Insight into structure and function of Dicer-related helicases from *Caenorhabditis elegans*

Li, Kuohan

2019

Li, K. (2019). Insight into structure and function of Dicer-related helicases from *Caenorhabditis elegans*. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/96068>

<https://doi.org/10.32657/10220/49480>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**Insight into structure and function of
Dicer-related helicases from *Caenorhabditis elegans***

Li Kuohan

SCHOOL OF BIOLOGICAL SCIENCES

2019

**Insight into structure and function of
Dicer-related helicases from *Caenorhabditis elegans***

Li Kuohan

SCHOOL OF BIOLOGICAL SCIENCES

A thesis submitted to the Nanyang Technological
University in partial fulfilment of the requirement for the
degree of Doctor of Philosophy

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

19 Jan 2019

.....
Date

.....
Li Kuohan

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

19 Jan 2019

.....

Date

.....

Luo Dahai

Authorship Attribution Statement

Please select one of the following; *delete as appropriate:

*(A) This thesis **does not** contain any materials from papers published in peer-reviewed journals or from papers accepted at conferences in which I am listed as an author.

(B) This thesis contains material from [x number] paper(s) published in the following peer-reviewed journal(s) / from papers accepted at conferences in which I am listed as an author.

[29 July 2019]

.....

Date

[

.....

[Li, Kuohan]

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors, Professor Luo Dahai and Professor Daniela Rhodes for giving me this chance to pursue my PhD in NTU and for their guidance, encouragement and trust throughout the past four years. I am deeply grateful to my thesis advisory committee Professor Francesc Xavier Roca Castella, Professor Katsutomo Okamura and Professor Wu Bin for their help and advice for this study.

My acknowledgements extend to Dr. Melissa Wirawan, Dr. Yong Huiyee, Dr. Wint Wint Phoo, Dr. Chen Ming Wei, Dr. Zhao Yongqian, Dr. Liew Chong Wai, Dr. Chia Wei Sheng and Dr. Simon Lattmann for their generous help and valuable advice throughout the time I worked with them.

I would like to thank Dr. El Sahili Abbas and Mr Jacky Tan for their help on X-ray data collection at synchrotrons. I would also like to thank Dr. Zheng Jie and Dr. Olga Fedorova whom I work closely with, in a collaboration project.

My sincere gratitude also goes to my lab mates Ms Zhang Zhenzhen, Ms Xiong Zhiling, Mr Tan Yawbia, Mr Chew Bing Liang Alvin and Mr Quek Jun Ping, Ms Nguyen Mai Trinh and Dr. Law Yee Song for their constant support and help.

Lastly, I am deeply indebted to my family, especially my husband Dr. Zhong Wenhe and my lovely daughter Letao, for their love, understanding and support.

Table of Contents

Statement of originality	ii
Supervisor declaration statement.....	iii
Acknowledgements	iv
Table of contents	v
List of figures	viii
List of tables	x
Abbreviations	xi
Abstract.....	xiv
Chapter 1: Introduction.....	1
1.1. Overview of helicases	1
1.1.1. Classification of helicases	1
1.1.2. The SF2 RNA helicases, the largest helicase group.....	5
1.1.3. Duplex RNA-dependent ATPases (DRAs)	7
1.2. RIG-I like receptors	10
1.2.1. The antiviral innate immunity	10
1.2.2. Key pattern recognition receptors in antiviral innate immunity....	11
1.2.3. Antiviral signaling pathways of RIG-I and MDA5	12
1.2.4. Structural overview of RLRs.....	14
1.2.4.1. <i>Helicase domain: a dsRNA-dependent ATPase</i>	14
1.2.4.2. <i>CTD and viral RNA structure recognition.</i>	18
1.2.4.3. <i>The CARD domain: a signaling motif.</i>	20
1.2.5. Functions and roles of RLRs	23
1.2.5.1. <i>RIG-I</i>	23
1.2.5.2. <i>MDA5</i>	25
1.2.5.3. <i>LGP2</i>	26
1.3. Dicer-related helicases.....	27
1.3.1. The discovery of RNA interference	27
1.3.2. Biogenesis and classification of small RNAs.....	29
1.3.3. Mechanism of RNAi.....	32
1.3.4. RNA-directed viral immunity.....	34
1.3.5. DRHs are worm-homolog of RLRs.....	37
1.3.6. DRH-1 and worm antiviral immunity	37
1.3.7. DRH-3 and RdRP-dependent secondary small RNA synthesis	38
1.4. Aims of this thesis	39
Chapter 2: Materials and Methods	41
2.1. Protein expression	41
2.1.1. Recombinant full-length DRH-1 and DRH-3.....	41
2.1.2. High-throughput screening of protein overexpression in <i>E. coli</i> ..	41
2.1.3. Expression of recombinant proteins	42
2.1.4. Selenomethionine incorporation protein expression	42
2.2. Protein purification.....	43
2.2.1. Preparation of cell lysates.....	43

2.2.2. Immobilized metal ion affinity chromatography (IMAC) purification	44
2.2.3. Optional Heparin column purification.....	44
2.2.4. Optional tag cleavage	45
2.2.5. Gel filtration chromatography	45
2.3. RNA Preparation	46
2.3.1. Synthetic RNAs	46
2.3.2. In vitro transcription	46
2.3.3. RNA purification	48
2.4. Biophysical and biochemical assays	50
2.4.1. ATPase assay.....	50
2.4.2. Thermal shift assay.....	53
2.4.3. Fluorescence polarization assay	54
2.4.4. Limited proteolysis	57
2.5. Crystallization.....	58
2.5.1. Crystal screening	58
2.5.2. Data collection.....	58
2.5.3. Structure determination and refinement	59
2.5.4. Structure modeling	60
2.6. Hydrogen Deuterium Exchange with Mass Spec (HDX-MS)	60
Chapter 3: Structural insights into the role of DRH-3 in <i>C. elegans</i> RNAi.	61
3.1. Overview	61
3.2. Protein and RNA production	61
3.2.1. DRH-3 is a worm homolog of RLRs.....	61
3.2.2. Full-length DRH-3 is a monomer in solution.....	66
3.2.3. Construct design for expressing DRH-3 domains	69
3.2.4. Expression, purification and crystallization of DRH-3 NTD	70
3.2.5. Expression, purification and crystallization of DRH-3 CTD	75
3.2.6. Synthesis and purification of RNAs	76
3.3. Structure determinations	79
3.3.1. Solving the phase problem for DRH-3 NTD.....	79
3.3.2. Structure determination of DRH-3 NTD.....	82
3.3.3. Structure determination of DRH-3 CTD	83
3.4. DRH-3 NTD is composed of tandem CARDS in parallel arrangement	86
3.4.1. DRH-3 NTD exists as a monomer.....	86
3.4.2. DRH-3 NTD is composed of two CARDS	88
3.4.3. The tandem CARDS are in parallel arrangement	91
3.5. Structural insight into the secondary siRNA recognition of DRH-3 CTD.....	93
3.5.1. Overall structure of DRH-3 CTD	93
3.5.2. The structure basis of 5'-ppp RNA recognition by DRH-3 CTD ..	95
3.5.3. Comparison of DRH-3 CTD and RLR CTDs	99
3.5.4. DRH-3 CTD binds 5'-ppp dsRNA with high affinity	103
3.6. RNA-induced conformational changes in full-length DRH-3 studied by HDX-MS	108
3.6.1. Protein dynamics of DRH-3	109
3.6.2. 5'-ppp RNA binding mechanism for DRH-3.....	110
3.6.3. DRH-3, RIG-I and MDA5 are activated by distinct mechanisms	112
3.7. Discussion.....	114
3.7.1. DRH-3 NTD acts as a signaling domain in the RNAi pathway ..	114
3.7.2. Secondary siRNA recognition by DRH-3	115

3.7.3. Is oligomerization required for a functional DRH-3?	117
3.8. Summary.....	118
Chapter 4: Duplex RNA-dependent ATPase DRH-1: Insight into its role in RNA directed viral immunity	
4.1. Overview	119
4.2. Expression and purification of DRH-1 constructs.....	120
4.3. DRH-1 ATPase activity is stimulated by RNA	128
4.4. RNA binding site mapping and allosteric effects upon 3p10L binding to DRH- 1 Δ NTD	132
4.5. DRH-1 CTD binds dsRNA bearing 5'-ppp and 3'-overhang with higher affinity	134
4.6. Discussion.....	138
4.6.1. Problems in expression of the full-length DRH-1	138
4.6.2. Problems in characterization of DRH-1NTD	139
4.6.3. DRH-1 recognizes primary vsiRNA	140
4.6.4. Antiviral functions of <i>C. elegans</i> DRHs and mammalian RLRs.	141
4.7. Summary	143
Chapter 5: Conclusion and Future directions	144
References	152
Appendix	165

List of Figures

Figure 1.1 Superfamilies of helicases.	2
Figure 1.2 Conserved motifs in SF1 and SF2.	4
Figure 1.3 The SF2 family.....	6
Figure 1.4 Phylogenetic tree analysis of RIG-I like RNA helicase family.	7
Figure 1.5 Domain architecture of RIG-I like RNA helicases.	9
Figure 1.6 Antiviral Signaling Pathways of RIG-I-like Receptors (RLR).	13
Figure 1.7 A typical RNA-binding domain of RIG-I-like receptors.	19
Figure 1.8 The type I, II, III interfaces.	22
Figure 1.9 Dicer Related Helicases (DRHs) are involved in the antiviral RNA interference (RNAi) pathway in <i>C. elegans</i>	36
Figure 2.1 Template design for <i>in vitro</i> transcription.	46
Figure 2.2 Schematic diagram for NADH coupled ATPase assay.	51
Figure 2.3 The Michaelis-Menten model	53
Figure 2.4 Log (Inhibitor) vs response curve.	57
Figure 3.1 Sequence alignment of DRHs and RLRs.	63
Figure 3.2 Expression trials for full-length DRH-3 and DRH-3 Hel-CTD.	66
Figure 3.3 Purification of His ₆ -SUMO-DRH-3 by IMAC.	67
Figure 3.4 Purification of full-length DRH-3 by Heparin column purification.	68
Figure 3.5 Purification of DRH-3 by gel filtration chromatography.....	69
Figure 3.6 Constructs of DRH-3 with soluble expression.	70
Figure 3.7 Expression and purification of His ₆ -SUMO-tagged DRH-3_N1.	71
Figure 3.8 Improvement of the DRH-3 NTD crystal quality.	73
Figure 3.9 Purification of His ₆ -SUMO-DRH NTD constructs.	74
Figure 3.10 Purification of the His ₆ -DRH-3 CTD construct.	76
Figure 3.11 Purification of RNA using FPLC.	77
Figure 3.12 Crystal of <i>SeMet</i> DRH-3_N2 and data collection.	82
Figure 3.13 Crystal structure of dimeric DRH-3 NTD.	87
Figure 3.14 Illustration of the interactions involved in dimer interface of DRH-3 NTD.	87
Figure 3.15 Cylinder depiction of DRH-3 NTD structure with helices labelled.	88

Figure 3.16 Comparison of the death fold-like subdomains of DRH-3 NTD with top 5 structural hits from Dali search.	90
Figure 3.17 DRH-3 NTD is composed of tandem CARDs in parallel arrangement. ..	92
Figure 3.18 Electrostatic surface of DRH-3 NTD.	93
Figure 3.19. Overall structure of DRH-3 in complex with 5'-ppp 12-mer dsRNA.	94
Figure 3.20 Electrostatic surface of DRH-3 CTD.	96
Figure 3.21 Molecular details of DRH-3 CTD-dsRNA interactions.	97
Figure 3.22 Overall structure of DRH-3 in complex with 5'-ppp 8-mer ssRNA.	99
Figure 3.23 Comparison of C-terminal domain of DRH-3 and RLRs.	101
Figure 3.24 Direct FP binding assay to determine K_d of the DRH-3 CTD.	103
Figure 3.25 Validation of the RNA binding affinity of DRH-3 CTD by FP competition assay.	104
Figure 3.26 Effect of RNAs on CTDs limited proteolysis.	106
Figure 3.27 Stabilization of CTDs by RNA binding.	108
Figure 3.28 Conformational dynamics of DRH-3 NTD and full-length DRH-3.	110
Figure 3.29 Protein dynamics of DRH-3 in response to 3p10L binding.	112
Figure 4.1 Expression trials for full-length DRH-1.	121
Figure 4.2 Constructs of DRH-1 with soluble expression.	122
Figure 4.3 SDS-PAGE analysis of purified of DRH-1 Δ NTD_1 and DRH-1 Δ NTD_2.	123
Figure 4.4 Purification profile of DRH-1 Δ NTD_3.	124
Figure 4.5. Purification of DRH-1_N1 protein with C-terminal His6-tag.	126
Figure 4.6. Purification of DRH-1_N4 protein.	127
Figure 4.7 Purification of DRH-1_C1 by gel-filtration chromatography.	128
Figure 4.8 ATPase activity of recombinant DRH-1 Δ NTD.	129
Figure 4.9 Steady state ATP hydrolysis by DRH-1 Δ NTD_3 stimulated with varying concentrations of RNAs.	131
Figure 4.10 Protein dynamic of DRH-1 Δ NTD in response to 3p10L binding.	133
Figure 4.11 Effect of RNAs on DRH-1 CTD proteolysis.	135
Figure 4.12 Thermal shift assay results for DRH-1 CTD.	136
Figure 4.13 Direct FP binding assay to determine K_d of the DRH-1 CTD.	137
Figure 4.14 FP competition assay for RNA binding to DRH-1 CTD.	138

Figure 5.1 Proposed roles of Dicer-related helicases in RNAi responses in <i>C. elegans</i>	144
--	-----

List of Tables

Table 1.1 Current available structures of RIG-I like receptors.	16
Table 3.1 Pairwise protein sequence comparisons of DRHs and RLRs showing amino- acid identities.	65
Table 3.2 List of RNAs.	78
Table 3.3 Data collection and refinement statistics.	84
Table 3.4 Multiple structural alignment of DRH-3 C1 and C2 with other known structures using the DALI server.	91
Table 4.1 Kinetic constants of DRH-1 in complex with RNAs.	131

Abbreviations

5'-OH	5 prime hydroxy
5'-p	5 prime monophosphate
5'-ppp	5 prime diphosphate
5'-ppp	5 prime triphosphate
ADAR1	adenosine deaminase acting on RNA 1
ATP	adenosine triphosphate
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
CARDs	caspase activation and recruitment domains
CTD	carboxy-terminal domain
DC	dendritic cell
DD	death domain
DRA	double-stranded RNA-dependent ATPase
DRH	Dicer-related RNA helicase
ds	double-stranded
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
endo-	endogenous
exo-	exogenous
FADD	Fas-associated protein with death domain
FAM	fluorescein amidite
FP	fluorescence polarization
GTP	guanosine triphosphate
HDX-MS	hydrogen deuterium exchange with mass spectrometry
HEL	helicase domain
Hel2i	insertion domain of the helicase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IC ₅₀	half maximal inhibitory concentration
IFN	interferon
IKK	IκB kinase
IPTG	isopropyl β-D-1-thiogalactopyranoside

IRF	interferon regulatory factor
K63-Ubn	lysine63-linked poly-ubiquitin chains
K_d	dissociation constant
k_{cat}	turnover number
kDa	kilodalton
K_M	Michaelis constant
LB	Luria Broth
LGP2	laboratory of genetics and physiology 2
MAD	multiple-wavelength anomalous diffraction
MAVS	mitochondrial antiviral-signaling protein
MDA5	melanoma differentiation-associated gene 5
MIR	multiple isomorphous replacement
miRNA	microRNA
MR	molecular replacement
MWCO	molecular weight cut off
N-terminal	amino-terminal
NADH	nicotinamide adenine dinucleotide reduced
NK	natural killer
NLR	nucleotide-binding oligomerization domain-like receptor
NTD	amino-terminal domain
O/N	overnight
OD	optical density
PACT	Protein activator of the interferon-induced protein kinase
PAMP	pathogen-associated molecular pattern
piRNA	piwi-interacting RNA
poly (I:C))	polyinosinic-polycytidylic acid
PRR	pattern recognition receptors
PTGS	post-transcriptional gene silencing
RdRP	RNA-dependent RNA polymerase
RFN	ring finger protein
RIG-I	retinoic acid-inducible gene I
RIP1	receptor-interacting serine/threonine-protein kinase 1

Riplet	ring finger protein leading to RIG-I activation
RISC	RNA-induced silencing complex
RLRs	RIG-I like receptors
RNA	ribonucleic acid
RNAi	RNA interference
rpm	revolution per minute
SAD	single-wavelength anomalous replacement
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SF	superfamily
SIR	single isomorphous replacement
siRNA	small interfering RNA
ss	single-stranded
STING	stimulator of interferon genes
TBK1	Tank binding kinase-1
TEV	tobacco etch virus
TLR	Toll-like receptor
T _m	melting temperature
TNF	tumor necrosis factor
TOM70	outer membrane 70
TRADD	tumor necrosis factor receptor type-1-associated DEATH domain
TRAF	tumor necrosis factor receptor-associated factor
TRBP	trans-activating response RNA-binding protein
Trim25	tripartite motif-containing protein 25
ViRC	viral recognition complex
β-ME	β-mercaptoethanol

Abstract

RNA interference (RNAi) is a major antiviral defense mechanism in nematodes. In *Caenorhabditis elegans* (*C. elegans*), Dicer-related helicase 1 (DRH-1) acts to enhance the production of primary virus-derived small interfering RNA (siRNA) in antiviral RNAi pathway. However, Dicer-related helicase 3 (DRH-3), is involved in the biogenesis of secondary siRNA in both endogenous and exogenous RNAi pathways. DRH-1 and DRH-3 are orthologs to RIG-I like receptors (RLRs), based on their structural and sequential conservation of the helicase domain (HEL) and the C-terminal RNA binding domain (CTD). The structure and function of worm-specific N-terminal domains (NTDs) of DRH-1 and DRH-3 are currently unknown.

This thesis describes biochemical, biophysical and structural studies of DRH-1 and DRH-3 (DRHs) to provide new insights into the RNAi mechanism at the molecular level. We first established the methods for expressing and purifying the full-length and various domains of DRHs. The purified recombinant proteins were then fully characterized to understand their functions in RNA recognition. For DRH-3, we applied X-ray crystallography and solved the crystal structure of DRH-3 NTD, as well as two crystal structures of DRH-3 CTD in complex with 5'-triphosphate (5'-ppp) RNAs. The NTD of DRH-3 adopts a novel fold of tandem CARDS that is different from the CARDS of RIG-I. This suggests DRH-3 may recruit an unknown protein partner in the endogenous RNAi pathway. Crystal structures and RNA binding assay confirmed that CTD preferentially recognizes RNAs with 5'-ppp, which is the typical feature of secondary siRNA transcript. Furthermore, the full-length DRH-3 displays unique structural dynamics and RNA differentiation patterns that are different from RIG-I and MDA5, as

revealed by the hydrogen/deuterium exchange coupled to mass spectrometry experiment (HDX-MS). For DRH-1, we provided the first evidence that the Hel-CTD of DRH-1 has a dsRNA-dependent ATPase function. Moreover, we demonstrated that DRH-1 preferred short dsRNAs, especially those with 5'-ppp and 3' overhang for its ATPase hydrolysis activity and the protein stability. Collectively, our findings reveal unique molecular features of DRHs in RNAi and provide new perspectives for advancing our understanding of small RNA processing and antiviral defense mechanism in worms.

Chapter 1: Introduction

1.1. Overview of helicases

Helicases, first discovered in *Escherichia coli* (*E. coli*) in 1976, have been found in virtually all the organisms and are almost involved in all facets of nucleic acid metabolism (Abdel-Monem, Durwald, & Hoffmann-Berling, 1976; Lohman & Bjornson, 1996; Singleton, Dillingham, & Wigley, 2007; Venkatesan, Silver, & Nossal, 1982). Helicases are ubiquitous and the corresponding genes constitute nearly 1 % of the total genes in numerous eukaryotes (E. Jankowsky, 2011). Helicases require the energy from nucleoside triphosphates such as adenosine triphosphate (ATP), to remodel nucleic acids (Abdel-Monem et al., 1976). As the name implies, helicases were initially identified as a protein that separates the double-stranded (ds) DNA in the presence of ATP. However, a growing number of studies showed that helicases can translocate, rewind or anneal nucleic acids rather than unwind nucleic acids (Byrd & Raney, 2012; Y. Wu, 2012). Helicases are involved in various cellular activities such as DNA replication and repair, transcription, RNA maturation and splicing, ribosome synthesis, translation, nuclear export, RNA interference and chromatin remodeling. Functional defects of helicases may cause various diseases such as neurodegenerative diseases, cancers and developmental defects in human (Hanada & Hickson, 2007).

1.1.1. Classification of helicases

Based on the protein sequence homology and the oligomeric state, helicases can be divided into six superfamilies (SF1-6) (Singleton et al., 2007). Superfamily 1 (SF1) and SF2 function as monomers or dimers known as the non-ring forming helicases

(Dillingham, 2011; Lohman, Tomko, & Wu, 2008). SF3 to SF6 form oligomeric (mostly hexameric) rings (**Figure 1.1**) (Patel & Picha, 2000). The hexamer rings were thought to enhance the processivity of the helicase. SF1 and SF2 appear to be the largest helicase groups among the six superfamilies (Gorbalenya & Koonin, 1993). Their helicase cores contain two identical protein domains that resemble the folding of the recombination protein RecA. In contrast, the SF3-6 consist of fewer conserved motifs and each monomer has only one RecA-like domain (Singleton et al., 2007).

Figure 1.1 Superfamilies of helicases. The noted PDB coordinate files were used to generate the figures using PyMOL (DeLano Scientific LLC, San Francisco, CA).

To date, many characteristic sequence motifs such as Q, I, Ia, Ib, and II to VI motifs, have been identified in helicases (Fairman-Williams, Guenther, & Jankowsky, 2010). The motifs I and II (also referring to Walker A and B motifs) involved in ATP binding and hydrolysis are found in all helicases (Walker, Saraste, Runswick, & Gay, 1982). In contrast, other motifs are less conserved, and their levels of conservation diminish throughout the respective superfamilies (Gorbalenya & Koonin, 1988; Koonin, 1995). For SF1 and SF2, at least 12 sequence motifs found in the two Rec-A like domains were shared but not all are present in one family (Fairman-Williams et al., 2010) (**Figure 1.2**).

Helicases can be grouped into DNA helicases and RNA helicases according to their substrate preferences. DNA helicases can function in a wide range of DNA metabolic processes such as unwinding duplex or alternative DNA structures (G-quadruplex, triplex), chromatin remodeling, and stripping protein bound to DNA (Beyer, Ghoneim, & Spies, 2013; Dillingham, 2011; Pyle, 2008; Y. Wu, 2012). RNA helicases in contrast, catalyze the ATP-dependent separation of RNA duplexes, as well as the structural rearrangement of RNA and RNA-protein complexes in almost every aspect of RNA metabolism. This includes transcription, mRNA splicing and translation, RNA modification and transport, ribosome biogenesis, RNA-protein complex assembly, and RNA degradation (E. Jankowsky, 2011). Most of the RNA helicases are members of SF1 and SF2 and a small number are classified as SF3. SF1 and SF2 are further divided into two groups, namely A and B based on the translocation polarity (5'-3', 3'-5') (Singleton et al., 2007). The other SFs have only 5'-3' translocation polarity.

Figure 1.2 Conserved motifs in SF1 and SF2. Positions of the conserved motifs in three-dimensional structure model of helicase domains 1 and 2 are shown. The domain shown here are from the structure of MDA5 (PDB code: 4GL2). The RNA is colored in green and ATP is colored in orange. The functions of the domains are highlighted accordingly: blue, nucleic acid binding sites; yellow, the binding sites of ATP and RNA; red, coordination sites between nucleic acid and ATP binding sites.

1.1.2. The SF2 RNA helicases, the largest helicase group

The SF2 is the largest superfamily (Gorbalenya & Koonin, 1993). The SF2 enzymes are typically named helicases; however, many of them have other biological functions such as strand annealing, translocation, and protein displacement (Pyle, 2008). Notably, additional enzymatic activities for helicases are attributed to insertion domains as well as amino- (N-) or carboxy- (C-) accessory domains. In some cases, the terminal domains

can be used as a platform to recruit cellular partners (e.g. caspase activation and recruitment domains (CARDs)), or to target substrates (e.g. zinc finger domains) (Ciccia, McDonald, & West, 2008; Kowalinski et al., 2011; Lulchev & Klostermeier, 2014; Shereda, Reiter, Butcher, & Keck, 2009; B. Wu et al., 2013). Defects of SF2 proteins are associated with predisposition in a variety of diseases like immunodeficiency, mental retardation, premature aging and cancer (Fairman-Williams et al., 2010).

Jankowsky et al. identified 10 distinct protein families in SF2 according to typical sequence, mechanistic, and structural features (E. Jankowsky, 2011). Five of these are RNA helicase families known as DEAD-box, Ski2-like, DEAH/RHA, NS3/NPH-II and RIG-I-like (Fairman-Williams et al., 2010; E. Jankowsky, 2011) (**Figure 1.3**). DEAD-box helicases are relatively conserved from bacteria to human and they form the largest family in SF2. Importantly, the DEAD-box proteins are involved in almost all RNA processes from transcription to RNA decay (Cordin, Banroques, Tanner, & Linder, 2006). Ski2 helicases, which are the core components of the Ski complex display functions in 3' to 5' RNA degradation by the exosome (Khemici & Linder, 2018). DEAH/RHA helicases are associated with mRNA splicing and ribosome biogenesis (Xin, Chen, Rety, Dai, & Xi, 2018). NS3/NPH-II helicases, also termed viral DExH proteins, are made by positive strand RNA viruses (Fairman-Williams & Jankowsky, 2012; Tai, Chi, Chen, & Hwang, 1996). NS3/NPH-II helicases unwind substrates with a 3' overhang and displace RNA or DNA oligonucleotides by processive, unidirectional translocation on the RNA or DNA templates. Last but not the least, RIG-I like family is one of the smallest SF2 families, including Dicer proteins, RIG-I like receptors (RLRs), worm-based Dicer-like RNA helicases (DRHs), and a DNA-repair protein FANCM (Fairman-Williams et al., 2010; Luo, Kohlway, & Pyle, 2013; Pyle, 2008). They confer

similar function as the RNA remodeling proteins in cellular antiviral mechanisms and RNA interference pathways.

Figure 1.3 The SF2 family. (A) Phylogenetic tree of SF2 proteins. The figure is modified from reference (A. Jankowsky, Guenther, & Jankowsky, 2011). The cladogram was generated from a structure-aided sequence alignment of all SF2 proteins. Families containing RNA helicases are marked with a circled R and the proposed roles of the helicase cores are indicated. (B) A helicase motif alignment comparing SF2 RNA helicase proteins. The diagram was modified from the alignment of SF2 proteins according to reference (Fairman-Williams et al., 2010). Amino acids conserved at least 80% of the time are shown as capital letters while those conserved 50%–79% of the time are in lower case; the height of each letter is proportional to its frequency;

1.1.3. Duplex RNA-dependent ATPases (DRAs)

The RIG-I-like/Dicer RNA helicases are also known as duplex RNA-dependent ATPases (DRAs), unlike the classical helicases, they bind to dsRNA and hydrolyze ATP but do not show unwinding ability (Fairman-Williams et al., 2010; Kowalinski et al., 2011; Luo et al., 2013). Apart from the conserved structural features for dsRNA and ATP binding shared in SF2, DRAs contain an insertion domain (Hel2i), which is required for dsRNA binding and the ATPase activities of DRAs.

Figure 1.4 Phylogenetic tree analysis of RIG-I like RNA helicase family. Multiple sequence alignment of DExD/H box RNA helicase domain of RIG-I, MDA5, LGP2, Dicer and *C. elegans* Dicer-related helicases was generated by CLUSTALW(J. D. Thompson, Higgins, & Gibson, 1994). The phylogenetic tree was constructed using the neighbor-joining method within the Mega3.1(Kumar, Tamura, & Nei, 2004).

Phylogenetic tree analysis shows that DRHs are more similar to RLRs than Dicers (**Figure 1.4**). Both RLRs and DRHs have three domains (**Figure 1.5**): the specific N-terminal domain (NTD), the conserved helicase domain and a C-terminal domain (CTD) for dsRNA binding (Kang et al., 2002; Yoneyama & Fujita, 2007; Yoneyama et al., 2005; Yoneyama et al., 2004). Both retinoic acid-inducible gene I (RIG-I) and melanoma differentiation-associated protein 5 (MDA5) have N-terminal tandem CARDs. The CARDs are for signaling, whereas the CTD is for RNA binding. LGP2 (laboratory of genetics and physiology 2) lacks the N-terminal CARDs, thus LGP2 does not achieve signaling (Yoneyama et al., 2005; Yoneyama et al., 2004). The N-terminal domains for DRHs have unknown functions and vary in different worms. The domain architecture for Dicer proteins on the contrary is distinct from RLRs and DRHs (**Figure 1.5**). For example, human Dicer in the order from N' to C'-termini contains a helicase domain, a domain with unknown function (DUF283), a Piwi-Argonaute-Zwille (PAZ) domain, two RNase III domains and a dsRNA binding domain (dsRBD) (Lau et al., 2012).

Figure 1.5 Domain architecture of RIG-I like RNA helicases. Domain boundaries are predicted in the Pfam database(El-Gebali et al., 2019; Finn et al., 2014). Note that the NTDs of DRHs are worm specific novel domains.

1.2. RIG-I like receptors

1.2.1. Antiviral innate immunity

The innate immune system acts in the front line of defense against pathogen invasion (Janeway & Medzhitov, 2002), which is activated by the recognition of the foreign molecules known as pathogen-associated molecular patterns (PAMPs). The PAMPs could be bacterial lipopolysaccharide, viral RNA, cytoplasmic DNA or bacterial flagellin (Iwasaki & Medzhitov, 2015; Roers, Hiller, & Hornung, 2016). The presence of microbial invasion is detected by various germline-encoded pattern recognition receptors (PRRs) (Iwasaki & Medzhitov, 2015; Roers et al., 2016). Detection of PRRs by PAMPs leads to the activation of host innate immune responses and the release of inflammatory cytokines. Furthermore, PRRs also help to induce co-stimulatory molecules, increase antigen-presenting capacity and guide adaptive immune response to antigens. The PRRs are expressed by different cells and can be discriminated by cellular localization, ligand specificity and downstream signaling pathways (Akira, Uematsu, & Takeuchi, 2006).

Viruses, one type of formidable pathogens express their proteins and replicate their genomes by hijacking the host system. Virus-induced signaling results in interferon regulatory factor 3 (IRF3) and IRF7 dependent type-1 interferon (IFN) responses and leads to the NF- κ B-dependent production of pro-inflammatory cytokines (Sen, 2001). Particularly, type-1 IFNs can enhance the maturation of dendritic cells (DCs) and the cytotoxicity of natural killer (NK) cells and the differentiation of virus-specific cytotoxic T lymphocytes, thereby promoting the adaptive immune responses (Honda, Takaoka, & Taniguchi, 2006).

1.2.2. Key pattern recognition receptors in antiviral innate immunity

Antiviral innate immune response largely depends on two types of key PRRs, the Toll-like receptors (TLRs) and the RIG-I like receptors (RLRs). However, it was also shown that NOD-like receptors (NLRs) can sense RNA virus (A. J. Thompson & Locarnini, 2007). TLRs are membrane proteins that act as the primary sensors to recognize specific PAMPs expressed by various pathogens. Among the ten TLRs found in humans, TLR3 and TLR7/8 are involved in viral RNA recognition. The PAMPs for TLR3 and TLR7/8 are dsRNAs and GU-rich single-stranded RNAs (ssRNAs), respectively (Jensen & Thomsen, 2012). Unlike other TLRs, they are located at the intracellular compartments such as endosomes, rather than the extracellular membranes (Jensen & Thomsen, 2012). TLR4 and TLR9 also participate in recognition of some viruses through envelope proteins and CpG sequences in viral DNAs (Jensen & Thomsen, 2012; Martinez-Campos, Burguete-Garcia, & Madrid-Marina, 2017).

Unlike TLRs which are largely dispensable for effective antiviral defense, RLRs were shown to be essential sensors of viral infections (Q. Sun et al., 2006). RLRs recognize specific foreign RNA features within the cytoplasm. RIG-I preferentially binds short dsRNA and 5'-triphosphate/diphosphate (5'-ppp/pp) dsRNA or ssRNA (Cui et al., 2008; C. Lu, Ranjith-Kumar, Hao, Kao, & Li, 2011; C. Lu et al., 2010; Y. Wang et al., 2010). Whereas, MDA5 recognizes long dsRNAs (>2000bp) (Kang et al., 2002). LGP2 senses dsRNA, independent of 5'-ppp (Yoneyama et al., 2005; Yoneyama et al., 2004). Additionally, RIG-I and MDA5 sense different sets of RNA viruses: *Flaviviridae*, *Paramyxoviridae*, *Orthomyxoviridae*, and *Rhabdoviridae* by RIG-I and *Picornaviridae* by MDA5 (Jensen & Thomsen, 2012; Kato et al., 2006). Due to the lack of the CARD

domains, LGP2 works as positive or negative regulator of RIG-I and MDA5 signaling, depending on the virus and the cell type involved (Bruns & Horvath, 2012; Bruns et al., 2013; Rothenfusser et al., 2005; Uchikawa et al., 2016; Venkataraman et al., 2007).

1.2.3. Antiviral signaling pathways of RIG-I and MDA5

Although RIG-I and MDA5 sense distinct RNAs, they use common signaling pathways to respond to various viruses (**Figure 1.6**). RIG-I and MDA5 are activated upon binding to their substrate viral RNAs. The activated RIG-I and MDA5 then assemble their CARD domains into lock-wash-like tetramer/oligomer. RIG-I and MDA5 interact with downstream Mitochondrial antiviral-signaling protein (MAVS) through CARD-CARD interactions (Kawai et al., 2005; Seth, Sun, Ea, & Chen, 2005). The MAVS CARDS propagate from the RIG-I/MDA5 CARD2 (second CARD domain) tetramer side, forming helical filaments composed of CARD domains. This huge complex then recruits various proteins such as tumor necrosis factor (TNF), receptor-associated factors 2 (TRAF-2), receptor-associated factors 6 (TRAF-6) and TNF receptor type-1-associated DEATH domain (TRADD) (Seth et al., 2005). TRAF-2 and TRAF-6 recruit TRAF-3 and TRAF family member associated NFKB activator (TANK) to trigger the activation of Tank binding kinase-1 (TBK1) and I κ B kinase epsilon (IKK epsilon), which then phosphorylate IRF3 (and IRF7). The MAVS-TRADD complex recruits Fas-associated protein with death domain (FADD) and receptor-interacting serine/threonine-protein kinase 1(RIP1) to activate IKKs that leads to the inhibitory phosphorylation of I κ B and release of NF- κ B. Finally, NF- κ B moves to the nucleus and induces the gene expression of its targets. Importantly, IRF3 and NF- κ B are able to cooperatively regulate the expression of the type-1 IFNs. Thereafter, IRF3-mediated

type-1 IFNs are released and *IRF7* gene expression is induced accordingly which enhances antiviral response.

Figure 1.6 Antiviral Signaling Pathways of RIG-I-like Receptors (RLR). RIG-I and MDA5 recognize viral RNA ligands. LGP2 plays a regulatory role in the signaling of RIG-I and MDA5. The activated RIG-I and the MDA5 trigger the CARD-CARD domain interaction with MAVS. MAVS then recruits its interaction partners such as TRAFs. TRAF-3 and TANK activates TBK1 and IKK which activates IRF3 and IRF7. The dimerized IRF3 and 7 then enter the nucleus to drive the expression of type-1 interferon genes. RIP1 and FADD activate I κ B kinase. The activated I κ B kinase phosphorylates the I κ B protein to release NF- κ B. NF- κ B then translocate into the nucleus and drives the expression of pro-inflammatory cytokines genes.

MAVS (known as IPS-1, VISA or Cardif) is a key effector in RIG-I and MDA5 signaling pathway. It contains an N-terminal CARD domain and a C-terminal transmembrane domain. Its N-terminal CARD serves as an adaptor for RIG-I and MDA5 CARD2. Whereas, the C-terminal domain anchors MAVSs to the cytosolic face of the outer mitochondrial membrane. MAVS-deficient mice have been shown to have impaired interferon induction during virus invasion in multiple cell types (Q. Sun et al., 2006). In addition to MAVS, two membrane adaptors, stimulator of interferon genes (STING) (located on the endoplasmic reticulum) and outer membrane 70 (TOM70) (located on the mitochondria), are reported to be required by specific viruses-induced RIG-I signaling pathway (X. Y. Liu, Wei, Shi, Shan, & Wang, 2010; Zevini, Olganier, & Hiscott, 2017).

1.2.4. Structural overview of RLRs

RLRs belong to the RIG-I like family of SF2 helicases, which are multipartite proteins with three domains (**Figure 1.5**): the N-terminal CARDS for signaling, the helicase for RNA-stimulated ATP hydrolysis and the CTD for RNA recognition (Kang et al., 2002; Yoneyama & Fujita, 2007; Yoneyama et al., 2005; Yoneyama et al., 2004). All of the resolved structures for RLRs are summarized in **Table 1.1**.

1.2.4.1. Helicase domain: a dsRNA-dependent ATPase.

The helicase core of RLRs comprises of two conserved Rec-A-like domains (Hel1 and Hel2) and a unique insertion domain (Hel2i) which is composed of mainly helices (Civril et al., 2011; Uchikawa et al., 2016). The Hel1 and Hel2 provide 12 sequential and structural conserved motifs for the binding of RNA and ATP (C. Lu et al., 2011; C.

Lu et al., 2010; B. Wu et al., 2013). Structural and biochemical studies revealed that motifs Q, I, Ia, II and III in Hel1 bind to ATP and the motif VI in Hel2 facilitates ATP binding and hydrolysis (Luo, Kohlway, Vela, & Pyle, 2012). The ATP binding site is conserved in RLRs and is required specifically for their enzymatic functions. Defects in the ATP binding site of RIG-I and MDA5 can cause autoimmune diseases (Funabiki et al., 2014; Hoffmann et al., 2015; Jang et al., 2015; Rice et al., 2014). On the other hand, the motifs Ia-Ic, IIa and IV-V interact with viral RNA backbone (Kowalinski et al., 2011). The helicase domain interacts predominantly with the backbone of RNA via these motifs in a sequence-independent manner (Luo et al., 2011). In addition, the Hel2i is also involved in RNA binding. Hel2i can ‘walk’ along dsRNA like a scanner and this movement of Hel2i may be related to ATP hydrolysis (Kohlway, Luo, Rawling, Ding, & Pyle, 2013).

Table 1.1 Current available structures of RIG-I like receptors.

Protein	PDB	Organism	CARDs	Helicase	CTD	ATP	RNA	Method	Reference
RIG-I	4A2W,	Duck	✓	✓	✓	-	-	X-ray Diffraction	(Kowalinski et al., 2011)
	4A2Q								
	4P4H	Human	✓	-	-	-	-	X-ray Diffraction	(B. Wu et al., 2014)
	4NQK	Human	✓	-	-	-	-	X-ray Diffraction	(Peisley, Wu, Yao, Walz, & Hur, 2013)
	2LWD,	Human	✓	-	-	-	-	Solution NMR	(Ferrage et al., 2012)
	2LWE								
	6BZH	Mouse	✓	-	-	-	-	X-ray Diffraction	(D'Cruz et al., 2018)
	4A2V	Duck	-	-	✓	-	-	X-ray Diffraction	(Kowalinski et al., 2011)
	2RMJ	Human	-	-	✓	-	-	Solution NMR	(Takahasi et al., 2008)
	2QFB,	Human	-	-	✓	-	-	X-ray Diffraction	(Cui et al., 2008)
	2QFD								
	3LRN	Human	-	-	✓	-	5'-ppp 14-mer dsRNA	X-ray Diffraction	(C. Lu et al., 2010)
	3LRR	Human	-	-	✓	-	5'-ppp 12-mer AU-rich dsRNA	X-ray Diffraction	(C. Lu et al., 2010)
	4A2X	Duck	-	-	✓	-	5'-OH 14-mer dsRNA	X-ray Diffraction	(Kowalinski et al., 2011)
	3OG8	Human	-	-	✓	-	5'-OH 14-mer dsRNA	X-ray Diffraction	(C. Lu et al., 2011)
	3NCU	Human	-	-	✓	-	5'-OH 14-mer dsRNA	X-ray Diffraction	(Y. Wang et al., 2010)
	4A2P	Duck	-	✓	-	-	-	X-ray Diffraction	(Kowalinski et al., 2011)
	4ON9	Human	-	✓	-	-	-	X-ray Diffraction	(Deimling, Cui, Lammens, Hopfner, & Witte, 2014)
	3TBK	Mouse	-	✓	-	ADP-PNP	-	X-ray Diffraction	(Civril et al., 2011)
	6GPG	Human	-	✓	-	-	5'-OH 14-mer RNA	X-ray Diffraction	(Lassig et al., 2015)
	4A36,	Duck	-	✓	-	ADP-ALF3-Mg,	5'-OH 19-mer dsRNA	X-ray Diffraction	(Kowalinski et al., 2011)
	4A2P					AMP.PNP			
	5F98	human	-	✓	✓	-	Cap 0 26-mer hairpin RNA	X-ray Diffraction	(Devarkar et al., 2016)
	4AY2	Human	-	✓	✓	ADP-Mg	5'-ppp 20-mer hairpin RNA	X-ray Diffraction	(Luo et al., 2012)
	5F9F	Human	-	✓	✓	-	5'-ppp 24-mer dsRNA	X-ray Diffraction	(Devarkar et al., 2016)
	5F9H	Human	-	✓	✓	-	5'-OH 24-mer dsRNA	X-ray Diffraction	(Devarkar et al., 2016)
	3ZD6,	Human	-	✓	✓	-	5'-OH 12-mer GC-rich dsRNA	X-ray Diffraction	(Kohlway et al., 2013)
	3ZD7								
	4BPB,	Human	-	✓	✓	-	5'-OH 12-mer GC-rich dsRNA	X-ray Diffraction	(Luo et al., 2011)
	2YKH								
	5E3H	Human	-	✓	✓	ADP•BeF3	5'-OH 14-mer RNA	X-ray Diffraction	(Jiang et al., 2011)
MDA5	2RQB	Human	-	-	✓	-	-	Solution NMR	(Takahasi et al., 2009)
	3GA3	Human	-	-	✓	-	-	X-ray Diffraction	(X. Li, Lu, et al., 2009)
	3TS9	Mouse	-	✓	-	-	-	Solution NMR	(Berke & Modis, 2012)
	5JC7	Chicken	-	✓	✓	ADP-ALF4-Mg	5'-p 24-mer dsRNA	X-ray Diffraction	(Uchikawa et al., 2016)
	5JC3,	Chicken	-	✓	✓	ADP-ALF4-Mg	5'-p 10-mer dsRNA	X-ray Diffraction	(Uchikawa et al., 2016)
	5JCF,								
	5JCH								
	4GL2	Human	-	✓	✓	ADPNP	5'-OH-12-mer dsRNA	X-ray Diffraction	(B. Wu et al., 2013)
LGP2	3EQT	Human	-	-	✓	-	5'-OH-8-mer dsRNA	X-ray Diffraction	(X. Li, Ranjith-Kumar, et al., 2009)
	2W4R	Human	-	-	✓	-	-	X-ray Diffraction	(Pippig et al., 2009)
	2RQA	Human	-	-	✓	-	-	Solution NMR	(Takahasi et al., 2009)
	5JB2	Chicken	-	✓	✓	ADP-ALF4-Mg	5'-ppp 26-mer hairpin RNA	X-ray Diffraction	(Uchikawa et al., 2016)
	5JBG	Chicken	-	✓	✓	ADP-ALF4-Mg	5'-ppp 10-mer dsRNA	X-ray Diffraction	(Uchikawa et al., 2016)
	5JBJ	Chicken	-	✓	✓	ADP-ALF4-Mg	5'-p 12-mer dsRNA	X-ray Diffraction	(Uchikawa et al., 2016)
	5JAJ	Chicken	-	✓	✓	ADP-ALF4-Mg	5'-p 10-mer dsRNA	X-ray Diffraction	(Uchikawa et al., 2016)

Hel1 and Hel2 are arranged in parallel with Hel2i located at the other side of Hel2. Furthermore, a V-shaped pincer domain contributed by an α -helical shaft extending from Hel2 links the CTD to Hel1 (Kowalinski et al., 2011; Luo et al., 2011). RNA binding and ATP hydrolysis can induce cooperative conformational changes of RIG-I. In the presence of RNA, the Hel1, Hel2i, Hel2 and CTD form a O-shape, the closed conformation (Civril et al., 2011; Jiang et al., 2011; Peisley et al., 2013). Upon ATP binding, the helicase domain adopts a semi-closed conformation. After ATP hydrolysis, the CTD and Hel2i move apart forming a C-shape open conformation, which leaves an exit for the bound RNA (Civril et al., 2011; Jiang et al., 2011; Kowalinski et al., 2011; Luo et al., 2011).

Notably, the insertion domain also contributes to inter- or intra- protein-protein interactions (Civril et al., 2011; Kowalinski et al., 2011; Z. Liu et al., 2018; Uchikawa et al., 2016; Wilson et al., 2015). Hel2i contributes to RIG-I auto-inhibition by interacting with the top end of CARD2 through electrostatic interactions and hydrophobic interactions (Civril et al., 2011; Kowalinski et al., 2011; Uchikawa et al., 2016). Thus, the CARDS are blocked from tetramer formation and the downstream signaling cannot be triggered. For Dicer, the interactions between Hel2i and RNA-binding proteins such as trans-activating response RNA-binding protein (TRBP), protein activator of the interferon-induced protein kinase (PACT) and adenosine deaminase acting on RNA 1 (ADAR1), involve in substrate selection, cleavage site selection as well as the regulation nuclease activity of Dicer (Z. Liu et al., 2018; Wilson et al., 2015).

1.2.4.2. CTD and viral RNA structure recognition.

In addition to the helicase core, the C-terminal RNA binding domains (CTDs) of RLRs are also essential for viral RNA recognition (Devarkar et al., 2016; Takahasi et al., 2008).

Despite the low sequence similarities within RLR CTDs, the overall structures of CTDs are highly conserved (X. Li, Lu, et al., 2009; X. Li, Ranjith-Kumar, et al., 2009; C. Lu et al., 2011; C. Lu et al., 2010) (**Figure 1.7**). A typical RNA-binding domain of RLR is composed of 10 β -strands (β 1-10), which are connected by loops and small helices: two antiparallel β -sheets (β -sheet1: β 1-2 and β 9-10, β -sheet2: β 5-8) are connected by a β -hairpin formed by strands β 3 and β 4 and a short α -helix α 1; a loop formed by β -strand 10 and a short α -helix α 2 located at the C-terminal end of CTD. Two loops connecting strands β 1-2 and β 6-7 forms a Zinc-finger binding site with four conserved cysteines accommodating a zinc ion. The zinc ion is important in maintaining the overall folding of the CTDs. Mutations of these cysteines abrogate RIG-I signaling *in vivo* (Cui et al., 2008). A saddle-like RNA binding surface which has a high positive charge consists of the β -sheet 2, the β -hairpin, the loops connecting β 5-6 and β 8-9, and the loop after β 10 exhibiting a high degree of shape complementarity to the A-form RNA. It has been shown that mutations of the conserved lysine residues K888 and K907 to glutamates on the surface of the predicted RNA binding surface in human RIG-I (equivalent to K983 & K1002 in human MDA5 and K634 & K653 in human LGP2) disrupts the RNA binding (Cui et al., 2008; X. Li, Lu, et al., 2009; X. Li, Ranjith-Kumar, et al., 2009). In addition, mutations of the conserved histidine residues H927 and H577 to glutamates can also abolish dsRNA binding by MDA5/LGP2 CTD (X. Li, Lu, et al., 2009; X. Li, Ranjith-Kumar, et al., 2009). The positively charged lysine and histidine residues that interact with the negatively charged backbone phosphates of RNA through electrostatic interactions are found in all RLR CTDs. RLR CTDs bind to a variety of

RNA structures in a sequence-independent manner, which allows them to sense a wide spectrum of viruses.

Figure 1.7 A typical RNA-binding domain of RIG-I-like receptors. The figure is based on human RIG-I CTD (PDB code: 5F9F). The antiparallel β -sheet1 (β 1-2 and β 9-10) and the β -sheet 2 β -sheet2 (β 5-8) are colored in light purple and pink, respectively. At the Zinc binding site, the 4 cysteines are shown in stick model in salmon color while the Zinc ion is shown in sphere model in light green color. The KWK RNA binding motif is highlighted in yellow.

In RLR CTDs, the loops connecting strands β 5 and β 6 are flexible and vary in length. The loops from LGP2 and RIG-I are short and contain aromatic hydrophobic residues (X. Li, Lu, et al., 2009; X. Li, Ranjith-Kumar, et al., 2009). These loops cap one side of the RNA-binding sites, making RIG-I and LGP2 CTDs not penetrable for long dsRNAs. In contrast, the loop from MDA5 CTD is long and is able to stack on β -sheet 2, leaving

enough space for dsRNA to pass through (Kang et al., 2002; X. Li, Lu, et al., 2009; Uchikawa et al., 2016). Therefore, LGP2 and RIG-I CTDs preferentially bind to the terminus of RNA while MDA5 CTD binds to the stem of long dsRNAs.

RIG-I CTD specifically recognizes RNA species with 5'-ppp/5'-pp (Cui et al., 2008; C. Lu et al., 2011; C. Lu et al., 2010; Takahasi et al., 2008; Y. Wang et al., 2010). A positively charged lysine-rich patch in RIG-I CTD, which is missing in MDA5 and LGP2, confers 5'-ppp-cap recognition through extensive electrostatic interactions. Four positively charged residues are responsible for the 5'-ppp termini binding: K858 interacts with γ - and probably β -phosphate, K861 interacts with the α - and β -phosphates, and K888 interacts with α -phosphate, facilitated by D872, which forms electrostatic interactions with the side chains of these two residues (C. Lu et al., 2010). The high binding affinity of RIG-I CTD to 5'-ppp termini (0.5 nM for a 5'-ppp 14-mer dsRNA vs 5.9 μ M for 5'-OH 14-mer dsRNA) ensures its specific and efficient binding (Vela, Fedorova, Ding, & Pyle, 2012). LGP2 CTD binds the end region of dsRNA independent of the 5'-ppp termini with a dissociation constant of 105 nM for a 8-mer dsRNA (X. Li, Lu, et al., 2009; Uchikawa et al., 2016). Unlike RIG-I or LGP2, MDA5 CTD recognizes the internal RNA backbone rather than the terminus of the dsRNA with a dissociation constant of 3 μ M for 24-mer dsRNA (X. Li, Lu, et al., 2009; Takahasi et al., 2009; Y. Wu, 2012). The difference in RNA affinity among the CTDs suggests that LGP2 potentially regulates the RIG-I and MDA5 signaling by competing for RNAs (Pippig et al., 2009) .

1.2.4.3. The CARD domain: a signaling motif.

Both RIG-I and MDA5 have tandem CARD domains at the N-terminal, which are responsible for signaling in the antiviral interferon effector pathway (Honda et al., 2006; Yoneyama & Fujita, 2007). CARD is one of the four subfamilies of the Death Domain

(DD) family (Kersse, Verspurten, Vanden Berghe, & Vandenabeele, 2011). CARD-containing proteins are found in fungi, plant, animals and prokaryotes and play an important role in many signaling pathways such as apoptosis, immune responses and cytokine processing. CARDS share a common tertiary structure which is characterized by a Greek-key like folding similar to the immunoglobulin-like domains (Steward, McDowell, & Clarke, 2009). However, CARDS are composed by 6 α -helices, instead of β -strands with a hydrophobic core found in the center of the helices (Bouchier-Hayes & Martin, 2002; Kao et al., 2015).

CARDS usually form oligomers via homotypic interactions such as the Apaf-1 CARD/caspase-9 CARD complex, which is the first reported CARD-CARD structure (Kersse et al., 2011; Qin et al., 1999). The oligomer formation among CARDS mainly depends on the complementary charged surfaces and is independent of the amino acid sequences. There are 16 structures (11 from X-ray diffraction, 5 from solution NMR) of CARD found in Protein Data Bank (PDB) and their amino-acid sequences share only 13-14 % identity (Kao et al., 2015). The large difference in the CARD domain sequences results in divergent interaction interfaces on the solvent-accessible surfaces of the CARDS. CARD domains interact with each other through three types of intermolecular interactions with six surrounding charged surfaces (type Ia/b, IIa/b and IIIa/b). The interfaces for CARDS of RIG-I and MAVS are shown in **Figure 1.8** (Ferraro & Wu, 2012; B. Wu et al., 2013; B. Wu et al., 2014).

Figure 1.8 The type I, II, III interfaces. The figure is derived from Figure 3 in reference (Kao et al., 2015). The interfaces were mapped on RIG-I CARD1 (A), RIG-I CARD2 (B) and MAVS CARD (C) and are colored as indicated below.

Similarly charged patches were also reported for the DD subfamily. The specificity and variety of the interfaces result in the formation of different types of oligomer. Increasingly, CARD-CARD interaction mechanisms have been identified and discussed, including homodimer, heterodimer, oligomer, ubiquitin-dependent oligomer as well as helical filament (Lin, Lo, & Wu, 2010; H. H. Park et al., 2007; Scott et al., 2009; Xiao, Towb, Wasserman, & Sprang, 1999).

1.2.5. Functions and roles of RLRs

1.2.5.1. *RIG-I*

It has been reported that RIG-I can recognize a wide range of viruses, including *Flaviviridae*, *Paramyxoviridae*, *Rhabdoviridae*, *Orthomyxoviridae* and *Arenaviridae* (Jensen & Thomsen, 2012). Rift Valley fever virus (from *Bunyaviridae*) and Ebola virus (from *Filoviridae*) are able to activate the RIG-I-mediated interferon production (Habjan et al., 2008). Some studies also suggested that RIG-I recognize virus-associated polyuridine-rich sequences in the ssRNA and 5'-phosphoryl group (Malathi et al., 2010; Saito, Owen, Jiang, Marcotrigiano, & Gale, 2008). The 5'-phosphoryl group serves as a viral PAMP since it is unlike the 5'-ends of host RNAs, which are either masked by post-translational modification or removed by cleavage (Devarkar et al., 2016). The post-transcriptional cap modifications are not only essential for the stability and translation of the self-mRNA, but also important for distinguishing self- and nonself-RNAs (Daffis et al., 2010). Cellular mRNA is normally capped with a 5' Cap (7mGpppN, Cap 0), which is subjected to 2'-O methylation (7mGpppNm, Cap 1). 2'-O-methylation then works together with m7G cap to reduce the binding affinity of RIG-I to RNA (Devarkar et al., 2016). In the case of self-tRNA, the 5' cleavage, the nucleotides

modification, and the ribosome binding prevent it from being recognized by RIG-I (Daffis et al., 2010; Devarkar et al., 2016).

ATP hydrolysis by RIG-I help remove low-affinity self-RNAs from its binding site on RIG-I, further increase the RIG-I specificity to viral RNA (Lassig et al., 2015; Takahasi et al., 2008). RIG-I mutants with low ATP hydrolysis activity continuously trigger IFN responses. This phenotype can be reverted by introducing mutations that abolish RNA binding (Louber, Brunel, Uchikawa, Cusack, & Gerlier, 2015). To confer selectivity, fast ATP turnover and dissociation of self-RNA provides insufficient time for RIG-I CARDS to interact with downstream signaling partners. Whereas, the tighter binding of foreign RNAs grants adequate time for CARDS to activate the downstream signaling.

The signaling activities of RIG-I are regulated via allosteric conformational changes (Civril et al., 2011; Kowalinski et al., 2011; Uchikawa et al., 2016). The hydrogen/deuterium exchange coupled with mass spectrometry (HDX-MS) experiments shows that allosteric effects in RIG-I are triggered by the sequential addition of RNA and ATP, leading to CARD release and activation (Zheng et al., 2015). In the absence of viral RNA binding, the insertion domain Hel2i binds to CARD2 to prevent the downstream signaling through the CARD-CARD interactions of RIG-I and MAVS. The 5'-ppp viral RNA recognition followed by ATP hydrolysis results in a conformation change in the helicase domain and releases CARD2. The CARDS then form lock-wash-like tetramer (B. Wu et al., 2014). MAVS CARDS nucleate and propagate in the same manner using the tetramer formed by RIG-I CARDS as a scaffold. MAVS filament then activates the cytosolic kinases IKK and TBK1, which in turn activate the type-1 INF and NF- κ B (S. Liu et al., 2013).

RIG-I signaling is also regulated by ubiquitin ligases (S. Liu & Chen, 2011). Ubiquitin ligases such as ring finger protein 125 (RNF125), ring finger protein 135 (Riplet/RNF135) and tripartite motif-containing protein 25 (TRIM25), regulate RIG-I through K63-linked ubiquitin ligation at the CARD domain (Gao et al., 2009; Munir, 2010; Oshiumi et al., 2010). In particular, TRIM25 interacts with the CARDS and assists tetramer formation (D'Cruz et al., 2018). Four K63-linked poly-ubiquitin chains (K63-Ub_n) wrap around four RIG-I CARDS, forming a signaling platform for MAVS CARDS (Peisley et al., 2013; B. Wu et al., 2013). Notably, the residues T170 and K172 in CARD2 are critical for RIG-I ubiquitination (Ferrage et al., 2012). The phosphorylation of residue Thr170 affects the interface between CARD1-CARD2, while the residue K172 is the ubiquitination site.

1.2.5.2. MDA5

Long dsRNA is another type of important PAMP produced by viruses, which can be recognized by MDA5 (Kang et al., 2002). Unlike RIG-I, MDA5 CTD recognizes the internal stem instead of the terminus of the dsRNA (Y. Wu, 2012). Known viral targets of MDA5 include picornaviruses, Sendai virus, WNV, rabies virus, Dengue virus, rotavirus, murine hepatitis virus and murine norovirus (Jensen & Thomsen, 2012). MDA5 can be selectively activated by transfection of polyinosinic-polycytidylic acid (poly (I:C)), which is the mimic of long dsRNA from viruses and widely used for RLR functional studies (Kato et al., 2006).

MDA5 helicase-CTD domains pack along the long dsRNA in head-to-tail manner through protein-protein interactions, resulting in a core filament structure (B. Wu et al., 2014). The unligated MDA5 lacks the Hel2i-CARD2 interaction which contributes to the auto-inhibition mechanism utilized by RIG-I (B. Wu et al., 2014). The tandem CARD domains flexibly hang away from the helicase-CTD core in MDA5 (Zheng et

al., 2015). The core filament formed by the helicase-CTD domains promotes the acquisition of CARDs domains. Therefore, unlike RIG-I activation which is regulated by CARD2 release, MDA5 activation is regulated by the need for cooperative binding to long dsRNA which results in filament formation of CARDs.

Similar to RIG-I, oligomerization of the MDA5 CARDs is required to activate MAVS, which is the downstream signaling effector of RLRs (Y. Wu, 2012). However, noncovalent ubiquitination is dispensable for MDA5 oligomerization. In contrast to RIG-I CARDs, MDA5 CARDs have lower affinity for K63-Ub_n, and form oligomer in a concentration-dependent instead of ubiquitin-dependent manner (B. Wu & Hur, 2015; B. Wu et al., 2013). Additionally, the linker between CARD2 and the helicase domain is ~50 amino acid (aa) long in RIG-I and ~100 aa long in MDA5. Instead of forming a tetramer like RIG-I CARDs, the longer linker allows MDA5 CARDs to form a larger oligomer such as an octamer (B. Wu & Hur, 2015; B. Wu et al., 2013).

1.2.5.3. LGP2

Since LGP2 lacks the tandem CARDs which is the signaling domain, it is unable to associate with MAVS and trigger the downstream activation of the type-1 IFN pathway (Rothenfusser et al., 2005). It is hypothesized that LGP2 is assembled in complex of MDA5-like helicase domain and RIG-I like CTD, acting as a modulator of RIG-I and MDA5 (Rothenfusser et al., 2005; Uchikawa et al., 2016; Yoneyama et al., 2005). LGP2 helicase shares higher amino acid sequence similarity with MDA5 helicase (40 %) than RIG-I (30 %) (Yoneyama et al., 2005). The helicase domain of LGP2 makes MDA5-like filaments (Uchikawa et al., 2016). Although the cooperative binding of LGP2 to dsRNA is not as strong as MDA5, it may facilitate the nucleation of MDA5 oligomerization on dsRNA and thereby enhance MDA5 activation. LGP2 CTD

recognizes terminus of RNA like RIG-I does and competes with RIG-I for RNA substrates, resulting in the inhibition of RIG-I signaling (Pippig et al., 2009).

Overexpression of LGP2 has been found to inhibit RIG-I signaling activated by Sendai virus and Newcastle disease virus (Rothenfusser et al., 2005). LGP2-deficient cells showed increased interferon responses to the infection of vesicular stomatitis virus (a target of RIG-I), but decreased response to the infection of encephalomyocarditis virus (a target of MDA5) (Venkataraman et al., 2007). These findings suggest that LGP2 positively regulates MDA5 but negatively regulates RIG-I (Uchikawa et al., 2016; Venkataraman et al., 2007). LGP2 may inhibit RIG-I signaling via dsRNA competition and enhance the MDA5 signaling by promoting MDA5 helicase assembling on dsRNA (Pippig et al., 2009; Uchikawa et al., 2016). However, titration of LGP2 expression indicated that at high expression level, LGP2 works as a negative regulator for both RIG-I and MDA5 (Bruns & Horvath, 2012; Bruns et al., 2013; Komuro & Horvath, 2006). Immunoprecipitation data revealed that LGP2 might directly interfere with MAVS, regardless of CARDs, thereby affecting the IKKi kinase recruitment (Komuro & Horvath, 2006).

1.3. Dicer-Related Helicases

1.3.1. The discovery of RNA interference

In the early 90s, a series of genetic and molecular biology explorations in plants, fungi and worms demonstrated that the expression of an endogenous gene can be suppressed by a homogenous transgene. This phenomenon often referred to as co-suppression, post-transcriptional gene silencing (PTGS) in plants or RNA interference (RNAi) in nematodes (Cogoni & Macino, 1999; Lee, Feinbaum, & Ambros, 1993; Napoli,

Lemieux, & Jorgensen, 1990; van der Krol, Mur, Beld, Mol, & Stuitje, 1990). Later on, the trigger of RNAi in *C. elegans* was confirmed to be a long dsRNA (Fire et al., 1998). This important discovery brought Andrew Z. Fire and Craig C. Mello the Nobel Prize in Physiology or Medicine in the year 2006. The long dsRNAs are usually derived from foreign or invasive nucleic acids such as transposons, transgenes, and viruses, and are processed into small interfering RNAs (siRNAs) by an endonuclease called Dicer (Ketting et al., 2001; Nakamura et al., 2007). The siRNAs are then incorporated with a multiprotein component complex referred to as RISC (RNA-induced silencing complex) which binds and cleaves the complementary mRNA (Czech & Hannon, 2011; Sijen, Steiner, Thijssen, & Plasterk, 2007). In 2001, Elbashir et al. successfully used synthetic siRNA to induce gene silencing in mammalian cells, providing the application of RNAi in gene regulation (Elbashir et al., 2001).

The process of RNA interference (RNAi) can also be moderated by a novel class of endogenous non-coding small RNAs termed microRNAs (miRNAs). The first reported miRNA is *lin-4* which regulates the developmental timing in *Caenorhabditis elegans* (*C. elegans*) (Lee et al., 1993). Lee et al. found that the *lin-4* locus does not encode a protein, yet short transcripts of *lin-4* interacted with the 3' UTR of *lin-14* mRNA through base-pairing, thereby suppressed the expression of *lin-14*. The second microRNA identified was named *let-7* which regulates *lin-4* and many other temporal control genes in *C. elegans* and was also found to be conserved among bilateral animals (Reinhart et al., 2000). Since then, numerous unique endogenous small RNAs have been reported in many animals and plants (Bartel, 2004). With the increasing knowledge of the fundamental mechanism of endogenous RNAi, RNAi has emerged into a useful tool in genomic analyses as well as gene-specific therapeutics against viral infections and cancer (Bora, Gupta, Mukkur, & Saini, 2012; Leung & Whittaker, 2005).

1.3.2. Biogenesis and classification of small RNAs

Small RNAs play crucial roles in gene expression regulation, chromosome structure modification, as well as antiviral immunity (Fischer, 2015). They guide target transcript selection via partial or perfect sequence complementation (Montgomery, Xu, & Fire, 1998). To date, three groups of small RNAs have been well characterized: miRNAs, siRNAs and Piwi-interacting RNAs (piRNAs).

The biogenesis of miRNA has been well studied (Choudhuri, 2010; Ghildiyal & Zamore, 2009). miRNAs are endogenous small non-coding RNAs which exist in introns of protein-coding genes, 3'-UTR of protein-coding genes, as well as introns and exons of noncoding genes. Most miRNAs are transcribed by RNA pol II or RNA pol III to yield primary miRNA transcripts. Primary miRNA transcripts are several hundred nt-long RNAs and contain one to six hairpins. The hairpins are the precursors of miRNAs. Each hairpin has a 70 bp dsRNA stem, flanked by ~11 nt-long sequences at their stem bases for efficient processing. In the nucleus, the hairpins are recognized by Pasha (or DGCR8) and cleaved by an RNase III-like endonuclease called Drosha. The resulting hairpins which are termed as precursor-miRNAs (pre-miRNAs), have 2-nt 3' overhangs and 5' phosphate groups. Some pri-miRNAs are directly spliced out from introns by spliceosome, bypassing Pasha and Drosha. Pre-miRNAs are then transported to the cytoplasm by exportin-5 and Ran-GTP. In the cytoplasm, the pre-miRNAs are further processed into 22 nt-long RNA duplexes known as mature miRNAs by another RNase III-like endonuclease Dicer. Dicer-produced 22-nt duplexes have 2-nt 3'-overhangs on both strands. The coordination of miRNA and complementary mRNA takes place in an effector complex called RNA-induced silencing complexes (RISCs) (Czech & Hannon, 2011). From a miRNA duplex, only one strand known as the “guide strand” is

incorporated into the RISC and loaded onto mRNA to suppress the translation. The other strand is the “passenger strand” which is eliminated and degraded.

MicroRNAs are made from single-stranded hairpin-shaped RNA precursors, whereas siRNAs derive from long dsRNA. The siRNA class can be further divided into endogenous siRNAs (endo-siRNAs) and exogenous siRNAs (exo-siRNAs) (Elbashir et al., 2001; Watanabe et al., 2008). The exo-siRNAs refer to those virus-produced or synthetic small dsRNA duplexes often used as a tool for genetic studies and RNAi-based gene therapy. Synthetic siRNAs (which are usually 21-bp in length with 2 nt 3' overhang) can trigger potent RNAi responses in human cell culture without inducing IFN response (Elbashir et al., 2001; Watanabe et al., 2008). Endo-siRNAs are derived from double-stranded intrinsic transcripts encoded by the genome. They are either processed by Dicer and or transcribed by RNA-dependent RNA polymerase (RdRP). The advance in small RNA sequencing technology has made it possible for us to characterize endo-siRNAs which are often produced at extremely low level (Ruby et al., 2006). Endo-siRNAs can be divided into several different populations: the 26G siRNAs (5' guanine 26-nt small RNAs), the 22G siRNAs (5' guanine 22-nt small RNAs) and 21U RNAs (5' uracil 21-nt small RNAs which were later confirmed to be the piRNAs of *C. elegans*) (Batista et al., 2008). 26G and 22G siRNAs are endo-siRNAs synthesized by different worm RdRPs (Claycomb et al., 2009; Gu et al., 2009; Vasale et al., 2010). 26G siRNAs are generated by one of worm RdRPs termed RRF-3 with Dicer modification whereas 22G siRNAs are *de novo* transcribed by RdRP EGO-1 or RRF-1 in a Dicer-independent manner. In the 22G siRNA pathway, the formation of functional RdRP module requires Dicer-related helicase-3 (DRH-3, an ortholog of human RIG-I) and enhancer of Ksr-1 Lethality (EKL-1, a dual Tudor domain protein).

The steps involved siRNA precursors processing by Dicer in the cytoplasm and formation of RISC resemble miRNAs of the RNA interference (RNAi) pathway. The siRNA-mediated RNAi pathway will be discussed later (**Section 1.3.3**). In contrast to pre-miRNA duplexes which only form incomplete pairings leading to formation of characteristic bulges, the pre-siRNA duplexes are perfectly paired (Choudhuri, 2010; Ghildiyal & Zamore, 2009). Usually, perfect complementarity between siRNA and mRNA results in mRNA cleavage and imperfect complementarity between miRNA and mRNA (typically the 3'-UTR of mRNA) often causes translational repression. This characteristic is essential for the specificity of different effector pathways.

Piwi-interacting RNAs are distinct from miRNAs and siRNAs with longer sequences (24-30nt) and engage themselves specifically in the silencing of transposons (Choudhuri, 2010). piRNAs are small non-coding RNAs found in Piwi complexes (Fu & Wang, 2014; Vodovar et al., 2012). Piwi proteins belong to a subclass of Argonaute proteins that are unique to the animal kingdom. In the germ cells, the Piwi subfamily of proteins together with piRNAs guard the germline genome by silencing retrotransposons and other repeat elements. Interestingly, piRNAs are produced through a ping-pong mechanism that does not involve Dicer (Brennecke et al., 2007). Majority of piRNA genes exist in the form of monodirectional clusters and the precursor piRNA transcripts are ssRNAs. In *Drosophila*, two Piwi proteins (Piwi and Aub) split the precursor sense piRNA transcripts into sense piRNAs. The sense piRNAs then associate with the third Piwi protein, Ago3, forming the Ago3 sense piRNA complex which directly cuts the antisense transposon transcripts and creates antisense piRNAs. Antisense piRNAs in turn bind to Piwi/Aub, forming the Piwi/Aub antisense piRNA complex which guides the cleavage of more sense transposon transcripts and further generates sense piRNAs. The above self-amplifying loop of piRNA is known as the ping-pong mechanism. A

typical piRNA usually has a 5' uracil and a 2'-O-methylated 3' end. Deep sequencing reveals that the 21U RNAs are the piRNAs of *C. elegans* (Batista et al., 2008). Notably, the siRNAi, miRNA and piRNA pathways were not distinct from each other, instead, they compete for and share substrates, effector proteins and cross-regulating each other.

1.3.3. Mechanism of RNAi

A typical RNAi-mediated regulatory event involves several consecutive enzymatic activities: biogenesis of siRNAs and miRNAs, guide strand separation and gene suppression by RNA-induced silencing complex (RISC). Major proteins involved in these pathways are Dicers which are involved in premature small RNA generation, RdRPs which are involved in secondary small RNA transcription, and Argonaute proteins which are the catalytic center of RISCs.

In an RNA response, miRNA or siRNA precursors, in the form of dsRNA, are first processed into short (21-23nt) dsRNAs bearing 5'-p and 3' overhang by Dicer in the cytoplasm (Chapman & Carrington, 2007; Czech & Hannon, 2011; Ghildiyal & Zamore, 2009). Some organisms have more than one Dicer. For example, the *Drosophila* genome encodes two different Dicers. Each of the Dicer proteins in *Drosophila* is responsible for the biogenesis of specific types of small RNAs (Kurzynska-Kokorniak et al., 2015; Song & Rossi, 2017). Mammals and worms, however, produce only one Dicer which requires specific protein cofactors and effectors to generate diverse small RNA species (Ketting et al., 2001; Lee et al., 1993; H. Li, Li, & Ding, 2002; Z. Liu et al., 2018; Wilson et al., 2015). Since worm Dicer is required for the biogenesis of both siRNAs and miRNAs which are essential for basic cellular function and development, loss of worm DCR-1 function leads to lethality and sterility due to abnormal germline development and heterochrony defects (Ketting et al., 2001; Knight & Bass, 2001).

Worms use two different mechanisms for siRNA biogenesis. In addition to DCR-1, RdRPs can produce siRNAs using mRNA-derived templates (Aoki, Moriguchi, Yoshioka, Okawa, & Tabara, 2007; Sijen et al., 2001; Smardon et al., 2000). These siRNAs exhibit a distinct 5' to 3' polarity and bear 5'-ppp ends which are not directly derived from the template dsRNA. In general, small RNAs generated by DCR-1 are called primary siRNAs whereas those produced by RdRPs are named secondary siRNAs. Some small RNA produced by RdRPs are triggered by Dicer-dependent small RNA molecules while some RdRPs are Dicer-independent. Interestingly, some of the RdRP products such as 26G siRNA generated by RRF-3 may require processing by Dicer. It was believed that secondary siRNAs work to amplify the silencing signal, as the secondary siRNAs are far more effective than primary siRNAs in the interference response (Sijen et al., 2001; Sijen et al., 2007).

Different types of small RNAs at different stages are then processed and loaded onto Argonaute proteins, the functional core protein of RISC which can perform a variety of functions (Hutvagner & Simard, 2008). Argonaute proteins are classified into three subgroups according to their substrate specificity and sequence similarity: the AGO subfamily which associates with Dicer processed dsRNA; the Piwi subfamily which loads ssRNA precursors (piRNAs), and the worm-specific WAGO subfamily which recruits RdRP synthesized RNAs (Hutvagner & Simard, 2008; Yigit et al., 2006). A typical Argonaute protein contains an N' domain, a PAZ domain, a MID domain and a Piwi domain (Willkomm, Zander, Gust, & Grohmann, 2015). Structural studies have revealed that the PAZ domain recognizes the 3' end of dsRNA, the MID domain confers the 5' nucleotide specificity, and the Piwi domain resembles an endonuclease. Not all Argonaute proteins have the slicer activity, and some Argonaute proteins only have cleavage function when coupled with specific guide RNA (M. S. Park et al., 2017).

Argonaute proteins also contribute to strand selection (Tahbaz et al., 2004). Only one strand of the siRNA is required for RNAi, while the other strand is removed during RISC assembly, either by cleavage or by unwinding. The remaining and removed strands are named guide strand and passenger strand respectively. Passenger strand displacement largely depends on the thermodynamic properties on the sequence difference between two strands such as G-C contents and base-mismatches.

Argonaute proteins silence gene expression through either slicer activity or non-cleavage translational repression (Hutvagner & Simard, 2008; M. S. Park et al., 2017). Guide strands with perfect complementary sequence to their target mRNAs mediate target transcript cleavage. On-cleavage translational repression is mediated by imperfect sequence matching which may occur at initiation or elongation stage where the canonical cap or poly (A) tail is targeted and the ribosome entry site is blocked (B. Wang, Love, Call, Doench, & Novina, 2006).

1.3.4. RNA-directed viral immunity

RNA interference (RNAi) serves as an antiviral mechanism in many eukaryotic hosts, ranging from plants to animals (Ding, 2010; Hamilton & Baulcombe, 1999; H. Li et al., 2002; Parameswaran et al., 2010). Virus RNAs can potentially trigger the host RNA interference or RNA-related silencing pathway. Notably, the antiviral RNAi pathway share key components with canonical RNAi pathways in many hosts (Gu et al., 2009; R. Lu, Yigit, Li, & Ding, 2009).

In the antiviral RNAi pathway (**Figure 1.7**), Dicer protein and its cofactors first recognize and cleave viral dsRNAs into virus-derived small siRNAs (vsiRNAs) with 5'-monophosphate group (5'-p) and 3'-hydroxyl group. In *C. elegans*, DCR-1 recruit its cofactors to form a Viral Recognition Complex (ViRC) in order to cleave the viral

replication intermediate, in the form of dsRNA, into 23 nt vsiRNA duplexes with a 2 nt 3' overhang (Tabara, Yigit, Siomi, & Mello, 2002). Here, ViRC plays a role in viral RNA sensing which parallels to the viral RNA recognition by RLRs in mammalian innate immunity. The Dicer-produced vsiRNA are then incorporated into an Argonaute protein termed RDE-1. Subsequently, the passenger strand is removed, and the guide strand is loaded to RDE-1 to form an active RISC. Amplification of RNAi response is achieved by RRF-1 and its cofactors via the synthesis of secondary vsiRNAs using viral genomic transcript as template. The RdRP-synthesized secondary vsiRNAs are 22nt single-stranded vsiRNAs carrying a 5'-ppp cap. After being loaded into secondary Argonaute protein, SAGO-2, secondary vsiRNAs then guide specific viral clearance by mediating the cleavage of viral transcripts or inhibiting virus replication (Snead & Rossi, 2010; Vasale et al., 2010). RNAi-based antiviral immunity in worm can spread non-cell-autonomously unlike the adaptive immunity in mammals which involve many immune effectors cells.

Figure 1.9 Dicer Related Helicases (DRHs) are involved in the antiviral RNA interference (RNAi) pathway in *C. elegans*. Upon viral infection, DCR-1 recruits its partner RDE-4 and DRH-1 to the viral dsRNA replication intermediate. DCR-1 cleaves the viral dsRNA into 23 nt dsRNA with a 2 nt 3' overhang. The DCR-1 produced viral siRNAs are then incorporated into the Argonaute protein RDE-1 and one strand is lost to give rise to activated RISC. The primary viral siRNAs and RDE-1 recruit an RdRP complex to the viral genome to synthesize secondary viral siRNAs which act to silence viral transcripts or inhibit virus replication. The antiviral RNAi pathway is dependent on the secondary Argonaute protein SAGO-2.

1.3.5. DRHs are worm-homolog of RLRs.

The worm specific Dicer-related helicase genes, *drh-1* and *drh-3* (and a pseudogene, *drh-2*) are named after the worm Dicer DCR-1 as each of them has a dicer-like DExD/H box helicase domain. Both DRHs and Dicer are found to contribute to RNAi and are involved in siRNA biogenesis. Interestingly, DRHs are thought to be the worm orthologs of RLRs, which are intracellular surveillance proteins for the detection of pathogenic RNAs in innate immunity, based on the observations that they have similar domain architecture: A N-terminal domain, a conserved helicase domain and a C-terminal RNA binding domain.

1.3.6. DRH-1 and worm antiviral immunity

DRH-1 (Dicer-related RNA helicase 1) was first reported as an important factor in antiviral RNAi pathway (Ashe et al., 2013; R. Lu et al., 2009; Tabara et al., 2002). Loss of function mutations in *drh-1* gene significantly affect the production of viral-derived small RNAs (Ashe et al., 2013). Although DRH-1 is critical for antiviral RNAi, it appears to be dispensable for exogenous RNAi triggered by artificial dsRNA (Guo, Zhang, Wang, Ding, & Lu, 2013; R. Lu et al., 2009). DRH-1 shares a high homology with RIG-I which is a cytosolic viral RNA sensor involved in vertebrate innate immune response. Interestingly, DRH-1 CTD in *C. elegans* can be replaced by the RIG-I CTD without affecting the antiviral RNAi function, suggesting that DRH-1 may play a similar role as RIG-I does in sensing viral dsRNA (Guo et al., 2013). In *C. elegans*, DRH-1 is engaged in physical interactions with DCR-1 (RNase III enzyme) and RDE-4 (dsRNA binding protein) (Duchaine et al., 2006; Tabara et al., 2002). DRH-1 is important for antiviral RNAi but dispensable for worm viability and may play an important role in facilitating the translocation of the Dicer complex along viral dsRNA (Coffman et al.,

2017). In contrast to the random distribution of vsiRNAs produced by wildtype DRH-1, DRH-1 mutants lead to the accumulation of highly abundant 23-nt vsiRNAs derived from the terminal regions of viral genome. However, to date, the molecular mechanism of DRH-1 in virus detection is still unclear.

1.3.7. DRH-3 and RdRP-dependent secondary small RNA synthesis

DRH-3 is an essential component of RdRP modules required for biogenesis of secondary small RNAs, including endogenous 22G and 26G siRNAs and exogenous secondary vsiRNAs (Aoki et al., 2007; Gu et al., 2009; Vasale et al., 2010). DRH-3 is also important for germline development, RNA-directed viral immunity and chromosome segregation (Claycomb et al., 2009; Duchaine et al., 2006; Gu et al., 2009; Nakamura et al., 2007; Vasale et al., 2010). DRH-3 was first found as a Dicer interacting protein (Duchaine et al., 2006). Aoki and colleagues reported immunoprecipitation data showing that DRH-3 interacts with RRF-1 (a worm RdRP) but not DCR-1, indicating DRH-3 could be involved in the RdRP-modules (Aoki et al., 2007). The *drh-3* mutants are sterile and exhibit defects in germline proliferation and gametogenesis. Deep sequencing analysis suggests that DRH-3 probably acts to promote sequential initiation of polymerization, relax template secondary structure, assist RdRP translocation along mRNA and promote secondary siRNA dissociation from the template. Biochemical and biophysical assays revealed that DRH-3 prefers longer dsRNA (at least 22bp) for its optimal enzymatic function. DRH-3 mutagenesis studies further suggested that the DRH-3 helicase domain is essential for RNA binding and ATP hydrolysis (Fitzgerald, Vela, & Pyle, 2014).

1.4. Aims of this thesis

RNAi refers to a complex process exploiting short dsRNA to guide sequence-specific gene silencing events at both transcriptional and post-transcriptional levels. In addition to several key enzymes such as Dicers, RdRPs and Argonaute proteins, many cofactors involved in distinct RNAi pathways have been identified. However, the molecular details and the specific functions for these cofactors have not been studied in depth. DRHs are vital proteins involved in the biogenesis of siRNAs. DRH-1 requires worm Dicer (DCR-1) as cofactor and is specifically involved in the antiviral RNAi pathways. DRH-3 which interacts with RdRPs, is involved in all of the secondary siRNA production events. The mechanism by which DRHs contribute to antiviral RNAi remains unclear. Although structures of RLRs and Dicer has been comprehensively studied, the structures of worm DRHs are still unknown. Study of the molecular details for DRHs would allow for a better understanding of their functional roles in RNAi. DRHs share similar domain structure with RIG-I like virus sensors but mediates interferon-independent antiviral immunity. This observation suggests that in-depth study of function and mechanism of worm DRHs would provide a solid foundation to understand the difference and the similarity of the RNA-dependent antiviral mechanisms in diverse organisms.

The aim of this study is to investigate the roles of the *C. elegans* DRHs by using crystallographic, biochemical and biophysical methods. First, the preparation of active recombinant DRHs proteins and their domains was optimized. Secondly, purified domains were functionally characterized with the emphasis being placed on the examination of the ATPase activity in the helicase domain and the RNA preference of the C-terminal RNA binding domain. Furthermore, conformational changes induced by

RNA-binding were studied. Findings from this study help clarify the roles and modes of action for worm DRHs in diverse RNAi pathways.

Chapter 2: Materials and methods

2.1. Protein expression

2.1.1. Recombinant full-length DRH-1 and DRH-3

The full-length target genes encoding *C. elegans* protein DRH-1 (UniProt G5EDI8) and DRH-3 (UniProt Q93413) were cloned into pET-SUMO bacterial expression vector created previously by Dr. Dahai Luo. The pET-SUMO expression system utilizes a small ubiquitin-related modifier (SUMO) fusion to enhance the soluble expression of fusion proteins. The hexa-histidine tag (His₆-tag) was inserted at the N-terminal of SUMO-tag (His₆-SUMO-tag) to aid in downstream purification. The His₆-SUMO-tag can be cleaved by the highly specific and active SUMO protease at the carboxyl terminal, producing a native protein.

2.1.2. High-throughput screening of protein overexpression in *E. coli*

The plasmids with full-length gene were submitted to the protein production platform (PPP) at SBS/NTU to generate multiple constructs for each domain. The truncation sites were designed based on the predicted functional domains of the protein and PPP performed all the cloning work, as well as a small-scale analytical expression screen to assess the likelihood of successful large-scale protein expression and purification. The expression vector used was pNIC28-Bsa4 (GenBank accession EF198106). Target genes were fused with a Tobacco Etch Virus (TEV) protease cleavage site for optional tag removal. Other vectors such as pNIC-CH2 (C-terminal His₆-tag without cleavage site) and pNIC-GST (separately cleavable GST fusion in front of cleavable His₆-tag) were also used as an alternative to optimize protein overexpression and solubility. Only

constructs with good overexpression and solubility are selected for large scale expression and purification.

2.1.3. Expression of recombinant proteins

The recombinant constructs were used for transformation of chemically competent *E. coli* BL21(DE3) Rosetta II or Rosetta T1R cell line (Novagen, Madison, USA). Colonies of the transformed cells harboring the target gene constructs were picked from Luria-Bertani (LB) agar plates containing 35 µg/ml chloramphenicol and 50 µg/ml kanamycin and used to inoculate 20 ml LB Broth containing 35 µg/ml chloramphenicol and 50 µg/ml kanamycin. Seed cultures were grown overnight at 37 °C with agitation (200 rpm). Ten milliliters of the overnight seed culture were used to inoculate one-liter (1:100 dilution) of LB broth media supplemented with 2.5 % glycerol, 50 mM phosphate buffer pH7.4, 35 µg/ml chloramphenicol and 50 µg/ml kanamycin. The inoculated culture was grown to OD₆₀₀ 0.6 at 37 °C, 200rpm, before cooling down to 18 °C. Expression was induced at 18 °C by adding Isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM and growth was maintained overnight (about 20 hours). The cells were then pelleted down at 4000 g in a JLA-9,1000 rotor for 20 minutes at 4 °C. Cell pellets (around 8-12 g from one-liter culture) were quickly flash frozen in liquid nitrogen and stored at -80 °C.

2.1.4. Selenomethionine incorporation protein expression

The protocol for protein expression with SelenoMethionine (SeMet) incorporation is slightly different from the normal cell culture methods. Plasmid pNIC28-Bsa4 harboring DRH-3NTD was introduced into B834(DE3) competent cells. Colonies were selected by LB agar plates containing 35 µg/ml chloramphenicol and 50 µg/ml kanamycin. Positive colonies harboring the pNIC28-Bsa4 DRH-3NTD plasmid were cultivated in

20 ml L-methionine medium (SelenoMethionine Medium Complete, md12-500) overnight. The cells were then pelleted, washed with sterile water before being resuspended into a liter of pre-warmed (37 °C) minimal medium containing L-SelenoMethionine (SelenoMethionine Medium Complete, md12-500). Subsequently, the cells were cultured at 37 °C till OD₆₀₀ reached 0.6 and the expression was induced by 0.5 mM IPTG at 18 °C. Cells were pelleted down after growing at 18 °C overnight (approximately 20 hours) and flash frozen in liquid nitrogen and stored at -80 °C.

2.2. Protein purification

Target protein purification is typically carried out in two steps, involving an initial affinity chromatography (the Immobilized Metal Ion Affinity Chromatography (IMAC)) followed by a size exclusion chromatography (SEC or ‘gel filtration’). Additional heparin chromatography or an extensive high salt buffer washing step during IMAC purification was used for certain constructs to increase protein purity. All purification steps were performed on ÄKTA pure FPLC system (GE Healthcare) at 4 °C unless stated otherwise.

2.2.1. Preparation of cell lysates

The frozen cells were thawed on ice and resuspended in 10 ml per 1g ice-cold Lysis Buffer (50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 8.0, 500 mM sodium chloride (NaCl), 10 % glycerol, and 5 mM β -mercaptoethanol (β -ME)). Cells were lysed using homogenizer (GEA Niro Soavi) at 900 bar over 3 cycles on ice. Cell debris was removed by centrifugation at 20000 rpm in a JA-25.50 rotor for 45 min at 4 °C. The supernatant was filtered through a 0.22 μ M syringe filter and kept on ice for the purification procedure.

2.2.2. Immobilized metal ion affinity chromatography (IMAC) purification

The Ni-NTA agarose beads were pre-equilibrated with lysis buffer. Two milliliter of pre-equilibrated Ni-NTA beads were added into 100 ml supernatant containing His₆-tagged protein and the mixture was incubated on roller shaker for 0.5 to 2 hours at 4 °C. The beads were loaded to Econo-Pac column (Biorad) by gravity-flow and washed with 10 column volumes of lysis buffer with 20 mM imidazole, followed by 10 column volumes of lysis buffer with 1M NaCl and 2 column volumes of lysis buffer with 40 mM imidazole. Target protein was eluted with 5 column volumes of lysis buffer with 250 mM imidazole. Protein concentration was determined using NanoDrop® spectrophotometer (Thermo Fisher Scientific, USA). Protein concentration was determined by the measurement of the absorbance at 280 nm, using the molar absorption coefficients and the Beer-Lambert law: $A = \epsilon \times c \times l$, where A is the absorbance at a particular wavelength, ϵ is the absorption coefficient, and l is the path length of the cuvette (or cell) which holds the solution of concentration. Protein purity was validated by 10-15 % sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

2.2.3. Optional Heparin column purification

Eluted protein solution was diluted 4 folds in Buffer A (25 mM HEPES pH 8.0, 10 % glycerol, 5 mM β -ME) to reduce the salt and imidazole concentration. Diluted proteins were then loaded onto a 5 ml HiTrap Heparin HP (GE Healthcare) that was pre-equilibrated on an ÄKTA Pure purification system (GE Healthcare). The protein of interest was eluted with Buffer B (25 mM HEPES pH 8.0, 10 % glycerol, 2 M NaCl, 5 mM BME) with a linear salt gradient. Protein concentration was determined by NanoDrop® spectrophotometer (Thermo Fisher Scientific, USA) and protein purity was

validated by SDS-PAGE. Fractions containing the protein of interest were pooled and concentrated to less than 5 ml using Vivaspin® Turbo 15 ultracentrifuge unit of suitable molecular weight cut-off ready for gel filtration chromatography.

2.2.4. Optional tag cleavage

Protein was dialyzed, overnight at 4 °C, with dialysis buffer (50 mM HEPES pH 8.0, 150 mM-300 mM NaCl, 10 % glycerol, and 5 mM β -ME) containing TEV protease (1 mg of protease for every 40 mg of protein) or SUMO protease (also known as Ulp1 protease) (1 mg of protease for every 100 mg of protein). Reverse IMAC for His₆-tagged protein was conducted to remove un-cleaved proteins and cleaved His₆-tags.

2.2.5. Gel filtration chromatography

Fractions containing the protein of interest were loaded onto a HiLoad 16/600 Superdex 200 pg column (or HiLoad 16/600 Superdex 75 pg for proteins with molecular weight lower than 30 kDa) that has been pre-equilibrated in gel filtration buffer [25 mM HEPES, 150 mM NaCl, 5 % glycerol, 2 mM Dithiothreitol (DTT)]. Fractions containing the protein of interest were pooled and concentrated to approximately 15 mg/ml using Vivaspin® Turbo 15 ultracentrifuge unit with suitable molecular weight cut-off. Purified proteins were analysed by SDS-PAGE and then quantified using NanoDrop® spectrophotometer (Thermo Fisher Scientific, USA). The purified protein was flash frozen and stored in small aliquot in -80 °C freezer.

2.3. RNA Preparation

2.3.1. Synthetic RNAs

The 5'-hydroxy (5'-OH) and 5'- monophosphate (5'-p) RNAs were purchased from IDT. The palindromic 5'-triphosphate (5'-ppp) dsRNA oligonucleotides were prepared by our collaborator Dr. Olga Fedorova. As described previously (Panjekar, Parthasarathy, Lamzin, Weiss, & Tucker, 2009), the 5'-ppp 12-mer dsRNAs were synthesized on a MerMade 12 DNA-RNA synthesizer (BioAutomation). Synthesized 5'-ppp 12-mer RNAs were deprotected and gel purified. The synthetic RNAs were further analysed for purity by mass spectrometry (Novatia LLC).

2.3.2. *In vitro* transcription

The 5'-ppp RNA oligonucleotides were also prepared by *in vitro* transcription (IVT) which allows for RNA synthesis in large scales (μg to mg quantities). When the DNA template is incubated in the presence of T7 RNA polymerase and rNTPs, a transcript is made as indicated with a triphosphate at the 5'-end. RNA transcripts obtained with the following method is mainly used for structural studies.

Figure 2.1 (A) Consensus sequence of class III T7 RNA polymerase promoter (Bold; corresponds to promoter sequence). (B) DNA template (Underline; corresponds to the nucleotides in the transcript). (C) RNA transcript with a triphosphate at the 5'-end.

Prepare 100 mM rNTP stock.

Nucleotide triphosphates were purchased from Hongene. Each rNTP was dissolved in RNase free sterile water to a concentration of 100 mM. The pH of rNTP stocks was quickly adjusted to 7-7.5 by adding 5M potassium hydroxide. The rNTP stocks were aliquoted into 0.5 ml per tube and stored in -20 °C.

Prepare T7 RNA polymerase.

The gene encoding bacteriophage T7 RNA polymerase (GenBank: V01146.1) was cloned into pNIC-28 vector. The T7 RNA polymerase was purified in two steps involving an IMAC followed by a gel filtration chromatography. The purified T7 RNA polymerase was stored in 5x IVT reaction buffer (300 mM HEPES pH 7.8 (by KOH), 10 mM spermidine, 0.05 % Triton X-100, 75 mM DTT) at -80 °C.

Prepare DNA template for IVT.

The complementary DNA oligonucleotides with T7 promoter (5'-GTAATACGACTCACTATA.....) and the corresponding RNA sequence were chemically synthesized by IDT. Two dried complementary DNA were resuspended into 200 µM stocks in autoclaved water. Templets were prepared by heating the two DNA strands to 94 °C for 5 minutes and then cooling the DNA down to room temperature.

In vitro transcription of short double-stranded RNA.

The transcription reaction was set up by adding the components in order shown below at room temperature.

Transcription recipe:

Stocks	Final conc.
10X Transcription buffer	1X
1 M DTT	5 mM
1 M MgCl ₂ **	22 mM
GTP (100 mM)	6 mM
CTP (100 mM)	4 mM
ATP (100 mM)	4 mM
UTP (100 mM)	4 mM
T7 pol (5 mg/ml)	2-2.5 μ M
DNA template	1 μ M

The mixture was incubated at 37 °C overnight. *In vitro* transcription of long double-stranded RNA (>60bp) was used MEGAscript T7 Transcription Kit (Thermo Fisher Scientific).

2.3.3. RNA purification

Phenol/Chloroform extraction of nucleic acids.

A phenol/chloroform extraction is a liquid-liquid extraction which separate nucleic acids from proteins. It is typically easy to carry the extraction out in 1.7–2 ml Eppendorf tubes. A step-by-step protocol for Phenol/Chloroform extraction was shown below:

1. Add samples into tubes such that no more than 700 μ L per tube.

2. Mix Phenol: Chloroform: Isoamyl Alcohol 25:24:1 (Sigma) stock vigorously to mix the phases before adding to samples.
3. Add an equal volume of Phenol/Chloroform to the tube, vortex vigorously to mix the phases.
4. Spin in a microfuge at top speed for 1–2 min to separate the phases.
5. Transfer the aqueous phase to a new tube, being careful not to transfer any of the protein at the phase interface.
6. Repeat the phenol extraction one more time.

Ethanol precipitation of RNA.

RNA in aqueous solution was concentrated by adding 3 volume ethanol and 0.1 volume 3M sodium acetate, pH 5.3. After being kept at -80 °C overnight, the RNA samples were spun at 20,000 ×g at 4 °C for 1 hour, the collected RNA was washed with pre-chilled 80 % ethanol to remove remaining salts. After drying at 55 °C, the RNA was dissolved in an appropriate volume of ME50 buffer (10 mM MOPS pH7.0, 10 mM NaCl, 1 mM EDTA).

Purification of transcripts by Ion exchange column.

Extracted RNA was purified using an AKTA pure FPLC system equipped with one 5 ml HiTrap Q FF column (GE Healthcare). The Q columns were equilibrated with buffer A (25 mM HEPES pH7.0, 5 % glycerol). The RNA was loaded to the pre-equilibrated Q column and eluted with Buffer B (25 mM HEPES pH 8.0, 5 % glycerol, 2 M NaCl) with a linear salt gradient. Fractions are analyzed using 20 % denaturing PAGE gel with 10 µl loaded for each fraction.

Gel purification of transcripts.

For the RNA samples that have alternative conformations, a denaturing polyacrylamide gel was used. In most of cases, a 20 % polyacrylamide gel is used. The RNAs were visualized by UV₂₅₄-shadowing over Xerox paper. The full-length transcript, usually the one on the top, was cut out and chopped into fine particles using a scalpel. Every 0.5 ml of gel slab was eluted with 0.3 M sodium acetate pH 5.2 followed by ethanol precipitation.

2.4. Biochemical and biophysical assays

2.4.1. NADH coupled ATPase assay

DRH-1 ATPase activity was measured using an established absorbance-based coupled assay system (**Figure 2.2**). In principle, DRH-1 works as an ATPase which hydrolyzes ATP into ADP. One molecule of hydrolyzed ATP was regenerated by pyruvate kinase which converts one phosphoenolpyruvate to pyruvate. The pyruvate is subsequently converted to lactate by lactate dehydrogenase resulting in the oxidation of one nicotinamide adenine dinucleotide (reduced form) (NADH) molecule. Therefore, the rate of steady-state ATP hydrolysis is proportional to the oxidation of NADH which can be measured by absorbance decrease at 340 nm. Providing sufficient pyruvate kinases, PEP, LDH and NADH, the constant regeneration of ATP allows monitoring the ATP hydrolysis rate over the entire course of the assay.

Figure 2.2 Schematic diagram for NADH coupled ATPase assay.

The DRH-1 protein of interest was diluted into ATPase assay buffer (25 mM 3-(N-morpholino) propanesulfonic acid (MOPS) pH 7.4, 150 mM NaCl, 5 mM DTT) to a final concentration of 20 nM for $K_{ATP, RNA}$ experiments in the presence of a coupled assay mix (1 mM NADH, 100 U/ml lactic dehydrogenase, 500 U/ml pyruvate kinase, 2.5 mM phosphoenol pyruvic acid).

For $K_{m,ATP}$ experiments, saturating amount of RNA was added at the concentration of 1 μ M and incubated with the protein for 2 hours prior to the initiation of the reaction. The reaction was initiated by the addition of 1 to 1 molar ratio of ATP and $MgCl_2$ diluted to 8 different concentrations ranging from 5 μ M to 5 mM.

For $K_{m,RNA}$ experiments, the RNA of interest was diluted into buffer (25 mM 2-(N-morpholino) ethanesulfonic acid (MES) pH 6.0, 0.5 mM EDTA) over a 12-point concentration series and added to the protein/NADH sample mix resulting in RNA concentrations varying from approximately 0.5–500 nM. Proteins were incubated with RNA for at least 2 hours at room temperature. The reaction was initiated by the addition of 5 mM ATP/5 mM $MgCl_2$ to all wells.

The rate of ATP hydrolysis was determined indirectly by monitoring the conversion of NADH to NAD⁺ which results in a loss of sample absorbance at 340 nM. The assay was performed in 96-well format and absorbance was measured over a 10 min time course using a Biotek Cytation 3 Plate Reader.

The initial velocity was determined based on the ATPase activity measured as described by Kiiianitsa et al (Kiiianitsa, Solinger, & Heyer, 2003). The path length for the reaction volume was calculated using Beer-Lambert law after measuring a known concentration of NADH in the same assay volume. Initial velocities were calculated using the following equation:

$$Initial\ velocity = -\frac{dA_{340}}{dt} \times K^{-1}_{path} \quad \text{Equation 1}$$

where K^{-1}_{path} is the molar absorption coefficient for NADH at a given optical length.

The initial velocities at various RNA concentrations and ATP concentrations were plotted using Michaelis-Menten equation with GraphPad Prism® version 6 program (GraphPad Software, Inc.) (**Figure 2.3**).

k_{cat} can be fit instead of Vmax when analyzing a substrate vs. velocity curve:

$$y = \frac{Et \times k_{cat} \times x}{K_M + x} \quad \text{Equation 2}$$

where y = the enzyme velocity, x = the total concentration of RNA (nM) or ATP (mM), Et = the concentration of enzyme catalytic sites, k_{cat} is the turnover number and K_m is the Michaelis constant for the variable substrate.

Figure 2.3 The Michaelis-Menten model. X is the substrate concentration. Y is enzyme velocity. k_{cat} is the turnover number, the number of times each enzyme site converts substrate to product per unit time. K_m is the Michaelis-Menten constant, in the same units as X. It is the substrate concentration needed to achieve a half-maximum enzyme velocity. Et is the concentration of enzyme catalytic sites. Vmax is the maximum enzyme velocity in the same units as Y. It is not directly shown in the model above. It is computed by multiplying Et times k_{cat} . The figure is adapted from GraphPad Curve Fitting Guide (https://www.graphpad.com/guides/prism/7/curve-fitting/index.htm?reg_kcat.htm).

2.4.2. Thermal shift assay

The ThermoFluor assay is a quick, temperature-based assay to assess the stability of proteins. The SYPRO Orange dye (Invitrogen) binds to hydrophobic denatured protein and fluoresces. This assay measures the changes in the fluorescence signal of SYPRO Orange dye (Invitrogen) as it interacts with a protein undergoing thermal unfolding. Each 50 μ l sample contains 1x thermal shift reaction buffer (25 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 10 % glycerol), 5x SYPRO Orange dye and 5 μ M unligated protein or RNA-protein complex. Samples were load into one well in a 96-well PCR plate (BioRad), sealed with optical quality sealing tape (BioRad). The plate was heated in an i-Cycler iQ5 real-time PCR detection system (BioRad) from 20 $^{\circ}C$ to 80 $^{\circ}C$ in increments of 1 $^{\circ}C$. When the temperature increases, the protein unfolds, the

fluorescence of SYPRO Orange dye increases. Fluorescence changes in the wells were monitored simultaneously with a charge-coupled (CCD) camera. The wavelengths for excitation and emission were 485 and 575 nm, respectively. The temperature midpoint for the protein unfolding transition, T_m , was calculated using the BioRad iQ5 software. The results represent three technical replicates.

2.4.3. Fluorescence polarization assay

Fluorescence polarization (FP) binding assay measures the amount of binding or affinity between two molecules, a small fluorophore and its binding partner, by measuring the polarized fluorescence signal. In the solution, the molecular rotation of the fluorophore results in depolarization of the plane of polarized incident light. Upon binding, the molecular rotation slows down, and the degree of depolarization of the incident light is reduced. The FP assay was widely used in protein-ligand interaction to obtain K_d as well as competition assays to derive IC_{50} or K_i . This method allows high throughput screening with low reagent consumption.

The competition assay depends on measuring the fluorescence polarization of a fluorescent labelled RNA in the presence and absence of an unlabeled competitor RNA molecule. RNA interaction with the protein can be detected by measuring the decrease of FP signal when the RNA competes the fluorescent RNA from the protein. If there is no interaction, then the FP signal from the fluorescent RNA bound to the protein will remain unchanged.

Fluorescence polarization is defined by:

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \quad \text{Equation 3}$$

P = Fluorescence polarization
 $I_{\parallel}$ = Intensity with polarizer parallel
 $I_{\perp}$ = Intensity with polarizer perpendicular.

The total intensity is calculated to confirm the true binding:

$$I_{tot} = I_{\parallel} + 2I_{\perp} \quad \text{Equation 4}$$

I_{tot} = Total intensity
 $I_{\parallel}$ = Intensity with polarizer parallel
 $I_{\perp}$ = Intensity with polarizer perpendicular.

Fluorescence polarization (FP) binding assay.

The 5'-OH hairpin RNA was purchased from IDT and fluorescein amidite (FAM) was labelled at the loop region of the hairpin RNA. Samples containing 5 nM FAM-labelled RNA and proteins of a range of concentrations (0, 5, 10, 20, 39, 78, 156, 313, 6325, 1250, 2500, 5000 nM) were equilibrated in 25 mM HEPES, 2.5 % glycerol, 75 mM NaCl, 0.5 mM DTT, 0.01 % Triton-X in a Greiner Bio-One 384-well Black Polystyrene Microplate for 1 hour at room temperature. Fluorescence polarization was read using a BioTek® Synergy™ H1 Microplate Reader with a Green FP (485/528) Filter Cube. A bandpass filter of 485/20 nm was used for excitation and a bandpass filter of 528/20 nm was used to measure the fluorescence emission. Polarization was calculated using Equation 3.

K_d Determination.

For the equilibrium binding using the FAM-labelled RNA as the ligand, the binding data were fit to the following quadratic equation with consideration of ligand depletion:

$$B = \frac{(T_{RNA} + T_{Protein} + K_d) - \sqrt{(T_{RNA} + T_{Protein} + K_d)^2 - 4 \times T_{RNA} \times T_{Protein}}}{2}$$

Equation 5

Where B is the concentration of the Protein-RNA complex, $T_{Protein}$ is the total protein concentration, T_{RNA} is the total concentration of RNA and K_d is the equilibrium dissociation constant.

FP competition assay.

For FP competitive assays, RNA competitors were titrated in a 1:1 serial dilution and the concentration range starting from 10 μ M to 2.4 nM. DRH-3 CTD (400 nM) or DRH-1 CTD (300 nM), the FAM-labelled RNA (5 nM) and the RNA competitors were mixed and distributed equally in a black, 384-well, low-flange microplate (Corning, UK). Polarization was calculated using Equation 2.

IC₅₀ Equation.

The titration curve was fitted with log (inhibitor) vs. response – variable slope (four parameters) using GraphPad Prism 5 (GraphPad software, San Diego California USA, www.graphpad.com).

$$Y = \frac{Bottom + (Top - Bottom)}{(1 + 10^{(LogIC_{50} - X) \times HillSlope})} \quad \text{Equation 6}$$

X = Concentration of compound

Bottom = Polarization of free FAM-RNA

Top = Polarization of free FAM-RNA bound to protein

Hill Slope = Slope factor As shown in **Figure 2.4**.

Figure 2.4 Log (Inhibitor) vs response curve. X values are logarithms of concentrations. Y are the normalized response data running from 0 % to 100 %. IC_{50} is the concentration of agonist that gives a response half way between Bottom and Top. This figure is adapted from GraphPad Curve Fitting Guide (https://www.graphpad.com/guides/prism/7/curve-fitting/index.htm?reg_kcat.htm).

2.4.4. Limited proteolysis

Trypsin stock (1 mg/ml) solution was prepared in trypsin digestion reaction buffer (25 mM HEPES, 5 % glycerol, 150 mM NaCl, 1 mM DTT). Unligated protein and protein that loaded with RNAs were incubated with trypsin protease at a ratio of 1:500 or 1:2000 (for DRH-3 CTD) for 0-2 hours at 25 °C. The reactions were quenched by the addition of SDS loading dye and boiling for 5 min. Samples were analyzed on SDS-PAGE.

2.5. Crystallization

2.5.1. Crystal screening

To identify crystallization condition for the target protein, small scale sitting drop vapor diffusion crystallization screening was carried out using sparse matrix crystallization screening kits (Hampton Research), including Index, Crystal Screen·Crystal Screen 2, PEGRx1· PEGRx 2, SaltRx1· SaltRx 2, PEG/Ion1· PEG/Ion2. High-throughput screenings were carried out in 96-well plate format via Phoenix (Art Robbins Instruments) or Mosquito (TTP Labtech). The plates were incubated at two temperatures (20 °C and 4 °C). The drops were checked to follow the precipitation/crystallization of the protein over time.

The native DRH-3 NTD crystals were grown at 0.2 M ammonium fluoride (pH 6.2), 20 % w/v polyethylene glycol 3350, 5 mM DTT at 4 °C. The selenomethionine (SeMet) incorporated DRH-3 NTD crystals were grown in 0.2 M potassium sodium tartrate tetrahydrate (pH 7.4), 20 % w/v polyethylene glycol 3350, 5 mM DTT at 4 °C. The DRH-3 CTD was mixed with self-complementary 5'-ppp 12-mer GC-rich RNA at 2:1 molar ratio. Crystals were grown from 0.1 M Tris pH 7.5, 30 % w/v polyethylene glycol 6,000 at 18 °C. The DRH-3 CTD was mixed with 5'-ppp 8-nt ssRNA at 1:1 ratio. Crystals were grown from 0.2 M ammonium acetate, 0.1 M Tris pH 8.5, 25 % w/v polyethylene glycol 3350 at 18 °C. Crystals were flash-frozen with the reservoir solution in addition of 25 % (v/v) glycerol as cryo-protectant.

2.5.2. Data collection

Diffraction data for native DRH-3 NTD and SeMet DRH-3 NTD were collected in the Swiss Light Source (SLS) at Paul Scherrer Institute in Switzerland. Diffraction intensities were collected at SLS-PXIII using the PILATUS 2M-F detector and the

multi-axis PIRGo goniometer. Integration, scaling and merging of intensities were carried out using program iMOSFLM (Battye, Kontogiannis, Johnson, Powell, & Leslie, 2011) and SCALA (P. R. Evans & Murshudov, 2013) from the CCP4 suite (McNicholas, Potterton, Wilson, & Noble, 2011). Diffraction intensities for DRH-3 CTD and RNA complexes were recorded at NE-CAT beamline ID-24 at the Advanced Photon Source (Argonne National Laboratory, Argonne, IL, USA). Integration, scaling, and merging of the intensities were carried out with the programs XDS (Kabsch, 2010) and SCALA (P. Evans, 2006). Diffraction data for native DRH-3 NTD were collected by Dr. El Sahili Abbas, diffraction data for SeMet DRH-3 NTD were collected by Mr Jacky Tan and diffraction data for DRH-3 CTD were collected by Dr. Dahai Luo.

2.5.3. Structure determination and refinement

Structure of DRH-3 NTD was determined using the single-wavelength anomalous dispersion (SAD) method with online server Auto-Rickshaw (Panjikar, Parthasarathy, Lamzin, Weiss, & Tucker, 2005; Panjikar et al., 2009). The strategy of SAD phasing followed by MRSAD model enhancement was used to build the model. The model was further built using rounds of manual building in COOT (Emsley, Lohkamp, Scott, & Cowtan, 2010) and refined using Phenix (Adams et al., 2010). The structure of DRH-3 CTD in complex with the single-strand 5'-ppp 8-nt RNA was solved by SAD method using SHELX (Sheldrick, 2010). The structure of DRH-3 CTD in complex with the 5'-ppp dsRNA12 was solved by molecular replacement using Phaser (McCoy et al., 2007). The models were further built using rounds of manual building in COOT (Emsley et al., 2010) and refined using Phenix (Adams et al., 2002). Statistics of data collection and structure refinement for all three structures are summarized in **Table 3.3**. The structure of DRH-3 CTD were determined and refined by Dr. Luo Dahai.

2.5.4. Structure modelling

The homology model of DRH-3 proteins was generated using SWISS-MODEL (Waterhouse et al., 2018). DRH-3 NTD, RIG-I helicase domain and DRH-3 CTD crystal structures were used as input reference models.

2.6. Hydrogen Deuterium Exchange with Mass Spec (HDX-MS)

Proteins were exposed to deuterated water for 1 hour followed by a quenching step (low pH on ice) to stop hydrogen/deuterium exchange. Proteins were digested into peptide fragments by a low temperature and low pH compatible pepsin. The resulting peptides were subsequently separated by capillary-flow liquid chromatography and measured by mass spectrometer. The deuterium incorporation was then determined by examining the change in the mass centroid. HDX Workbench (Pascal et al., 2012) was utilized to analyze data and generate graphs. HDX-MS results were shown as graphical heat map revealed the level of deuterium incorporation %D for detected peptides. The heat map of HDX was subsequently mapped onto the structure models to probe the molecular structure and protein dynamics of the target protein. The HDX-MS experiments were performed by Dr. Jie Zheng.

Chapter 3: Structural insights into the role of DRH- 3 in *C. elegans* RNAi.

3.1. Overview

DRH-3 is essential for the biogenesis of secondary siRNAs by RdRP modules and plays an important role in germ-line development, DNA double-stranded break repair and chromosomal integrity (Aoki et al., 2007; Duchaine et al., 2006; Eki, Ishihara, Katsura, & Hanaoka, 2007; Gu et al., 2009; Nakamura et al., 2007; van Haaften et al., 2006). Previous biochemical studies have indicated that DRH-3 serves as a duplex RNA-dependent ATPase (Matranga & Pyle, 2010). DRH-3 preferentially binds to dsRNA (~22-bp) for its optimal enzymatic function. Duplex RNA binding may induce conformation changes in DRH-3 proteins. However, the molecular details of RNA ligand recognition by DRH-3 remain unknown. Therefore, structure studies of DRH-3 are important to understand the functional role of DRH-3 in *C. elegans* RNAi. In this chapter, the bacterial expression system and protein purification strategies for producing the recombinant full-length DRH-3 and its domains were described. The purified proteins were characterized by structural, biochemical and biophysical methods, including crystallography, HDX-MS, thermal shift assay, limit proteolysis assay and FP assay.

3.2. Protein and RNA production

3.2.1. DRH-3 is a worm homolog of RLRs

The DExD/H-box-containing RNA helicase, DRH-3, has been identified as an essential component of the RdRP complex which generates secondary siRNAs. DRH-3 encodes a worm-specific domain at the N terminus and an RNA binding domain at the C terminus, in addition to an RNA helicase domain. The domain architecture of DRH-3 is similar to RLRs. In addition, the helicase domain of DRH-3 is more similar to that of RLRs based on their primary sequence, despite the fact that DRH-3 is functionally related to Dicer. Comparison of the amino acid sequence of DRH-3 with those of RLRs revealed overall predicted amino acid sequence identities ranging from 18 to 22 % (**Figure 3.1** and **Table 3.1**). The helicase domain (amino acids 379 - 915) and the CTD (amino acids 915 - 1119) of DRH-3 are homologous corresponding domains in human RLRs, exhibiting 24-25 % and 21-23 % amino acid identities respectively. The conserved motifs in the helicase domain as well as the C4 type zinc finger motif and the KWK motif in the CTD were present in DRH-3 (**Figure 3.1**). Striking differences were found in the N-terminal region of DRH-3 (amino acids 1 - 350). The overall amino acid sequence identities between DRH-3 NTD and the N-terminal CARDS in RLRs ranging from 9-12 %. Diversification of the NTD might be fundamental to the functional diversity of DRH-3 and RLRs.

Figure 3.1 Sequence alignment of DRHs and RLRs. The figure was prepared using the program Geneious (Kearse et al., 2012). Black indicates high-consensus residues (> 90 % conserved), and light grey indicates low-consensus residues (50 % – 90 % conserved). Predicted secondary structures of DRH-3 were shown on the top of the aligned sequences. The NTD, helicase domain and CTD were indicated in red, green and blue boxes, respectively. The conserved helicase motifs, the Zinc finger motif and the KWK motif were indicated by asterisks. *CeDRH-3* (UniProtKB: Q93413); *CeDRH-1* (UniProtKB: G5EDI8); *HsRIG-I* (UniProtKB: O95786); *HsMDA5* (UniProtKB: Q9BYX4); *HsLGP2* (UniProtKB: Q96C10); *Hs*, *Homo sapiens*; *Ce*, *Caenorhabditis elegans*

(Continued)

Table 3.1 Pairwise protein sequence comparisons of DRHs and RLRs showing amino- acid identities.

Full length					
	<i>Ce</i> DRH-3	<i>Ce</i> DRH-1	<i>Hs</i> RIG-I	<i>Hs</i> MDA5	<i>Hs</i> LGP2
<i>Ce</i> DRH-3	100 %	19 %	18 %	22 %	18 %
<i>Ce</i> DRH-1		100 %	22 %	23 %	20 %
<i>Hs</i> RIG-I			100 %	30 %	28 %
<i>Hs</i> MDA5				100 %	41 %
<i>Hs</i> LGP2					100 %
Helicase domain					
	<i>Ce</i> DRH-3	<i>Ce</i> DRH-1	<i>Hs</i> RIG-I	<i>Hs</i> MDA5	<i>Hs</i> LGP2
<i>Ce</i> DRH-3	100 %	25 %	24 %	24 %	24 %
<i>Ce</i> DRH-1		100 %	28 %	26 %	25 %
<i>Hs</i> RIG-I			100 %	36 %	33 %
<i>Hs</i> MDA5				100 %	44 %
<i>Hs</i> LGP2					100 %
CTD					
	<i>Ce</i> DRH-3	<i>Ce</i> DRH-1	<i>Hs</i> RIG-I	<i>Hs</i> MDA5	<i>Hs</i> LGP2
<i>Ce</i> DRH-3	100 %	23 %	22 %	21 %	22 %
<i>Ce</i> DRH-1		100 %	26 %	25 %	23 %
<i>Hs</i> RIG-I			100 %	36 %	33 %
<i>Hs</i> MDA5				100 %	44 %
<i>Hs</i> LGP2					100 %
NTD					
	<i>Ce</i> DRH-3	<i>Ce</i> DRH-1	<i>Hs</i> RIG-I	<i>Hs</i> MDA5	
<i>Ce</i> DRH-3	100 %	9 %	11 %	12 %	
<i>Ce</i> DRH-1		100 %	12 %	10 %	
<i>Hs</i> RIG-I			100 %	20 %	
<i>Hs</i> MDA5				100 %	

The pairwise sequence analysis was obtained from the program ClustalW2 (J. D. Thompson et al., 1994). Values are the percent sequence identities for the full-length proteins, the helicase domains, the NTDs and the CTDs. The domain boundaries were shown in **Figure 1.5**. *Ce*DRH-3 (UniProtKB: Q93413); *Ce*DRH-1 (UniProtKB: G5EDI8); *Hs*RIG-I (UniProtKB: O95786); *Hs*MDA5 (UniProtKB: Q9BYX4); *Hs*LGP2 (UniProtKB: Q96C10); *Hs*, *Homo sapiens*; *Ce*, *Caenorhabditis elegans*

3.2.2. Full-length DRH-3 is a monomer in solution

To characterize the protein function, large quantities of highly pure and active DRH-3 are required. The results of expression trials were analyzed by SDS-PAGE and summarized in **Figure 3.2**. The protein band which is corresponding to the tagged full-length DRH-3 (143.5 kDa) was observed after induction by IPTG. About 4 grams (wet weight) of cell pellets were harvested from 1-liter cell culture. Plasmid pET-SUMO-DRH-3_Hel-CTD which expresses a truncated DRH-3 (residues 358-1119) containing only the helicase domain and CTD was also prepared. Unfortunately, all the expressed DRH-3_Hel-CTD was contained within the inclusion bodies.

Figure 3.2 Expression trials for full-length DRH-3 and DRH-3 Hel-CTD. Samples of cell lysates were analyzed by 12 % SDS-PAGE and visualized through Coomassie staining. The band with ‘*’ next to it is the His₆-SUMO-tagged full-length DRH-3 (residues 1-1119) which has a molecular weight of 143.5 kDa; the band indicated with ‘***’ is the His₆-tagged DRH-3_Hel-CTD (residues 358-1119) with a molecular weight of 100.2 kDa.

The clarified cell lysates containing soluble His₆-SUMO-tagged full-length DRH-3 were purified by IMAC. The purified proteins were analyzed on a 10 % SDS-PAGE where His₆-SUMO-DRH-3 migrated as a single protein band (142.5 kDa). A few other faint protein bands indicate the presence of small amounts of contaminants (**Figure 3.3**).

Figure 3.3 Purification of His₆-SUMO-DRH-3 by IMAC. A 10 % SDS-PAGE with Coomassie staining is shown to analyze the flow-through, high-salt wash and elution fractions.

Heparins are negatively charged polysaccharides which not only serve as affinity ligands binding to nucleic acid-binding proteins, but also as ion exchangers with high charge density and distribution (Xiong, Zhang, & He, 2008). HiTrap Heparin HP column (GE) is designed for high-resolution purification of nucleic acid-binding proteins. DRH-3 has two RNA binding domains: the helicase domain and the CTD. Thus, DRH-3 has strong interaction with the heparins in low-salt buffer condition and

purity of greater than 95 % was further confirmed using 10 % SDS-PAGE (**Figure 3.5B**).

The purified proteins were used for further characterizations and assays.

Figure 3.5 Purification of DRH-3 by gel filtration chromatography. (A) The elution profile of full-length DRH-3 from HiLoad 16/600 Superdex 200 pg (GE, Healthcare). (B) The 10 % SDS-PAGE with Coomassie staining is shown here to compare the gel filtration elution and the heparin elution of DRH-3.

3.2.3. Construct design for expressing DRH-3 domains

A series of plasmid constructs for expressing DRH-3 domains in *E. coli* system were carefully designed. The truncation sites were determined base on the predicted functional domain boundaries and the predicted secondary structures of DRH-3. The purpose of testing several constructs with variable lengths was to identify constructs capable of expressing soluble proteins that can be crystallized. The constructs that allowed for soluble protein expressions were shown in **Figure 3.6**.

Figure 3.6 Constructs of DRH-3 that express soluble proteins. Upper diagram shows the domain organization of DRH-3. The N-terminal domain (NTD) was colored green, the helicase domain was indicated in grey and the C-terminal domain (CTD) was in blue. The lower diagram shows the series of constructs of DRH-3 NTD and CTD that were prepared for crystallization screening. SUMO-tag was colored in yellow and His6-tag was in blue. * indicated that the construct DRH-3_N2 was crystallized.

3.2.4. Expression, purification and crystallization of DRH-3 NTD

The boundary of DRH-3 NTD was predicted to be at the region around the residue T350. However, the recombinant His₆-tagged DRH-3_N1 was not soluble. To enhance the soluble expression of the protein, a SUMO-tag was inserted in between His₆-tag and the N-terminus of DRH-3_N1. The SUMO-tag significantly improved the solubility of the target protein and no protein precipitation was observed after tag cleavage (**Figure 3.7**). The His₆-SUMO-tagged DRH-3_N1 was purified by a two-step purification strategy: IMAC followed by gel filtration chromatography. The His₆-SUMO-tag was cleaved by TEV protease after the IMAC elution. The untagged DRH-3_N1 (41.3 kDa) were eluted in monomeric form in the gel-filtration chromatography (**Figure 3.7A**). Twenty milligrams of untagged DRH-3_N1 was obtained from 1 liter of bacterial cell culture. DRH-3_N1 proteins were then subjected to crystallization screening, where cube-like crystals were grown after 2 days at 4 °C in the condition containing 0.2 M ammonium fluoride (pH 6.2), 20 % w/v polyethylene glycol 3350, 5 mM DTT (**Figure 3.8A and B**). Unfortunately, the crystals of DRH-3_N1 diffracted to no better than 10 Å and the resolution could not be further improved by optimizing the crystallization conditions.

A

B

C

Figure 3.7 Expression and purification of His₆-SUMO-tagged DRH-3_N1. (A) Gel filtration chromatography for purifying untagged DRH-3_N1 using HiLoad 16/600 Superdex 200 pg column. (B) Protein production and purity were analyzed on the SDS-PAGE. (C) MALDI-TOF mass spectrometry examination of molecular weight of untagged DRH-3_N1. The theoretical molecular weight of DRH-3 is 41285.9 Da and the MS molecular weight is 41363.1 Da.

To obtain the NTD domains that were suitable for crystallization, the DRH-3 NTD constructs were further optimized by removing amino acids from its C-terminus. Both of the new constructs DRH-3_N2 (M1-S335) and DRH-3_N3 (M1-S304) were expressed in the soluble fractions in *E. coli* Rosetta (DE3) cells. The purification procedures for these two constructs were similar to DRH-3_N1 (**Figure 3.9A and B**). The yield for DRH-3_N3 (25 mg per 1 liter) was higher than DRH-3_N2 (5 mg per 1 liter). But only the DRH-3_N2 proteins were able to form crystal.

The DRH-3_N2 crystals were optimized to be grown at 4 °C using the hanging drop vapor diffusion method. The best reservoir solution for DRH-3 NTD contained. Crystals were transferred to a cryoprotecting solution containing 0.2 M ammonium 0.2 M ammonium fluoride (pH 6.2), 20 % w/v polyethylene glycol 3350, 5 mM DTT fluoride (pH 6.2), 25 % w/v polyethylene glycol 3350, 20 % glycerol, before being flash frozen in 100K liquid nitrogen. The crystals with single layer (0.05 mm×0.3 mm) were shown to have better diffracting quality by preliminary scanning and were chosen for X-ray data collection in the synchrotron to determine the crystal structure of DRH-3 NTD (**Figure 3.8C and D**). The statistics of data collection and the processing were summarized in **Table 3.3**. Two CARD-like domains were identified in DRH-3 NTD (which will be discussed in **section 3.4**). According to the crystal structure of DRH-3 NTD, another construct DRH-3_N4 (1-152) was designed representing the first CARD-like domain (**Figure 3.6**). About 10 mg of DRH-3_N4 proteins was produced from 1 liter cell culture (**Figure 3.9C**). Unfortunately, no crystal hit for DRH-3_N4 was observed from the crystallization screening.

Figure 3.8 Improvement of the DRH-3 NTD crystal quality. (A, C) Light-microscope images of DRH-3_N1 (A) and DRH-3_N2 (C) crystals. The images were taken at same scale. Crystals of DRH-3_N1 in (A) were in cube-like shape and the crystals of DRH-3_N2 in (C) were elongated plates. (B, D) The oscillation images during data collection from DRH-3_N1 and DRH-3_N2 crystals. The images shown here correspond to the diffraction patterns used for structure determinations. The edges of the oscillation images are indicated by blue circles and the corresponding resolutions are shown in the lower left corner.

A DRH-3_N2

B DRH-3_N3

C DRH-3_N4

Figure 3.9 Purification of His₆-SUMO-DRH NTD constructs. (A) SDS-PAGE gels with Coomassie staining were shown to analyze the purity of the target proteins at each purification step for DRH-3_N2. (B) SDS-PAGE gels were shown to analyze the purity of the target proteins at each purification step for DRH-3_N3. (C) SDS-PAGE were shown to analyze the purity of the target proteins at each purification step for DRH-3_N4.

3.2.5. Expression, purification and crystallization of DRH-3 CTD

Plasmid pET15b_DRH-3_C1 (residues 940-1108) and plasmid pNIC28-Bsa4_DRH-3_C2 (residues 943-1104) that express N-terminal His₆-tagged DRH-3 CTD were constructed and sequence verified. The same expression and purification protocols were used for DRH-3 CTD and full-length DRH-3. Around 15 grams and 4 grams of cell pellets for DRH-3_C1 and DRH-3_C2 were harvested from 1-liter cell culture, respectively. A typical SDS-PAGE data of samples at different purification stages was shown in **Figure 3.10**. The DRH-3 CTD proteins were eluted in a monomeric form in the gel-filtration chromatography. Both constructs were subjected to RNA binding assay and crystallization screening. RNA binding assay indicated that both products displayed similar thermal stability and similar RNA affinity (which will be discussed later). Only the DRH-3_C1 proteins were successfully co-crystallized with RNAs, while no crystal was obtained for DRH-3_C2 under any of the available conditions. Crystallization of DRH-3_C1 was done by Dr. Dahai Luo.

Figure 3.10 Purification of the His₆-DRH-3 CTD construct. (A) SDS-PAGE analysis of DRH-3_C1. (B) Gel filtration chromatogram (HiLoad 16/600 Superdex 200 pg column) of untagged DRH-3_C1 (19 kDa). The retention volume for each peak was labelled.

3.2.6. Synthesis and purification of RNAs

A series of short RNAs were designed and used as potential substrates for DRH proteins to test the enzymatic activity as well as to crystallize DRH-3. RNAs used in this study are listed in **Table 3.2**. As large amount of RNAs were required for the crystallization screenings of DRH-3, several 5'-ppp RNAs were generated in-house by *in vitro* transcription. A purification example of synthetic RNAs is shown in **Figure 3.11**. Notably, an 8-mer ssRNA (RNA 1) and a 12-mer dsRNA (RNA 2) were able to co-crystallize with DRH-3 CTD. The crystal structures of DRH-3 in complex with RNAs are shown in **Section 3.5**. A 10-mer hairpin RNA (RNA 6) was used for HDX-MS studies of DRH-3 (**Section 3.6**). To determine the optimal RNA length for the activation of DRH proteins' ATPase activity, 5'-ppp dsRNAs with sizes ranging from 8 to 20 bp

(RNA 2-5) were used for ATPase assays. For the FP competition assay, five 12-mer dsRNAs with different 5'-modification were used: RNA 7 is a blunt-ended 12-mer dsRNA with 5'-ppp; RNA 8 is a blunt-ended 12-mer dsRNA with 5'-OH; RNA 9 is a blunt-ended 12-mer dsRNA with 5'-p; RNA 10 is a 5'-ppp 12-mer dsRNA with 5' guanine overhang and RNA 11 is a 5'-ppp 12-mer dsRNA with 3' cytosine overhang. The FAM-labelled RNA (RNA 12) which is known to bind DRH-3 CTD is available in the lab.

Figure 3.11 Purification of RNA using FPLC (A) Elution profile of the 12 bp RNA (RNA 2), purified from a 2 ml in vitro transcription reaction using ion-exchange chromatography (5 ml HiTrap Q FF column, GE Healthcare). The rNTPs were eluted at lower NaCl concentration and the RNA of interest was eluted from a smaller peak at higher NaCl concentration. (B) Gel filtration of the 12bp RNA (Superdex 200 10/300, GE Healthcare).

Table 3.2 List of RNAs

In vitro transcription			
	Name	RNA sequence	DNA sequence
1	8-nt ssRNA	5' ppp-GGCCGCCC	5'-G TAATACGACTCACTATA GGCCGCCC
2	12-mer dsRNA	5' ppp-GGCGCGCGCGCC CCGCGCGCGCGG-5' ppp	5'-G TAATACGACTCACTATA GGCGCGCGCGCC
3	8-mer dsRNA	5' ppp-GGCGCGCGCGCC CCGCGCGCGCGG-5' ppp	5'-G TAATACGACTCACTATA GGCGCGCGCGCC
4	10-mer dsRNA	5' ppp-GGCGCGCGCGCC CCGCGCGCGCGG-5' ppp	5'-G TAATACGACTCACTATA GGCGCGCGCGCC
5	20-mer dsRNA	5' ppp-GGCGCGCGCGCC CCGCGCGCGCGG-5' ppp	5'-G TAATACGACTCACTATA GGCGCGCGCGCC
6	3P10L	5' PPP-GGACGCGUGC U CCUGCGCACG C G	5'-GTAATACGACTCACTATA GGAGAGGUGCUUCGGCACGCGUCC
Chemical synthesis			
	Name	RNA sequence	
7	R1	5' ppp-GGCGCGCGCGCC CCGCGCGCGCGG-5' ppp	
8	R1OH	5' OH-GGCGCGCGCGCC CCGCGCGCGCGG-5' OH	
9	R1P	5' p-GGCGCGCGCGCC CCGCGCGCGCGG-5' ppp	
10	R2	5' pppGGGCGCGCGCGCC CCGCGCGCGCGGG-5' ppp	
11	R3	5' pppGGGCGCGCGCGCCC CCCGCGCGCGCGG-5' ppp	
12	R0	5' OH-UUUUAUAUAUAGGACCGC \ AAAUUAUAUAUCCUGGCG C G	U-FAM

3.3. Structure determinations

3.3.1. Solving the phase problem for DRH-3 NTD

The phase problem. To determine a crystal structure, both the diffraction intensities and the phases of the X-ray data are required (Drenth, 1994). The intensities are directly measured with an X-ray detector, while the phase cannot be directly measured which is known as the ‘phase problem’. Several phasing methods are available, including molecular replacement (MR), single/multiple isomorphous replacement (SIR/MIR), single/multiple-wavelength anomalous diffraction (SAD/MAD), etc. (Pike, Garman, Krojer, von Delft, & Carpenter, 2016).

The most common and convenient method to solve the phase problem is the molecular replacement method (MR) which uses the phase information from a homologous protein structure as a model (Scapin, 2013). MR is an indirect method to solve the phase problem. Since no homologous structure for DRH-3 NTD was available, the potential functional homolog of DRH-3 NTD, the RIG-I CARDS domain (PDB code: 4A2W), was used as the search model (Kowalinski et al., 2011). Unfortunately, MR for DRH-3 NTD failed due to the low sequence homology. MR is usually only recommended when the sequence identity between the model and the target protein is at least 40 % (Ramachandran & Srinivasan, 1961; Rossmann & Blow, 1962; Terwilliger, 2004). The amino acid sequence identity between DRH-3 NTD and RIG-I CARDS is only 26 % which is significantly lower than the recommended value (>40 % is preferred) (Ramachandran & Srinivasan, 1961; Rossmann & Blow, 1962; Terwilliger, 2004).

For proteins that lack useful structure model for MR, the initial phase will have to be obtained experimentally. X-ray is diffracted by the electron cloud around the nucleus, therefore, the more electrons an atom has, the more X-ray will be diffracted. Heavy atoms (HA) (Pt, Hg, Se, Au, Pb, etc.) have more electrons than those normally found in

proteins (H, O, N, C, S, etc.) (Boggon & Shapiro, 2000; Pike et al., 2016). Thus, the HA-incorporated crystal will result in a different diffraction pattern compared with the native crystal, giving phase information.

Hendrickson and colleagues first reported the possibility of using SelenoMethionine (SeMet) as a phasing tool in 1990 and the ease of incorporating selenium in recombinant system. This makes selenium to be the first choice for multiple-wavelength (and single-wavelength) anomalous dispersion experiments (Pahler, Smith, & Hendrickson, 1990; Walden, 2010). Incorporating SeMet to the recombinant protein during the protein expression in cells does not cause the structural disturbance which is commonly associated with heavy-atom incorporation. SeMet incorporation during protein production removes the need for time consuming and challenging screening for heavy-atom derivatives. It has been reported in prokaryotic systems that the incorporation of SeMet is often close to 100 %, especially when a methionine auxotrophic strain such as B834 is used (Leahy, Hendrickson, Aukhil, & Erickson, 1992). It is noteworthy that selenium only has 34 electrons which is less than most of other HA, for example, mercury (80 electrons). Thus, selenium provides lower X-ray scattering power per HA. The content of Se per protein directly affects the scattering power. Sequence analysis of DRH-3_N2 shown that there were 7 methionine residues out of 335 amino acids. The selenium content is sufficient to provide a strong X-ray scattering for DRH-3_N2. Therefore, SeMet incorporation for DRH-3 NTD was used as a strategy to solve the phase problem.

Selenomethionine incorporation for DRH-3 NTD. The methionine auxotrophic bacteria strain B834 (DE3) was used to express SeMet-incorporated DRH-3 NTD_N2 (*SeMet*DRH-3_N2). The bacteria were grown in the minimal medium with SeMet as the only source of methionine and the protein production was induced using the same

method as described before for the native protein. Although the overexpression of SeMet-incorporated DRH-3_N2 (*SeMet*DRH-3_N2) were not as good as the native proteins, sufficient SeMet proteins could be obtained and was purified using the same purification method as that for native protein. *SeMet*DRH-3_N2 crystals were grown at 4 °C by the hanging drop vapor diffusion method. The reservoir solution for *SeMet*DRH-3_N2 contained 0.2 M potassium sodium tartrate tetrahydrate (pH 7.4), 20 % w/v polyethylene glycol 3350, 5 mM DTT. Crystals were transferred to a cryoprotectant solution containing 0.2 M potassium sodium tartrate tetrahydrate (pH 7.4), 25 % w/v polyethylene glycol 3350, 20 % glycerol, before being flash frozen in 100K liquid nitrogen. The *SeMet*DRH-3_N2 crystal (**Figure 3.12**) was diffracted to 3.11 Å.

To determine the SeMet incorporation, fluorescence scanning was applied to the SeMet crystal at synchrotron to directly detect the signal of Se. Thus, the incorporation of Se with desired scattering power was confirmed. Another purpose of fluorescence scan is to determine the wavelength of X-ray used for *SeMet* DRH-3_N2 data collection. Selenoneine atoms have viable absorption edges accessible to energies at typical synchrotron beamline. Experimental fluorescence scan was required to verify the presence and location of the absorption edge for the crystal. The experimental peak energy was 12659.3eV and the peak wavelength was 0.9794 Å. X-ray diffraction data for *SeMet* DRH-3NTD_N2 was collected at the selenium peak wavelength. Data collection information was shown in **Table 3.3**.

Figure 3.12 Crystal of *SeMet* DRH-3_N2 and data collection. (A) Light-microscope image of *SeMet* DRH-3_N2. (B) Diffraction image of *SeMet* DRH-3_N2.

3.3.2. Structure determination of DRH-3 NTD

Auto-Rickshaw, an automated crystal structure determination server, was used to build a model from SAD phases (Panjikar et al., 2005, 2009). A model containing 314 of 662 residues in the asymmetric unit was obtained. This partial model was used as an input for the MRSAD protocol and was directly fed into Phaser for SAD phasing and substructure completion (McCoy et al., 2007). The phasing and building were repeated for 6 cycles. In the last cycle, 682 residues were built, and refinement of the model gave

R_{work} and R_{free} values of 0.46 and 0.51, respectively. In the fourth cycle, 917 residues were built, and 633 residues were docked into the sequence. The side chains in the model were then manually adjusted using COOT (Emsley et al., 2010), followed by several cycles of restrained refinement using Phenix (Adams et al., 2010; Adams et al., 2002). When appropriate, water molecules were added to the structure using COOT (Emsley et al., 2010). After additional cycles of restrained refinement and manual adjustments to side chains and water molecules, refinement of the model resulted in R_{work} and R_{free} values of 0.23 and 0.28, respectively (**Table 3.3**).

3.3.3. Structure determination of DRH-3 CTD

Data collection and structure refinement of DRH-3 CTD were performed by Dr. Dahai Luo and the statistics of data collection and structure refinement for the DRH-3 CTD structures are summarized in **Table 3.3**.

Table 3.3 Data collection and Refinement statistics

Data collection				
Macromolecule	DRH-3 NTD (Native)	DRH-3 NTD (SeMet)	DRH-3 CTD: 5'ppp 8-nt ssRNA	[DRH-3 CTD: 5'ppp 12-bp dsRNA]2
Resolution (Å)	85.56 - 2.80 (2.95 - 2.80)	47.63-3.11 (3.28-3.11)	43.04 - 1.89 (1.93 - 1.89)	43.96 - 1.60 (1.63 - 1.6)
Space group	P 4 ₁ 2 ₁ 2	P 4 ₁ 2 ₁ 2	P 4 ₃ 2 ₁ 2	P 2 2 ₁ 2 ₁
Unit cell a/b/c(Å)	93.18/93.18/215.96	89.925/89.925/215.68	56.843/56.843/131.75	33.775/93.090/133.823
$\alpha/\beta/\gamma$ (°)	90/90/90	90/90/90	90/90/90	90/90/90
Unique reflections	24012 (3401)	16670 (2324)	17289 (984)	56358 (2845)
Multiplicity	11.1	12.8	9.2	3.1
Completeness (%)	99.3 (99.0)	99.7 (98.3)	96.3 (88.4)	99.4 (99.9)
I/sigma(I)	10.9 (2.2)	11.0 (2.1)	12.0 (1.7)	14.9 (1.6)
R _{merge}	0.134 (0.896)	0.241 (1.352)	0.142 (1.715)	0.042 (0.761)
Wilson B-factor	66.91	59.3	26.7	20.1
CC1/2	1 (0.727)	0.996 (0.664)	0.996 (0.378)	0.998 (0.586)
Refinement Statistics(F >0)				
Reflections used in refinement	22582		17263	56281
Number of atoms	5390		1678	3650
macromolecules	5252		1521	3188
ligands	0		33	130
water	138		124	332
Protein residues	659		167	322
R _{work}	0.2337		0.1777	0.1692
R _{free}	0.286		0.2213	0.1988

(Continued)

Table 3.3 Data collection and Refinement statistics (continued)

Data collection				
Macromolecule	DRH-3 NTD (Native)	DRH-3 NTD (SeMet)	DRH-3 CTD: 5'ppp 8-nt ssRNA	[DRH-3 CTD: 5'ppp 12-bp dsRNA]2
Rmsd from ideal bond length(Å)	0.008		0.008	0.017
Rmsd from ideal bond angle(°)	1.19		0.991	1.88
Ramachandran statistics				
Favored (%)	97.6		95.73	98.2
Allowed (%)	100		100	100
Number of outliers	0		0	0
Clashscore	6.53		2.66	5.53
Average B-factor	86.3		37.06	29.77
macromolecules	86.8		36.32	28.96
solvent	67.43		38.98	36.60

a. Values in parentheses indicated values in the highest resolution shell

b. $R_{\text{merge}} = \sum |I_{\text{obs}} - I_{\text{avg}}| / \sum I_{\text{avg}}$.

c. $R_{\text{free}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum I_{\text{calc}}$. R_{free} was calculated with 5% of reflections.

3.4.DRH-3 NTD is composed of tandem CARDS in parallel arrangement

3.4.1. DRH-3 NTD exists as a monomer.

The solved DRH-3 NTD crystal structure contained two monomers in one asymmetric unit. Both monomers were refined without applying non-crystallographic symmetry (NCS) restraints, and the resulting structures were identical with a root-mean-square deviation (RMSD) of 0.367 Å for 305 superimposed Ca atoms (**Figure 3.13**). Previous studies have shown that the NTD may mediate DRH-3 dimer assembly (Fitzgerald et al., 2014). To examine whether the NTD dimer assembly shown in the crystal structure is biologically relevant, the interface of the dimer was analyzed by PDBePISA (Krissinel & Henrick, 2007) (**Figure 3.14**). Seven charged or hydrophobic residues contributed to the hydrophobic and electrostatic interactions (**Figure 3.14**). The interface area (540 Å²) is only about 10 % of the total solvent-accessible area (5273 Å²), suggesting that the dimer formation is most likely the result of crystal packing. Consistent with the size-exclusion chromatography profiles shown previously, both DRH-3 NTD and full-length DRH-3 existed in monomeric form in solution (**Figure 3.5 and 3.7**).

Figure 3.13 Crystal structure of dimeric DRH-3 NTD. Two DRH-3 NTD molecules in the asymmetric unit are shown as cartoons. Chain A and Chain B are indicated by different colors.

Figure 3.14 Illustration of the interactions involved in dimer interface of DRH-3 NTD. We evaluated DRH-3 NTD for its ability to form dimer by the PDBePISA server (Krissinel & Henrick, 2007). Hydrophobic and electrostatic interactions involved in the dimer interface are shown. The residues involved in the interface interactions are shown as sticks. The yellow dotted lines represent atomic distances of the bonds.

3.4.2. DRH-3 NTD is composed of two CARDs

DRH-3 NTD adopts a flattened triangle shape, resembling a boomerang (**Figure 3.15**). This domain is composed of primarily α helices ($\alpha 1$ – $\alpha 12$) that further form two helical bundles. Visual inspections of the relative orientation of the helices reveal two Greek-key-like folds linked by loops and small helices (αb -c). Interestingly, RIG-I and MDA5, the human homologs of DRH-3, also have two Greek-key-like 6- α -helical bundles at their N-terminal regions known as the tandem CARDs domains. For convenience, we will refer to two of the helical bundles in DRH-3 NTD as CARD-related domain 1 (C1, helix $\alpha 1$ -6) and CARD-related domain 2 (C2, helix $\alpha 7$ -12).

Figure 3.15 Cylindrical depiction of DRH-3 NTD structure with helices labelled. DRH-3 NTD contains two CARDs, highlighted in red (NTD_C1) and salmon (NTD_C2) colors, respectively. Cylinders represent the helices and the arrow represents direction of rotation of the polypeptide chain.

We compared the three-dimensional structures of DRH-3 C1 and C2 with their structure homologs suggested by DALI search (Holm & Rosenstrom, 2010) (**Table 3.3**). As expected, all the hits for the homology search indicate high similarity to the CARD subfamily which participates either in apoptosis [nucleolar protein 3 (NOL3),

procaspase-9 (pc-9) and apoptotic protease activating factor 1 (Apaf-1)] or inflammation [caspase recruitment domain family member 11(CARD11), ICEBERG, RIG-I, B Cell CLL/Lymphoma 10 (Bcl10) and Bcl10-interacting CARD protein (BinCARD)]. Despite the low amino acid sequences identity (10 %) shared by C1 and C2 with their homologous hits, the relative orientation of the helices is similar (**Figure 3.16**). The helix $\alpha 1$ in C1 has a kink which is a common feature of CARDs (Bouchier-Hayes & Martin, 2002; Kao et al., 2015). The third and fourth helices in DRH-3 CARDs are longer and more flexible. This is unique to DRH-3 CARDs.

Figure 3.16 Comparison of the death fold-like subdomains of DRH-3 NTD with top 5 structural hits from Dali search. Two α -helical bundles of DRH-3 NTD are highlighted in red (C1) and salmon (C2) colors, respectively. Only the aligned regions of the 5 hits for each subdomain are shown in orange color. The hits were superimposed with DRH-3 C1 and C2, respectively. The six helices are labelled from N' terminus to C' terminus as H1-H6. The structures are listed, from left to right, in descending order of structure similarity by Z score, with the PDB code under the structure. Structural superimposition was done by pairwise DALI and represented by PyMOL (DeLano Scientific LLC, San Francisco, CA). For clarity, the structure models are rotated to give a similar orientation. The results of pairwise DALI are summarized in Table 3.4.

Table 3.4 Multiple structural alignment of DRH-3 C1 and C2 with other known structures using the DALI server

TOP 5 HITS FOR DRH3C1							
No:	Chain	Z	rmsd	lali	nres	%id	Description
1	4uz0-A	5.9	2.9	73	87	8	NUCLEOLAR PROTEIN 3
2	4i16-A	4.7	2.7	67	90	7	CASPASE RECRUITMENT DOMAIN-CONTAINING PROTEIN 11
3	4dwn-A	4.6	2.9	68	97	6	BCL10-INTERACTING CARD PROTEIN
4	2mb9-A	4.5	2.7	69	106	10	B-CELL LYMPHOMA/LEUKEMIA 10
5	5juy-B	4.2	3.1	69	1234	7	APOPTOTIC PROTEASE-ACTIVATING FACTOR 1
TOP 5 HITS FOR DRH3C2							
No:	Chain	Z	rmsd	lali	nres	%id	Description
1	5juy-A	6.4	2.1	69	1234	8	APOPTOTIC PROTEASE-ACTIVATING FACTOR 1
2	1dgn-A	5.5	2.3	67	89	7	ICEBERG (PROTEASE INHIBITOR)
3	4i16-A	5.4	2.3	66	90	6	CASPASE RECRUITMENT DOMAIN-CONTAINING PROTEIN 11
4	4a2w-A	5.4	2.4	68	678	10	RETINOIC ACID INDUCIBLE PROTEIN I
5	5wvc-B	5.2	2.4	71	103	7	CASPASE-9

Hits are ranked by Z-score with the best hits at the top of the table. Chain: the PDB code with a chain ID; rmsd: root-mean-square deviation of C α atoms of superimposed proteins in Angstroms; lali: number of structural equivalent positions; nres: number of structurally equivalent aligned residues; %id: percentage of amino acid identity in aligned positions.

3.4.3. The tandem CARDs are in parallel arrangement

The two CARD-related domains in DRH-3 are different in sequence and structure (RMSD of 4.9 Å for all C α atoms). Both of them are stabilized by hydrogen bonds between helices α 1- α 10 and helices α 4- α 7 (**Figure 3.17A**) and are arranged side-by-side. In contrast, the two CARD domains (CARD 1 and CARD 2) in RIG-I/MDA5 are identical and arranged in a head-to-tail arrangement (**Figure 3.17B**).

Figure 3.17 DRH-3 NTD is composed of tandem CARDs in parallel arrangement. (A) The interface of the CARD-like domains. A close view of the interface is shown as cartoons. The interface residues are shown as sticks. The yellow dotted lines represent atomic distances of the polar bonds. (B) Comparison of the tandem CARDs in DRH-3 and RIG-I.

Functionally, the charged patches of death fold proteins are important for protein-protein bindings via electrostatic interactions (Kao et al., 2015). This prompts us analyze the electrostatic potential at the solvent-accessible surface of DRH-3 NTD (**Figure 3.18**). The surface of DRH-3 NTD contains several highly charged patches. The positive patches fall within the helix stem composed of helices $\alpha 4$ and $\alpha 3'$, the convex formed by helix $\alpha 10$, and the loop between $\alpha 7$ and $\alpha 8$; the negative patches present at the helix stem composed of helix $\alpha 2$ and $\alpha 3$, the convex formed by helices $\alpha 1$ and $\alpha 5$, the loop between $\alpha 5$ and $\alpha 6$, and the helix $\alpha 6$ tail region.

Figure 3.18 Electrostatic surface of DRH-3 NTD. Positively charged surfaces are colored in blue and negatively charged surfaces are colored in red, while neutral surfaces are colored in white. Large positive and negative patches on the surfaces are circled in yellow and black, respectively.

3.5. Structural insight into the secondary siRNA recognition by DRH-3 CTD

3.5.1. Overall structure of DRH-3 CTD

To elucidate the structural basis of 5'-ppp dsRNA recognition by DRH-3, we determined the crystal structure of DRH-3 CTD in complex with palindromic 5'-ppp 12-mer GC-rich dsRNA (5'-ppp GGCGCGCGCGCC) at 1.6 Å resolution (**Figure 3.19A**; **Table 3.3**). The crystallographic asymmetric unit contains a complex of DRH-3 CTD and the dsRNA with a ratio of 2:1, where the 5'-ppp dsRNA adopts a standard A-form double

helical structure and recruits two CTD molecules. The complex exhibits pseudo 2-fold symmetry and the RMSD (for all C α atoms) between the two DRH-3 CTD molecules is 0.495 Å.

Figure 3.19 Overall structure of DRH-3 CTD in complex with 5'-ppp 12-mer dsRNA. The crystallographic asymmetric unit contains a 2:1 complex between DRH-3 CTD and the dsRNA. Chain B and the 5' strand are colored. The 5'-ppp is shown as orange spheres and the zinc molecule is shown as a purple sphere. (B) The schematic diagram of contacts between DRH-3 CTD and 5'-ppp 12-mer dsRNA.

DRH-3 CTD is composed of 9 β -strands (β 1-9) connected by loops and small helices (**Figure 3.19A**). Two antiparallel β -sheets (β -sheet1: β 1 and β 8-9, β -sheet2: β 5-7) are connected by a β -hairpin formed by strands β 3 and β 4 and a short α -helix α 1; a long helix α 4 is located at the C-terminal end of CTD. A saddle-like RNA binding surface consists of the β -sheet2, the β -hairpin and 3 loops (the loops that connect β 5-6, β 7-8, and α 2), exhibiting a high degree of shape complementarity to the A-form dsRNA. The long helix (α 4) at the C-terminus is a unique feature of DRH-3. The helix α 4 at the C-terminus stacks on the back side of the 5'-ppp binding site. The fact that the CTD caps the dsRNA indicates that DRH-3 may preferentially bind the 5'-terminus like RIG-I and LGP2 do.

The two loops connecting strands β 1-2 and β 5-6 forms a C4 (C952, C955, C1007 and C1010) type zinc finger binding motif. The four thiol groups of these cysteines capture a zinc ion in each CTD molecule. In addition to the C4 type zinc finger nucleotide binding motif, the helix α 1 appears to support the C4 type pocket. The zinc ion is important in maintaining the overall fold of the CTDs (Cui et al., 2008) and this C4 type zinc finger binding motif is conserved in both DRH-3 and RLR CTDs (X. Li, Lu, et al., 2009; X. Li, Ranjith-Kumar, et al., 2009; C. Lu et al., 2011; C. Lu et al., 2010).

3.5.2. The structure basis of 5'-ppp RNA recognition by DRH-3 CTD

DRH-3 CTD recognizes the 5'-ppp dsRNA through extensive electrostatic interactions (**Figure 3.19B** and **3.20A**). Positively charged residues are involved in 5'-ppp RNA binding: K992 and R1016 with γ -phosphate; K990 and K993 with γ - and probably β -phosphates; and K988 with α - and probably β -phosphates. The residues K1048 and S1050 contribute to a second positive patch on the CTD surface, binding to the

phosphate backbone of the 5'-strand (**Figure 3.19B** and **3.20C**). In addition, the 5'-strand also interacts with CTD through hydrogen bonds, where residues N971, Y1018 and Q1027 make contacts with G1, and F969 and D970 interact with G2 (**Figure 3.19B** and **3.20B**). In contrast to the extensive interactions at 5'-strand, the 3'-strand RNA makes relatively less interaction with DRH-3. The 3'-strand RNA interacts with residues Y994, N999, N971 and F969 through either stacking interactions or hydrogen bonds. The electrostatic potential at the solvent-accessible surface of DRH-3 CTD is shown on **Figure 3.21**. Two positively charged patches corresponding to the 5'-ppp recognition site and the phosphate backbone binding site were identified respectively.

Figure 3.20 Molecular details of DRH-3 CTD-dsRNA interactions. (A) Interactions between the 5'-ppp of the RNA and DRH-3 CTD. The first nucleotide 5'-ppp G₁ of 5'-strand and the key residues involved in RNA binding are shown as sticks. The dotted lines represent atomic distances of the polar bonds. LYS988, LYS990, SER922, LYS933 and ARG1016 are the residues that recognize the 5'-ppp cap. (B) Interactions between the 5'G₁G₂ of the RNA and DRH-3 CTD. The first two nucleotides G₁ and G₂ of 5'-strand, the last nucleotides C₈ or C₁₂ of the 3'-strand and the key residues involved in RNA binding are shown as sticks. The dotted lines represent atomic distances of the polar bonds. 5'-G₁ is recognized mainly by TYR1018, GLN1027 and ASA971. ASA971 also recognizes 3'-C₁ in dsRNA. TYR994 and PHE969 contribute to the dsRNA stabilization by providing two hydrophobic interactions. (C) Interactions between the phosphate backbone of the RNA and the KWK motif of DRH-3 CTD. The 5'-strand RNAs and the key residues involved in RNA binding are shown as sticks. The dotted lines represent atomic distances of the polar bonds. The C4 phosphate forms interactions with both LYS1048 and SER1050.

Figure 3.21 Electrostatic surface of DRH-3 CTD. DRH-3 and its RNA ligand 5'-ppp 12-mer dsRNA are shown as cartoons. Surface of DRH-3 is represented by different colors corresponding to the electrostatic charges. Positively charged surfaces are colored in blue and negatively charged surfaces are colored in red, while neutral surfaces are colored in white.

To understand the binding mechanism for 5'-ppp ssRNA, we solved another crystal structure of DRH-3 CTD in complex with 5'-ppp 8-mer GC-rich ssRNA at 1.89 Å resolution (**Figure 3.22A; Table 3.3**). In contrast to the structure for dsRNA, the ssRNA-bound CTD was solved in a different space group (P22₁2₁) where one CTD molecule and one ssRNA were determined in an asymmetric unit (**Figure 3.22A**). The recognitions by DRH-3 CTD for ssRNA and dsRNA are highly similar (**Figure 3.22C**). Superposition of the two DRH-3 CTD structures gave a RMSD value 0.495 Å for all Cα atoms (**Figure 3.22B**). The 5'-ppp binding site is conserved between the two structures. The 5'-ppp ssRNA preserves the interaction with the KWK motif but the residue K1048 interacted with the G5 phosphate at the phosphate backbone in ssRNA-

DRH-3 CTD complex instead of C4 in the dsRNA-DRH-3 CTD complex. Moreover, the 5'-ppp ssRNA binds RNAs in a slightly different orientation due to the lack of the 3'-strand whose 5'-nucleotides interact with DRH-3 CTD. The strong interactions of 5'-ppp dsRNA causes the RNA backbone to tilt away from the KWK domain and the binding at the RNA backbone may become dispensable. Thus, the laying position of the 5'-ppp dsRNA may make it more suitable for the helicase hole than the ssRNA.

Important insights into the function of DRH-3 will likely come from identifying the RNA species that reside in or are processed by this protein. Our structure observation suggests the 5'-ppp ds/ssRNAs are potential substrates or products for DRH-3.

Figure 3.22 Overall structure of DRH-3 in complex with 5'-ppp 8-mer ssRNA. (A) The crystallographic asymmetric unit contains a 1:1 complex between DRH-3 CTD and the ssRNA. The symmetry mate structure is shown in grey. The 5'-ppp is shown as orange spheres and the zinc molecule is shown as a purple sphere. (B) Superposition of DRH-3 CTD structures bound to the 5'-ppp 12-mer dsRNA and the 5'-ppp 8-mer ssRNA. (C) The schematic diagram of contacts between DRH-3 CTD and 5'-ppp 8-mer ssRNA.

3.5.3. Comparison of DRH-3 CTD and RLR CTDs

Previous studies have suggested that the DRH-3 CTD is structurally related to RLR CTDs and is proposed to confer the binding affinity to RNA ligands (Binder et al., 2011;

Cui et al., 2008). To further provide structural insight into the RNA recognition by CTDs, we compared the DRH-3 CTD with available RLRs structures. Superposition of the DRH-3 CTD with human RLR CTDs gives a RMSD value ranging from 0.9 to 2.3 Å for all C α atoms, indicating a high structural similarity (**Figure 3.23 A-C**). Notably, a major structural difference is observed at the 5'-ppp cap binding site, where the structures of β 4 and the following turn found in DRH-3 CTD are different from the corresponding structures in RLRs (**Figure 3.23 A-C**).

A lysine residue (K1048) located at the loop connecting β 9- α 2 mediates the interaction between DRH-3 CTD and the RNA sugar phosphate backbone. This lysine residue is conserved among RLR CTDs and this binding site is referred as the KWK motif or RNA binding loop (Guo et al., 2013; Takahasi et al., 2009). The KWK motifs in RLRs and DRH-1 are critical for RNA recognition (Guo et al., 2013; Takahasi et al., 2009). Interestingly, MDA5 and LGP2 CTDs require additional lysine residues in the KWK motif to stabilize the 5'-strand RNA. The two lysine residues in MDA5 and LGP2 CTDs recognize the minor groove of dsRNA, revealing the canonical mode of RNA binding by dsRNA binding domains (Gleghorn & Maquat, 2014). Unlike the CTDs in MDA5 and LGP2, DRH-3 and RIG-I lack of the second lysin residue and bind to both dsRNAs and ssRNAs.

Figure 3.23 Comparison of the C-terminal domains in DRH-3 and RLRs. (A) Superposition of the RNA bound CTDs in DRH-3 (yellow) and LGP2 (PDB code: 3EQT, orange). (B) Superposition of the RNA bound CTDs in DRH-3 (yellow) and MDA5 (PDB code: 4GL2, green). (C) Superposition of the RNA bound CTDs in DRH-3 (yellow) and RIG-I (PDB code: 5F9F, blue). Polypeptide chains are shown as cartoons. Electrostatic surfaces of LGP2 CTD (D), MDA5 CTD (E) and RIG-I CTD (F) are also shown, respectively. Positively charged surfaces are colored in blue and negatively charged surfaces are colored in red. The key charged residues are indicated.

The electrostatic potential at the solvent-accessible surface of each CTD was also determined to compare the difference in surface charge distribution (**Figure 3.21** and **Figure 3.23 D-F**). In general, the positive charged groove around the KWK motif region is conserved among the CTDs. However, the 5'-ppp binding site found in DRH-3 and RIG-I CTDs is missing in MDA5 and LGP2. The high affinity of RIG-I CTD to 5'-ppp termini (0.5 nM for a 5'-ppp 14-mer dsRNA vs 5.9 μM for 5'-OH 14-mer dsRNA), which ensures its specific and efficient binding to different forms of RNAs, has been reported previously (Vela et al., 2012). Therefore, the way DRH-3 CTD recognizes

RNA termini and interacts with the 5'-ppp may be similar to RIG-I CTD (Ciccia et al., 2008; Lulchev & Klostermeier, 2014; Shereda et al., 2009).

3.5.4. DRH-3 CTD binds 5'-ppp dsRNA with high affinity

To better understand the RNA binding preference, fluorescence polarization (FP) based competition assay was used to quantitatively assess DRH-3 CTD binding affinity to different types of RNA ligands. A FAM-labelled hairpin RNA (RNA 12, **Table 3.2**) was used as the fluorescent probe and its binding affinity to DRH-3 CTD was determined with a K_d of 325 nM (**Figure 3.24**).

Figure 3.24 Direct FP binding assay to determine K_d of the DRH-3 CTD. To characterize the baseline FP signal, the proteins were serially diluted in the concentration range of 1–5000 nM. The FP signals were measured and plotted as a function of protein concentration. Average values $\pm$ standard error (SE) from three measurements per condition are shown in the graph.

For the FP competition assay, we titrated each RNA competitor (**Figure 3.25A**) and calculated concentrations that resulted in reduction of polarity to 50 % (IC_{50} values) of the maximum value from the dose-response data (**Figure 3.25B**). In agreement with our findings in crystal structures where 5'-ppp RNA ligands make extensive interactions to

DRH-3 CTD, 5'-ppp RNAs including 5' or 3' overhang RNAs and blunt-ended RNAs (R1, R2 and R3 in **Figure 3.25A**) show the highest competition with the probe (IC_{50} values ranging from 80 nM to 234 nM), indicating that DRH-3 CTD preferentially recognizes 5'-ppp RNAs. In contrast, the 5'-OH and 5'-p RNAs (R1P and R1OH in **Figure 3.25A**) compete with the RNA probe poorly, showing relatively high IC_{50} values (631 nM to 752 nM).

A

B

Figure 3.25 FP competition assay of DRH-3.(A) RNA oligonucleotides with different 5' end modifications. (B) Validation of the RNA binding affinity of DRH-3 CTD by FP competition assay. Percentage inhibition of the FAM-labelled 5'-OH 18-mer hairpin RNA binding versus RNA competitor concentration response curves. The titration curves were fitted using GraphPad prism and the IC_{50} values were calculated from the titration curves. All data are mean $\pm$ SD for three replicates.

To test whether the RNA binding is affected by the 5'-modification, the free DRH-3 CTD (19.9 kDa) and the DRH-3 CTD loaded with various RNAs (RNA 7-11, **Table 3.2**) were digested by trypsin at different time points. Trypsin liberates two major polypeptide fragments from free DRH-3 CTD. It was expected that the dsRNA binding could protect DRH-3 CTD from trypsin digestion. However, DRH-3 CTD pre-incubated with the RNAs resulted in the same digestion pattern as the free protein. A possible reason could be that DRH-3 CTD was highly sensitive to protease even at a mild trypsin digestion condition (1:2000 of trypsin to target protein ratio) and during a short period of time. As a comparison and an experimental control, the trypsin digestion assay was performed on RIG-I CTD incubated with RNA ligands. In contrast, all of the tested 12-mer dsRNAs, especially the blunt-ended 5'-ppp RNA (R1 in **Figure 3.25A**), were shown to bind to RIG-I CTD and protect the protein from trypsin digestion (**Figure 3.26**).

A. DRH-3 CTD

B. RIG-I CTD

Figure 3.26 Effect of RNAs on CTDs limited proteolysis. The limited proteolysis analysis of His₆-tagged DRH-3 CTD (A) and SUMO-tagged RIG-I CTD (B). Proteins with different RNAs (Protein: RNA molar ratio = 1: 2) were digested with trypsin at different duration and quenched by adding SDS protein loading dye. Denatured proteins were resolved on reducing SDS-PAGE.

To test the stability of the protein-RNA complexes, thermal shift assay was performed. In this assay, we quantified the effects of RNAs on the melting temperatures (T_m) of a protein by gradually increasing the temperature to unfold the protein. The thermal stability curve and its midpoint value were obtained by measuring the fluorescence at each temperature point using a fluorescence-enabled thermal cycler. When an RNA binds and stabilizes the protein, an increase in T_m could be observed. We adopted the thermal shift assay as described in Materials and Methods, to determine the T_m of CTDs and to evaluate how their thermal stabilities are impacted by RNAs. This assay can be applied to characterize stable CTD-RNA complexes which can be used for crystallization. We first verified our method on RIG-I CTD proteins. All the 12-mer RNAs increased the thermal stability of RIG-I CTD, especially the 5'-ppp with 3'-overhang (ΔT_m up to +11 °C) (**Figure 3.27**).

Free DRH-3 CTD exhibits a T_m of 47 °C under the experimental condition. Two truncated DRH-3 CTD proteins (namely DRH-3_C1 and DRH-3_C2) have the same T_m . The initial result shows that the tested RNA ligands do not change the melting temperature of DRH-3 CTD ($T_m = 47$ °C), indicating that binding of these RNAs failed to increase the thermal stability of DRH-3 CTD (**Figure 3 31.27**). Since the T_m of the DRH-3 CTD may be dependent on the concentration of RNA ligands we increased the RNA concentration to 5-fold (50 μ M) of the protein concentration. However, the T_m of the CTD-RNA remained the same.

Together, although crystal structures and the FP competition assay confirm the 5'-ppp RNA preference for DRH-3 CTD, both limited proteolysis and thermal stability methods failed to confirm that this preference is true for DRH-3 CTD.

Figure 3.27 Stabilization of CTDs by RNA binding. Thermal shift assay results for DRH-3 CTD and RIG-I CTD in the absence (buffer as negative control) and presence of test RNAs. The melting temperature T_m values (°C) are shown above the bars. All data are mean $\pm$ SEM for two independent experiments done in triplicate.

3.6.HDX-MS study of RNA-induced conformational changes in full-length DRH-3

In order to solve the structure of full-length DRH-3, we set up a series of crystal screening to crystallize the full-length DRH-3 in the presence of various 5'-ppp RNAs. Unfortunately, we were unsuccessful in finding any crystal hit. As an alternative strategy, we applied the HDX-MS method to explore the protein dynamics and RNA binding signatures of full-length DRH-3. HDX-MS has emerged as a sensitive way to study protein folding, protein-ligand interaction and protein intra- and inter-molecular interactions in solution, providing dynamic structural data (Englander, 2006; Lisal et al., 2005; West et al., 2013; J. Zhang et al., 2010). The rate of backbone amide hydrogen exchanges with deuterium is measurable using mass spectrometer (MS) (Konermann,

Rodriguez, & Sowole, 2014). Given solvent accessibility and structure stability, different regions on a protein or the same protein region in different states can become deuterated at different rates and to different extents. The differential hydrogen deuterium exchange events measured by MS can be mapped to the primary sequence and structure model of the protein for analysis.

In this study, HDX-MS was combined with X-ray crystallography to dissect the overall conformation details of DRH-3 with and without the presence of distinct RNA species (Konermann, Pan, & Liu, 2011; Konermann et al., 2014; Zheng et al., 2015). The full-length DRH-3 model was generated from the predicted helicase domain (I-TASSER) (Y. Zhang, 2008) and the crystal structural models of NTD and CTD solved in this study (**Figure 3.28**).

3.6.1. Protein dynamics of DRH-3

To investigate the domain-domain interaction within DRH-3, we compared the HDX data of full-length DRH-3 protein and DRH-3 NTD (DRH-3_N1). The majority of peptides on DRH-3 NTD (86 %) exhibit a magnitude of deuterium uptake less than 50 % (**Figure 3.27A**). This relatively low H/D exchange suggests a rigid conformation of DRH-3 NTD which is consistent with the crystal structure. The NTD in full-length DRH-3 deuterated at similar rates and to similar extents to the free NTD (**Figure 3.27B**), indicating that DRH-3 NTD does not interact with the other domains within DRH-3. The linker region (I351-E368) between NTD and helicase domain (HEL) is more solvent exposed (70 % D), suggesting that NTD is attached to the helicase domain through a flexible linker. In conclusion, the NTD as a rigid domain may swing around instead of interacting with the rest part of the protein.

A DRH-3 NTD

B full length DRH-3

HDX level (%)

Figure 3.28 Conformational dynamics of DRH-3 NTD and full-length DRH-3. The full-length DRH-3 is shown by assembling all three domains via flexible loops. Deuterium uptake profiles of DRH-3 NTD (A) and full-length DRH-3 (B) are presented in a front and back view. The % deuterium exchange of each peptide is colored according to the scaled bar. Blue indicates the lowest and orange indicates the highest exchange rate.

3.6.2 The 5'-ppp RNA binding mechanism for DRH-3

To explore RNA binding mechanism on full-length DRH-3, we also performed the HDX-MS experiment for DRH-3 in the presence of a short hairpin RNA, 3p10L (5'-ppp GGACGUACGUUUCGACGUACGUCC), which has been shown to bind and activate RIG-I *in vitro* and in cell-based IFN production assay (Jonsson et al., 2017; X. Sun et al., 2016). The difference in H/D incorporation for DRH-3 with and without 3p10L was

mapped to the primary sequence and the structure model of full-length DRH-3 (**Figure 3.29**). In the presence of 3p10L, several conserved motifs of the helicase (Motif Ia, II, IVa, Vb and VI) and the CTD (the 5'-ppp binding site and the KWK motif) domains displayed significantly reduced deuteration as expected, suggesting their involvement in RNA binding. A helix in Hel2i domain and the Pincer region were also affected by 3p10L binding. The observed HDX changes in Hel2i and Pincer could be due to the conformational structure changes after RNA binding. In addition, there was no difference in H/D incorporation in the NTD region was observed, suggesting that the binding of 3p10L does not affect the NTD. In contrast, the N-terminal CARD region of RIG-I has been shown the occurrence of deuterium uptake variation due to RNA-induced allosteric effects (Zheng et al., 2015). Unlike RIG-I, DRH-3 may not utilize the auto-inhibitory mechanism to regulate signaling activity.

Figure 3.29 Protein dynamics of DRH-3 in response to 3p10L binding. Cartoon representation of the DRH-3 structure bound to the 3p10L from a front (left) and back (right) view. Regions that show statistically significant reduced D uptake at 1 hour of incubation in deuterated buffer upon 3p10L binding were mapped on to the cartoon model.

3.6.3 DRH-3, RIG-I and MDA5 are activated by distinct mechanisms

The HDX patterns for RIG-I-3p10L complex and MDA5-poly (I:C) complex were discussed by Zheng and the colleagues in a previous study (Zheng et al., 2015). Here, I compared the HDX profiles of DRH-3-3p10L complex with published data.

As expected, the footprints of RNA binding on the CTDs are observed in all of these three proteins. The KWK motifs in all three CTDs are involved in RNA binding. In addition, the DRH-3 and RIG-I regions corresponding to the 5'-ppp recognition also display a high degree of protection against HDX upon binding to 3p10L. This is in agreement with the structural study of the 5'-ppp recognition by CTD in both DRH-3 and RIG-I. The conserved motifs for RNA and ATP binding in the helicases domain of these three proteins are also observed (Motif Ia, II and IVa for DRH-3; Motif Ia, Ic and Iv for RIG-I; and Motif Ic, IIa, III for MDA5).

Intra- or inter- molecular interactions can also be identified based on the different deuterium uptakes on the peptides. RIG-I adopts an auto-inhibited conformation in solution, while MDA5 does not have an auto-inhibited state. Therefore, an RNA-induced deuterium uptake difference was observed in the CARD2 and the Hel2i interface region, indicating that the binding to RNA allosterically destabilized the CARD2-HEL2i interaction by releasing the RIG-I CARD2 (Zheng et al., 2015). The interface of CARD2 and Hel2i in unligated RIG-I exhibited the greatest magnitude of protection to HDX and such a protection was greatly compromised in the presence of 3p10L. In contrast, the deuterium uptake magnitude at the NTD and Hel2i regions of DRH-3 remained unchanged before and after RNA binding. Unlike RIG-I CARDs, NTD and Hel2i in DRH-3 are not involved in the intra-molecular auto-inhibitory interactions.

The HDX-MS experiments identified several HDX resistant peptides which are located at the surface of Hel-CTD but not required for RNA binding in MDA5. Trans interaction via these peptides allows the Hel-CTDs of MDA5 to stack to each other, indicating the Hel-CTDs cooperatively bind to long dsRNA. However, the inter-protein interactions observed in MDA5 Hel-CTDs are absent in DRH-3 proteins.

The CARD2 in RIG-I plays a role in allosteric auto-inhibition through its interaction with the Hel2i domain, while MDA5 cooperatively assembles into filaments along the long dsRNA in order to activate the downstream signaling upon its helicase-CTD binding to a long dsRNA. Neither intramolecular interactions as that observed for RIG-I nor trans homotypic interactions as that for MDA5 were observed for DRH-3 in the HDX-MS experiments. Notably, HDX-MS has the limitation to identify weak interactions that are mediated entirely through side chains. Besides, different allosteric effects in RIG-I, MDA5 and DRH-3 can result in complex protection patterns. Therefore, we conclude that, in contrast to mammalian RLRs, DRH-3 may use a different mechanism in *C. elegans* RNAi pathway.

3.7. Discussion

3.7.1. DRH-3 NTD acts as a signaling domain in the RNAi pathway

In this study, we found that DRH-3 NTD is structurally similar to the death fold proteins, especially the CARD domains. Structurally, the DRH-3 NTD shares a common tertiary structure of CARD domain which is characterized as a Greek-key like folding consisting of 6 α -helices (H1 to H6) (Steward et al., 2009). The similar topology was found in the immunoglobulin-like domain, a domain involved in protein-protein interaction (Kao et al., 2015). In mammals, CARD domains play important roles in apoptosis, immune responses and cytokine processing signaling pathways, while CARDS in worms are restricted to apoptotic proteins (Kersse et al., 2011). The DRH-3 NTD is the first CARD that has been found to be involved in *C. elegans* RNAi pathway to date.

Our study indicates that it is unlikely that DRH-3 proteins can form oligomers via homotypic CARD-CARD interactions. DRH-3 exists as a monomer in solution and its

NTD does not interact with the Hel-CTD domains in the presence or absence of 3p10L binding. No auto-inhibition conformation (as seen in RIG-I) nor oligomerization (as seen in MDA5) was observed in the HDX-MS experiments of DRH-3. Since the CARD-like domain in DRH-3 are different from RLR CARDS in their sequences and their structures, DRH-3 may interact with its binding partners in a different way. One explanation could be that DRH-3 CARDS may bind to another death fold-containing domain via unknown mechanism. The heterotypic CARD-CARD interaction between DRH-3 and its binding partner may help to enhance translocation rates and provide a basis for cooperativity translocation during secondary siRNA and miRNA biogenesis.

Does DRH-3 display cooperative binding as seen in MDA5? To address this question, we performed HDX-MS experiment for DRH-3 in the presence of poly (I:C). Unfortunately, no H/D difference was found between unligated DRH-3 and DRH-3 with poly (I:C), indicating that the poly (I:C) may not bind to DRH-3. Previous studies have shown that DRH-3 displays cooperativity for a duplex RNA (about 25bp) and forms dimer (Matranga & Pyle, 2010). Small regulator RNAs range from 18 to 30 bp. We could not exclude the possibility that the length of RNA is critical to DRH-3 dimerization and too long [such as poly(I:C) as ligand for MDA5] or too short (such as 3p10L as ligand for RIG-I) would not trigger the dimerization.

3.7.2. Secondary siRNA recognition by DRH-3

Our structural studies show a highly positively charged binding site on DRH-3 CTD for 5'-ppp. This 5'-ppp binding site is also conserved in RIG-I CTD (**Figure 3**). In humans, the 5'-ppp dsRNA, as a viral PAMP, is recognized by RIG-I which activates downstream type-1 interferon response (Daffis et al., 2010; Devarkar et al., 2016). The 5'-ppp binding site of RIG-I excludes the cap structure of self RNA, thereby preventing

downstream activation (Cordin et al., 2006). In addition, RIG-I differentiates viral RNA from cellular RNA via ATP turnover. The binding affinities of 5'-OH dsRNA and 5'-ppp ssRNA are significantly low in comparison with 5'-ppp dsRNA (C. Lu et al., 2010; Y. Wang et al., 2010). ATP hydrolysis facilitates the removal of lower-affinity self-RNAs from RIG-I binding site thereby to prevent downstream signaling (Lassig et al., 2015; Takahashi et al., 2008). It is also possible that DRH-3 discriminates different types of RNAs via a similar mechanism. Our FP-based competition assays suggest that DRH-3 prefers RNAs with 5'-ppp (RdRP-generated RNAs) to RNAs with 5'-p and 5'-OH (Dicer-produced and exogenous RNAs). In *C. elegans*, RdRPs perform *de novo* transcription to produce secondary small regulatory RNAs using mRNA-derived templates (Aoki et al., 2007; Sijen et al., 2001; Smardon et al., 2000). DRH-3 together with a dual Tudor domain protein is required for the biogenesis of secondary small RNAs, including endogenous 22G and 26G siRNAs and exogenous secondary vsiRNAs (Aoki et al., 2007; Claycomb et al., 2009; Duchaine et al., 2006; Gu et al., 2009). 26G RNAs are generated by RRF-3 (RdRP), DRH-3, and ERI-5 (the dual Tudor domain protein) and require Dicer modification. Additionally, immunoprecipitation experiment showed that ERI-5 can be co-purified with DCR-1, indicating an association of the RdRP complex with Dicer. It is possible that DRH-3 is involved in Dicer activities by facilitating the dissociation of Dicer-modified small dsRNAs with 5'-p and 3'-end overhang. DRH-3 may contribute to target RNA selection via its CTD, which specifically binds 5'-ppp RNA, and hydrolyze ATP using its helicase domain like RIG-I does in mammalian innate immunity.

The RNA binding specificity of DRH-3 may also play a role in regulating the efficiency and accuracy of antiviral RNAi pathway. DRH-3 participates in the secondary vsiRNA production downstream of DRH-1 which directly interacts with the

viral-derived dsRNAs in the RNA-based antiviral RNAi pathway. It would be interesting to fully understand the function of DRH-3 in RNA-based worm antiviral immunity. Further studies should elaborate whether DRH-3 is involved in the induction of an adaptive immunity through the amplification of secondary vsiRNAs.

3.7.3. Is oligomerization required for DRH-3 function?

RIG-I binds to dsRNA termini as a monomer in the absence of ATP. However, in the presence of ATP, RIG-I assembles into a filament that propagates from the end to the interior region of the dsRNA target (Peisley et al., 2013). In contrast, MDA5 forms oligomers on long duplex dsRNA (Zheng et al., 2015). In the HDX-MS experiment, we chose RNA ligands that have been validated in RIG-I and MDA5. The HDX-MS data show that DRH-3 binds to 3p10L and the protein-RNA complex exists as a monomer in solution. The purified DRH-3 in this study exhibited rather weak ATPase activities on substrates 5'-ppp dsRNAs (8 bp to 20 bp), 3p10L and poly (I:C). Previous ATPase hydrolysis and binding assays show that a 22 bp dsRNA is able to stimulate DRH-3 ATP hydrolysis activity and assemble DRH-3 dimer (Fitzgerald et al., 2014). It seems that the oligomerization (dimerization) of DRH-3 may depend on ATP hydrolysis and binding to specific RNA ligands. One possible explanation to this discrepancy is that the RNA ligands were too short to activate the ATPase activity of DRH-3 or DRH-3 does not prefer GC-rich sequences. Future studies could be to screen the RNA substrates and identify the essential RNA features that are capable to activate the ATPase activity of DRH-3.

3.8. Summary

In this study, crystal structures of DRH-3 NTD and DRH-3 CTD with 5'-ppp RNAs were solved. The NTD of DRH-3 adopted a novel fold of tandem CARs in a parallel arrangement, suggesting that it is a signaling domain in the RNAi pathway. Crystal structures and the RNA binding assays confirmed that DRH-3 CTD preferred RNAs with 5'-ppp, which is a typical feature of secondary siRNA transcripts. The full-length DRH-3 displayed unique structural dynamics, as revealed by HDX-MS, that were distinct from RIG-I and MDA5. Collectively, the unique molecular features for DRH-3 proteins identified in this study may be important for DRH-3 to fulfil its role in worm RNAi.

Chapter 4: Duplex RNA-dependent ATPase DRH-1: insight into its role in RNA directed viral immunity

4.1. Overview

Antiviral defense is one of the major biological functions of RNAi. The RNAi directed viral immunity (RDVI) is critical for *C. elegans* to keep virus infection under control. To sufficiently curb viral infection, RDVI initiates a series of linked biochemical events that involves a large number of protein components such as Dicer which processes the viral derived dsRNA into siRNA, Argonaute protein which is the key component of the RISC, RdRP which enhances the antiviral activity via amplifying secondary siRNAs, and many types of RNA binding proteins. Although considerable efforts in identifying the different proteins involved in RDVI have been made, how those proteins contribute to antiviral RNAi remains poorly understood. Functional and structural studies of those RDVI proteins and their cofactors will give us a better understanding of the worm immunity system.

DRH-1 was first identified as a Dicer binding partner. DRH-1 forms physical interactions with DCR-1 (the only worm dicer) and RDE-4 (a dsRNA binding protein) (Duchaine et al., 2006; Tabara et al., 2002) and is essential for the primary viral small interfering RNA (vsiRNA) production. It is intriguing that this protein is only involved in antiviral RNAi pathway but not required for the endogenous and exogenous RNAi pathways (Guo et al., 2013; R. Lu et al., 2009). As a novel factor of worm RDVI, DRH-1 may play a role in guiding the single worm Dicer to specifically target the replicating viral RNAs.

DRH-1 contains helicase and C-terminal domains that are homologous to RIG-I-like cytosolic immune receptors. Previous studies have shown that the helicase and C-terminal domains of the human RIG-I can functionally replace the corresponding domains of DRH-1 in *C. elegans* in antiviral RNAi pathway, suggesting that DRH-1 acts as a virus sensor in RDVI, a role similar to that of RIG-I in humans (Guo et al., 2013). Thus, mechanistic study on the unique function of DRH-1 in the antiviral immunity will allow us to gain insights into the molecular mechanism underlying virus detection by RLRs.

So far, our knowledge about DRH-1 is still limited mainly due to the lack of molecular details about this protein and its cofactors. Previous success in studying DRH-3 using biochemical and structural strategies (see **Chapter 3 for more details**) has open a door to study DRH-1 in depth. Producing active DRH-1 proteins for enzymatic activity studies as well as structural and cofactor recognition studies will be important to understand the viral RNA sensing mechanism by DRH-1 in RDVI.

4.2. Expression and purification of DRH-1 protein

The secondary structure analysis was used as a guide for the design of my constructs. Constructs expressing DRH-1 domains with different lengths were created and tested for protein expression and solubility. The soluble constructs of DRH-1 are indicated in the **Figure 4.1**.

Figure 4.1 DRH-1 constructs with soluble expression. Upper diagram shows the domain organization of DRH-1. The N-terminal domain (NTD) was colored orange, the helicase domain was indicated in grey and the C-terminal domain (CTD) was in blue. The lower diagram shows the series of constructs were prepared for enzymatic assays and crystalization screenings. SUMO-tag was colored in yellow and His6-tag was in blue.

The *drh-1* gene was cloned into either pET-SUMO (Invitrogen) or pNIC28-Bsa4 (Addgene). The N-terminal SUMO-tag was expected to increase the protein solubility. Extensive optimizations for full-length DRH-1 overexpression in *E. coli* were carried out including cell lines, growth conditions (e.g. media, temperatures) and IPTG induction levels (**Figure 4.2**). Unfortunately, proteins expressed using all of the DRH-1 constructs were in inclusion bodies and all attempts at protein refolding were unsuccessful.

Figure 4.2 Expression trials for full-length DRH-1. Cell extract samples were analyzed on 10 % SDS-PAGE and visualized by Coomassie staining. * indicates the band corresponding to His₆-SUMO-tagged full-length DRH-1 with a molecular weight of 136.2 kDa.

Secondary structure prediction revealed that the worm specific NTD of DRH-1 contains many flexible regions and might affect the solubility of the full-length protein. I then attempted to obtain a truncated DRH-1 (DRH-1 Δ NTD) without NTD. The helicase and CTD domains of DRH-1 are closely related to the corresponding domains in RLRs and DRH-3, which are important domain for RNA recognition and ATP hydrolysis (Guo et al., 2013). Several DRH-1 Δ NTD constructs with different lengths were designed (**Figure 4.1**). The soluble expressions were obtained from two of the five DRH-1 protein constructs (DRH-1 Δ NTD_1 and DRH-1 Δ NTD_2). However, the yields were relatively low: 0.1 mg from 1 liter culture for DRH-1 Δ NTD_1 and 0.5 mg from 1 liter culture for DRH-1 Δ NTD_2. The purified DRH-1 Δ NTD proteins were analyzed on SDS-PAGE (**Figure 4.3**).

Figure 4.3 SDS-PAGE analysis of purified DRH-1ΔNTD_1 and DRH-1ΔNTD_2. *, the band correspond to DRH-1ΔNTD_1, molecular weight of 87.9 kDa. **, the band corresponding to DRH-1ΔNTD_2, molecular weight of 85.7 kDa.

Interestingly, changing the His6-tag of DRH-1ΔNTD_2 from the N-terminal to the C-terminus (DRH-1ΔNTD_3) substantially improved the expression and solubility of the protein. The yield of active DRH-1ΔNTD_3 could achieve 10 mg from 1 liter cell culture. Precipitation and degradation have been observed from this protein. The protein purification profiles of DRH-1ΔNTD_3 are shown in **Figure 4.4**. DRH-1ΔNTD_3 was eluted as a single peak with a retention volume of 76.2 ml during the gel filtration purification, indicating that this protein is in a monomeric form (84.0 kDa).

A

B

C

Figure 4.4 Purification profiles of DRH-1ΔNTD₃. (A) SDS-PAGE analysis of DRH-1ΔNTD₃ fractions in each purification step. Shown here is a Coomassie stained 10 % SDS-PAGE gel result. (B) The elution profile of DRH-1ΔNTD₃ from the Heparin column. (C) Gel filtration chromatogram of DRH-1ΔNTD₃.

The NTD is indispensable for the antiviral function of DRH-1, but the function of this domain is not clear so far (Guo et al., 2013). As shown in **Chapter 3**, the NTD of DRH-3 is composed of tandem CARDs. The NTD of DRH-1 could also fold into a similar structure due to sequence similarity. Therefore, several constructs for DRH-1 NTD were designed base on the secondary structure prediction (**Figure 4.1**). The protein DRH-1_N1 was eluted in a single monodispersed peak as a monomer in the gel filtration chromatography step and the purified monomeric DRH-1_N1 was subjected for crystallization screening (**Figure 4.5**). In contrast, the gel filtrations profile of the shorter variants (DRH-1_N2-4) contained broad peaks. The purification profile of DRH-1_N4 was shown as an example (**Figure 4.6**). The gel filtration elution fractions were analyzed by SDS-PAGE. The target proteins were present in all of the gel filtration fractions with negligible contaminations. The fractions of different oligomeric states were collected and confirmed by negative staining electron microscopy (**Figure 4.6**).

Figure 4.5. Purification of DRH-1_N1 protein with C-terminal His₆-tag. (A) Gel filtration chromatogram of DRH-1_N1. Protein absorbance at 280 nm is represented by the blue lines. The expected molecular weight of DRH-1_N1 is 30.06kDa. (B) Coomassie stained SDS-PAGE of the elution fraction from gel filtration chromatography.

Figure 4.6. Purification of DRH-1_N4 protein. (A) Gel filtration chromatogram of DRH-1_N4 after SUMO-tag cleavage. The absorbances at 280 nm and 260 nm are represented by the blue and red lines, respectively. (B) Coomassie stained SDS-PAGE of fractions from DRH-1_N4 gel filtration purification. The molecular weight of DRH-1_N4 without SUMO-tag is 22.34kDa. DRH-1_N4 eluted from mainly two peaks with retention volumes of 50 ml and 61 ml, respectively. (C-E) Transmission electron microscopy. Negative-stain transmission electron micrographs of gel filtration chromatography elution fractions of DRH-1 (C) soluble aggregation fraction, (D) oligomer fraction, (E) monomeric fraction.

Two DRH-1 CTD constructs were expressed (**Figure 4.1**). Products from both constructs were soluble with high yield and were subsequently subjected to RNA-protein binding studies and crystallization screenings. The expression and purification strategies for DRH-1 CTD were the same as the those for DRH-3 CTD. The purification profile of DRH-1_C1 was shown as an example (**Figure 4.7**). DRH-1_C1 was eluted as a monomer from the gel filtration purification (19.5 kDa). The proteins were used for further characterizations and crystallization screenings.

Figure 4.7 Purification of DRH-1_C1 by gel filtration chromatography. (A) The elution profile of full-length DRH-3 from Superdex 200 10/300 GL (GE, Healthcare). (B) The 12 % SDS-PAGE with Coomassie staining showing the purity of DRH-1_C1.

4.3. DRH-1 ATPase activity is stimulated by RNA

To investigate whether the purified recombinant DRH-1 Δ NTD possesses an ATPase activity when not bound to RNA, NADH coupled ATPase assay was performed using 20 nM of DRH-1 Δ NTD protein in the absence of RNA. The result showed that without RNA binding the free DRH-1 Δ NTD did not exhibit ATPase activity (**Figure 4.8**). It is thus possible that the ATPase activity of DRH-1 requires the binding of cofactor such

as a nucleotide substrate. Therefore, the stimulation of ATPase activity of DRH-1 Δ NTD by a 5'-ppp 12-bp RNA (RNA8, **Table 3.2**) was tested. Double-stranded and 5'-ppp are the characteristics of viral RNAs, which are the potential natural RNA ligands for DRH-1. The ATPase assay carried out by titrating substrate ATP in the presence of saturating pppGC12 revealed a $k_{cat} = 20.84 \pm 0.72 \text{ s}^{-1}$ and an apparent $K_{M,ATP} = 233.6 \pm 32.6 \text{ }\mu\text{M}$ (**Figure 4.8**). These results suggest that the ATPase activity of DRH-1 is directly linked to the RNA binding and is RNA-dependent.

Figure 4.8 ATPase activity of recombinant DRH-1 Δ NTD. The ATPase activity of free DRH-1 Δ NTD in the presence and absence of a 5'-ppp 12-bp RNA. The data were fitted to the Michaelis-Menten equation and the $K_{M,ATP}$ and k_{cat} were determined. All data are mean $\pm$ SE for three replicates.

To find out how RNA is recognized by DRH-1 and to better understand RNA-regulated allosteric mechanism of DRH-1, I set out crystallization screens of DRH-1 Δ NTD to obtain crystal structure of unligated DRH-1 Δ NTD and DRH-1 Δ NTD bound

to various 5'-ppp RNAs. Unfortunately, no promising crystal hits were found. The failure to obtain a crystal structure could be attributed to the instability of the DRH-1 Δ NTD which might precipitate or degrade before crystallization, or the instability of the complex where RNA become dissociated from the helicase core during incubation.

To determine the optimal RNA length for the activation of DRH-1 ATPase activity, we titrated 5'-ppp dsRNAs with sizes ranging from 8 to 20 base pairs in the presence of saturating ATP and measured the ATPase activity of DRH-1 Δ NTD (**Figure 4.9 and Table 4.1**). The result revealed a $K_{M,RNA}$ of 497.2 nM, 7.4 nM, 21.1 nM and 41.7 nM for 8-bp, 10-bp, 12-bp and 20-bp, respectively. Thus, 10-bp and 12-bp 5'-ppp dsRNAs have the highest affinity to DRH-1 and they both stimulated high ATPase activities, with k_{cat} = 18.94 s⁻¹ and 24.49 s⁻¹, respectively. These results indicate that DRH-1 may prefer a 5'-ppp dsRNA with 10-12 bp in length for achieving an optimal ATP hydrolysis activity. Considering that ATP hydrolysis by DRH-1 may be required for RNA differentiation, the NADH coupled ATPase assay established here could be used as a tool to study the role of DRH-1 ATPase activity in viral dsRNA stimulation.

Figure 4.9 Steady state ATP hydrolysis by DRH-1 Δ NTD_3 stimulated with RNAs of different concentrations. To determine the RNA preference of DRH-1, the apparent $K_{m,RNA}$ and k_{cat} for each RNA were calculated in the presence of saturating substrate ATP. RNAs with different length were tested. All data are mean $\pm$ SEM for three replicates.

Table 4.1 Kinetic constants of DRH-1 in complex with RNAs. The data were fitted to the Michaelis-Menten equation and k_{cat} and $K_{M,RNA}$ was determined with saturating amount of ATP.

Steady-state kinetic parameters			
RNA	k_{cat} (s ⁻¹)	$K_{M,RNA}$ (nM)	$k_{cat}/K_{M,RNA}$
8-bp	12.33 $\pm$ 1.95	479 $\pm$ 163	0.026
10-bp	18.94 $\pm$ 0.65	7.38 $\pm$ 1.30	2.566
12-bp	24.49 $\pm$ 0.85	21.10 $\pm$ 3.23	1.611
20-bp	17.91 $\pm$ 1.16	41.65 $\pm$ 9.27	0.430

4.4. DRH-1 Δ NTD RNA binding site mapping and the allosteric effects upon binding with 3p10L

To provide a mechanistic explanation for enzymatic activity stimulated by 5'-ppp RNA binding, I used HDX-MS to analyze the structural dynamics of DRH-1 Δ NTD_3 complexed with 3p10L (**Figure 4.10**). The D % uptake was reduced at various regions: the motif Ia-b, II, III which are the conserved RNA and ATP binding regions for RIG-like RNA helicases; a helix on Hel2i which also showed occurrence of deuterium reduction in RIG-I due to allosteric effects upon RNA binding; the loop connecting the motif IV and Iva which is involved in MDA5 helicase-CTD filament formation; the pincer region; the 5'-end RNA binding loop and the KWK motif in DRH-1 CTD. The HDX-MS experiment mapped the RNA binding sites where the most significant deuterium reduction rates were observed such as the motif Ia-b, II, III. Interestingly, similar to RIG-I, RNA binding also induced conformational changes (allosteric effects) of the protein structure where we found the moderate deuterium reduction on Hel2i. In addition, several loops also undergo conformational changes upon RNA binding, which are probably required for the ATP hydrolysis activity. In conclusion, 3p10L binding may induce conformational changes in DRH-1 Δ NTD_3. Notably, the observed HDX changes could not only due to shielding by the RNA but may be also caused by the conformational changes after adding RNA. Therefore, the HDX-MS data of DRH-1 should be interpreted together with other experimental data such the RNA binding assays and gel filtrations

Figure 4.10 Protein dynamic of DRH-1 Δ NTD in response to 3p10L binding. Map of the DRH-1 regions that show statistically significant reduced D uptake at 1 hour of incubation in deuterated buffer upon 3p10L binding (blue: regions with heavily reduced D % uptake; Green: regions with mild reduced D % uptake; yellow: regions with increased D % uptake). (A-B) Cartoon representation of DRH-1 Δ NTD structure bound to the 3p10L from a front (left, A) and back (right, B) view. (C-D) Surface representation of DRH-1 Δ NTD structure bound to 3p10L from a front (left, C) and back (right, D) view, as in panel.

4.5. DRH-1 CTD binds dsRNA bearing 5'-ppp and 3'-overhang with higher affinity

I have shown that DRH-1 binds to 5'-ppp dsRNA that is required for protein conformational changes and the optimal enzyme activity. To further understand the function of *C. elegans* DRH-1 in RDVI and its stimulation by RNA, I characterized its biochemical and biophysical properties *in vitro* using the methods such as limited proteolysis assay, thermal shift assay and fluorescence polarization assay, established previously for DRH-3 and RIG-I CTDs. As a homolog of viral sensor RIG-I, DRH-1 was expected to bind RNA transcripts bearing 5'-ppps. As a cofactor of Dicer, DRH-1 might also bind Dicer-produced RNAs which have 5'-monophosphate and 3' overhang. Therefore, five 12-bp RNAs with different 5'-modifications (RNA 7-11, **Table 3.2**) were selected for these studies.

Firstly, I checked the trypsin digestion pattern of DRH-1 CTD with or without RNA being loaded. The trypsin at a moderate concentration (1: 500) liberated two fragments of DRH-1 CTD in the absence of any RNA. Short dsRNA binding protected DRH-1 CTD from trypsin digestion. DRH-1 CTD loaded with 12-bp RNAs bearing 5'-ppp, 5'-p or 5'-ppp and 3' cytosine overhang is resistant to the digestion by trypsin even under the condition of more than 2-hour incubation (**Figure 4.11**). DRH-1 CTD loaded with the 12-bp RNA bearing 5'-OH (R1OH) or 5'-ppp and 5' guanine (R2) was degraded at lower efficiency compared to the free DRH-1 CTD. Therefore, the limited proteolysis results show that short dsRNAs carrying different modifications bind to DRH-1 CTD with different affinities and stabilize the proteins, making DRH-1 CTD proteins less accessible to proteolysis.

Figure 4.11 Effect of RNA binding on DRH-1 CTD proteolysis. (A) RNA oligonucleotides with different 5' end modifications. (B) Tryptic digest pattern of DRH-1 CTD proteins with dsRNA with different 5'-modifications. Limited proteolysis was performed with unloaded DRH-1CTD and DRH-1 CTD loaded with RNAs. Samples were taken at the indicated time points and subjected to SDS-PAGE.

Secondly, the thermal stability of DRH-1 CTD in complex with different RNAs was determined as a way to evaluate the RNA binding preference (**Figure 4.12**). All of the 12-bp RNAs used in this study increased the melting temperature (T_m) of DRH-1 CTD, ranging from 1 °C for R2 to the maximum of 9 °C for R3, suggesting that the short dsRNA binding, especially the binding of 5'-ppp short dsRNAs, could stabilize the

thermal stability of DRH-1 CTD. Importantly, the observed increase in the melting temperature suggests that 5'-ppp dsRNA with 3' overhang has highest binding affinity to DRH-1 CTD. These results are consistent with the trypsin digestion data.

Figure 4.12 DRH-1 CTD thermal shift assay. Thermal shift assay results for DRH-1 CTD in the absence (buffer as negative control) and presence of test RNAs. The buffer condition in this sample contains 50 mM HEPES (pH7.4), 150 mM NaCl and 0.01 % tween-20. The melting temperature T_m values (°C) are shown above the bars. All data are mean $\pm$ SEM for two independent experiments done in triplicate.

Lastly, to quantify the binding kinetics for the observed protein-RNA interaction, fluorescence polarization (FP) competition assay was performed. The equilibrium binding of the FAM-labelled RNA (as a probe) with DRH-1 CTD protein was carried out and the K_d was determined to be 203 nM (**Figure 4.13**). For the competition assay, I titrated all of the 12-bp RNAs and calculated the IC_{50} values from the curves (**Figure**

4.14). Consistent with the previous conclusion, 5'-ppp dsRNAs have higher affinity to DRH-1 CTD compared to dsRNAs with 5'-OH and 5'-p. In particular, the 5'-ppp dsRNA with 3' overhang showed the highest competition with the probe ($IC_{50} = 39$ nM), indicating DRH-1 CTD preferentially recognizes 5'-ppp RNAs with 3' overhang.

Dissection of the interactions between DRH-1 CTD and RNAs would be informative to elucidate how the RNA is recognized and may also help explain how DRH-1 selectively binds to viral RNAs. Since short 5'-ppp dsRNAs with 3' overhang exhibited better binding affinity to DRH-1 CTD, I tried to obtain a crystal structure of this stable DRH-1 CTD-RNA complex. The DRH-1 CTD proteins were first pre-incubated with the RNAs for at least 1 hour on ice. Crystallization trials using commercial screening kits were then set up. Unfortunately, no crystal hits were obtained. Future work will focus on optimizing the purity and the homogeneity of the protein-RNA complexes.

Figure 4.13 Direct FP binding assay to determine the K_d of the DRH-1 CTD to the FAM-labelled RNA probe. To characterize the baseline FP signal, the proteins were serially diluted in the concentration range of 1–5000 nM. The FP signals were measured and plotted as a function of protein concentration. Average values are shown $\pm$ SE from three measurements per condition.

Figure 4.14 FP competition assay for RNAs binding to DRH-1 CTD. The FAM-labelled RNA (5 nM) was first mixed with DRH-1 CTD (300 nM). The 2-fold serial dilutions (0-10000 nM) of the unlabeled RNAs were then added and FP signals were recorded. In all experiments, average FP values are shown $\pm$ SE from three measurements per condition.

4.6. Discussion

4.6.1. Problems in expression of the full-length DRH-1

DRH-1 contains three domains: NTD, helicase domain and CTD. There are many advantages to study full-length DRH-1 compared to truncation variants. However, in general, it is more difficult to express the full-length proteins. Many strategies are available to improve soluble protein expression in *E. coli*, including codon optimization, selection of expression strains, manipulating expression conditions, using fusion tags and chaperone co-expression (Schrodel & de Marco, 2005; Schrodel, Volz, & de Marco, 2005; Sorensen & Mortensen, 2005). In the case of expressing recombinant full-length DRH-1 in *E. coli*, I first co-expressed the recombinant protein with N-terminal SUMO-tag, which is known to enhance the solubility of their fusion partners (D. Kuo, Nie, &

Courey, 2014). Unfortunately, the expressed protein was only detected in the insoluble fraction as inclusion bodies. I further modified the growth condition and changed the expression cell lines. Unfortunately, all of my attempts failed to increase the soluble protein expression. The NTD of DRH-1 contains many flexible regions which may make it difficult to fold the full-length DRH-1 properly in bacteria. By removing the NTD, a truncated DRH-1 (named DRH-1 Δ NTD_3 in Figure 4.2), which contained only the helicase domain and the CTD, was soluble and expressed at high level in *E. coli* expression system.

4.6.2. Problems in characterization of DRH-1 NTD

Sequence analyses grouped DRH-1 to the RIG-I-like/Dicer RNA helicase family. My characterization of DRH-1 NTD was mainly focused on its unique features as compared to the rest family members. Since sequence-based secondary structure prediction reveals that DRH-1 NTD only contains α helices, I speculated that DRH-1 NTD may fold into tandem CARD domains, similar to its homologs DRH-3 and RLRs. In order to better understand the function of DRH-1 NTD, I designed and created full-length NTD or a single CARD domain of DRH-1 with different lengths for further crystallographic structural studies (**Figure 4.2**). Unfortunately, all of the recombined NTD proteins were in the soluble aggregated state according to the gel filtration chromatography analysis. The obstacle to obtain soluble DRH-1 NTD proteins was the boundary prediction. The identification of domain boundary could possibly increase the chance to get a soluble DRH-1 NTD after successful expression in *E. coli*. In addition, the C-terminal boundary for DRH-1 NTD was edited or shortened to allow for a more stable NTD product. Notably, with the availability of DRH-3 NTD structure and the growing power of computational modelling, a reliable DRH-1 NTD model could be built and used as a guide for construct design of DRH-1.

4.6.3. DRH-1 recognizes primary vsiRNA

The CTD domain is conserved among DRH-1, DRH-3 and RLRs. Multiple sequence alignments reveal that a C4 type zinc binding site and the KWK motif are also present in DRH-1 CTD, indicating that DRH-1 CTD could bind to RNA. The zinc binding site is essential for RNA binding and the KWK motif is indispensable for the antiviral activity of DRH-1 (Guo et al., 2013). Unfortunately, crystallization of DRH-1 CTD has been unsuccessful given much effort has been made. Nevertheless, I successfully carried out several protein-RNA binding assays which suggest that short dsRNAs, especially those with 5'-ppp and 3' overhang, can bind to DRH-1 CTD and stabilize the protein. Notably, the 5'-ppp is a typical molecular pattern of viral RNA and the 3' overhang is a feature of Dicer-produced RNA. Unlike plants and insects that process distinct siRNA and miRNA precursors with dedicated Dicers, *C. elegans* produces only one Dicer (Ketting et al., 2001; Lee et al., 1993). It is therefore possible that that DRH-1, which shares many protein characteristics of RLRs, may play a role in viral RNA selection before or after Dicer processing.

The ATPase activity is important for the recognition of viral dsRNA by DRH-1 and the production of vsiRNA. In this Chapter, I showed that the ATPase activity of DRH-1 is RNA-dependent and DRH-1 displays distinct ATP turnover rate in the presence of RNAs with different 5'-modifications. RIG-I, the human homolog of DRH-1, controls the viral RNAs/host RNA recognition with ATP turnover rate. The specific binding of viral RNA is coupled with a slower ATP hydrolysis, which can cause a conformation change in RIG-I and lead to downstream signaling activation. Although the exact role of DRH-1 ATPase activity in viral dsRNA processing remains unclear, it is possible that DRH-1 may utilize a RIG-I-like mechanism in RNA selection.

4.6.4. Antiviral functions of *C. elegans* DRHs and mammalian RLRs

In *C. elegans*, the viral RNAs are processed into siRNAs in stages and the siRNAs are used directly for viral gene suppression. During viral infection, DRH-1 acts to specifically enhance the production of viral primary siRNAs by facilitating the recruitment of DCR-1 and RDE-4 to the viral dsRNA (Tabara et al., 2002). DRH-3 coordinates with Dicer (DCR-1) and RdRP (RRF-1) to amplify secondary siRNAs (Aoki et al., 2007). The secondary siRNAs are then loaded to mRNA by Argonaute complex to silence or inhibit viral replication. In mammalian cells, RIG-I and MDA5 recognize distinct RNA viruses. The CARD-CARD interaction between RIG-I/MDA5 and MAVS triggers downstream activation of transcription factors NF- κ B and IRF3, which coordinately regulate the expression of type-1 interferons and pro-inflammatory cytokines (Kawai et al., 2005; Seth et al., 2005). My studies have shown that the activities of RNA detection of DRHs in *C. elegans* and RLRs in mammals seem conserved regardless of their different effector mechanisms.

Antiviral RNAi is not a perfect solution for the viral infection. Viruses can encode suppressors of RNAi to escape this antiviral defense system (Schuster, Miesen, & van Rij, 2019; van Cleef, van Mierlo, van den Beek, & van Rij, 2011). Therefore, mammals utilize a protein-based signaling pathway known as the type-1 interferon system (IFN), to fight against viral infection. Mammals also have the antiviral RNAi system. When the IFN system is suppressed, RNAi is able to participate in mammalian antiviral defense. Recent studies show that antiviral RNAi has been observed in a diverse range of cell types under specific conditions (Y. Li, Lu, Han, Fan, & Ding, 2013; Qiu et al., 2018). It seems that mammalian antiviral RNAi can provide IFN-independent protection in cancer cells via killing of specific oncolytic virus (Bastin et al., 2018). Unlike IFN signaling pathways, RNAi-based virus clearance pathway does not cause cell death.

Since RNAi machinery is conserved in mammals, perhaps RLRs also participate in mammalian RNAi like DRHs. Organisms such as mammals and *C. elegans* have only one Dicer protein (Ketting et al., 2001; Lee et al., 1993; H. Li et al., 2002; Z. Liu et al., 2018; Wilson et al., 2015). RLRs may interact with Dicer like their worm homologs and play a role in regulation of mammalian RNAi. Indeed, a recent study shows that LGP2 suppresses RNAi through inhibiting Dicer-dependent processing of long dsRNA (Maillard et al., 2016; Schuster et al., 2019; van der Veen et al., 2018).

Interestingly, DRHs and RLRs have similar domain architecture despite that they are found in different organisms and have low protein sequence similarity. DRHs specifically present in worms while RLRs predominantly exist in vertebrates. Unlike *MDA5* and *LGP2* genes which emerged from vertebrates, *RIG-I-like* genes have been found in sea urchin indicating that *RIG-I* may have appeared before the appearance of vertebrates (Hibino et al., 2006). Phylogenetic studies indicate that *C. elegans* DRHs are more close to sea urchin RIG-I-like proteins (Zou, Chang, Nie, & Secombes, 2009). These findings suggest RIG-I and DRHs might have been evolved from a same ancestor before the appearance of invertebrates. However, not all RIG-Is have tandem CARDs, xenopus and zebrafish RIG-Is contain a single CARD domain corresponding to CARD2 in human RIG-I (Zou et al., 2009). The tandem CARDs architecture of RIG-I and *MDA5* arose in the deuterostome lineage. It is still an open question that whether the CARD domains in RLRs and DRHs are coevolved with the helicase domain or grafted and assembled independently during different evolutionary events.

In addition, a recent study shows that the CARDs in *MDA5* interact with a viral RdRP (R. L. Kuo et al., 2019). It is also known that DRH-3 forms physical contacts with worm RdRP. Eukaryotic RdRPs are acquired horizontally from RNA viruses (Dalmay, Hamilton, Rudd, Angell, & Baulcombe, 2000) and RNAi is the only known function for

eukaryotic RdRPs. Only worm RNAi necessitates RdRPs (Sijen et al., 2001; Smardon et al., 2000). In contrast, *Drosophila* and mammals elicit RNAi without the involvement of an RdRP (Ghildiyal & Zamore, 2009). One possible explanation could be that while RdRP in metazoans is lost during the evolutionary process, its cofactors remain. The ancestors of RLRs are adopted by protein-based viral immunity, namely the innate immune system.

4.7. Summary

In this chapter, I described the expression and purification of several DRH-1 domains and conducted biochemical and biophysical characterization of DRH-1 Δ NTD and DRH-1 CTD. I found that ATP hydrolysis by DRH-1 was exclusively dependent on RNA and the optimal activity was achieved when binding to 10-12 bp short dsRNA. The HDX-MS analysis further suggests that RNA binding could induce a series of conformation changes in DRH-1 Δ NTD, allowing the ATPase activity to be induced. By performing limited proteolysis assay, thermal shift assay and fluorescence polarization competition assay, I found that DRH-1 CTD exhibits a higher binding affinity to dsRNAs with 5'-ppp and 3' overhang. These findings bring new insights to our understanding of how DRH-1 recognizes dsRNAs thereby to contribute to RDVI. Taken together, my results suggest that DRH-1 not only shares similar domains with RLRs but also confers similar enzymatic function such as dsRNA-dependent ATPase activity and 5'-ppp dsRNA binding. DRH-1 could potentially detect the unique viral dsRNA in a similar way as RLRs do in mammals. Moreover, the engagement of DRH-1 in worm RDVI may ensure effective DCR-1 activity by facilitating substrate selection or product release.

Chapter 5: Conclusions and future directions

Figure 5.1 Proposed roles of Dicer-related helicases in RNAi responses in *C. elegans*. Endogenous pre-siRNA and pre-miRNA are required to trigger endogenous RNAi pathway. The synthetic dsRNAs trigger the exogenous RNAi pathway while viral RNA and viral RNA replicates trigger an antiviral RNAi pathway. Various dsRNAs are recognized and cleaved by a DCR-1 into ~23-bp dsRNAs bearing 5'-p and 3'-overhang. DRH-1 facilitates DCR-1 in primary vsiRNAs production in a hierarchical RNAi pathway. DRH-1 recognizes the features e.g. 5'-ppp of viral RNAs and recruits DCR-1 and its partner RDE-4 to the viral dsRNA replication intermediate. DRH-1 also helps the release of primary siRNA by recognizing the dsRNA with 3'-overhang. Duplex siRNA/miRNAs complex with the first Argonaute protein and the passenger strand of the duplex is lost. These complexes recruit distinct RdRPs to either host or virus mRNA to synthesize ~22-bp secondary siRNA/miRNAs with 5'-ppp. DRH-3 is essential for amplification of secondary siRNA/miRNAs. The siRNA/miRNAs are loaded to the second Argonaute protein, forming different RISC complexes. Gene silencing is achieved by either mRNA degradation and/or translation silencing.

In this study, I performed extensive research on the function and mechanism of DRH-3 and DRH-1 using biochemical, biophysical and structural strategies. Findings from my study not only generated novel insights into the roles of worm DRHs in small RNA generation but also paved the way for future in-depth study on RLR functions and mechanisms (**Figure 5.1**).

In Chapter 3, I identified structure features for the NTD and CTD domains of DRH-3, which shed light on the functional role of DRH-3 in RNAi in *C. elegans*. To better understand the function of DRH-3 NTD, I solved the crystal structure of DRH-3 NTD after overcoming the phasing problem. The DRH-3 NTD is composed of tandem CARDs, suggesting that it may facilitate protein-protein interaction in RNAi, similar to what other CARD domains do. The CARD-containing proteins either participate in inflammation or apoptosis (Steward et al., 2009). My study, for the first time, revealed that a CARD-containing protein can also contribute to the RNAi pathway.

Importantly, the tandem CARDs of DRH-3 are different from those of RIG-I in that the DRH-3 CARDs are in parallel arrangement instead of the head-to-tail arrangement found in RIG-I. In addition, the DRH-3 NTD exists as a monomer in solution while the

RLR CARDS fold into homotypic oligomers. Probably, DRH-3 interacts with partner proteins to form a heterotypic complex, via its N-terminal CARDS to achieve signaling function in RNAi.

The preferred RNA target of the DRH-3 was not clear although it has been shown that dsRNA instead of ssRNA was required to activate the ATPase activity in a sequence independent manner (Fitzgerald et al., 2014). To identify the type of RNAs recognized by DRH-3, I provided structural insights into the RNA recognition by DRH-3 CTD. Crystallization of DRH-3 CTD revealed a conserved C4 type RNA binding domain. A 5'-ppp binding site contains numerous positively charged residues that form extensive electrostatic interactions with the 5'-ppp RNA. DRH-3 CTD (via its conserved KWK motif) interacts with the RNA backbone through limited interactions and in a sequence independent manner. To better understand the RNA preference of DRH-3 to 5'-ppp RNAs I compared the DRH-3 CTD crystal structure to those of RLR CTDs. I found that the 5'-ppp binding site in DRH-3 CTD is similar to the one in RIG-I but appears absent in LGP2 and MDA5. Thus, my structural studies on DRH-3 CTD revealed the importance of 5'-ppp for the selectivity of RNAs. The 5'-ppp modification is an important feature of the secondary siRNA. My structural studies on DRH-3 help explain the mechanism of secondary siRNA recognition by DRH-3.

To better understand how RNA binding stabilizes DRH-3 I selected several 12-bp RNAs carrying different 5'-modifications and tested the RNA binding effects on DRH-3 CTD stability. Probably due to the intrinsic instability of DRH-3 CTD even when forming complex with RNA ligands, both thermal shift assay and limited proteolysis assay failed to show any difference between the unligated CTD and the RNA bounded CTDs. Therefore, instead of testing protein stability, I performed FP competition assay as an alternative strategy to elucidate the RNA binding to DRH-3 CTD and quantify the

binding affinity. In agreement with the crystal structures, the FP competition results confirmed that DRH-3 CTD bound to 5'-ppp RNAs with higher affinity than 5'-p and 5'-OH RNAs. I believe that further studies on the role of DRH-3 CTD should be pursued, considering that deletions and substitution of the key residues at the RNA binding sites could alter the binding activity of the CTD.

It was hypothesized that DRH-3 may facilitate the translocation of RdRP along the template RNA and the RNA-dependent ATPase function of DRH-3 is essential for this role (Aoki et al., 2007). To test this hypothesis, it is important to conduct enzyme kinetic analysis for DRH-3 ATPase activity. The ATPase activity of DRH-3 is largely dependent on the type of RNA ligands employed. Future works should consider choosing dsRNAs with 5' or 3' overhang, mimicking RNA transcript intermediates or dsRNA in a length of 22-bp, mimicking the secondary siRNA products as ligands to trigger the ATPase activity of DRH-3.

My HDX-MS study suggests that full-length DRH-3 is an allosteric enzyme regulated by RNA binding. DRH-3 conformational changes in response to ATP hydrolysis is still unclear. Domains of DRH-3 may cooperate with each other to facilitate the interaction with RNA and substrate ATP for the enzymatic activity. Dissecting the allosteric regulation mechanism of DRH-3 may allow better understanding of DRH-3 activation mechanism. In this regard, protein dynamic studies of full-length DRH-3 in the presence and absence of natural RNA ligand and substrate ATP would be particularly interesting. A crystal structure of the full-length DRH-3 protein in complex with ATP or RNA would also be important to achieve this goal.

I also checked the oligomeric state of free full-length DRH-3 or full-length DRH-3 complexed with a short hairpin RNA. I found both of them were in monomeric states in

solution. Oligomerization is important for the signaling function of CARD-containing proteins (Steward et al., 2009). Currently it remains possible that other CARD-containing proteins also contribute to RNAi. It is thus likely that DRH-3 interacts with other CARD-containing proteins via heterotypical CARD-CARD interactions.

Although physical interaction between DRH-3 and the worm RdRPs has been demonstrated previously (Aoki et al., 2007; Gu et al., 2009; Vasale et al., 2010), currently we know very little about the macromolecular organization of the RdRP-DRH-3 complex. Therefore, to better understand the signaling mechanism of DRH-3 and the secondary siRNA generation mechanism of the RdRP complex, future studies should be focused on the identification of DRH-3 binding partners. A pull-down assay is feasible to isolate the binding partners of DRH-3 from *C. elegans* cell lysates using NTD or full-length DRH-3 as bait. Subsequently the protein identification can be done by mass spectrometry. The protein-protein interaction characteristics can be revealed by HDX-MS and the protein crystallography. Last but not least, it is also important to confirm all the findings *in vivo*, including identify the signaling pathway examined by the binding partners of DRH-3.

In chapter 4, I characterized another RNA helicase, DRH-1. Although DRH-1 is the worm ortholog of DRH-3, it plays a totally different role in RNAi (**Figure 5.1**). DRH-1 is important for the antiviral RNAi pathway and identified as a Dicer binding partner (Tabara et al., 2002). The mechanism by which DRH-1 facilitates the worm Dicer, DCR-1, in the biogenesis of vsiRNAs remains largely unclear. We sought to understand the function of DRH-1 by investigating the RNA binding and enzymatic activities of DRH-1 by conducting a series of biochemical and biophysical analysis.

I first conducted a comparative sequence analysis on DRH-1 and its homologs and

identified several structural motifs essential for ATP hydrolysis and RNA binding. DRH-1 is a bona fide member of the RIG-I-like/Dicer RNA helicase family and a likely duplex RNA-dependent ATPase. To test this hypothesis, I cloned and expressed recombinant DRH-1 protein without the N-terminal domain. In the absence of RNA, DRH-1 Hel-CTD did not exhibit any ATPase activity. However, in the presence of dsRNAs its ATPase activity was observed, suggesting that DRH-1 is indeed an RNA-dependent ATPase and activation by RNA appears to be a feature conserved in the RIG-I-like/Dicer RNA helicase family. No protein in this family has been shown to have duplex-unwinding capability so far. It will be interesting to test if DRH-1 unwind dsRNA in the future studies.

To test RNA binding specificity, thermal shift assay, limited proteolysis assay and FP competition assay were carried out to profile various distinct RNA ligands. The 5'-ppp RNAs were found to have higher binding affinities to DRH-1 CTD compared to 5'-p and 5'-OH RNAs. In addition, the 5'-ppp 12-bp RNA with 3' overhang exhibited the most dramatic changes in both polarization signal and melting temperature, indicating strong RNA recognition by DRH-1 CTD. Moreover, the stimulation of DRH-1 ATPase activity using 5'-ppp dsRNAs as substrates was also validated.

DCR-1, the single Dicer in *C. elegans* is required for processing distinct types of small RNAs (Ketting et al., 2001; Lee et al., 1993). DRH-1 may be one of the binding partners of worm Dicer, which aids in the specific viral RNA recognition and vsiRNA processing. Little is known about how DRH-1 coordinates its ATPase activity with viral RNA recognition. Although my study has been demonstrated that DRH-1 exhibits higher binding affinities and ATPase activities to short dsRNAs with 5'-ppp and probably 3' overhang *in vitro* the RNA ligands bound by DRH-1 *in vivo* are unclear. A

possible RNA ligand for DRH-1 could be the viral dsRNAs bearing 5'-ppp and the Dicer processed siRNA bearing 5'-p. Further biochemical analysis combined with mutagenesis analysis may help explain the mechanism for RNA differentiation and ATP hydrolysis by DRH-1. Assays established in this study can be used to identify true RNA ligands for DRH-1.

The crystal structure of DRH-3-CTD complexed with 5'-ppp RNAs successfully uncovered the RNA recognition by DRH-3 at molecular level. However, the structure of DRH-1 has yet to be determined. A short-term goal would be to unravel the molecular details of DRH-1. It would be interesting to see if DRH-1 exhibits similar NTD structure and RNA preference compared with DRH-3. More effort should be focused on the design of additional DRH-1 expressing constructs and crystallization screen of full-length DRH-1, DRH-1 NTD and DRH-1 CTD in the presence or absence of RNAs.

My study on DRH-3 NTD suggests that the NTD domain of DRH-1 could also be a CARD-like domain. Moreover, the NTD of DRH-1 has been shown to be essential for antiviral activity (Guo et al., 2013). The contribution of DRH-1 NTD in antiviral RNA probably relies on the protein-protein interactions in the signaling cascade. Although no other binding partners of DRH-1, in addition to DCR-1 and RDE-4 (Duchaine et al., 2006; Tabara et al., 2002), have been reported so far. However, it remains possible that DRH-1 interacts with yet unknown proteins thereby contribute to antiviral RNAi. Thus, identification of novel DRH-1 partner proteins has the promise to better understand the function and mechanism of DRH-1. Co-immunoprecipitation (Co-IP) can be used as an *in vitro* strategy to detect protein-protein interactions between DRH-1 and its partner proteins.

Like that found for DRH-3, crystal structures of DRH-1 CTD complexed with small RNAs could be helpful to understand the RNA recognition by DRH-1. Protein

purification protocols and the protein-RNA interaction assays for DRH-1 CTD have been established in this study. Currently I am working on crystalizing DRH-1 CTD in the presence of various 5'-ppp RNA ligands. Systematically testing the conditions to crystallize the DRH-1 CTD in the presence RNAs will be my next move.

RNA binding causes a series of conformational changes in DRH-1 helicase and CTD domains as revealed by the HDX-MS assay. The allosteric regulation mechanism of DRH-1 may be regulated by the RNA binding during ATP hydrolysis. The allosteric regulation is known to be critical to RLRs and my study suggests that DRH-3 utilizes a different mechanism for allosteric regulation. Further studies including extensive HDX-MS analyses and crystal structure solving will be needed to resolve this puzzle.

DRHs are important RNA helicases implicated in the RNAi mechanism. Findings in this study have helped define biological roles for worm DRHs in RNA silencing pathways. My successes in this study suggest that the roles and mode of action for worm DRHs can be fully characterized by studying their enzymatic activities, their preferred RNA substrates and their crystal structures.

Reference

- Abdel-Monem, M., Durwald, H., & Hoffmann-Berling, H. (1976). Enzymic unwinding of DNA. 2. Chain separation by an ATP-dependent DNA unwinding enzyme. *Eur J Biochem*, 65(2), 441-449.
- Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., . . . Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 2), 213-221.
- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., . . . Terwilliger, T. C. (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr*, 58(Pt 11), 1948-1954.
- Akira, S., Uematsu, S., & Takeuchi, O. (2006). Pathogen recognition and innate immunity. *Cell*, 124(4), 783-801.
- Aoki, K., Moriguchi, H., Yoshioka, T., Okawa, K., & Tabara, H. (2007). In vitro analyses of the production and activity of secondary small interfering RNAs in *C. elegans*. *Embo j*, 26(24), 5007-5019.
- Ashe, A., Belicard, T., Le Pen, J., Sarkies, P., Frezal, L., Lehrbach, N. J., . . . Miska, E. A. (2013). A deletion polymorphism in the *Caenorhabditis elegans* RIG-I homolog disables viral RNA dicing and antiviral immunity. *Elife*, 2, e00994.
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116(2), 281-297.
- Bastin, D., Aitken, A. S., Pelin, A., Pikor, L. A., Crupi, M. J. F., Huh, M. S., . . . Ilkow, C. S. (2018). Enhanced susceptibility of cancer cells to oncolytic rhabdovirotherapy by expression of Nodamura virus protein B2 as a suppressor of RNA interference. *J Immunother Cancer*, 6(1), 62.
- Batista, P. J., Ruby, J. G., Claycomb, J. M., Chiang, R., Fahlgren, N., Kasschau, K. D., . . . Mello, C. C. (2008). PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol Cell*, 31(1), 67-78.
- Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R., & Leslie, A. G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr*, 67(Pt 4), 271-281.
- Berke, I. C., & Modis, Y. (2012). MDA5 cooperatively forms dimers and ATP-sensitive filaments upon binding double-stranded RNA. *Embo j*, 31(7), 1714-1726.
- Beyer, D. C., Ghoneim, M. K., & Spies, M. (2013). Structure and Mechanisms of SF2 DNA Helicases. *Adv Exp Med Biol*, 767, 47-73.
- Binder, M., Eberle, F., Seitz, S., Mucke, N., Huber, C. M., Kiani, N., . . . Bartenschlager, R. (2011). Molecular mechanism of signal perception and integration by the innate immune sensor retinoic acid-inducible gene-I (RIG-I). *J Biol Chem*, 286(31), 27278-27287.
- Boggon, T. J., & Shapiro, L. (2000). Screening for phasing atoms in protein crystallography. *Structure*, 8(7), R143-R149.
- Bora, R. S., Gupta, D., Mukkur, T. K., & Saini, K. S. (2012). RNA interference therapeutics for cancer: challenges and opportunities (review). *Mol Med Rep*, 6(1), 9-15.
- Bouchier-Hayes, L., & Martin, S. J. (2002). CARD games in apoptosis and immunity. *EMBO Rep*, 3(7), 616-621.

- Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*, 128(6), 1089-1103.
- Bruns, A. M., & Horvath, C. M. (2012). Activation of RIG-I-like receptor signal transduction. *Crit Rev Biochem Mol Biol*, 47(2), 194-206.
- Bruns, A. M., Pollpeter, D., Hadizadeh, N., Myong, S., Marko, J. F., & Horvath, C. M. (2013). ATP hydrolysis enhances RNA recognition and antiviral signal transduction by the innate immune sensor, laboratory of genetics and physiology 2 (LGP2). *J Biol Chem*, 288(2), 938-946.
- Byrd, A. K., & Raney, K. D. (2012). Superfamily 2 helicases. *Front Biosci (Landmark Ed)*, 17, 2070-2088.
- Chapman, E. J., & Carrington, J. C. (2007). Specialization and evolution of endogenous small RNA pathways. *Nat Rev Genet*, 8(11), 884-896.
- Choudhuri, S. (2010). Small noncoding RNAs: biogenesis, function, and emerging significance in toxicology. *J Biochem Mol Toxicol*, 24(3), 195-216.
- Ciccia, A., McDonald, N., & West, S. C. (2008). Structural and functional relationships of the XPF/MUS81 family of proteins. *Annu Rev Biochem*, 77, 259-287.
- Civril, F., Bennett, M., Moldt, M., Deimling, T., Witte, G., Schiesser, S., . . . Hopfner, K. P. (2011). The RIG-I ATPase domain structure reveals insights into ATP-dependent antiviral signalling. *EMBO Rep*, 12(11), 1127-1134.
- Claycomb, J. M., Batista, P. J., Pang, K. M., Gu, W., Vasale, J. J., van Wolfswinkel, J. C., . . . Mello, C. C. (2009). The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. *Cell*, 139(1), 123-134.
- Coffman, S. R., Lu, J., Guo, X., Zhong, J., Jiang, H., Broitman-Maduro, G., . . . Ding, S. W. (2017). *Caenorhabditis elegans* RIG-I Homolog Mediates Antiviral RNA Interference Downstream of Dicer-Dependent Biogenesis of Viral Small Interfering RNAs. *MBio*, 8(2).
- Cogoni, C., & Macino, G. (1999). Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase. *Nature*, 399(6732), 166-169.
- Cordin, O., Banroques, J., Tanner, N. K., & Linder, P. (2006). The DEAD-box protein family of RNA helicases. *Gene*, 367, 17-37.
- Cui, S., Eisenacher, K., Kirchhofer, A., Brzozka, K., Lammens, A., Lammens, K., . . . Hopfner, K. P. (2008). The C-terminal regulatory domain is the RNA 5'-triphosphate sensor of RIG-I. *Mol Cell*, 29(2), 169-179.
- Czech, B., & Hannon, G. J. (2011). Small RNA sorting: matchmaking for Argonautes. *Nat Rev Genet*, 12(1), 19-31.
- D'Cruz, A. A., Kershaw, N. J., Hayman, T. J., Linossi, E. M., Chiang, J. J., Wang, M. K., . . . Nicholson, S. E. (2018). Identification of a second binding site on the TRIM25 B30.2 domain. *Biochem J*, 475(2), 429-440.
- Daffis, S., Szretter, K. J., Schriewer, J., Li, J., Youn, S., Errett, J., . . . Diamond, M. S. (2010). 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature*, 468(7322), 452-456.
- Dalmay, T., Hamilton, A., Rudd, S., Angell, S., & Baulcombe, D. C. (2000). An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell*, 101(5), 543-553.
- Deimling, T., Cui, S., Lammens, K., Hopfner, K. P., & Witte, G. (2014). Crystal and solution structure of the human RIG-I SF2 domain. *Acta Crystallogr F Struct Biol Commun*, 70(Pt 8), 1027-1031.

- Devarkar, S. C., Wang, C., Miller, M. T., Ramanathan, A., Jiang, F., Khan, A. G., . . . Marcotrigiano, J. (2016). Structural basis for m7G recognition and 2'-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I. *Proc Natl Acad Sci U S A*, 113(3), 596-601.
- Dillingham, M. S. (2011). Superfamily I helicases as modular components of DNA-processing machines. *Biochem Soc Trans*, 39(2), 413-423.
- Ding, S. W. (2010). RNA-based antiviral immunity. *Nat Rev Immunol*, 10(9), 632-644.
- Drenth, J. (1994). *Principles of Protein X-ray Crystallography*.: New York: Springer-Verlag.
- Duchaine, T. F., Wohlschlegel, J. A., Kennedy, S., Bei, Y., Conte, D., Jr., Pang, K., . . . Mello, C. C. (2006). Functional proteomics reveals the biochemical niche of C. elegans DCR-1 in multiple small-RNA-mediated pathways. *Cell*, 124(2), 343-354.
- Eki, T., Ishihara, T., Katsura, I., & Hanaoka, F. (2007). A genome-wide survey and systematic RNAi-based characterization of helicase-like genes in Caenorhabditis elegans. *DNA Res*, 14(4), 183-199.
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., . . . Finn, R. D. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res*, 47(D1), D427-d432.
- Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., & Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, 411(6836), 494-498.
- Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 4), 486-501.
- Englander, S. W. (2006). Hydrogen exchange and mass spectrometry: A historical perspective. *J Am Soc Mass Spectrom*, 17(11), 1481-1489.
- Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr*, 62(Pt 1), 72-82.
- Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr*, 69(Pt 7), 1204-1214.
- Fairman-Williams, M. E., Guenther, U. P., & Jankowsky, E. (2010). SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol*, 20(3), 313-324.
- Fairman-Williams, M. E., & Jankowsky, E. (2012). Unwinding initiation by the viral RNA helicase NPH-II. *J Mol Biol*, 415(5), 819-832.
- Ferrage, F., Dutta, K., Nistal-Villan, E., Patel, J. R., Sanchez-Aparicio, M. T., De Ioannes, P., . . . Aggarwal, A. K. (2012). Structure and dynamics of the second CARD of human RIG-I provide mechanistic insights into regulation of RIG-I activation. *Structure*, 20(12), 2048-2061.
- Ferrao, R., & Wu, H. (2012). Helical assembly in the death domain (DD) superfamily. *Curr Opin Struct Biol*, 22(2), 241-247.
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., . . . Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Res*, 42(Database issue), D222-230.
- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature*, 391(6669), 806-811.
- Fischer, S. E. (2015). RNA Interference and MicroRNA-Mediated Silencing. *Curr Protoc Mol Biol*, 112, 26 21 21-25.
- Fitzgerald, M. E., Vela, A., & Pyle, A. M. (2014). Dicer-related helicase 3 forms an obligate dimer for recognizing 22G-RNA. *Nucleic Acids Res*, 42(6), 3919-3930.

- Fu, Q., & Wang, P. J. (2014). Mammalian piRNAs: Biogenesis, function, and mysteries. *Spermatogenesis*, 4, e27889.
- Funabiki, M., Kato, H., Miyachi, Y., Toki, H., Motegi, H., Inoue, M., . . . Fujita, T. (2014). Autoimmune disorders associated with gain of function of the intracellular sensor MDA5. *Immunity*, 40(2), 199-212.
- Gao, D., Yang, Y. K., Wang, R. P., Zhou, X., Diao, F. C., Li, M. D., . . . Chen, D. Y. (2009). REUL is a novel E3 ubiquitin ligase and stimulator of retinoic-acid-inducible gene-I. *PLoS One*, 4(6), e5760.
- Ghildiyal, M., & Zamore, P. D. (2009). Small silencing RNAs: an expanding universe. *Nat Rev Genet*, 10(2), 94-108.
- Gleghorn, M. L., & Maquat, L. E. (2014). 'Black sheep' that don't leave the double-stranded RNA-binding domain fold. *Trends Biochem Sci*, 39(7), 328-340.
- Gorbalenya, A. E., & Koonin, E. V. (1988). One more conserved sequence motif in helicases. *Nucleic Acids Res*, 16(15), 7734.
- Gorbalenya, A. E., & Koonin, E. V. (1993). Helicases: amino acid sequence comparisons and structure-function relationships. *Current Opinion in Structural Biology*, 3(3), 419-429.
- Gu, W., Shirayama, M., Conte, D., Jr., Vasale, J., Batista, P. J., Claycomb, J. M., . . . Mello, C. C. (2009). Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Mol Cell*, 36(2), 231-244.
- Guo, X., Zhang, R., Wang, J., Ding, S. W., & Lu, R. (2013). Homologous RIG-I-like helicase proteins direct RNAi-mediated antiviral immunity in *C. elegans* by distinct mechanisms. *Proc Natl Acad Sci U S A*, 110(40), 16085-16090.
- Habjan, M., Andersson, I., Klingstrom, J., Schumann, M., Martin, A., Zimmermann, P., . . . Weber, F. (2008). Processing of genome 5' termini as a strategy of negative-strand RNA viruses to avoid RIG-I-dependent interferon induction. *PLoS One*, 3(4), e2032.
- Hamilton, A. J., & Baulcombe, D. C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*, 286(5441), 950-952.
- Hanada, K., & Hickson, I. D. (2007). Molecular genetics of RecQ helicase disorders. *Cell Mol Life Sci*, 64(17), 2306-2322.
- Hibino, T., Loza-Coll, M., Messier, C., Majeske, A. J., Cohen, A. H., Terwilliger, D. P., . . . Rast, J. P. (2006). The immune gene repertoire encoded in the purple sea urchin genome. *Dev Biol*, 300(1), 349-365.
- Hoffmann, F. S., Schmidt, A., Dittmann Chevillotte, M., Wisskirchen, C., Hellmuth, J., Willms, S., . . . Rothenfusser, S. (2015). Polymorphisms in melanoma differentiation-associated gene 5 link protein function to clearance of hepatitis C virus. *Hepatology*, 61(2), 460-470.
- Holm, L., & Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res*, 38(Web Server issue), W545-549.
- Honda, K., Takaoka, A., & Taniguchi, T. (2006). Type I interferon [corrected] gene induction by the interferon regulatory factor family of transcription factors. *Immunity*, 25(3), 349-360.
- Hutvagner, G., & Simard, M. J. (2008). Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol*, 9(1), 22-32.
- Iwasaki, A., & Medzhitov, R. (2015). Control of adaptive immunity by the innate immune system. *Nat Immunol*, 16(4), 343-353.
- Janeway, C. A., Jr., & Medzhitov, R. (2002). Innate immune recognition. *Annu Rev Immunol*, 20, 197-216.

- Jang, M. A., Kim, E. K., Now, H., Nguyen, N. T., Kim, W. J., Yoo, J. Y., . . . Ki, C. S. (2015). Mutations in DDX58, which encodes RIG-I, cause atypical Singleton-Merten syndrome. *Am J Hum Genet*, 96(2), 266-274.
- Jankowsky, A., Guenther, P., & Jankowsky, E. (2011). The RNA helicase database. *Nucleic Acids Res*, 39(Database issue), D338-341.
- Jankowsky, E. (2011). RNA helicases at work: binding and rearranging. *Trends Biochem Sci*, 36(1), 19-29.
- Jensen, S., & Thomsen, A. R. (2012). Sensing of RNA viruses: a review of innate immune receptors involved in recognizing RNA virus invasion. *J Virol*, 86(6), 2900-2910.
- Jiang, F., Ramanathan, A., Miller, M. T., Tang, G. Q., Gale, M., Jr., Patel, S. S., & Marcotrigiano, J. (2011). Structural basis of RNA recognition and activation by innate immune receptor RIG-I. *Nature*, 479(7373), 423-427.
- Jonsson, K. L., Laustsen, A., Krapp, C., Skipper, K. A., Thavachelvam, K., Hotter, D., . . . Jakobsen, M. R. (2017). IFI16 is required for DNA sensing in human macrophages by promoting production and function of cGAMP. *Nat Commun*, 8, 14391.
- Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 2), 133-144.
- Kang, D. C., Gopalkrishnan, R. V., Wu, Q., Jankowsky, E., Pyle, A. M., & Fisher, P. B. (2002). mda-5: An interferon-inducible putative RNA helicase with double-stranded RNA-dependent ATPase activity and melanoma growth-suppressive properties. *Proc Natl Acad Sci U S A*, 99(2), 637-642.
- Kao, W. P., Yang, C. Y., Su, T. W., Wang, Y. T., Lo, Y. C., & Lin, S. C. (2015). The versatile roles of CARDs in regulating apoptosis, inflammation, and NF-kappaB signaling. *Apoptosis*, 20(2), 174-195.
- Kato, H., Takeuchi, O., Sato, S., Yoneyama, M., Yamamoto, M., Matsui, K., . . . Akira, S. (2006). Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature*, 441(7089), 101-105.
- Kawai, T., Takahashi, K., Sato, S., Coban, C., Kumar, H., Kato, H., . . . Akira, S. (2005). IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nat Immunol*, 6(10), 981-988.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.
- Kersse, K., Verspurten, J., Vanden Berghe, T., & Vandenabeele, P. (2011). The death-fold superfamily of homotypic interaction motifs. *Trends Biochem Sci*, 36(10), 541-552.
- Ketting, R. F., Fischer, S. E., Bernstein, E., Sijen, T., Hannon, G. J., & Plasterk, R. H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev*, 15(20), 2654-2659.
- Khemici, V., & Linder, P. (2018). RNA helicases in RNA decay. *Biochem Soc Trans*, 46(1), 163-172.
- Kiiianitsa, K., Solinger, J. A., & Heyer, W. D. (2003). NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities. *Anal Biochem*, 321(2), 266-271.
- Knight, S. W., & Bass, B. L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in *Caenorhabditis elegans*. *Science*, 293(5538), 2269-2271.

- Kohlway, A., Luo, D., Rawling, D. C., Ding, S. C., & Pyle, A. M. (2013). Defining the functional determinants for RNA surveillance by RIG-I. *EMBO Rep*, 14(9), 772-779.
- Komuro, A., & Horvath, C. M. (2006). RNA- and virus-independent inhibition of antiviral signaling by RNA helicase LGP2. *J Virol*, 80(24), 12332-12342.
- Konermann, L., Pan, J., & Liu, Y. H. (2011). Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem Soc Rev*, 40(3), 1224-1234.
- Konermann, L., Rodriguez, A. D., & Sowole, M. A. (2014). Type 1 and Type 2 scenarios in hydrogen exchange mass spectrometry studies on protein-ligand complexes. *Analyst*, 139(23), 6078-6087.
- Koonin, E. V. (1995). Multidomain organization of eukaryotic guanine nucleotide exchange translation initiation factor eIF-2B subunits revealed by analysis of conserved sequence motifs. *Protein Sci*, 4(8), 1608-1617.
- Kowalinski, E., Lunardi, T., McCarthy, A. A., Louber, J., Brunel, J., Grigorov, B., . . . Cusack, S. (2011). Structural basis for the activation of innate immune pattern-recognition receptor RIG-I by viral RNA. *Cell*, 147(2), 423-435.
- Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, 372(3), 774-797.
- Kumar, S., Tamura, K., & Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform*, 5(2), 150-163.
- Kuo, D., Nie, M., & Courey, A. J. (2014). SUMO as a solubility tag and in vivo cleavage of SUMO fusion proteins with Ulp1. *Methods Mol Biol*, 1177, 71-80.
- Kuo, R. L., Chen, C. J., Wang, R. Y. L., Huang, H. I., Lin, Y. H., Tam, E. H., . . . Shih, S. R. (2019). Role of Enteroviral RNA-Dependent RNA Polymerase in Regulation of MDA5-Mediated Beta Interferon Activation. *J Virol*, 93(10).
- Kurzynska-Kokorniak, A., Koralewska, N., Pokornowska, M., Urbanowicz, A., Tworak, A., Mickiewicz, A., & Figlerowicz, M. (2015). The many faces of Dicer: the complexity of the mechanisms regulating Dicer gene expression and enzyme activities. *Nucleic Acids Res*, 43(9), 4365-4380.
- Lassig, C., Matheisl, S., Sparrer, K. M., de Oliveira Mann, C. C., Moldt, M., Patel, J. R., . . . Hopfner, K. P. (2015). ATP hydrolysis by the viral RNA sensor RIG-I prevents unintentional recognition of self-RNA. *Elife*, 4.
- Lau, P. W., Guiley, K. Z., De, N., Potter, C. S., Carragher, B., & MacRae, I. J. (2012). The molecular architecture of human Dicer. *Nat Struct Mol Biol*, 19(4), 436-440.
- Leahy, D. J., Hendrickson, W. A., Aukhil, I., & Erickson, H. P. (1992). Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science*, 258(5084), 987-991.
- Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, 75(5), 843-854.
- Leung, R. K., & Whittaker, P. A. (2005). RNA interference: from gene silencing to gene-specific therapeutics. *Pharmacol Ther*, 107(2), 222-239.
- Li, H., Li, W. X., & Ding, S. W. (2002). Induction and suppression of RNA silencing by an animal virus. *Science*, 296(5571), 1319-1321.
- Li, X., Lu, C., Stewart, M., Xu, H., Strong, R. K., Igumenova, T., & Li, P. (2009). Structural basis of double-stranded RNA recognition by the RIG-I like receptor MDA5. *Arch Biochem Biophys*, 488(1), 23-33.

- Li, X., Ranjith-Kumar, C. T., Brooks, M. T., Dharmaiah, S., Herr, A. B., Kao, C., & Li, P. (2009). The RIG-I-like receptor LGP2 recognizes the termini of double-stranded RNA. *J Biol Chem*, 284(20), 13881-13891.
- Li, Y., Lu, J., Han, Y., Fan, X., & Ding, S. W. (2013). RNA interference functions as an antiviral immunity mechanism in mammals. *Science*, 342(6155), 231-234.
- Lin, S. C., Lo, Y. C., & Wu, H. (2010). Helical assembly in the MyD88-IRAK4-IRAK2 complex in TLR/IL-1R signalling. *Nature*, 465(7300), 885-890.
- Lisal, J., Lam, T. T., Kainov, D. E., Emmett, M. R., Marshall, A. G., & Tuma, R. (2005). Functional visualization of viral molecular motor by hydrogen-deuterium exchange reveals transient states. *Nat Struct Mol Biol*, 12(5), 460-466.
- Liu, S., Chen, J., Cai, X., Wu, J., Chen, X., Wu, Y. T., . . . Chen, Z. J. (2013). MAVS recruits multiple ubiquitin E3 ligases to activate antiviral signaling cascades. *Elife*, 2, e00785.
- Liu, S., & Chen, Z. J. (2011). Expanding role of ubiquitination in NF-kappaB signaling. *Cell Res*, 21(1), 6-21.
- Liu, X. Y., Wei, B., Shi, H. X., Shan, Y. F., & Wang, C. (2010). Tom70 mediates activation of interferon regulatory factor 3 on mitochondria. *Cell Res*, 20(9), 994-1011.
- Liu, Z., Wang, J., Cheng, H., Ke, X., Sun, L., Zhang, Q. C., & Wang, H. W. (2018). Cryo-EM Structure of Human Dicer and Its Complexes with a Pre-miRNA Substrate. *Cell*, 173(5), 1191-1203 e1112.
- Lohman, T. M., & Bjornson, K. P. (1996). Mechanisms of helicase-catalyzed DNA unwinding. *Annu Rev Biochem*, 65, 169-214.
- Lohman, T. M., Tomko, E. J., & Wu, C. G. (2008). Non-hexameric DNA helicases and translocases: mechanisms and regulation. *Nat Rev Mol Cell Biol*, 9(5), 391-401.
- Louber, J., Brunel, J., Uchikawa, E., Cusack, S., & Gerlier, D. (2015). Kinetic discrimination of self/non-self RNA by the ATPase activity of RIG-I and MDA5. *BMC Biol*, 13, 54.
- Lu, C., Ranjith-Kumar, C. T., Hao, L., Kao, C. C., & Li, P. (2011). Crystal structure of RIG-I C-terminal domain bound to blunt-ended double-strand RNA without 5' triphosphate. *Nucleic Acids Res*, 39(4), 1565-1575.
- Lu, C., Xu, H., Ranjith-Kumar, C. T., Brooks, M. T., Hou, T. Y., Hu, F., . . . Li, P. (2010). The structural basis of 5' triphosphate double-stranded RNA recognition by RIG-I C-terminal domain. *Structure*, 18(8), 1032-1043.
- Lu, R., Yigit, E., Li, W. X., & Ding, S. W. (2009). An RIG-I-Like RNA helicase mediates antiviral RNAi downstream of viral siRNA biogenesis in *Caenorhabditis elegans*. *PLoS Pathog*, 5(2), e1000286.
- Lulchev, P., & Klostermeier, D. (2014). Reverse gyrase--recent advances and current mechanistic understanding of positive DNA supercoiling. *Nucleic Acids Res*, 42(13), 8200-8213.
- Luo, D., Ding, S. C., Vela, A., Kohlway, A., Lindenbach, B. D., & Pyle, A. M. (2011). Structural insights into RNA recognition by RIG-I. *Cell*, 147(2), 409-422.
- Luo, D., Kohlway, A., & Pyle, A. M. (2013). Duplex RNA activated ATPases (DRAs): platforms for RNA sensing, signaling and processing. *RNA Biol*, 10(1), 111-120.
- Luo, D., Kohlway, A., Vela, A., & Pyle, A. M. (2012). Visualizing the determinants of viral RNA recognition by innate immune sensor RIG-I. *Structure*, 20(11), 1983-1988.
- Maillard, P. V., Van der Veen, A. G., Deddouche-Grass, S., Rogers, N. C., Merits, A., & Reis e Sousa, C. (2016). Inactivation of the type I interferon pathway reveals

- long double-stranded RNA-mediated RNA interference in mammalian cells. *Embo j*, 35(23), 2505-2518.
- Malathi, K., Saito, T., Crochet, N., Barton, D. J., Gale, M., Jr., & Silverman, R. H. (2010). RNase L releases a small RNA from HCV RNA that refolds into a potent PAMP. *Rna*, 16(11), 2108-2119.
- Martinez-Campos, C., Burguete-Garcia, A. I., & Madrid-Marina, V. (2017). Role of TLR9 in Oncogenic Virus-Produced Cancer. *Viral Immunol*, 30(2), 98-105.
- Matranga, C., & Pyle, A. M. (2010). Double-stranded RNA-dependent ATPase DRH-3: insight into its role in RNA silencing in *Caenorhabditis elegans*. *J Biol Chem*, 285(33), 25363-25371.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *J Appl Crystallogr*, 40(Pt 4), 658-674.
- McNicholas, S., Potterton, E., Wilson, K. S., & Noble, M. E. (2011). Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr D Biol Crystallogr*, 67(Pt 4), 386-394.
- Montgomery, M. K., Xu, S., & Fire, A. (1998). RNA as a target of double-stranded RNA-mediated genetic interference in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A*, 95(26), 15502-15507.
- Munir, M. (2010). TRIM proteins: another class of viral victims. *Sci Signal*, 3(118), jc2.
- Nakamura, M., Ando, R., Nakazawa, T., Yudazono, T., Tsutsumi, N., Hatanaka, N., . . . Eki, T. (2007). Dicer-related drh-3 gene functions in germ-line development by maintenance of chromosomal integrity in *Caenorhabditis elegans*. *Genes Cells*, 12(9), 997-1010.
- Napoli, C., Lemieux, C., & 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(4), 279-289.
- Oshiumi, H., Miyashita, M., Inoue, N., Okabe, M., Matsumoto, M., & Seya, T. (2010). The ubiquitin ligase Riplet is essential for RIG-I-dependent innate immune responses to RNA virus infection. *Cell Host Microbe*, 8(6), 496-509.
- Pahler, A., Smith, J. L., & Hendrickson, W. A. (1990). A probability representation for phase information from multiwavelength anomalous dispersion. *Acta Crystallogr A*, 46 (Pt 7), 537-540.
- Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S., & Tucker, P. A. (2005). Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr D Biol Crystallogr*, 61(Pt 4), 449-457.
- Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S., & Tucker, P. A. (2009). On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr D Biol Crystallogr*, 65(Pt 10), 1089-1097.
- Parameswaran, P., Sklan, E., Wilkins, C., Burgon, T., Samuel, M. A., Lu, R., . . . Fire, A. Z. (2010). Six RNA viruses and forty-one hosts: viral small RNAs and modulation of small RNA repertoires in vertebrate and invertebrate systems. *PLoS Pathog*, 6(2), e1000764.
- Park, H. H., Logette, E., Raunser, S., Cuenin, S., Walz, T., Tschopp, J., & Wu, H. (2007). Death domain assembly mechanism revealed by crystal structure of the oligomeric PIDDosome core complex. *Cell*, 128(3), 533-546.

- Park, M. S., Phan, H. D., Busch, F., Hinckley, S. H., Brackbill, J. A., Wysocki, V. H., & Nakanishi, K. (2017). Human Argonaute3 has slicer activity. *Nucleic Acids Res*, 45(20), 11867-11877.
- Pascal, B. D., Willis, S., Lauer, J. L., Landgraf, R. R., West, G. M., Marciano, D., . . . Griffin, P. R. (2012). HDX workbench: software for the analysis of H/D exchange MS data. *J Am Soc Mass Spectrom*, 23(9), 1512-1521.
- Patel, S. S., & Picha, K. M. (2000). Structure and function of hexameric helicases. *Annu Rev Biochem*, 69, 651-697.
- Peisley, A., Wu, B., Yao, H., Walz, T., & Hur, S. (2013). RIG-I forms signaling-competent filaments in an ATP-dependent, ubiquitin-independent manner. *Mol Cell*, 51(5), 573-583.
- Pike, A. C., Garman, E. F., Krojer, T., von Delft, F., & Carpenter, E. P. (2016). An overview of heavy-atom derivatization of protein crystals. *Acta Crystallogr D Struct Biol*, 72(Pt 3), 303-318.
- Pippig, D. A., Hellmuth, J. C., Cui, S., Kirchhofer, A., Lammens, K., Lammens, A., . . . Hopfner, K. P. (2009). The regulatory domain of the RIG-I family ATPase LGP2 senses double-stranded RNA. *Nucleic Acids Res*, 37(6), 2014-2025.
- Pyle, A. M. (2008). Translocation and unwinding mechanisms of RNA and DNA helicases. *Annu Rev Biophys*, 37, 317-336.
- Qin, H., Srinivasula, S. M., Wu, G., Fernandes-Alnemri, T., Alnemri, E. S., & Shi, Y. (1999). Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. *Nature*, 399(6736), 549-557.
- Qiu, Y., Xu, Y., Zhang, Y., Zhou, H., Deng, Y. Q., Li, X. F., . . . Zhou, X. (2018). Human Virus-Derived Small RNAs Can Confer Antiviral Immunity in Mammals. *Immunity*, 49(4), 780-781.
- Ramachandran, G. N., & Srinivasan, R. (1961). An Apparent Paradox in Crystal Structure Analysis. *Nature*, 190(4771), 159-161.
- Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., . . . Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, 403(6772), 901-906.
- Rice, G. I., Del Toro Duany, Y., Jenkinson, E. M., Forte, G. M., Anderson, B. H., Ariaudo, G., . . . Crow, Y. J. (2014). Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. *Nat Genet*, 46(5), 503-509.
- Roers, A., Hiller, B., & Hornung, V. (2016). Recognition of Endogenous Nucleic Acids by the Innate Immune System. *Immunity*, 44(4), 739-754.
- Rossmann, M. G., & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica*, 15(1), 24-31.
- Rothenfusser, S., Goutagny, N., DiPerna, G., Gong, M., Monks, B. G., Schoenemeyer, A., . . . Fitzgerald, K. A. (2005). The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I. *J Immunol*, 175(8), 5260-5268.
- Ruby, J. G., Jan, C., Player, C., Axtell, M. J., Lee, W., Nusbaum, C., . . . Bartel, D. P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*, 127(6), 1193-1207.
- Saito, T., Owen, D. M., Jiang, F., Marcotrigiano, J., & Gale, M., Jr. (2008). Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. *Nature*, 454(7203), 523-527.
- Scapin, G. (2013). Molecular replacement then and now. *Acta Crystallogr D Biol Crystallogr*, 69(Pt 11), 2266-2275.

- Schrodel, A., & de Marco, A. (2005). Characterization of the aggregates formed during recombinant protein expression in bacteria. *BMC Biochem*, 6, 10.
- Schrodel, A., Volz, J., & de Marco, A. (2005). Fusion tags and chaperone co-expression modulate both the solubility and the inclusion body features of the recombinant CLIPB14 serine protease. *J Biotechnol*, 120(1), 2-10.
- Schuster, S., Miesen, P., & van Rij, R. P. (2019). Antiviral RNAi in Insects and Mammals: Parallels and Differences. *Viruses*, 11(5).
- Scott, F. L., Stec, B., Pop, C., Dobaczewska, M. K., Lee, J. J., Monosov, E., . . . Riedl, S. J. (2009). The Fas-FADD death domain complex structure unravels signalling by receptor clustering. *Nature*, 457(7232), 1019-1022.
- Sen, G. C. (2001). Viruses and interferons. *Annu Rev Microbiol*, 55, 255-281.
- Seth, R. B., Sun, L., Ea, C. K., & Chen, Z. J. (2005). Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. *Cell*, 122(5), 669-682.
- Sheldrick, G. M. (2010). Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 4), 479-485.
- Shereda, R. D., Reiter, N. J., Butcher, S. E., & Keck, J. L. (2009). Identification of the SSB binding site on E. coli RecQ reveals a conserved surface for binding SSB's C terminus. *J Mol Biol*, 386(3), 612-625.
- Sijen, T., Fleenor, J., Simmer, F., Thijssen, K. L., Parrish, S., Timmons, L., . . . Fire, A. (2001). On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell*, 107(4), 465-476.
- Sijen, T., Steiner, F. A., Thijssen, K. L., & Plasterk, R. H. (2007). Secondary siRNAs result from unprimed RNA synthesis and form a distinct class. *Science*, 315(5809), 244-247.
- Singleton, M. R., Dillingham, M. S., & Wigley, D. B. (2007). Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem*, 76, 23-50.
- Smardon, A., Spoerke, J. M., Stacey, S. C., Klein, M. E., Mackin, N., & Maine, E. M. (2000). EGO-1 is related to RNA-directed RNA polymerase and functions in germ-line development and RNA interference in *C. elegans*. *Curr Biol*, 10(4), 169-178.
- Snead, N. M., & Rossi, J. J. (2010). Biogenesis and function of endogenous and exogenous siRNAs. *Wiley Interdiscip Rev RNA*, 1(1), 117-131.
- Song, M. S., & Rossi, J. J. (2017). Molecular mechanisms of Dicer: endonuclease and enzymatic activity. *Biochem J*, 474(10), 1603-1618.
- Sorensen, H. P., & Mortensen, K. K. (2005). Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J Biotechnol*, 115(2), 113-128.
- Steward, A., McDowell, G. S., & Clarke, J. (2009). Topology is the principal determinant in the folding of a complex all-alpha Greek key death domain from human FADD. *J Mol Biol*, 389(2), 425-437.
- Sun, Q., Sun, L., Liu, H. H., Chen, X., Seth, R. B., Forman, J., & Chen, Z. J. (2006). The specific and essential role of MAVS in antiviral innate immune responses. *Immunity*, 24(5), 633-642.
- Sun, X., Xian, H., Tian, S., Sun, T., Qin, Y., Zhang, S., & Cui, J. (2016). A Hierarchical Mechanism of RIG-I Ubiquitination Provides Sensitivity, Robustness and Synergy in Antiviral Immune Responses. *Sci Rep*, 6, 29263.

- Tabara, H., Yigit, E., Siomi, H., & Mello, C. C. (2002). The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in *C. elegans*. *Cell*, 109(7), 861-871.
- Tahbaz, N., Kolb, F. A., Zhang, H., Jaronczyk, K., Filipowicz, W., & Hobman, T. C. (2004). Characterization of the interactions between mammalian PAZ PIWI domain proteins and Dicer. *EMBO Rep*, 5(2), 189-194.
- Tai, C. L., Chi, W. K., Chen, D. S., & Hwang, L. H. (1996). The helicase activity associated with hepatitis C virus nonstructural protein 3 (NS3). *J Virol*, 70(12), 8477-8484.
- Takahasi, K., Kumeta, H., Tsuduki, N., Narita, R., Shigemoto, T., Hirai, R., . . . Inagaki, F. (2009). Solution structures of cytosolic RNA sensor MDA5 and LGP2 C-terminal domains: identification of the RNA recognition loop in RIG-I-like receptors. *J Biol Chem*, 284(26), 17465-17474.
- Takahasi, K., Yoneyama, M., Nishihori, T., Hirai, R., Kumeta, H., Narita, R., . . . Fujita, T. (2008). Nonspecific RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses. *Mol Cell*, 29(4), 428-440.
- Terwilliger, T. C. (2004). Using prime-and-switch phasing to reduce model bias in molecular replacement. *Acta Crystallogr D Biol Crystallogr*, 60(Pt 12 Pt 1), 2144-2149.
- Thompson, A. J., & Locarnini, S. A. (2007). Toll-like receptors, RIG-I-like RNA helicases and the antiviral innate immune response. *Immunol Cell Biol*, 85(6), 435-445.
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, 22(22), 4673-4680.
- Uchikawa, E., Lethier, M., Malet, H., Brunel, J., Gerlier, D., & Cusack, S. (2016). Structural Analysis of dsRNA Binding to Anti-viral Pattern Recognition Receptors LGP2 and MDA5. *Mol Cell*, 62(4), 586-602.
- van Cleef, K. W., van Mierlo, J. T., van den Beek, M., & van Rij, R. P. (2011). Identification of viral suppressors of RNAi by a reporter assay in *Drosophila* S2 cell culture. *Methods Mol Biol*, 721, 201-213.
- van der Krol, A. R., Mur, L. A., Beld, M., Mol, J. N., & Stuitje, A. R. (1990). Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell*, 2(4), 291-299.
- van der Veen, A. G., Maillard, P. V., Schmidt, J. M., Lee, S. A., Deddouche-Grass, S., Borg, A., . . . Reis e Sousa, C. (2018). The RIG-I-like receptor LGP2 inhibits Dicer-dependent processing of long double-stranded RNA and blocks RNA interference in mammalian cells. *Embo j*, 37(4).
- van Haaften, G., Romeijn, R., Pothof, J., Koole, W., Mullenders, L. H., Pastink, A., . . . Tijsterman, M. (2006). Identification of conserved pathways of DNA-damage response and radiation protection by genome-wide RNAi. *Curr Biol*, 16(13), 1344-1350.
- Vasale, J. J., Gu, W., Thivierge, C., Batista, P. J., Claycomb, J. M., Youngman, E. M., . . . Conte, D., Jr. (2010). Sequential rounds of RNA-dependent RNA transcription drive endogenous small-RNA biogenesis in the ERGO-1/Argonaute pathway. *Proc Natl Acad Sci U S A*, 107(8), 3582-3587.
- Vela, A., Fedorova, O., Ding, S. C., & Pyle, A. M. (2012). The thermodynamic basis for viral RNA detection by the RIG-I innate immune sensor. *J Biol Chem*, 287(51), 42564-42573.

- Venkataraman, T., Valdes, M., Elsby, R., Kakuta, S., Caceres, G., Saijo, S., . . . Barber, G. N. (2007). Loss of DExD/H box RNA helicase LGP2 manifests disparate antiviral responses. *J Immunol*, 178(10), 6444-6455.
- Venkatesan, M., Silver, L. L., & Nossal, N. G. (1982). Bacteriophage T4 gene 41 protein, required for the synthesis of RNA primers, is also a DNA helicase. *J Biol Chem*, 257(20), 12426-12434.
- Vodovar, N., Bronkhorst, A. W., van Cleef, K. W., Miesen, P., Blanc, H., van Rij, R. P., & Saleh, M. C. (2012). Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. *PLoS One*, 7(1), e30861.
- Walden, H. (2010). Selenium incorporation using recombinant techniques. *Acta Crystallogr D Biol Crystallogr*, 66(Pt 4), 352-357.
- Walker, J. E., Saraste, M., Runswick, M. J., & Gay, N. J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo j*, 1(8), 945-951.
- Wang, B., Love, T. M., Call, M. E., Doench, J. G., & Novina, C. D. (2006). Recapitulation of short RNA-directed translational gene silencing in vitro. *Mol Cell*, 22(4), 553-560.
- Wang, Y., Ludwig, J., Schuberth, C., Goldeck, M., Schlee, M., Li, H., . . . Patel, D. J. (2010). Structural and functional insights into 5'-ppp RNA pattern recognition by the innate immune receptor RIG-I. *Nat Struct Mol Biol*, 17(7), 781-787.
- Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., . . . Sasaki, H. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, 453(7194), 539-543.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., . . . Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*, 46(W1), W296-W303.
- West, G. M., Pascal, B. D., Ng, L. M., Soon, F. F., Melcher, K., Xu, H. E., . . . Griffin, P. R. (2013). Protein conformation ensembles monitored by HDX reveal a structural rationale for abscisic acid signaling protein affinities and activities. *Structure*, 21(2), 229-235.
- Willkomm, S., Zander, A., Gust, A., & Grohmann, D. (2015). A prokaryotic twist on argonaute function. *Life (Basel)*, 5(1), 538-553.
- Wilson, R. C., Tambe, A., Kidwell, M. A., Noland, C. L., Schneider, C. P., & Doudna, J. A. (2015). Dicer-TRBP complex formation ensures accurate mammalian microRNA biogenesis. *Mol Cell*, 57(3), 397-407.
- Wu, B., & Hur, S. (2015). How RIG-I like receptors activate MAVS. *Curr Opin Virol*, 12, 91-98.
- Wu, B., Peisley, A., Richards, C., Yao, H., Zeng, X., Lin, C., . . . Hur, S. (2013). Structural basis for dsRNA recognition, filament formation, and antiviral signal activation by MDA5. *Cell*, 152(1-2), 276-289.
- Wu, B., Peisley, A., Tetrault, D., Li, Z., Egelman, E. H., Magor, K. E., . . . Hur, S. (2014). Molecular imprinting as a signal-activation mechanism of the viral RNA sensor RIG-I. *Mol Cell*, 55(4), 511-523.
- Wu, Y. (2012). Unwinding and rewinding: double faces of helicase? *J Nucleic Acids*, 2012, 140601.
- Xiao, T., Towb, P., Wasserman, S. A., & Sprang, S. R. (1999). Three-dimensional structure of a complex between the death domains of Pelle and Tube. *Cell*, 99(5), 545-555.

- Xin, B. G., Chen, W. F., Rety, S., Dai, Y. X., & Xi, X. G. (2018). Crystal structure of Escherichia coli DEAH/RHA helicase HrpB. *Biochem Biophys Res Commun*, 504(1), 334-339.
- Xiong, S., Zhang, L., & He, Q. Y. (2008). Fractionation of proteins by heparin chromatography. *Methods Mol Biol*, 424, 213-221.
- Yigit, E., Batista, P. J., Bei, Y., Pang, K. M., Chen, C. C., Tolia, N. H., . . . Mello, C. C. (2006). Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell*, 127(4), 747-757.
- Yoneyama, M., & Fujita, T. (2007). Function of RIG-I-like receptors in antiviral innate immunity. *J Biol Chem*, 282(21), 15315-15318.
- Yoneyama, M., Kikuchi, M., Matsumoto, K., Imaizumi, T., Miyagishi, M., Taira, K., . . . Fujita, T. (2005). Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. *J Immunol*, 175(5), 2851-2858.
- Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., . . . Fujita, T. (2004). The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat Immunol*, 5(7), 730-737.
- Zevini, A., Olagnier, D., & Hiscott, J. (2017). Crosstalk between Cytoplasmic RIG-I and STING Sensing Pathways. *Trends Immunol*, 38(3), 194-205.
- Zhang, J., Chalmers, M. J., Stayrook, K. R., Burris, L. L., Garcia-Ordonez, R. D., Pascal, B. D., . . . Griffin, P. R. (2010). Hydrogen/deuterium exchange reveals distinct agonist/partial agonist receptor dynamics within vitamin D receptor/retinoid X receptor heterodimer. *Structure*, 18(10), 1332-1341.
- Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9, 40.
- Zheng, J., Yong, H. Y., Panutdaporn, N., Liu, C., Tang, K., & Luo, D. (2015). High-resolution HDX-MS reveals distinct mechanisms of RNA recognition and activation by RIG-I and MDA5. *Nucleic Acids Res*, 43(2), 1216-1230.
- Zou, J., Chang, M., Nie, P., & Secombes, C. J. (2009). Origin and evolution of the RIG-I like RNA helicase gene family. *BMC Evol Biol*, 9, 85.

APPENDIX

A: Heat map for unligated full-length DRH-3

B: Heat map for unligated DRH-3 NTD

C: Differential heat map for full-length DRH-3 with 3p10L.

D: Differential heat map for DRH-1 Hel-CTD with 3p10L.