~30° bending long α-helix (α4). A pair of anti-parallel short β-strands (β1, β2) is embedded within a partially disordered long loop connecting α5 to α6, whereas α6 is connected to α7 by another partially disordered long loop. The Middle domain is followed by the carboxyl-terminal C_2H_2 zinc finger domain via a short α-helix (α8). The presence of three disordered loops (residues 290-324, 366-380 and 435-444) in Middle domain implicates its role in protein-protein interaction with SE partners and/or substrates (Fig. 3.1.C.).

1.2. Features of carboxyl-terminal Zinc Finger domain

The magenta-colored carboxyl-terminal non-canonical C_2H_2 zinc finger domain (residues 471-543) adopts a variant C_2H_2 zinc finger fold with β-β-β-α architecture (β3, β4, β5 and α10) flanked by one additional α-helix (α9) at its N-terminus and one additional

atypically kinked long α-helix at its C-terminus (α11) (Fig. 3.2.B. left panel). The central β-β-β-α architecture has the similar structural fold of transcriptional factor Swi5 (Z score 3.9, r.m.s.d. 2.9Å, 46 Cα) (Fig. 3.3.B. right panel) (64) and zinc finger domain of human zinc finger BED domain containing protein 2 (Z score, 3.6, r.m.s.d. 4.2 Å, 60 Cα) (Fig. 3.3.B. middle panel) (PDB accession: 2DJR). Notably, the structural motif of the zinc finger domain, including the canonical β-β-α structure and the carboxyl-terminal long α helix–kink–helix (HkH) motif, resembles that of Isopentenyltransferase (IPTase) (Fig. 3.3.B. right panel) (65). The β-β-α-α(HkH) motif identified in IPTase putatively represents zinc finger proteins that specifically bind to dsRNA (65).

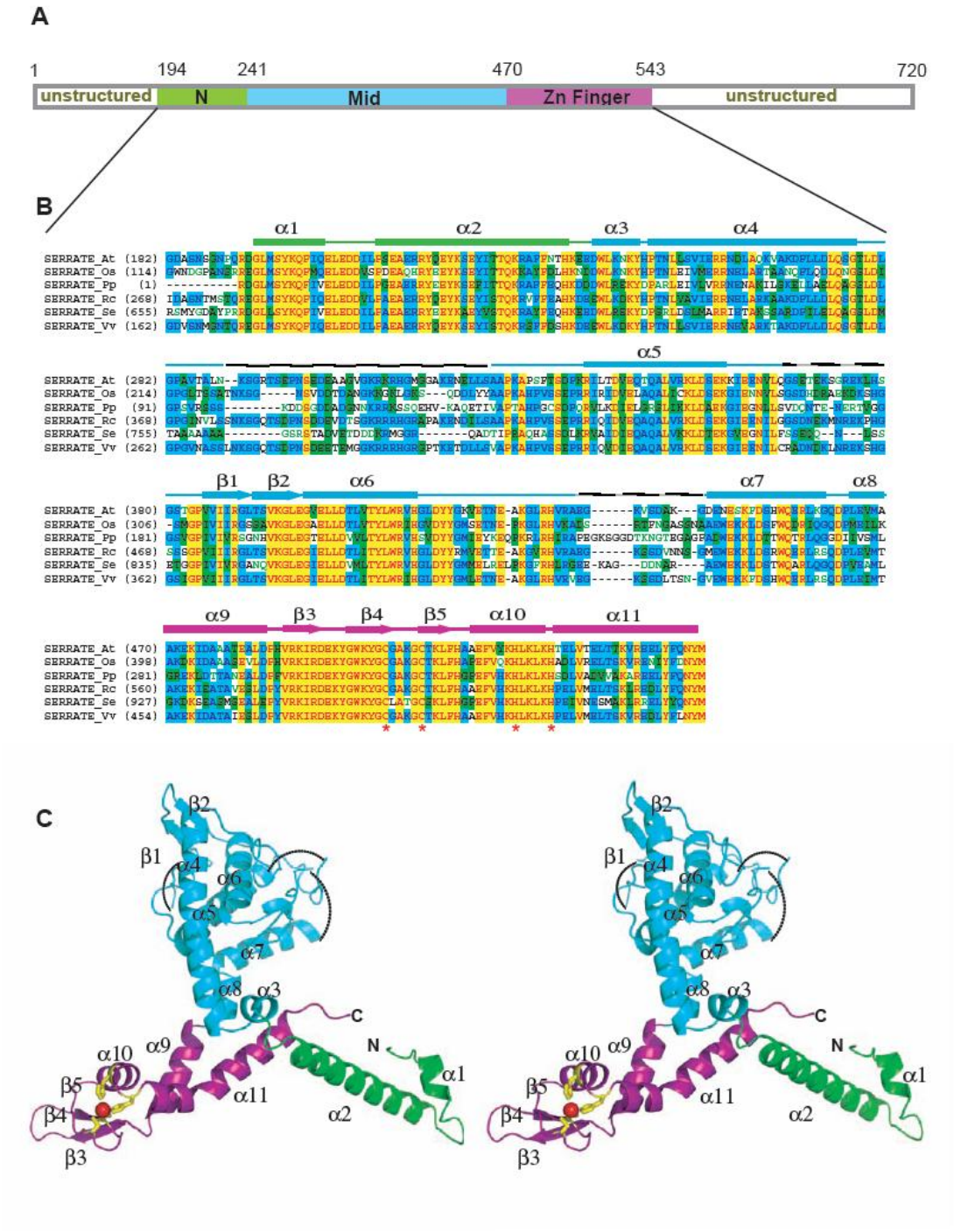


Figure 3.1. Domain architecture, sequence alignment and overall structure of *Arabidopsis* SE core.

Figure 3.1. Domain architecture, sequence alignment and overall structure of *Arabidopsis* SE core.
A: Schematic of the domain borders in *Arabidopsis* SE core, N- and C-terminal unstructured regions are indicated.
B: Sequence alignment and secondary structure of SE core. The aligned sequences are in the order of *Arabidopsis*, *Oryza*, *Physcomitrella*, *Ricinus*, *Selaginella* and *Vitis*. The secondary structure diagram for *Arabidopsis* SE is shown on the top. The α-helices are indicated as bricks, β-strands are indicated as arrows. Three domains are colorred in green, cyan and magenta, respectively from N- to C-terminus. Conserved residues are shaded in cyan (80% similarity) and green (60% similarity), whereas essentially invariant residues are shaded in yellow. The critical residues for zinc ion coordination (C500, C505, H518 and H523) are indicated by asterisks beneath the residues.
C: Stereo-view ribbon representation of SERRATE core showing the N-terminal domain (green), the middle domain (cyan) and the zinc-finger domain (magenta).

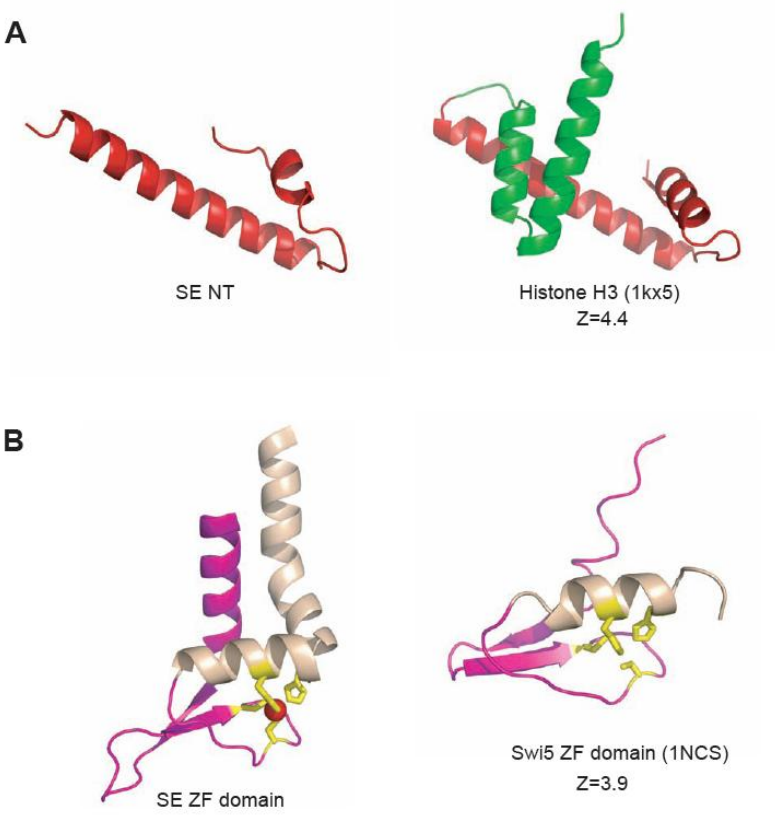


Figure 3.2. Comparison of N-terminal and zinc finger domains of SE with their closest matching counterparts in the DALI data-base.
A: Comparison of SE N-terminal domain (left panel) with the first two α-helices of histone H3 protein (1kx5) (right panel, colorred in red). The comparative Z score = 4.4, with r.m.s.d. = 2.0 Å for 43 Cα atoms.
B: Comparison of SE zinc finger domain (left panel) with the yeast Swi5 zinc finger domain (1ncs) (right panel). The comparative Z score = 3.9, with r.m.s.d. = 2.9 Å for 46 Cα atoms. The C_2H_2 zinc binding motif is colorred in yellow, whereas the bound zinc ion is colorred in red.

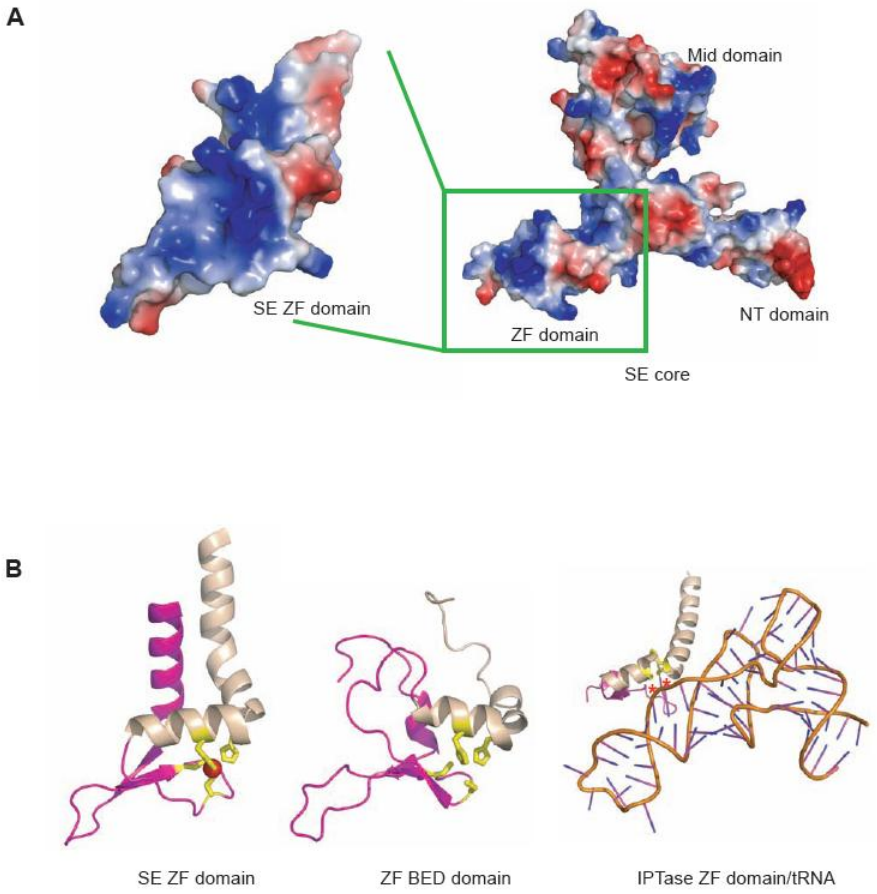


Figure 3.3. Implication of SE zinc-finger domain in RNA binding.
A: Electrostatic potential surface representation of SERRATE core. The zinc-finger domain is boxed and highlighted at left panel with ~45° rotation along *y*-axis. The SERRATE zinc finger has a highly positive charged surface (colorred in blue) for potential nucleic acid binding.
B: Structure-based sequence alignments of SE zinc-finger domain (left panel), C_2H_2 type zinc-finger domain of human zinc-finger BED domain containing protein 2 (middle panel) and eukaryotic IPTase zinc-finger domain/RNA compelx (right panel). Cysteines and histidines coordinated to zinc are highlighted in yellow and zinc ion is highlighted in red. Residues that form hydrogen bonds with RNA bases via their backbone functional groups in IPTase structure are indicated by asterisks and colorred in red. The unique C-terminal HkH motifs within zinc fingers are highlighted in wheat. The views are positioned at the same orientation followed by superimposition.

2. Role of SE core in SE-HYL1 and SE-DCL1 interactions

HYL1 and SE are known to bind DCL1 to promote preciseness of pre-miRNA processing (66). However, the domain contacts within the hetero-trimer are yet to be detailed. With the SE core structure in hand, we attempted on mapping of domain interactions between SE, HYL1 and DCL1 by *in vitro* pull down assay. DCL1 RNaseIIIa+IIIb domains could not be tested in this assay due to low solubility of this fragment expressed in bacterium. GST-fused DCL1 helicase and PAZ domains showed weak interactions with hexahistidine-tagged SE core, detected by polyclonal anti-SE core antibody raised in rabbit (Fig. 3.4. A, B.). In addition, both GST-fused HYL1 RBD1+2 and GST-fused HYL1 RBD2 were able to pull down SE core, whereas GST-fused HYL1 RBD1 failed to pull down SE core (Fig. 3.4. A, B). The observation agrees with the notion proposed by Yang *et al* (2010) (67) that HYL1 RBD2, not RBD1, is solely responsible for protein-protein interaction. Next, we asked whether a certain domain within SE core specifically participates in or eludes SE-HYL1 interaction. As expected, both GST-fused SE N-terminal (SE-NT) and SE Middle (SE-Mid) domains were able to pull down HYL1 RBD2 but not HYL1 RBD1 (Fig. 3.4. C.). SE carboxyl-terminal zinc-finger domain alone could not be expressed in bacterium and so was excluded from this assay.

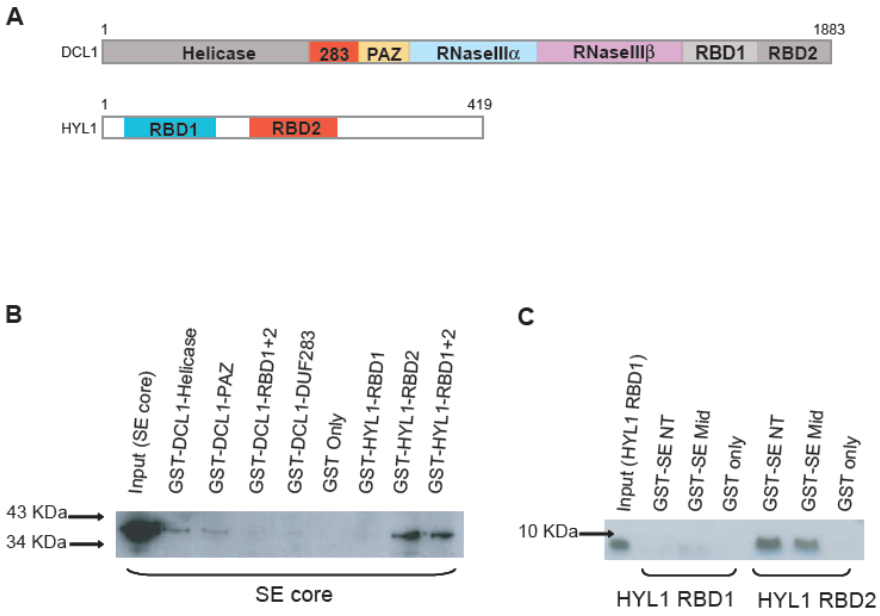


Figure 3. 4. Domain interaction network amongst SE, HYL1 and DCL1.
A: Domain architectures of DCL1 (upper panel) and HYL1 (lower panel).
B: *In vitro* pull-down assay showing *Arabidopsis* SE core binds to *Arabidopsis* DCL1 helicase and PAZ domains as well as *Arabidopsis* HYL1 RBD2 domain directly.
C: *In vitro* pull-down assay showing *Arabidopsis* SE core binds to HYL1 RBD2 via its N-terminal and middle domains.

3. S*e-1* phenotype rescue by expression of SE fragments

3.1. The critical role of carboxyl-terminal tail

The protein-protein interactions described above were tested within SE core covering only ~50% of full-length SE sequence. The amino- and carboxyl-terminal tails outside SE core were predicted to be unstructured and turned out, in fact, not to crystallize. Nevertheless, these putatively unstructured tails may have important functional impacts for miRNA processing. To investigate this possibility, we generated transgenic lines expressing HA- and myc-tagged full length SE and SE core under 35S promoter in the *se-1* background, respectively. Notably, full-length SE, instead of SE core can considerably rescue the morphological phenotype of s*e-1* mutant (Fig.3.5. A, B.). Typical phenotypes of *se-1* mutant, such as leaf serration and abnormal cluster of flowers and siliques, went unobserved by expression of full-length SE protein. By contrast, transgenic *Arabidopsis* expressing myc-tagged SE core in the *se-1* background showed pleiotropic phenotypes characteristic of s*e-1*mutant line. Surprisingly, the expression of longer fragment including SE core and carboxyl-terminal tail (SE core+C) was sufficient to rescue the miRNA morphology phenotype (Fig. 3.5. A, B.).

3.2. The minimal fragment to rescue s*e-1* phenotype

To narrow down the minimum requirement of SE domains capable of phenotype rescue, we made a truncated fragment consisting only of the zinc finger domain and the carboxyl-terminal tail (SE ZF+C). Remarkably, the expression of SE ZF+C in the *se-1* mutant background was able to rescue serration phenotype (Fig. 3.6. A). The expression of miRNAs was rescued (Fig. 3.6. B). Consistent to the notion of the importance of carboxyl-terminal ZF domain, the introduction of point mutations at the two cysteine residues involved

in Zinc coordination (C500S/ C505S double mutant) disrupted the rescue ability and caused reversion of serration phenotype and extremely retarded growth (Fig 3.6. A.).

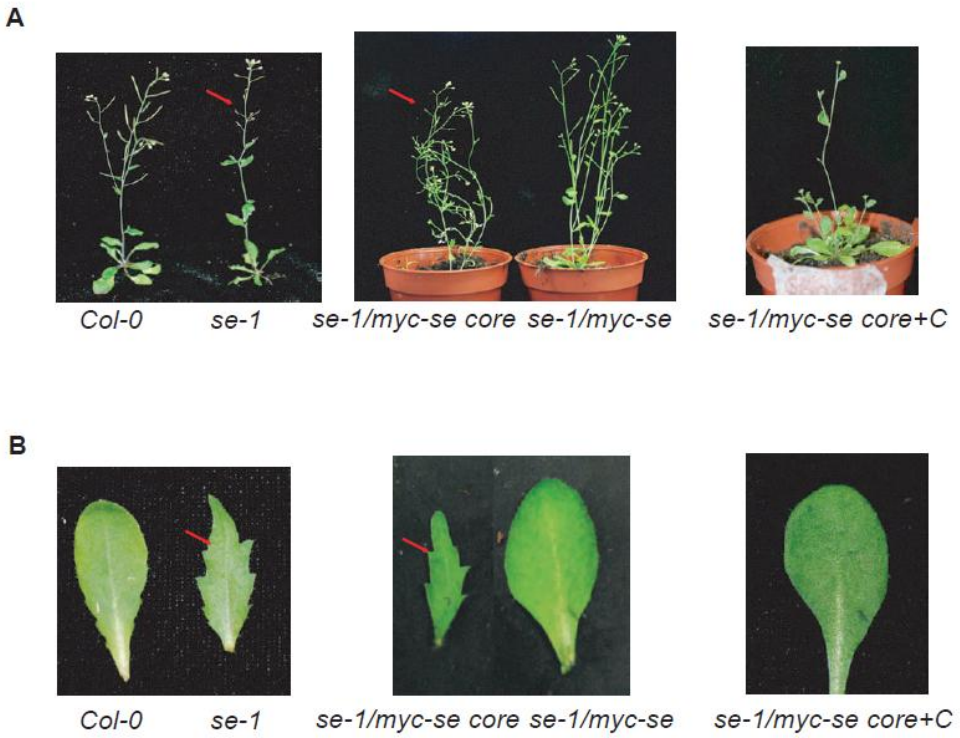


Figure 3.5. SE C-terminal region is essential to rescue *se-1* mutant phenotype.
A: Comparison of the morphological phenotype between *Col-0*, *se-1*, *se-1/35S::myc-se*, *se-1/35S::myc-se core*, *se-1/35S::myc-se core+C* plants.
B: Close-up of the rosette leaves. The arrow points the leaf serration phenotype typical of *se-1* line.

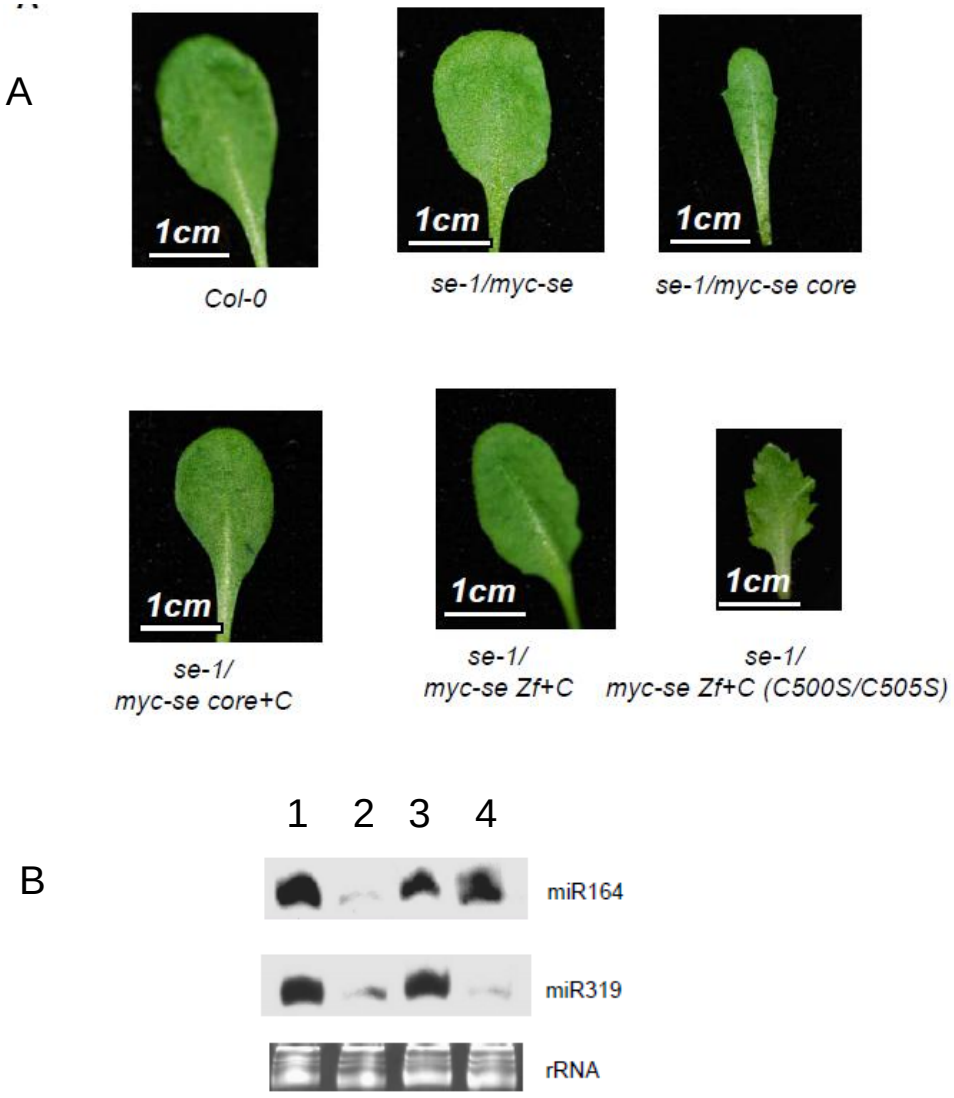


Figure 3.6. Both SE zinc-finger domain and C-terminal region is required to rescue *se-1* mutant phenotype.
A: Comparison of the morphological phenotype between *Col-0*, *se-1/35S::myc-se*, *se-1/35S::myc-se core*, *se-1/35S::myc-se core+C*, *se-1/myc-se Zf+C*, *se-1/myc-se Zf+C (C500S/C505S)* plants.
B: Accumulation of miR164 and miR319 in *Col-0*, *se-1*, *se-1/35S::myc-se*, *se-1/35S::myc-se ZF+C* plants. Each lane contained 12 µg RNA. rRNAs were used as a loading control.

5. SE core+C binds pre-miRNA *in vitro*

The phenotype rescue experiment indicates that SE core together with some undetermined residues located somewhere in carboxyl-terminal tail retains *in vivo* function. Next, we asked whether the functional role in miRNA processing displayed by SE *in vivo* is due to the RNA binding ability of SE core+C. To this end, we tested substrate-binding ability of SE core+C by electrophoretic mobility shift assay (EMSA) *in vitro*, and found that SE core+C can stably associate with biotin-labeled pre-miR164c (Fig. 3.7. A.). To further narrow down the necessary residues within carboxyl-terminal tail that enables SE zinc-finger domain to bind miRNA precursors, we made a series of SE constructs gradually lengthening the tail from SE core toward C-terminus and tested substrate binding by EMSA *in vitro*. We found out that the fragment comprising SE core plus 36 residues (194-579) is enough to stably associate with internally digoxigenin-labled pre-miR164c (Fig. 3.7. B.). These results show that SE is an RNA-binding protein and that substrate recognition by SE probably plays a role in phenotype rescue.

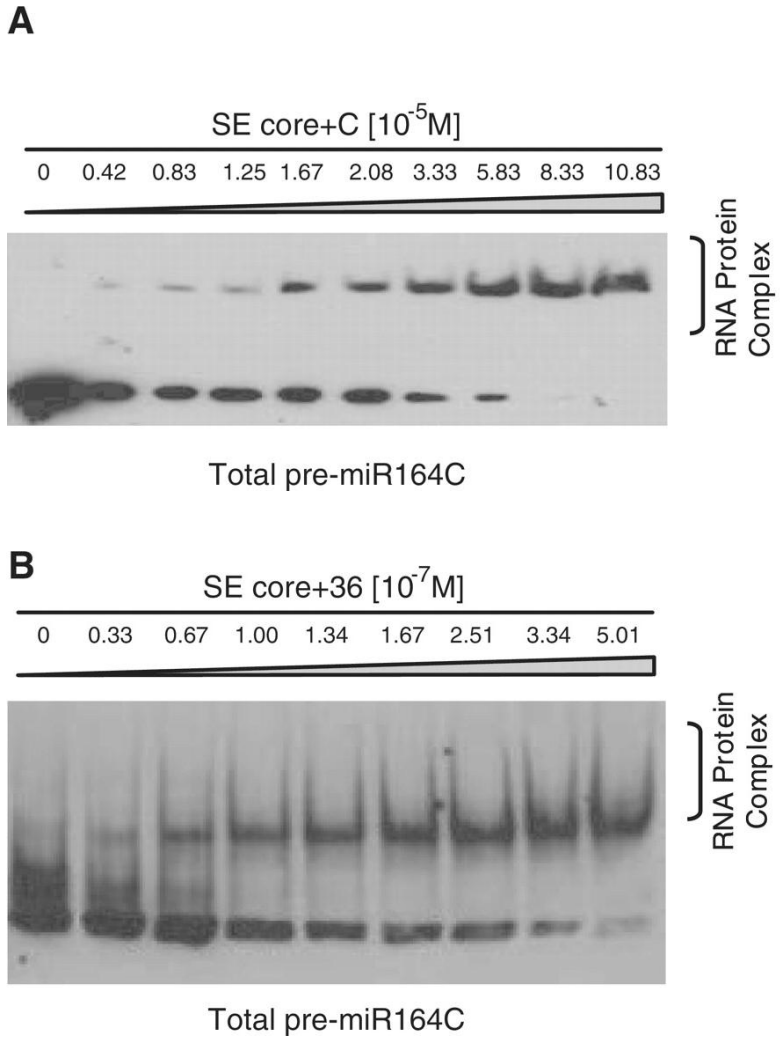


Figure 3.7. SE core+C binds pre-miR164c *in vitro*.
A: EMSA results showing that SE core+C (residues 194–720) binds biotin-labelled pre-miR164c. B: EMSA results showing that SE core+36 amino acid (residues 194–579) binds digoxigenin-labelled pre-miR164c.

6. Crystal structure of DDL FHA domain

To address the molecular basis for substrate recognition by DDL, we determined the crystal structure of the free form DDL FHA domain at 2.0Å resolution, using the fragment corresponding to minimum boundary of the FHA domain, residues 180-314 with carboxyl-terminal hexahistidine tag. The overall structure reveals that DDL-FHA forms a 7-stranded β-sandwich architecture. The additional α-helix on carboxyl terminus is an artefact that results from hexahistidine tagging utilized for protein purification (Fig. 3.8.). The strand-dominant architecture features prominently long strands β3 and β4 connected with short loop in anti-parallel orientation. The first and the last strand in the sandwich lie in adjacent anti-parallel, maintaining the modular nature of conserved FHA domain. The residues expected to recognize a target epitope are found on the long loops connecting anti-parallel β2/3 and β4/5, emanating in amino terminal direction, which agrees with sequence-based prediction of target epitope recognition site identified by sequence alignment (Fig.3.9. A, B.).

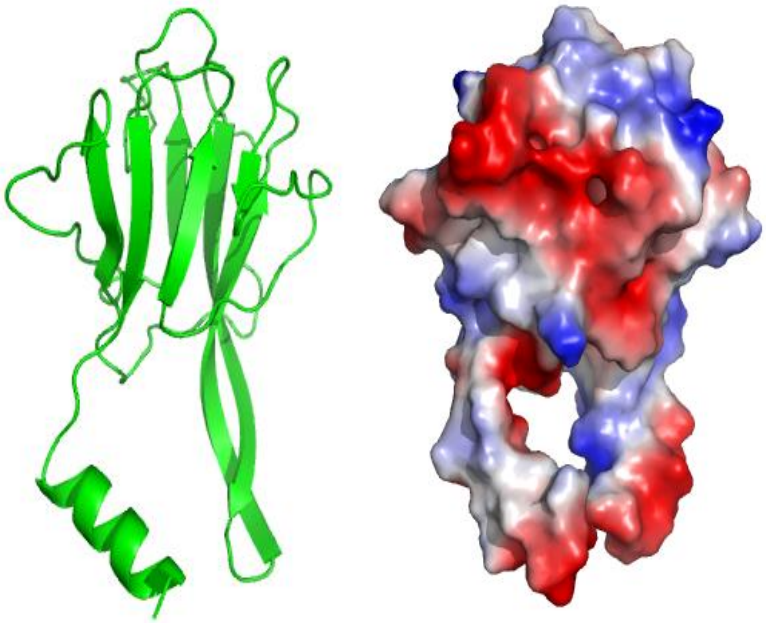


Figure 3.8. Overall fold of free form DDL FHA domain and potential map.
Overall fold of DDL FHA domain is presented in Richardson diagram (left) and electrostatic potential map (right). The features are prominently long strands β3 and β4 connected with short loop in anti-parallel orientation. The first and the last strand in the sandwich lie in adjacent anti-parallel, maintaining the modular nature of conserved FHA domain. The residues expected to recognize a target epitope are found on the long loops connecting anti-parallel β2/3 and β4/5, emanating in amino terminal direction

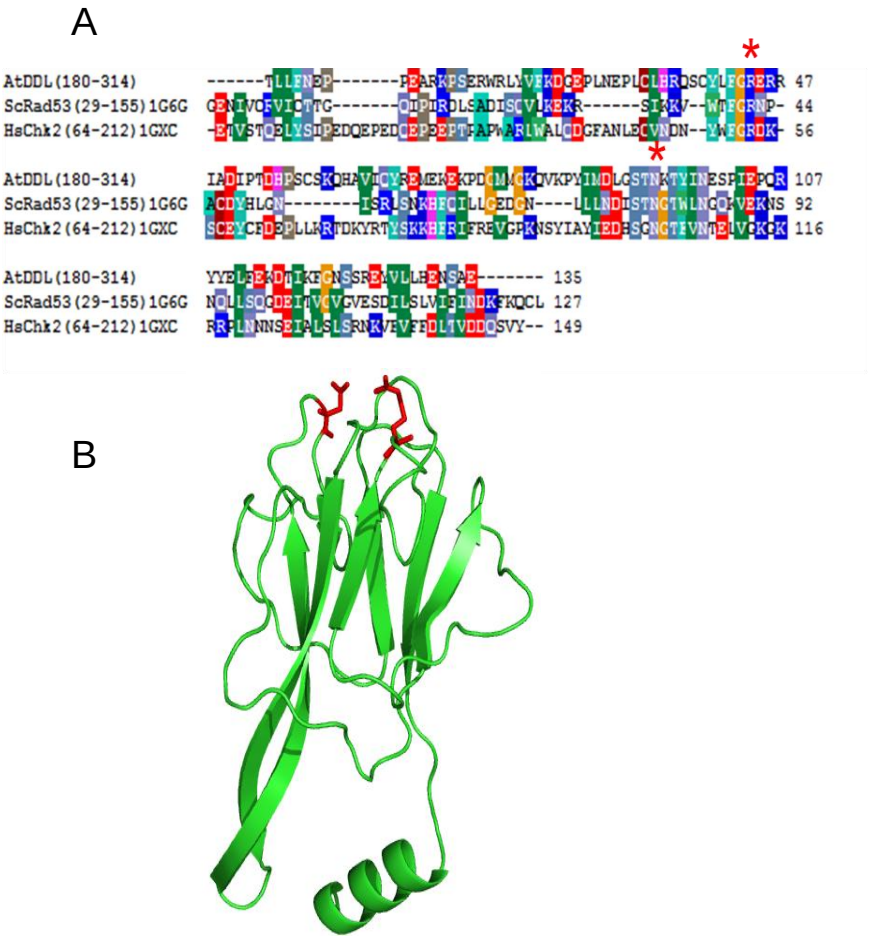


Figure 3.9. The identification of putative phopho-recognition cleft by sequence alignment.
A: Sequence alignment of FHA domains from *Arabidosis* DDL, *homo sapiens* Chk2 and Rad53. Sequences of DDL FHA, Chk and Rad53 were aligned using ClustalW to identify the conserved residue in DDL, putatively responsible for phospho-threonine recognition. The conserved arginine and asparagine residues marked with asterisk(*) in Chk2 and Rad53 are reported to be structurally conserved and point-mutation at these residues abolish phospho-threonine binding of Chk2 and Rad53.
B: Orientation of putative phospho-recognition cleft identified by sequence alignment
The conserved arginine and asparagine marked out in Fig.3.9.A. are presented as red sticks. Those conserved residues are located on the well-defined long loops (β2-β3) and (β4-β5), suggesting that phospho-threonine-recognzing cleft is also oriented toward this surface.

7. Putative phospho-serine/threonine recognition cleft

7.1. Structural homology with other FHA domain-containing proteins

To have clues as to the function of DDL, three-dimensional homologues were searched for in Protein Data Bank by Dali server (68). The structural search resulted in 321 non-spurious hits with Z-score higher than 2.0 despite paucity of sequence similarity detectable with BLAST search (69). Significant hits among them were Serine/Threonine Protein Kinase Chk2 (PDB code 1GXC, Z-score = 12.5, RMSD = 2.0 Å) and Protein Kinase Rad53 (PDB code 1G6G, Z-score = 12.6, RMSD = 2.2 Å) (Fig. 3.10. A, B.). Superimposition of Rad53 and Chk2 on our DDL FHA domain structure demonstrates that, while similar arrangements of β-sandwich architecture are observed among the three, the conspicuous extension of two anti-parallel strands β3 and β4 are peculiar to DDL FHA domain alone. We find it intriguing that *Arabidopsis* kinase interacting FHA domain (KI-FHA) from Receptor Kinasase Associated Protein Phosphatase shares decent structural homology (PDB code 1MZK, Z-score = 11.5, RMSD = 2.4 Å) with DDL FHA domain but lacks the long strands (70).

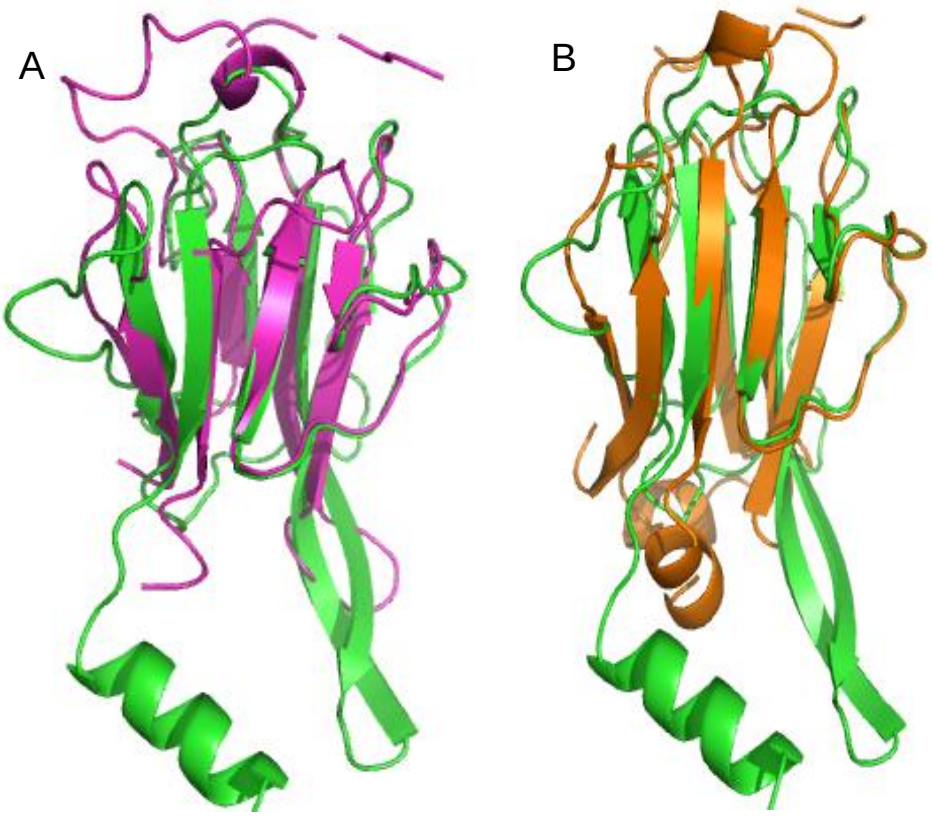


Figure 3.10. Superimposition of DDL FHA domain on structural homologues.
A: Superimposition of DDL FHA domain on Chk2 FHA domain. Chk2 FHA domain (magenta) is superimposed on DDL FHA domain (green). Chk2 FHA domain is comprised of 11-stranded β-sandwich but lacks the long strands seen in β3 and β4 in DDL FHA and has a short helix (β3-β4) inserted between loops. Note that the putative phospho-binding cleft of DDL FHA is located on the same side as that of Chk2 consisting of relatively rigid loop.
B: Superimposition of DDL FHA domain on Rad53 FHA domain
Rad53FHA domain (orange) is superimposed on DDL FHA (green). Rad53 has two helices intersecting (β2-β3) and (β10-β11) as well as the carboxyl-terminal helix.

7.2. Phospho-recognition cleft found in the asymmetric unit

Plus to the structure, the crystal system provided us with a glimpse into phospho-peptide recognition by DDL FHA domain. The molecules in an asymmetric unit pack against each other with an acidic residue inserted into the putative phospho-epitope binding cleft. It further supports the notion that DDL recognizes the target by phospho-dependent electrostatic interaction and that the molecules might form either homo-dimer or hetero-dimer with its homologues in solution (71-74).

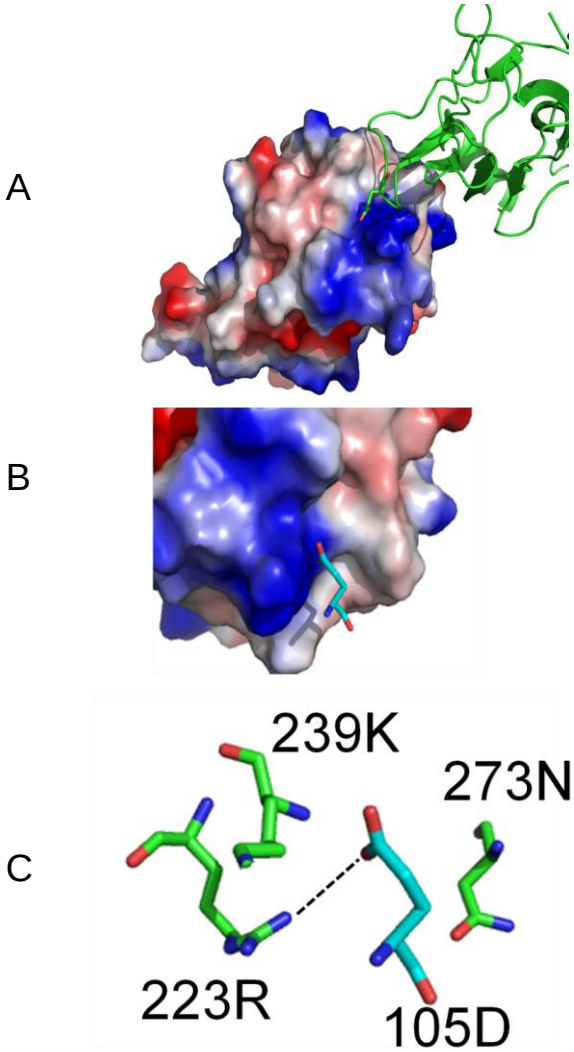


Figure 3.11. Insertion of glutamate 105 into the putative phospho-recognition cleft.
A:The unit cell contained one molecule per asymmetric unit. Symmetrically related molecules packed against each other. A glutamate (105D) is inserted into the putative phospho-threonine-recognizing cleft formed by two long loops connecting , β2/3 and β4/5.
B: The side chain carboxyl group is arranged on the positively charged patch of the rigid loop.
C: The packing is interpreted as mimicking negatively charged phospho-dependent interaction between DDL and DCL1. The highly conserved arginine (223R) is within the hydrogen bond distance from (105D).

Chapter Four: Discussion

1. Computational docking of SE core to tRNA precursor

The biochemical analysis of *Arabidopsis* SE demonstrated that SE is an RNA binding protein contributed from the non-canonical C_2H_2 zinc finger domain and carboxyl-terminal tail. The topological similarity between SE C_2H_2 zinc finger motif and isopentenyltransferase (IPTase) Helix-kink-Helix zinc finger domain prompted us to superimpose SE core structure with IPTase/tRNA complex aligned on zinc finger domain and examine the possible arrangement of dsRNA bound along the SE core structure (75) (Fig. 4.1.). In our computational model, the N-terminal α-helix of the HkH motif packs against the minor groove of the bound dsRNA, whereas the carboxyl-terminal helix runs across the major groove. Several highly conserved residues (K522, T524 and T528) are positioned to a place near the junction of stem and D-loop. Interestingly, the architecture of tRNA, especially at the junction of stem and D-loop resembles the ssRNA-dsRNA junction within pre-/pri-miRNA, which prompts us to speculate that the recently discovered tRNA-derived small RNAs may adopt a cleavage mechanism similar to that of miRNAs (75,76).

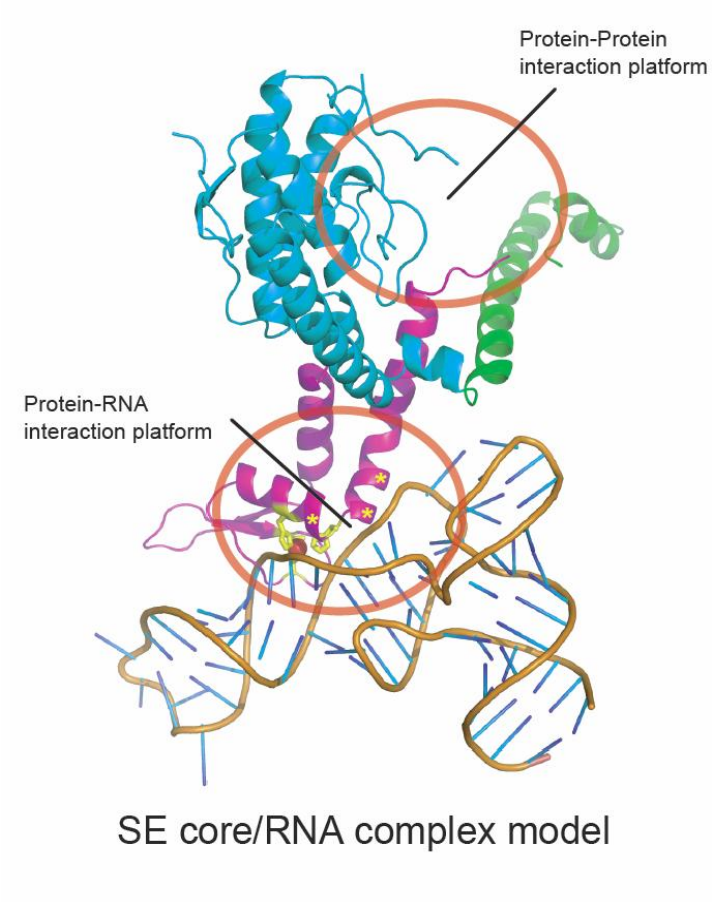


Figure 4.1. Hypothetical model of SE bound to dsRNA with ss-dsRNA junction.
A cartoon representation of SE/RNA complex model. N-terminal α-helix of the HkH motif of SE zinc finger domain packs against the minor groove of the bound dsRNA, whereas the C-terminal helix runs across the major groove. Several highly conserved residues (K522, T524 and T528) are positioned to a place near the ss-dsRNA junction, which represents the unique structural architecture of miRNA precursors. The C_2H_2 zinc binding motif is colorred in yellow and the bound zinc ion is colorred in red, whereas the conserved residues probably involved in dsRNA binding are labeled by asterisks in yellow. Protein-protein interaction platform is highlighted in green, whereas protein-nucleic acid interaction platform within SE core is highlighted by red circles.

2. Significance of SE zinc finger plus carboxyl-terminal tail

The RNA recognition model presented here is shared among other dsRNA-specific zinc finger proteins (65). However, our SE sequence lacks a highly conserved Lys/Arg residue located one helical turn after the second zinc-coordinating histidine residue (65) (Fig.3.1.B., Fig. 4.2.), reminding us that SE C_2H_2 zinc finger domain is essential but not sufficient for ssRNA-dsRNA junction recognition. Therefore, other critical residues, most likely from the conserved carboxyl-terminal tail of SE, are expected to recognize ssRNA-dsRNA junction of pre-miRNA. This hypothesis is consistent with our observation that SE core was not sufficient to rescue the *se-1* mutant phenotype *in vivo*. Our SE core structure supplemented by rescue experiment *in vivo* and RNA binding assay *in vitro* suggest that SE works both in recognition of RNA substrate by ZF domain together with carboxyl-terminal tail and in guidance of RNA into the miRNA processing machinery through protein-protein interaction, like a scaffold bridging RNA, HYL1 and DCL1, and guarantee a synergistic stimulating effect on DCL1 catalyzed miRNA processing (66).

Zhou et al (2008)

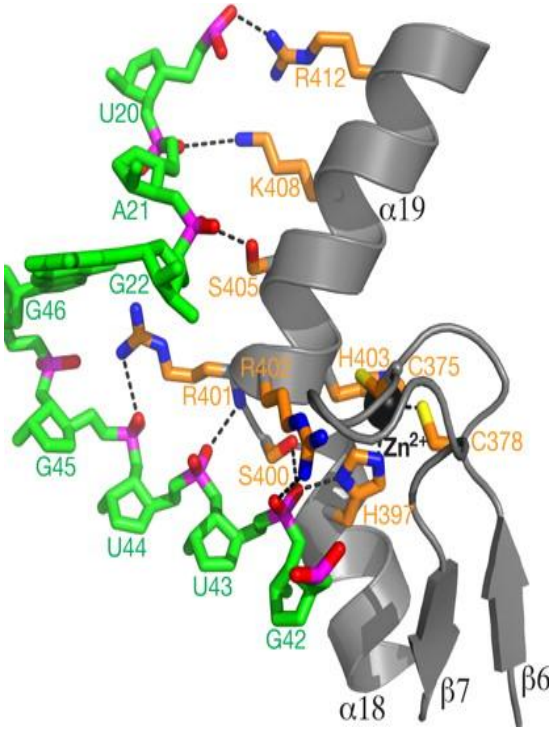


Figure 4.2. Major contacts between IPTase carboxyl-terminal zinc-finger and tRNA. The main chain of IPTase is represented as gray ribbons. The side chains of IPTase and tRNA are depicted as sticks. Hydrogen bonds are depicted as dotted lines in black. The carbon atoms of IPTase and tRNA are colorred orange and green, respectively. Hetero-atoms are colored individually, with nitrogen in blue, oxygen in red, phosphorus in magenta, and sulfur in yellow. The zinc ion is depicted as a small sphere in dark gray. The lysine residue located at 408, one helical turn after the second coordinating histidine, is well conserved in IPTase but not observed in our docking model, which may account for lack of SE core's affinity to pre-miR164c *in vitro*.

3. The proposed mechanism of cleavage site recognition

Based on the *in vitro* cleavage assay performed by Dong et al (46) (Fig. 1.2.) and our results at hand, our speculation follows that SE and HYL1 guide DCL1 in search for the correct cleavage site on pri-miRNA. The mechanisms whereby endo-ribonuclease type III enzymes zero in on the cleavage site on double-strand RNAs aided by subunits have not been kinetically described. As to sequence-specific cleavage of double-strand DNA, ample information is available on restriction enzymes. A restriction endo-deoxyribo-nuclease EcoRI searches for the cleavage site by combination of hopping and sliding. It has long been believed that restriction endonucleases' rapid sequence-recognition and precise cleavage were achieved solely by sliding on double-helical DNA simply because reduction in motion dimensionality makes the search faster by one-dimensional diffusion. Even though the sliding distance was recently shown to be not so long, theoretical and experimental results agree that combination of hopping and sliding is the most efficient. In fact, HIV reverse transcriptase searches for the terminal start site by sliding (77). EcoRI slides to stop by interaction between the specific palindrome and the contacting amino-acid side chains, and allosterically bends the double stranded DNA to cleave. Though it is inconspicuous, this process of endonucleotic catalysis takes two steps in DNA binding, namely, sequence-non-specific electrostatic interaction in sliding and sequence-specific recognition of palindrome followed by allostery. With these examples taken into account, the plausible model of cleavage site recognition by DCL1 will involve non-specific electro-static interaction between DCL1 and pri-miRNA, next sliding and hopping on the RNA backbone to search, and finally its halt on the specific secondary structure aided by SE and HYL1.

4. Examples of RNA structure recognition facilitated by protein subunits

There are nuclease complexes whose subunits are essential for specificity. Among the informative examples is the sliding clamp of Proliferating Cell Nuclear Antigen (PCNA) (78,79). PCNA itself does not have a catalytic activity. Instead, it serves to support catalysis as it slides on replicating DNA under Okazaki fragment processing until PCNA carboxyl-terminal HhH_2 (helix-hairpin-helix-helix) domain recognizes ss/dsDNA hinge and cause DNA to be distorted. The major conformational change of DNA, including a 90^o kink of the DNA duplex and organization of the single-stranded flap offers XPF nuclease access to the cleavage site. Another example is RNaseP among the enzymes involved in the maturation of transfer-RNAs (tRNAs). In bacteria, pre-tRNA is excised from the initial transcript of tRNA gene, 3^o and 5^o termini are cleave off, and the nucleotides go through modifications to be a mature tRNA. Ribonuclease-P (RNaseP) catalyzes cleavage of leader sequence on 5^o terminus of pre-tRNA. Prokaryotic RNaseP holozyme is a ribozyme complex consisting of catalytic RNA part and a small protein subunit, while in eukaryotes the protein subunits form a larger complex. RNaseP proteins are directly involved in the recognition and cleavage of pre-tRNA at specific ribo-nucleotides preceding 5^o-leader sequence (80,81). The recognition of ssRNA-dsRNA junction by the protein subunits is essential to excise mature tRNA from pre-tRNA. The junction recognition is likewise expected to signify the substrate specificity of IPTase. Although IPTase is not ribonuclease and not proven to slide on RNA either, the way helix-kink-helix get stuck at ssRNA-dsRNA junction to induce allostery can be well likened to the above examples. Taken together, we propose our model in which SE zinc-finger domain plus carboxyl-terminal tail fits onto the ss/dsRNA junction as a part of DCL1 complex to aid substrate recognition and perhaps allostery.

5. The possibility of precursor-specific involvement of SE domains

Having noted that SE ZF+C fragment lacking SE-NT and Mid is capable of phenotype rescue, a question arises of whether SE-NT plus Mid possibly evades a role in DCL1 complex formation, contradicting our afore statement that SE constitutes the ternary complex together with HYL1 mediated by SE-NT and Mid domains. We speculate that although SE ZF+C alone is able to largely rescue the miRNA deficient phenotype *in vivo*, SE-NT, Mid domains and/or N-terminal tail might serve unidentified protein-protein interactions involved in pri-miRNA processing. This line of speculation was inspired by the observation that rescue of leaf serration with *in vivo* expression of ZF+C failed to entail recovered accumulation of miR319 (Fig. 3.6.B. lane 4), despite apparently normal leaf morphology and vigor of the plant. More rigorous morphological and genetic analyses would be required to explicate the discrepancy and assess the possibility of SE domain-specific involvement in different miRNA maturation and of having other SE-requiring miRNAs whose deficiency inflicts more significant impacts on leaf morphology than miR319 does. Another explanation is that the molecular phenotype did not recover miR319 because of its unique mechanism of excision from pri-miRNA. As mentioned in Chapter One, pri-miRNA319 goes through multiple-step cleavage beginning with loop-proximal cut followed by gradual shortening from loop proximity. SE ZF+C without Nt and Mid domains is most likely devoid of interaction with HYL1 and as a result mitigates maturation of miR319, strongly suggesting that one or both of SE amino-terminal region and HYL1 participate(s) in recognition and multiple processing of structurally unique miR319. If the processing pathway of pri-miR319 is proven to be non-canonical, HYL1 might play the non-canonical path guide dispensable for certain miRNAs.

6. Potential roles of SE and DDL in bridging miRNA and other pathways

Furthermore, our unexpected discovery of minimal SE-fragment capable of serration rescue further raises intriguing questions of how others parts of SE than ZF+C participate in and facilitate RNA metabolism at the bifurcation of micro- and messenger-RNA processing pathways (82). It is yet to be seen whether SE homolog, Ars2, adopts similar structural principles to recognize miRNA precursors in flies and in mammals. In Ars2-deficient phenotype, the accumulation of pri-miRNA as well as miRNA declines while *se-1* renders only miRNA reduced. This molecular phenotype of Ars2 is more similar to that of DDL which, Yu et al (2008) proposes, works at the upper stream of miRNA processing than HYL1 and SE do (48). Therefore, the entire functions of SE might be better explored in relation to DDL. Elucidation of DDL function awaits further efforts both from structural and functional standpoints in order to identify the phosphorylated residue in DCL1 and co-crystallize DDL FHA with phospho-serine/threonine containing fragment of DCL1. The connection between miRNA processing and kinase signalling pathway deserves pursuit. For instance, TRBP, a binding partner of animal Dicer is known to cross-talk with MAPK/Erk signaling pathway. The phosphorylation pattern of DCL1 might account for why DCL1, not other Dicer-Like homologues, mainly mediates miRNA biogenesis.

7. Phospho-dependent protein-protein interation in miRNA pathway

Another question of interest is how DCL1 winnows out hundreds of thousands of similar RNAs from miRNA precursors in sequence-dependent or independent manners. The recent updates from mammalian systems illuminate one of the precursor selection processes likely applicable to plant systems. Human miRNA processing by Drosha involves Smad proteins with MH1 and MH2 domain. Smad proteins are not homologous to FHA-containing proteins in amino acid sequence but share structural homology according to Dali search. MH2 domain

of Smad proteins can be well superimposed on FHA domains from mammalian phospho-signalling pathways, among which Smad Nuclear Interacting Protein 1 (SNIP1) alone shares sequence homology with *Arabidopsis* DDL FHA. Drosha-mediated miRNA processing uses amino-terminal MH1and carboxyl-terminal MH2 domains interacting with pre-miRNAs and p68 respectively (83,84). This mechanism makes a striking resemblance with DDL that recruits pri-miRNA by flexible amino-terminus and bind DCL1 by carboxyl-terminus where FHA domain is located (48). Furthermore, each Smad protein recognizes unique short consensus sequence in pri-miRNA, and recruits it to Drosha for cleavage. It makes a paradigmatic sense that ribo-nuclease III Drosha and Dicer, although their catalytic activities are known to be independent of substrate nucleotide sequence, actually choose specific sequences by the microprocessor subunits, and that the choice of pri-miRNA to be cleaved might depend on growth factor stimulation.

8. Phospho-dependent interaction by DDL

Our structural insight provides a clue that DDL FHA domain interacts with its binding partner, reportedly DCL1, through the relatively rigid loops connecting the strands that form β-sandwich. To corroborate our point, different modes of DDL-DCL1 interaction is considered for possibility. What could be the recognition mode if the interaction is independent of phospho-threonine and thus of signaling pathway? Tong et al (2010) recently showed that phospho-independent FHA domain interaction does exist (85). Kinesin-Like Protein KIF13B is among the Dali search results by DDL FHA domain (PDB code 8FM8, Z-score = 12.6, RMSD = 2.0 Å). Figure 4.3. shows that FHA domain of KIF13B is a typical 11-stranded β-sandwich fold common to many FHA proteins, but interacts with CENTA1 through multiple hydrophobic and polar contacts on the sides of β-strands instead of strands-connecting loops. Their point is supported by the absence of conserved arginine after glycine

residues identified by alignment in the sequence of KIF13B FHA domain (Fig. 4.4.A). Based on the residue after the invariant glycine (G473 in KIF13B), the FHA domains are classified into two groups. One group contains a lysine or an arginine right after the invariant glycine, while the rest lack a positively charged residue at the same position. All phospho-dependent FHA domains selected have either lysine or arginine after the invariant glycine. So does DDL FHA (Fig. 4.4.B). Our crystal structure demonstrates that homo-dimerization of DDL FHA inserts a glutamate residue into a positively charged cleft made by loops on the top of β-sandwich instead of the side of β-strands (Fig. 3.11.). Together with the presence of conserved arginine after glycine (Fig. 4.4.B.), we propose that DDL-DCL1 interaction occurs at the well defined loop on FHA β-sandwich that resembles canonical phospho-dependent FHA domains and likely recognize a phosphorylated threonine on DCL1.

Tong et al (2010)

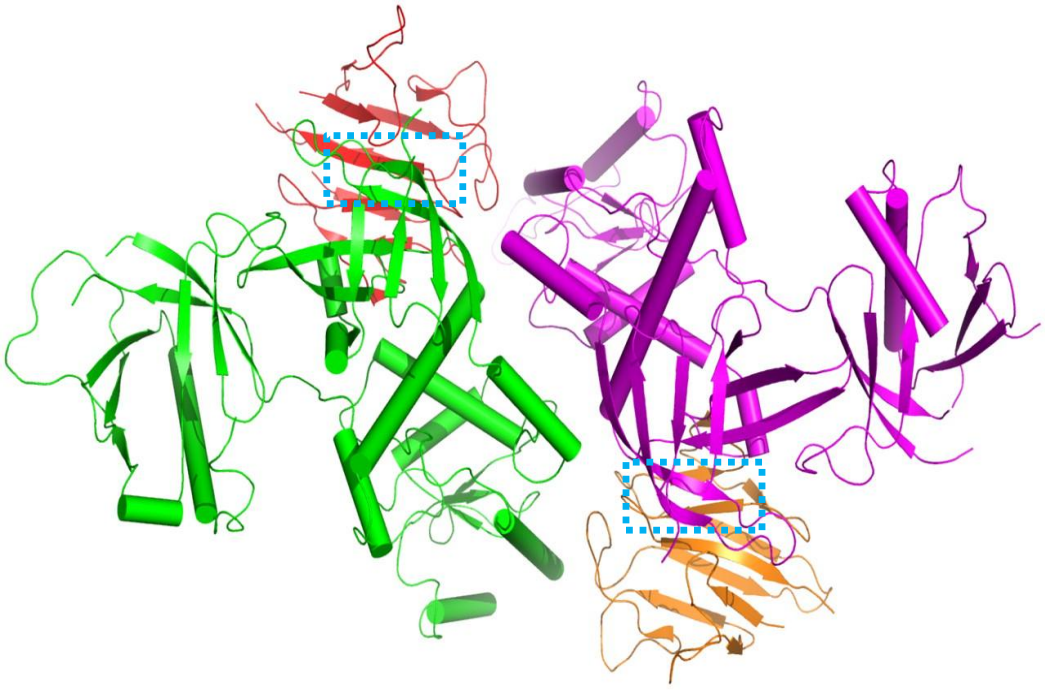
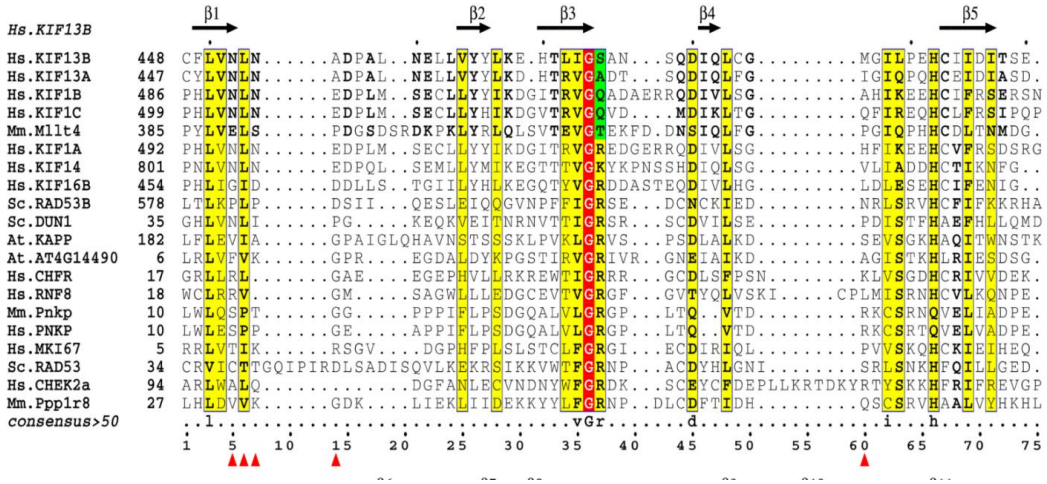


Figure 4.3. Structure of CENTA1 bound to KIF13B-FHA domain in the asymmetric unit.
CENTA1 bound to KIF13B-FHA in the asymmetric unit through multiple hydrophobic and polar contacts on β-strands of FHA. The asymmetric unit contains two CENTA1 and two KIF13B-FHA molecules. The multiple hydrophobic and polar contacts surface made by KIF 13B-FHA is highlighted by blue dotted squares.

Tong et al (2010)

A



B

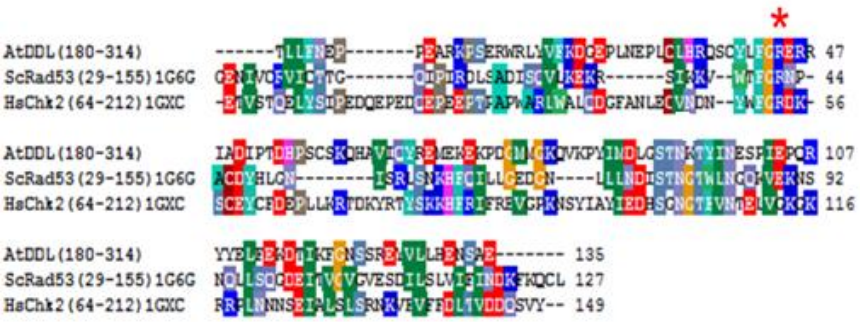


Figure 4.4. Structure-based sequence alignment of kinesin-3 family FHA domains with phospho-threonine-binding FHA domains.
A: KIF13B-FHA residues interacting with CENTA1 are marked with red triangles. Based on the residue after the invariant glycine (G473), the all FHA domains with a positive charge immediately following the glycine qualify for phospho-dependent interaction.
B: Sequences of DDL FHA and structurally conserved Rad53 and Chk2 FHA domains are aligned. The asterisk(*) underscores the conserved arginine together with the glycine shared among phospho-dependent FHA domains.

Bibliography

1. Mattick, J.S. (2007) A new paradigm for developmental biology. *J Exp Biol*, **210**, 1526-1547.
2. Beadle, G.W. and Tatum, E.L. (1941) Genetic Control of Biochemical Reactions in Neurospora. *Proc Natl Acad Sci U S A*, **27**, 499-506.
3. Ingram, V.M. (1956) Specific Chemical Difference between the Globins of Normal Human and Sickle-Cell Anaemia Haemoglobin. *Nature*, **178**, 792-794.
4. Mattick, J.S. and Makunin, I.V. (2006) Non-coding RNA. *Hum Mol Genet*, **15 Spec No 1**, R17-29.
5. Berget, S.M., Moore, C. and Sharp, P.A. (2000) Spliced segments at the 5 ' terminus of adenovirus 2 late mRNA - (adenovirus 2 mRNA processing/5 ' tails on mRNAs/electron microscopy of mRNA DNA hybrids) (Reprinted from Proc Natl Acad Sci USA, vol 74, pg 3171-3175, 1977). *Reviews in Medical Virology*, **10**, 356-362.
6. Chow, L.T., Gelinas, R.E., Broker, T.R. and Roberts, R.J. (2000) An amazing sequence arrangement at the 5 ' ends of adenovirus 2 messenger RNA (Reprinted from Cell, vol 12, pg 1-12, 1977). *Reviews in Medical Virology*, **10**, 362-369.
7. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, **31**, 147-157.
8. Visser, C.M. (1984) Evolution of Biocatalysis .1. Possible Pre-Genetic-Code Rna Catalysts Which Are Their Own Replicase. *Origins Life Evol B*, **14**, 291-300.
9. Altman, S., Baer, M.F., Bartkiewicz, M., Gold, H., Guerriertakada, C., Kirsebom, L.A., Lumelsky, N. and Peck, K. (1989) Catalysis by the Rna Subunit of Rnase-P - a Minireview. *Gene*, **82**, 63-64.
10. Taft, R.J., Pheasant, M. and Mattick, J.S. (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays*, **29**, 288-299.
11. Buchler, N.E., Gerland, U. and Hwa, T. (2003) On schemes of combinatorial transcription logic. *Proc Natl Acad Sci U S A*, **100**, 5136-5141.
12. Orgel, L.E. and Crick, F.H.C. (1980) Selfish DNA - the Ultimate Parasite. *Nature*, **284**, 604-607.
13. Collins, F.S., Lander, E.S., Rogers, J., Waterston, R.H. and Conso, I.H.G.S. (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931-945.
14. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304-1351.
15. Napoli, C., Lemieux, C. and Jorgensen, R. (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in Trans. *Plant Cell*, **2**, 279-289.
16. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E. and Mello, C.C. (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, **391**, 806-811.
17. Lorkovic, Z.J. (2009) Role of plant RNA-binding proteins in development, stress response and genome organization. *Trends Plant Sci*, **14**, 229-236.
18. Ronemus, M., Vaughn, M.W. and Martienssen, R.A. (2006) MicroRNA-targeted and small interfering RNA-mediated mRNA degradation is regulated by argonaute, dicer, and RNA-dependent RNA polymerase in Arabidopsis. *Plant Cell*, **18**, 1559-1574.
19. Xie, Z., Khanna, K. and Ruan, S. (2010) Expression of microRNAs and its regulation in plants. *Semin Cell Dev Biol*, **21**, 790-797.
20. Cuperus, J.T., Montgomery, T.A., Fahlgren, N., Burke, R.T., Townsend, T., Sullivan, C.M. and Carrington, J.C. (2010) Identification of MIR390a precursor processing-defective mutants in Arabidopsis by direct genome sequencing. *Proc Natl Acad Sci U S A*, **107**, 466-471.

21. Wu, S.H., Chen, H.M., Chen, L.T., Patel, K., Li, Y.H. and Baulcombe, D.C. (2010) 22-nucleotide RNAs trigger secondary siRNA biogenesis in plants. *Proc Natl Acad Sci U S A*, **107**, 15269-15274.
22. Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A. and Carrington, J.C. (2005) Expression of Arabidopsis MIRNA genes. *Plant Physiol*, **138**, 2145-2154.
23. Zhou, X., Ruan, J., Wang, G. and Zhang, W. (2007) Characterization and identification of microRNA core promoters in four model species. *PLoS Comput Biol*, **3**, e37.
24. Merchan, F., Boualem, A., Crespi, M. and Frugier, F. (2009) Plant polycistronic precursors containing non-homologous microRNAs target transcripts encoding functionally related proteins. *Genome Biol*, **10**, R136.
25. Rajagopalan, R., Vaucheret, H., Trejo, J. and Bartel, D.P. (2006) A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev*, **20**, 3407-3425.
26. Park, W., Li, J., Song, R., Messing, J. and Chen, X. (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr Biol*, **12**, 1484-1495.
27. Bernstein, E., Caudy, A.A., Hammond, S.M. and Hannon, G.J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, **409**, 363-366.
28. Finnegan, E.J., Margis, R. and Waterhouse, P.M. (2003) Posttranscriptional gene silencing is not compromised in the Arabidopsis CARPEL FACTORY (DICER-LIKE1) mutant, a homolog of Dicer-1 from Drosophila. *Curr Biol*, **13**, 236-240.
29. Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T. and Zamore, P.D. (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science*, **293**, 834-838.
30. Boutet, S., Vazquez, F., Liu, J., Beclin, C., Fagard, M., Gratias, A., Morel, J.B., Crete, P., Chen, X. and Vaucheret, H. (2003) Arabidopsis HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. *Curr Biol*, **13**, 843-848.
31. Yang, Z., Ebright, Y.W., Yu, B. and Chen, X. (2006) HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res*, **34**, 667-675.
32. Ruiz-Ferrer, V. and Voinnet, O. (2009) Roles of Plant Small RNAs in Biotic Stress Responses. *Annu Rev Plant Biol*, **60**, 485-510.
33. Clarke, J.H., Tack, D., Findlay, K., Van Montagu, M. and Van Lijsebettens, M. (1999) The SERRATE locus controls the formation of the early juvenile leaves and phase length in Arabidopsis. *Plant J*, **20**, 493-501.
34. Grigg, S.P., Canales, C., Hay, A. and Tsiantis, M. (2005) SERRATE coordinates shoot meristem function and leaf axial patterning in Arabidopsis. *Nature*, **437**, 1022-1026.
35. Lobbes, D., Rallapalli, G., Schmidt, D.D., Martin, C. and Clarke, J. (2006) SERRATE: a new player on the plant microRNA scene. *EMBO Rep*, **7**, 1052-1058.
36. Fujioka, Y., Utsumi, M., Ohba, Y. and Watanabe, Y. (2007) Location of a possible miRNA processing site in SmD3/SmB nuclear bodies in Arabidopsis. *Plant Cell Physiol*, **48**, 1243-1253.
37. Kurihara, Y., Takashi, Y. and Watanabe, Y. (2006) The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *Rna*, **12**, 206-212.
38. Luo, Q.J., Samanta, M.P., Koksal, F., Janda, J., Galbraith, D.W., Richardson, C.R., Ou-Yang, F. and Rock, C.D. (2009) Evidence for antisense transcription associated with microRNA target mRNAs in Arabidopsis. *PLoS Genet*, **5**, e1000457.
39. Tagami, Y., Motose, H. and Watanabe, Y. (2009) A dominant mutation in DCL1 suppresses the hyl1 mutant phenotype by promoting the processing of miRNA. *Rna*, **15**, 450-458.
40. Yang, S.W., Chen, H.Y., Yang, J., Machida, S., Chua, N.H. and Yuan, Y.A. (2010) Structure of Arabidopsis HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. *Structure*, **18**, 594-605.

41. Lu, C. and Fedoroff, N. (2000) A mutation in the Arabidopsis HYL1 gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. *Plant Cell*, **12**, 2351-2366.
42. Pouch-Pelissier, M.N., Pelissier, T., Elmayan, T., Vaucheret, H., Boko, D., Jantsch, M.F. and Deragon, J.M. (2008) SINE RNA induces severe developmental defects in Arabidopsis thaliana and interacts with HYL1 (DRB1), a key member of the DCL1 complex. *PLoS Genet*, **4**, e1000096.
43. Song, L., Han, M.H., Lesicka, J. and Fedoroff, N. (2007) Arabidopsis primary microRNA processing proteins HYL1 and DCL1 define a nuclear body distinct from the Cajal body. *Proc Natl Acad Sci U S A*, **104**, 5437-5442.
44. Fang, Y. and Spector, D.L. (2007) Identification of nuclear dicing bodies containing proteins for microRNA biogenesis in living Arabidopsis plants. *Curr Biol*, **17**, 818-823.
45. Yang, L., Liu, Z., Lu, F., Dong, A. and Huang, H. (2006) SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. *Plant J*, **47**, 841-850.
46. Dong, Z., Han, M.H. and Fedoroff, N. (2008) The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. *Proc Natl Acad Sci U S A*, **105**, 9970-9975.
47. Morris, E.R., Chevalier, D. and Walker, J.C. (2006) DAWDLE, a forkhead-associated domain gene, regulates multiple aspects of plant development. *Plant Physiol*, **141**, 932-941.
48. Yu, B., Bi, L., Zheng, B., Ji, L., Chevalier, D., Agarwal, M., Ramachandran, V., Li, W., Lagrange, T., Walker, J.C. *et al.* (2008) The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. *Proc Natl Acad Sci U S A*, **105**, 10073-10078.
49. Chen, X. (2008) A silencing safeguard: links between RNA silencing and mRNA processing in Arabidopsis. *Dev Cell*, **14**, 811-812.
50. Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J.U., Ratsch, G. and Weigel, D. (2008) Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. *Proc Natl Acad Sci U S A*, **105**, 8795-8800.
51. Kim, S., Yang, J.Y., Xu, J., Jang, I.C., Prigge, M.J. and Chua, N.H. (2008) Two cap-binding proteins CBP20 and CBP80 are involved in processing primary MicroRNAs. *Plant Cell Physiol*, **49**, 1634-1644.
52. Gruber, J.J., Zatechka, D.S., Sabin, L.R., Yong, J., Lum, J.J., Kong, M., Zong, W.X., Zhang, Z., Lau, C.K., Rawlings, J. *et al.* (2009) Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. *Cell*, **138**, 328-339.
53. Nielsen, A.F., Gloggnitzer, J. and Martinez, J. (2009) Ars2 and the Cap-Binding Complex Team up for Silencing. *Cell*, **138**, 224-226.
54. Sabin, L.R., Zhou, R., Gruber, J.J., Lukinova, N., Bambina, S., Berman, A., Lau, C.K., Thompson, C.B. and Cherry, S. (2009) Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in Drosophila. *Cell*, **138**, 340-351.
55. Voinnet, O. (2009) Fly antiviral RNA silencing and miRNA biogenesis claim ARS2. *Cell Host Microbe*, **6**, 99-101.
56. Wilson, M.D., Wang, D., Wagner, R., Breyssens, H., Gertsenstein, M., Lobe, C., Lu, X., Nagy, A., Burke, R.D., Koop, B.F. *et al.* (2008) ARS2 is a conserved eukaryotic gene essential for early mammalian development. *Mol Cell Biol*, **28**, 1503-1514.
57. Dominski, Z. (2010) An RNA end tied to the cell cycle: new ties to apoptosis and microRNA formation? *Cell Cycle*, **9**, 1308-1312.
58. Kim, V.N., Han, J. and Siomi, M.C. (2009) Biogenesis of small RNAs in animals. *Nat Rev Mol Cell Biol*, **10**, 126-139.
59. Machida, S., Chen, H.Y. and Adam Yuan, Y. (2011) Molecular insights into miRNA processing by Arabidopsis thaliana SERRATE. *Nucleic Acids Res*.
60. Werner, S., Wollmann, H., Schneeberger, K. and Weigel, D. (2010) Structure determinants for accurate processing of miR172a in Arabidopsis thaliana. *Curr Biol*, **20**, 42-48.

61. Bologna, N.G., Mateos, J.L., Bresso, E.G. and Palatnik, J.F. (2009) A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *Embo J*, **28**, 3646-3656.
62. Addo-Quaye, C., Snyder, J.A., Park, Y.B., Li, Y.F., Sunkar, R. and Axtell, M.J. (2009) Sliced microRNA targets and precise loop-first processing of MIR319 hairpins revealed by analysis of the Physcomitrella patens degradome. *Rna*, **15**, 2112-2121.
63. Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W. and Richmond, T.J. (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 a resolution. *J Mol Biol*, **319**, 1097-1113.
64. Neuhaus, D., Nakaseko, Y., Schwabe, J.W. and Klug, A. (1992) Solution structures of two zinc-finger domains from SWI5 obtained using two-dimensional 1H nuclear magnetic resonance spectroscopy. A zinc-finger structure with a third strand of beta-sheet. *J Mol Biol*, **228**, 637-651.
65. Andreeva, A. and Murzin, A.G. (2008) A fortuitous insight into a common mode of RNA recognition by the dsRNA-specific zinc fingers. *Proc Natl Acad Sci U S A*, **105**, E128-129.
66. Dong, Z., Han, M.H. and Fedoroff, N. (2008) The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. *Proc Natl Acad Sci U S A*, **105**, 9970-9975.
67. Yang, L., Liu, Z., Lu, F., Dong, A. and Huang, H. (2006) SERRATE is a novel nuclear regulator in primary microRNA processing in Arabidopsis. *Plant J*, **47**, 841-850.
68. Holm, L. and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Research*, **38**, W545-W549.
69. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389-3402.
70. Lee, G.I., Ding, Z.F., Walker, J.C. and Van Doren, S.R. (2003) NMR structure of the forkhead-associated domain from the Arabidopsis receptor kinase-associated protein phosphatase. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 11261-11266.
71. Qin, B.Y., Chacko, B.M., Lam, S.S., de Caestecker, M.P., Correia, J.J. and Lin, K. (2001) Structural basis of Smad1 activation by receptor kinase phosphorylation. *Molecular Cell*, **8**, 1303-1312.
72. Wu, G., Chen, Y.G., Ozdamar, B., Gyuricza, C.A., Chong, P.A., Wrana, J.L., Massague, J. and Shi, Y.G. (2000) Structural basis of Smad2 recognition by the Smad anchor for receptor activation. *Science*, **287**, 92-97.
73. Hao, R., Chen, L., Wu, J.W. and Wang, Z.X. (2008) Structure of Drosophila Mad MH2 domain. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications*, **64**, 986-990.
74. Wang, C., Chen, L., Wang, L. and Wu, J.W. (2009) Crystal structure of the MH2 domain of Drosophila Mad. *Science in China Series C-Life Sciences*, **52**, 539-544.
75. Zhou, C. and Huang, R.H. (2008) Crystallographic snapshots of eukaryotic dimethylallyltransferase acting on tRNA: insight into tRNA recognition and reaction mechanism. *Proc Natl Acad Sci U S A*, **105**, 16142-16147.
76. Li, Y., Luo, J., Zhou, H., Liao, J.Y., Ma, L.M., Chen, Y.Q. and Qu, L.H. (2008) Stress-induced tRNA-derived RNAs: a novel class of small RNAs in the primitive eukaryote Giardia lamblia. *Nucleic Acids Research*, **36**, 6048-6055.
77. Liu, S.X., Abbondanzieri, E.A., Rausch, J.W., Le Grice, S.F.J. and Zhuang, X.W. (2008) Slide into Action: Dynamic Shuttling of HIV Reverse Transcriptase on Nucleic Acid Substrates. *Science*, **322**, 1092-1097.
78. Beattie, T.R. and Bell, S.D. (2011) The role of the DNA sliding clamp in Okazaki fragment maturation in archaea and eukaryotes. *Biochemical Society Transactions*, **39**, 70-76.
79. Pan, M., Kelman, L.M. and Kelman, Z. (2011) The archaeal PCNA proteins. *Biochemical Society Transactions*, **39**, 20-24.

80. Crary, S.M., Niranjanakumari, S. and Fierke, C.A. (1998) The protein component of Bacillus subtilis ribonuclease P increases catalytic efficiency by enhancing interactions with the 5 ' leader sequence of pre-tRNA(Asp). *Biochemistry*, **37**, 9409-9416.
81. Niranjanakumari, S., Stams, T., Crary, S.M., Christianson, D.W. and Fierke, C.A. (1998) Protein component of the ribozyme ribonuclease P alters substrate recognition by directly contacting precursor tRNA. *Proceedings of the National Academy of Sciences of the United States of America*, **95**, 15212-15217.
82. Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J.U., Ratsch, G. and Weigel, D. (2008) Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. *Proc Natl Acad Sci U S A*, **105**, 8795-8800.
83. Warner, D.R., Bhattacherjee, V., Yin, X.L., Singh, S., Mukhopadhyay, P., Pisano, M.M. and Greene, R.M. (2004) Functional interaction between Smad, CREB binding protein, and p68 RNA helicase. *Biochemical and Biophysical Research Communications*, **324**, 70-76.
84. Davis, B.N., Hilyard, A.C., Lagna, G. and Hata, A. (2008) SMAD proteins control DROSHA-mediated microRNA maturation. *Nature*, **454**, 56-U52.
85. Tong, Y.F., Tempel, W., Wang, H., Yamada, K., Shen, L.M., Senisterra, G.A., MacKenzie, F., Chishti, A.H. and Park, H.W. (2010) Phosphorylation-independent dual-site binding of the FHA domain of KIF13 mediates phosphoinositide transport via centaurin alpha 1. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 20346-20351.

Appendix

Publications

S Machida, H Y Chen and Y. Adam Yuan (2011) Molecular insights into miRNA processing by Arabidopsis thaliana SERRATE. Nucleic Acids Research, 1-9, doi:10.1093/nar/gkr428

Wang, M, T Soyano, S Machida, JY Yang, C Jung, NH Chua and YA Yuan. (2011) Molecular insights into plant cell proliferation disturbance by ***Agrobacterium*** protein 6b. Genes & Development 25: 64-76.

Qin, H, F Chen, X Huan, S Machida, J Song and YA Yuan. (2010) Structure of Arabidopsis thaliana DCL4 DUF283 domain reveals a non-canonical double-stranded RNA-binding fold for protein-protein interaction. RNA 16: 474-481.

Yang, SW, HY Chen, J Yang, S Machida, NH Chua and YA Yuan. (2010) Structure of ***Arabidopsis*** HYPONASTIC LEAVES1 and its molecular implications for miRNA processing. Structure 18: 594-605.

*S Machida., Shibuya, M., and Sano, T. (2008). Enrichment of viroid small RNAs by hybridization selection using biotinylated RNA transcripts to analyze viroid induced RNA silencing. *Journal of General Plant Pathology* 74, 203-207.

*S Machida., Yamahata, N., Watanuki, H., Owens, R. A., and Sano, T. (2007). Successive accumulation of two size classes of viroid-specific small RNA in potato spindle tuber viroid-infected tomato plants. *Journal of General Virology* 88, 3452-3457.

**These works were conducted outside NUS.*