STRUCTURAL BASIS FOR RNA RECOGNITION AND ACTIVATION BY HUMAN INNATE IMMUNE RECEPTOR RIG-I

by

FUGUO JIANG

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Chemistry and Chemical Biology

Written under the direction of

Joseph Marcotrigiano, Ph.D.

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2013

ABSTRACT OF THE DISSERTATION

STRUCTURAL BASIS FOR RNA RECOGNITION AND ACTIVATION BY HUMAN INNATE IMMUNE RECEPTOR RIG-I

By FUGUO JIANG

Dissertation Director:

Joseph Marcotrigiano, Ph.D.

Innate immunity provides the first line of host defense against pathogenic microbial and viral invasion. Activation of innate immune responses relies on the specific recognition of pathogen-associated molecular patterns (PAMPs) by pattern-recognition receptors (PRRs). Retinoic acid Inducible Gene- I (RIG-I) is a crucial PPR in the cytoplasm that induces antiviral and inflammatory immune responses against RNA viruses by selectively detecting PAMP RNAs. The RIG-I signaling pathway is highly regulated. Aberrant signaling can lead to apoptosis and altered cell differentiation, which have been implicated in the development of inflammation, autoimmune diseases including type 1 diabetes, and cancer.

We have collaborated with Dr. Michael Gale at University of Washington School of Medicine to identify the poly-uridine motif of the Hepatitis C virus (HCV) genome 3′ non-translated region and its replication intermediate as a PAMP substrate of RIG-I (Saito et al., 2008). To this end, I have developed efficient

expression and purification methods for human RIG-I, and characterized the protein using biochemical and biophysical methods. Highly purified RIG-I protein was then used to verify HCV PAMP RNA *in vitro* by gel shift assay and limited proteolysis.

RIG-I consists of two N-terminal caspase recruitment domains (CARDs), a central DExD/H box RNA helicase/ATPase domain, and a C- terminal repressor domain (RD). To understand how the RIG-I helicase binds RNA and leads to activation, I have determined the crystal structure of the human RIG-I helicase-RD domain bound to dsRNA and ADP•BeF_3 in collaboration with Dr. Smita Patel's group at UMDNJ. The structure of ternary complex reveals a major contribution from the helicase domain to RNA binding and a synergy between the helicase and RD in recognition of blunt-ended dsRNA (Jiang et al., 2011).

Furthermore, I have determined the crystal structures of RIG-I bound to panhandle-like short hairpin RNAs in the presence or absence of 5'-triphosphorylated modification, and chimeric RNA-DNA duplex at high resolution. These recent structures provide further insights into the molecular mechanics of RNA recognition and RIG-I activation upon viral infection.

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest and sincere appreciation and gratitude to my advisor Dr. Joseph Marcotrigiano. I am very grateful to Joe for giving me the great opportunity, freedom, patience to work on RIG-I project. He always guides me to see a big picture of the project and focus on the most important question to address. He always encourages me to have patience, persistence and perseverance in the research. My deep appreciation is beyond words for his tremendous support and excellent guidance throughout my graduate career.

I would like to thank my committee members, Dr. Eddy Arnold, Smita S. Patel, Richard H. Ebright and former committee member Dr. Jean Baum, for their time and thoughtful discussions of my work. I was attracted to Rutgers by the reputation of Eddy in HIV virology and protein crystallography. I did my first lab rotation in his lab and learned a lot. Eddy really taught me how to communicate efficiently with the labmates and collaborators. I also want to thank Eddy for his continuous support and encouragement, and generous advice and guidance through my Ph.D studies at CABM. In particular, I would like to thank Dr. Smita Patel. I have learned a lot from Smita in many ways. From her, I have learned to approach and solve a scientific question carefully yet without hesitation, to interpret data more thoroughly and critically, and to present ideas with better clarity and precision. I also appreciate Dr. Richard H. Ebright for his generosity to

be on my committee. And I really appreciate his time, insightful discussions, and constructive comments on my work throughout my committee meetings.

Particularly, I want to thank Dr. Matthew Miller, our X-ray lab manager. I am truly and deeply indebted to him so much. He always taught me with the great patience in many things, including how to collect and process the X-ray data, how to solve the structure, how to process SAXS data and build an *ab initio* model, etc. I can always find the answer when I ask him a technical question. And I really appreciate the wonderful time we have during a lot of scientific and non-scientific discussions in his office.

I also want to thank our brilliant collaborators: Anand Ramanathan, Swapnil Devarkar, Vaishnavi Rajagopal and Dr. Smita Patel (UMDNJ); Dr. Takeshi Saito, Dr. Helene Minyi Liu and Dr. Michael Gale (University of Washington School of Medicine); and Dr. Elizabeth Xu and Dr. Taekjip Ha (UIUC).

I would like to thank Dr. Ann Stock, Helen Berman and the late Dr. Aaron Shatkin for providing the support and helpful comments. I also thank Dr. Richard Gillilian (MacCHESS), Dr. Stephen K. Burley, Laura Morisco (Eli Lilly Company & APS) for the help and assistance in data collection.

I would like to thank all the lab members I have encountered throughout the years in Joe's lab: Jill Whidby, Gyehwa Shin, Samantha Yost, Ankita Basant, and

Abdul Khan, Lexi Zatorski and our wonderful undergraduates Alicja Cygan, Patricia Yau, Elizabeth Chang, and Tom Ehrlich. I also want to thank the rotation students who have made great efforts on the RIG-I project: Fei Wang, Xia Tang. Special thanks to Jill for teaching me the English with so much patience! I also want to thank Dr. Abdul Khan. Thank you very much for a lot of interesting discussions.

In addition, I thank Steve Tuske, Joe Bauman, Art Clark and all other members from Eddy's Lab, Yu Zhang, Yu Feng and Qumiao Xu from Prof. Richard Ebright's Lab, Chun Chu from Dr. Shatkin's lab, Rong Xiao, Rongjin Guan from Dr. Guy Montelione's lab, Haiyan Zheng, Meiqian Qian, and Caifeng Zhao from Mass Spectrometry Facility / Dr. Peter Lobel lab, for their enthusiastic help during my Ph.D studies.

Thanks to my friends outside the CCB/CABM department: Zhe Liang, Tian Sun, Yuehua Wei, Ying Cai, Lijuan Kang, Gao Shang, Kaiyuan He, Li Gu, Da Xu, Gaojie Hu, Qingda Zang, Sugang Li and Nan Wang. They made my life in New Jersey full of fun. I also thank my friends and classmates Yuan Cheng (UT Southwestern), Qiang Chen (University of Michigan), Zhike Zi (Universität Freiburg), Yushu Chen (Columbia university), Xianjin Ou (Tsinghua University). I really appreciate theirs great help, suggestions and advices through my studies in China and US.

I also want to thank my parents and younger brother. Without their unconditional love, unwavering support and encouragement, I would not have gone so far in pursuing science. Finally, I would like to thank the most important person in my life, my fiancée, Meng Chen. This dissertation would not have ever been finished without her, and the words are never enough to express my deepest and heartfelt love and gratitude to her.

DEDICATION

This thesis is dedicated to the memory of Dr. Aaron J. Shatkin. I met Dr. Shatkin when I started my first lab rotation in Dr. Eddy Arnold's lab. It's a collaborative project between Dr. Arnold's group and Dr. Shatkin's lab. I feel so lucky to have the opportunity to work in his lab under his supervision. Aaron had always made his time to review my experiments and given excellent guidance despite his busy schedule. Aaron revised my lab report so carefully that even punctuation had been corrected. He was always there to help whenever you needed advice. Aaron is someone you will instantly love and never forget once you meet him. I am immensely indebted to him for his unwavering encouragement, advice, guidance, and support throughout my graduate studies.

Table of Contents

List of Figures

1. Introduction

1.1 Antiviral innate immune response

The innate immune system functions as the first line of host defense against pathogen invasion by responding rapidly to infection, preventing the spread of infection, and relaying signals to the adaptive immune system (Janeway and Medzhitov, 2002). It provides an early and important response to microbial attack. Innate immunity depends on germline-encoded host receptors to recognize the distinct structural components that are common to many pathogens, including viruses (Kumar et al., 2009; Meylan et al., 2006; Vance et al., 2009). Often, viral infections of mammalian cells trigger strong innate immune defenses against the invading viruses through host pathogen-recognition receptors (PRRs). These receptors specifically detect pathogen-associated molecular patterns (PAMPs) present in virus-associated molecules, such as genomic DNA and RNA or dsRNA (Akira et al., 2006; Kawai and Akira, 2006; Saito and Gale, 2007). Following recognition of viral PAMPs, PPRs activate intracellular signaling cascades and eventually trigger antiviral responses, entailing production of type I interferon (IFN-α/IFN-β), induction of proinflammatory cytokines and adaptive immunity (Katze et al., 2008; Stetson and Medzhitov, 2006). Therefore, the effective sensing of viral PAMPs by the PRRs expressed on or in the host cells plays a crucial role in activating an appropriate immune response upon viral infection.

1.2 RIG-I and MDA-5: the key cytoplasmic virus sensors for antiviral innate immunity

To date, two major classes of PRRs, the Toll-like receptors (TLRs) and the RIG-I-like receptors (RLRs), are known to detect RNA viruses (Seth et al., 2006). The TLRs are expressed on the cell surface membrane or endosomes (Kawai and Akira, 2008). In contrast, the RLRs is a newly discovered family of long-sought intracellular viral RNA sensors for viral infection that includes RIG-I (retinoic aid inducible gene I, also known as DDX58) and MDA-5 (melanoma differentiation-associated antigen 5, also known as IFIH1 or Helicard) (Andrejeva et al., 2004; Imaizumi et al., 2002; Kang et al., 2002; Yoneyama et al., 2004). Amino acid sequence alignment indicates that RIG-I and MDA-5 are homologous cytosolic proteins belonging to superfamily 2 (SF2) helicases. Further functional studies revealed that full-length RIG-I and MDA-5 consist of three individual domains: (i) two repeats of the caspase recruitment domain (CARDs) at the N-terminus, (ii) the DExD/H RNA helicase domain in the middle, and (iii) a comparatively short but important repression domain (RD) at the C-terminus (Kovacsovics et al., 2002; Yoneyama et al., 2004) (Figure 1). The CARD domain acts as signaling domain that interacts with downstream molecule and thus relays the signal to induce IFN-α/β production, whereas RNA helicase domain is likely responsible for dsRNA recognition and ATPase activity (Yoneyama et al., 2004). It has been demonstrated that RD domain acts as internal repressor of RIG-I signaling activation; however, this kind of regulatory function has not been observed in MDA-5 (Saito et al., 2007).

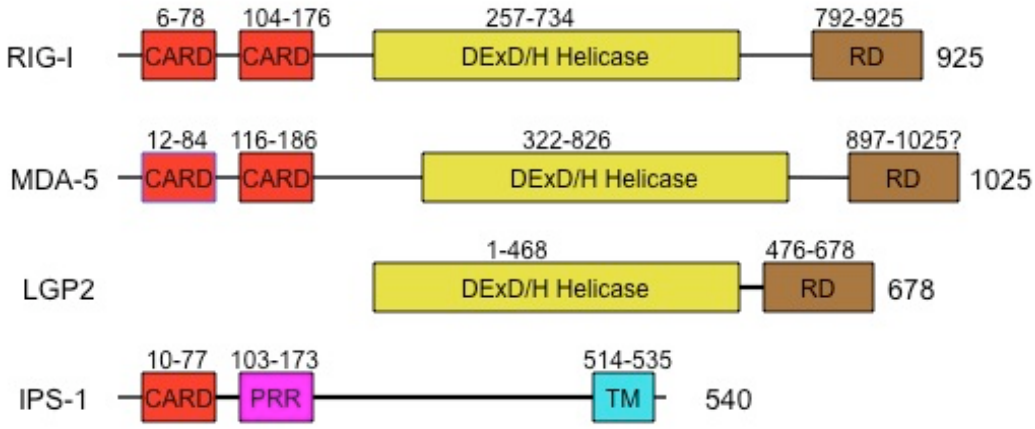


Figure 1. Domain architecture of RLRs based on sequence alignment and functional studies.

Despite the overall structural similarity between those two cytoplasmic viral RNA receptors, RIG-I and MDA-5 recognize different types of RNA viruses (Kato et al., 2006). RIG-I is essential for the recognition of Paramyxoviruses, Sendai virus, Influenza virus and Japanese encephalitis virus, whereas MDA-5 is critical for the recognition of Picornaviruses. Kato et al. (2006) showed that RIG-I and MDA-5 detect in vitro transcribed short dsRNA, the sythetic long dsRNA analog polyinosine-polycytidylic acid (poly (I:C)), respectively. RIG-I specifically senses the nascent 5′ triphosphate moiety, a modification that widely exists in many viral RNAs and in vitro transcribed long dsRNAs (Hornung et al., 2006; Pichlmair et al., 2006; Plumet et al., 2007). In addition, RIG-I also recognizes poly-U/UC rich region within the HCV genome, in which 5'triphosphate is necessary, but not sufficient, for RIG-I binding (Saito et al., 2008). Notably, RIG-I most efficiently recognizes RNA substrates marked with 5'ppp and containing a short partial double-stranded structure (Schlee et al., 2009b; Schmidt et al., 2009). Recently it has been found that RIG-I can recognize 5'-triphosphorylated dsRNA transcribed from AT-rich exogenous dsDNA by host DNA-dependent RNA polymerase III, resulting in the activation of RIG-I signaling pathway, thereby linking RIG-I signaling with DNA virus infection (Ablasser et al., 2009; Chiu et al., 2009).

Taken together, these results indicate that RIG-I is able to discriminate non-self RNA in the cytoplasm by a clever mechanism, since self RNA species do not possess free 5'-ppp due to its removal or masking by acquisition of a cap structure (Hornung et al., 2006; Pichlmair et al., 2006). However, the precise

mechanism underlying the sensing of a distinct spectrum of RNAs is still unclear. The previously solved crystal and NMR solution structures of RD domain of RIG-I reveal that RD is the specific recognition region within RIG-I for binding dsRNA (Cui et al., 2008; Takahasi et al., 2008). Recent crystal structures of RIG-I RD in complex with different RNA ligands, either with dsRNA (Lu et al., 2011) or 5'-ppp dsRNA (Lu et al., 2010; Wang et al., 2010), further demonstrated that RIG-I RD recognizes the termini of dsRNA and interacts with the 5'-ppp as well as the backbone phosphodiesters of the RNA. However, the role of helicase domain in dsRNA recognition was still unknown back that time, given that the helicase domain is still required for high binding capacity with RNA.

1.3 LGP2: a natural regulator in RIG-I signaling pathway

LGP2 (laboratory of genetics and physiology 2), originally identified as a highly expressed gene in mammary tissue (Cui et al., 2001), is another DExD/H RNA helicase that also bears a homologous RNA helicase domain and a similar repression domain (Saito et al., 2007). However, LGP2 lacks the CARD domain in contrast with RIG-I and MDA-5 (Yoneyama et al., 2005) (Figure 1). Because of its lack of CARD signaling domain, LGP2 is unable to stimulate IFN-α/β production (Rothenfusser et al., 2005). Overexpression of LGP2 in cell culture results in dominant inhibition of IFN promoter activation, thus suppressing IFN-α/β induction for antiviral response (Rothenfusser et al., 2005; Yoneyama et al., 2005). LGP2 appears to exert this regulatory activity at three molecular levels: (i) competitively sequestering viral RNA with RIG-I and MDA-5; (ii) binding C-

terminal CARD domain of IPS-1 and then preventing recruitment of the downstream signaling molecules; (iii) and/or binding directly to RIG-I CARD domain using its analogous RD domain (Komuro and Horvath, 2006; Murali et al., 2008; Saito et al., 2007). Current evidence suggests LGP2 mediates a regulatory role in RLR signaling pathway, either negatively or positively (Saito et al., 2007; Satoh et al., 2010). However, its exact physiological role in vivo still remains controversial.

1.4 IPS-1: the central adaptor in RIG-I signaling pathway

IPS-1 (IFN-β promoter stimulator-1, also known as MAVS, VISA, or Cardif), a newly identified adaptor protein triggering RIG-I- and MDA-5-meidated IFN-α/β induction, is located in the outer mitochondrial membrane (Kawai et al., 2005; Meylan et al., 2005; Seth et al., 2005; Xu et al., 2005). IPS-1 is composed of three domains: an amino-terminal CARD-like domain that mediates interaction with CARD domain within RIG-I and MDA-5, a proline-rich region (PRR) that recruits FADD (Fas-associated death-domain protein) and RIP1 (receptor interacting protein 1) to facilitate NF-κB (nuclear factor κB) activation, and a C-terminal mitochondrial membrane region (TM) (Johnson and Gale, 2006) (Figure 1). Overexpression of IPS-1 led to activation of IFN-α/β and NF-κB promoters (Kawai et al., 2005; Meylan et al., 2005; Seth et al., 2005). Studies of gene silencing further demonstrated that knockdown of IPS-1 by small interfering RNA (siRNA) blocked interferon activation in antiviral response to dsRNA stimulation and viral infection (Kawai et al., 2005). Collectively, these observations support

the idea that IPS-1 is an essential adaptor involved in RIG-I- and MDA-5-mediated antiviral immune response (Johnson and Gale, 2006).

1.5 A model for RIG-I signaling pathway

At that time, a model was proposed based on previous experimental results from biochemical, structural, and functional studies in order to try to explain how RIG-I/ MDA-5, IPS-1, and LGP2 regulate activation, signaling and transduction in RIG-I-mediated antiviral innate responses (Cui et al., 2008; Takahasi et al., 2008) (Figure 2). Under normal physiological conditions, RIG-I was in a monomeric and inactive form, and remained locked in a closed, auto-repressed conformation through binding the RD domain to CARD and RNA helicase domains. In this conformation, the function of CARD domain as a signaling transmitter was repressed. Upon viral infection, RD domain would bind the short dsRNA or 5'-triphosphate RNA produced by viruses, which therefore caused RIG-I to undergo a conformational shift and release the CARD domain from the constraint. Then, the CARD domain interacted with each other to form a RIG-I dimer, allowing it to interact with IPS-1 through CARD-CARD homotypic interactions; which, in turn, activates IPS-1 and, ultimately leads to IFN induction, transcription and production for antiviral state.

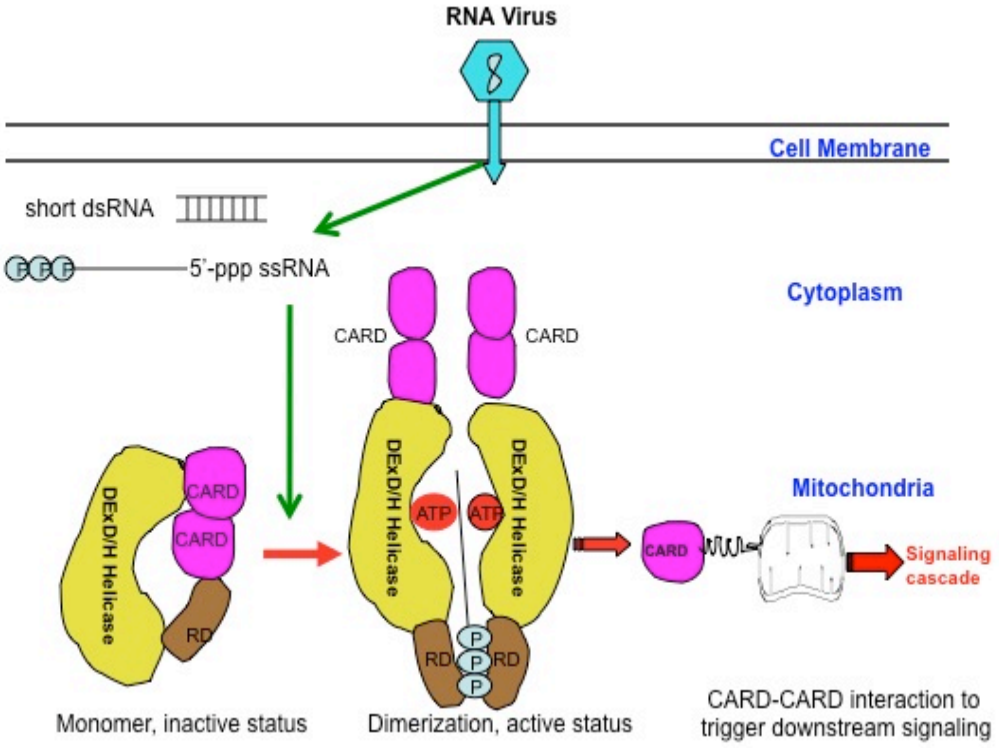


Figure 2. **Schematic representation of a previously proposed model for RIG-I activation.** In this model, it was proposed that in the absence of short dsRNA or 5'-triphosphate ssRNA, RIG-I keeps in a closed conformation, that is, in the monomeric and inactive form. Upon detecting and binding ligand RNA, RIG-I changes its conformation and releases CARD domain from constraint. The released CARD domains then interact with each other to form RIG-I dimer, thereby triggering the downstream signaling via the interaction with IPS-1 CARD domain.

1.6 Aims and Scope of the Thesis

RIG-I signaling pathway plays an important role in antiviral innate immune responses. Although intense research has been conducted to clarify the functions and signaling pathway of RLRs, the investigation into molecular structure and biochemical properties of RIG-I in vitro has fallen behind. Previous studies suggested that, in resting cells, RIG-I remained in a closed, auto-repressed conformation. Upon binding dsRNA or 5'-triphosphate ssRNA, RIG-I underwent a conformational change that promotes self-association and then interaction with IPS-1 to trigger IFNs activation and production. However, a clear molecular mechanism of how RIG-I domains interacts with each other in the absence and presence of RNA, and how RIG-I distinguishes non self viral RNA from self cellular RNA, and how RIG-I changes its conformation upon binding RNA, still remains largely unknown.

In this thesis, we proposed to use X-ray crystallography in combination with biochemical and biophysical approaches to investigate the structures of RIG-I apo form and in complex with RNA duplex. Our long-term goal was to dissect the functional cooperation of the three domains of RIG-I in ligand RNA recognition, and to shed new light on the signaling mechanism behind the host-virus interaction, thereby providing a structural foundation for developing novel therapeutic strategies to manipulate RIG-I-mediated antiviral responses.

2. Materials and Methods

2.1 Materials

Unless otherwise specified, all common chemicals were ordered from Sigma-Aldrich/Fluku. The restriction enzymes and T4 Polynucleotide Kinase (PNK) were purchased from New England Biolabs (NEB). For DNA amplification by PCR, Clontech Advantage HD polymerase was used. RIG-I cDNA template for molecular cloning of expression constructs was a gift from Dr. Michael Gale. DNA oligonucleotides for subcloning were ordered from Sigma-Aldrich. Bacteria strains used for making chemical competent cells are purchased from Novagen, and the antibiotics (Ampicilin, Kanamycin, Chloramphenicol) were ordered from Fisher Scientific. All chromatographic systems and columns were obtained from GE Healthcare. SUMO protease (Ulp1 protease) and PreScission protease were expressed and purified in the lab. Sequencing-Grade proteases used in limited proteolysis were purchased from Roche. Radioactive material (γ-32P-ATP) was ordered from Perkin-Elmer. Oligonucleotides for protein-RNA studies in RNA binding and crystallization trials were synthesized either from Dharmacon (Thermo Scientific), IDT, or Bio-Synthesis Company. All crystallization screens, crystallization grade reagents, cryo-tools, were obtained from Hampton Research, Qiagen and Molecular Dimensions Protein Crystallography Suppliers.

2. 2 Methods

2.2.1 Protein expression and purification

All protein constructs were subcloned into a modified pET28a-His-smt3 vector with a Clontech's In-Fusion cloning strategy, except that RIG-I N-terminal CARDs was cloned into pGEX-6P-1 vector (GE healthcare). RIG-I N-terminal CARDs domain (residues 1-228) was overexpressed as GST fusion protein in E.coli strain BL21 (DE3). After releasing the GST tag by PreScission protease, CARDs domain was further purified through HiTrap Q anion exchange column and gel filtration chromatography (Superdex200, GE Healthcare).

Human RIG-I Helicase-RD (residues 232-925) was overexpressed in *Escherichia coli* strain Rosetta 2 (DE3) (Novagen) as a soluble protein. Harvested cells are lysed by passing through Emusiflex homogenizer (Avestin, Canada) three times at 15,000 Psi and clarified by centrifugation at 20,000g for 45 min. The isolation of the pure protein involved three chromatographic steps: Ni^{2+}-nitrilotriacetate (Qiagen) column, hydroxyapatite column (CHT-II, Bio-Rad), and heparin sepharose column (GE Healthcare). Recombinant full-length RIG-I (residues 1-925), RIG-I ΔRD (residues 1-794), RIG-I ΔCARD1 (residues 96-925), RIG-I CARD1-Helicase (residues 96-794) and RIG-I Helicase (residues 232-794) were expressed and purified in a similar manner with an additional gel-filtration chromatography step (Hiload 16/26 Superdex200, GE Healthcare) in order to remove aggregates and protein impurities. RIG-I RD (794-925) was expressed in E. coli BL21 Star (DE3) cells and the soluble fraction was purified to homogeneity using a Ni^{2+}-nitrilotriacetate column, further by cation exchange (HiTrap SP, GE Healthcare) and gel filtration chromatography.

Protein concentrations were measured by Bradford assay (Bio-Rad), using bovine serum albumin (Pierce) as a standard. All purified RIG-I proteins were further dialyzed overnight at 4°C into 50 mM HEPES 7.5, 50 mM NaCl, 5mM DTT, 10% glycerol, snap frozen in liquid nitrogen, and stored at -80 °C.

The Selenomethionine (SeMet)–labeled RIG-I Helicase-RD protein was produced in Rosetta 2 (DE3) cells grown in M9 minimal medium supplemented with 60 mg ml^{-1} L-SeMet (Sigma) and specific amino acids (100mg L-Lys, Phe, and Thr; 50mg L-Ile, Leu, Val) to inhibit endogenous methionine synthesis(Van Duyne et al., 1993). The SeMet protein was then purified with the same procedure as the native protein.

2.2.2 ATPase activity

A time course (0, 5, 10, 15, 30 min) of the ATPase reactions were carried out using RIG-I full length protein, RIG-I ΔCARD1 (residues 96-925), and helicase-RD (10 nM), ATP (1 mM) spiked with traces of [γ-^{32}P] ATP with and without 14bp dsRNA (80 nM) in buffer containing 50mM MOPS-NaOH (pH7.4), 5mM $MgCl_2$, 5mM DTT, 0.01%Tween 20 at 37°C. The ATPase activity of the RIG-I ΔRD (residues 1-794), RIG-I CARD1-Helicase (residues 96-794) and RIG-I Helicase (residues 232-794) were measured using a higher concentration of protein (100 nM) and RNA (1 μM). Reactions were incubated for 0, 5, 10, 15, 30, 45, and 90 min at 37°C and stopped by adding 20 μl of 4 M formic acid. All the quenched reactions were then analyzed by PEI-Cellulose-F TLC plate (Merck) developed in

0.4 M potassium phosphate buffer (pH 3.4). The TLC plates were then exposed to a phosphorimager and quantified using ImageQuant software (GE Healthcare).

The fraction of hydrolyzed ATP was calculated according to following equation: Fraction of hydrolyzed ATP = amount of liberated [γ-^{32}Pi] / (amount of unhydrolyzed [γ-^{32}P] ATP + liberated [γ-^{32}Pi]). The ATPase rate was determined from the plots of fraction of free [Pi] produced versus time and the k_{cat} values were calculated by dividing the ATPase rate by the respective enzyme concentration using the curve-fitting software Prism 5 (GraphPad software).

2.2.3 5' End-labeling of the RNA oligonucleotides

The RNA oligonucleotides were purchased either from Dharmacon (Thermo Scientific) or IDT. The RNA oligos were then fully deprotected, desalted, annealed and ready-to-use for downstream assays. The RNA double-stranded oligonucleotides were then end-labelled with [γ-^{32}P] ATP using T4 Polynucleotide Kinase (PNK) (NEB) according to the manufacturers' instructions. Labeled oligos were further purified by a Micro Bio-Spin 6 column (Bio-Rad) equilibrated in 10 mM Tris-HCl pH 7.5, 50 mM NaCl, and quantified using a liquid scintillation counter.

2.2.4 Electrophoretic mobility shift assays

For the gel shift assay, 2.5 fmol labeled RNA was mixed with 0, 0.5, 1, 2, 5, 10, 20, 50, 100 nM protein in a 20 µl reaction containing 30 mM HEPES (pH 7.5), 50 mM NaCl, 2 mM DTT and 5% (v/v) glycerol. Binding reactions were performed at room temperature for 15 min, and the mixtures were further incubated 15 min at 4°C before adding 5 µl of 15% (w/v) Ficoli and 0.2% (w/v) Orange G dye dissolved in binding buffer. The samples were run on a 6% (w/v) non-denaturing Tris-glycine-polyacrylamide gel (37.5:1 acrylamide/bisacrylamide) at 4°C in 0.5xTB electrophoresis buffer**.** After electrophoresis, the gels were dried and visualized by phosphorimaging. Data were analyzed using the Prism 5 software (GraphPad Software) and the nonlinear curve fitting methods were used to determine the apparent equilibrium dissociation constant (Kd), with an estimate of Hill coefficient. It is noted that the sum of the bound complexes in each lane was considered as the total bound fraction and the data was fitted to a one-site binding model.

For competition experiments, nonradiolabeled RNA was added either 5 min before the end of the preincubation step, or premixed with [^{32}P]-labeled RNA at the same time. For analysis of RIG-I/HCV RNA complex formation, 0-60 pmol of purified full length RIG-I or RIG-I CARDs (residues 1-228) protein were incubated for 15 min at room temperature with 6-10 pmol of the in vitro transcribed RNA (containing 5'ppp) or chemically synthesized RNA (lacking 5'ppp but containing 5'OH) in 1xbinding buffer (20mM Tris-HCl pH8.0, 1.5mM $MgCl_2$, 2mM DTT). The

mixture was then subjected on a 2% RNA agarose gel. The gel was stained with Sybr Green II (Lonza) and visualized using a UV illuminator (302 nm).

2.2.5 Limited proteolysis

Limited proteolysis with trypsin, chymotrypsin, Glu-C and subtilisin, was performed using 120 μg of purified RIG-I proteins in the absence or presence of RNA ligands and incubated with trypsin at a protein:protease mass ratio of 300:1. The reactions were carried out at room temperature or 37°C. Aliquots were removed at different time points (0, 10, 15, 30, 60, and/or 120 min) and the reactions were quenched by the addition of equal volume of 2xSDS-PAGE loading buffer. Products were analyzed by SDS-PAGE and visualized by Coomassie Brilliant Blue staining.

2.2.6 Analytical gel filtration chromatography

Analytical gel filtration chromatography was carried out on AKTA FPLC system (GE Healthcare). Proteins were loaded on to 24 ml of Superdex200 10/300GL column (GE Healthcare) equilibrated with a buffer containing 50 mM HEPES 7.5, 50 mM NaCl, 5mM $MgCl_2$, 5 mM DTT. The eluate was monitored by ultraviolet absorbance at 280 nm and 260 nm. For making the RIG-I-RNA complex, RIG-I proteins were pre-incubated with 14bp dsRNA at the molar ratio of 1:1.5 on ice for 15 min before the sample was applied onto column. The Superdex200 column was calibrated with gel-filtration protein standards (Sigma).

2.2.7 RIG-I Helicase-RD-RNA complex preparation

The blunt-end self-complimentary RNA oligonucleotides were commercially synthesized from Dharmacon, and then deprotected and desalted into 20mM potassium phosphate, pH7.4. Annealing was performed by incubation at 95 °C for 1 min in the heating block followed by gradual cooling to room temperature. Hairpin RNA were purchased from IDT and directly resuspended in 50 mM HEPES 7.5 buffer. Chemically synthesized 5'-triphosphated RNA oligonucleotides (5'ppp-dsRNA or 5'ppp hairpin RNA) were purchased from Bio-Synthesis Company. The resulting RNA oligos were mixed with purified RIG-I Helicase-RD at a RNA:protein molar ratio of 1.5:1, incubated at room temperature for 15 min, and then purified by gel filtration column (Superdex200, GE Healthcare) with an elution buffer consisting of 50 mM HEPES 7.5, 50 mM NaCl, 5 mM DTT, 5mM $MgCl_2$.

2.2.8 Crystallization screen and optimization

Although nowadays crystallization techniques are becoming more standardized and automatic, crystallization of protein is still the bottleneck. Generally, crystallization can be divided into two stages: (i) initial crystallization screening to obtain any kind of crystals or promising hints such as microcrystalline precipitate and phase separation, (ii) optimization of initial conditions to improve crystal quality.

The purified protein or protein-RNA complex separated from gel filtration column were concentrated to 10~15 mg ml^{-1} in a buffer consisting of 50 mM HEPES pH7.5 50 mM NaCl, 5 mM DTT, with or without 5%, 10% glycerol and then used for crystallization trials. Preliminary crystallization screens were set up by sitting-drop vapor diffusion method using commercially available crystallization kits in combination with high-throughput Gryphon robot (Art Robbins Instruments). Screen kits included crystal screen I/II, Index I/II, crystal screen lite, PEG/Ion screen, MembFac, Natrix (Hampton Research) and PEG suits series from Qiagen. Typically, the sitting drop was mixed a size of 200 nl protein solution with 200 nl reservoir solution, and then sealed against 65 µl of reservoir solution with transparent film. Once establishing promising conditions, optimization of initial conditions was conducted by equilibrating 2.5 µl protein solution and 2.5 µl crystallization solution against 500 µl reservoir solution using the hanging-drop vapor diffusion method through screening a huge range of variables including temperature, pH, precipitant concentration, ionic strength, additives, buffer types and protein concentration in order to obtain single crystals of suitable size.

2.2.9 X-ray diffraction data collection and processing

Single crystals were flash-frozen in liquid nitrogen at 100 K with cryo-protectants to minimize radiation damage as free radicals form. Diffraction data were collected at several synchrontron sources: X29A beamline of the National Synchrotron Light Source (NSLS), LRL-CAT 31-ID beamline of the Advanced Photon Source (APS), and the Cornell High Energy Synchrotron Source

(CHESS) F1 beamline. A total of 360 images were collected with 1° oscillation per frame for all available crystals.

Once we obtain the crystals suitable for X-ray diffraction and collected a complete data set, we need to process the data and evaluate its quality to make sure the data is good enough for structure determination. All diffraction data were processed using iMosflm and scaled in SCALA (1994). Basically, data reduction and scaling involves four critical steps: indexing, cell refinement, integration, and scaling. Indexing determines the crystal parameters (a, b, c, α, β, γ) and the relative orientation of the crystals from detector spot positions, and then decides a most possible space group. Cell refinement is used to further tune the parameters of the unit cell and verify if we do not find the right space group. Subsequently, the intensity (***I***) of each reflection on each of the diffraction images or frames is determined by automatically summing up the pixel values under each spot on the detector and subtracting the detector background in a local neighboring area. During the scaling, the SCALA software embedded in CCP4 first tries to put all frames on the same scale so that the values from images with different background intensity can be compared. Then scaling adds up partial reflections from adjacent frames and merges all symmetry-related reflections (***I*** (*h, k, l*) = ***I*** (*-h, -k, -l*) for non-anomalous data). It is also worth noting that the option “separate anomalous pairs for merging statistics” needs to be activated in SCALA in order to get anomalous difference, where the Friedel pairs cannot be merged because ***I*** (*h, k, l*) ≠ ***I*** (*-h, -k, -l*) in this case. In the end, a MTZ file is

generated with a list of reflections (h, k, l), intensity *I*, error σ, and calculated statistic R factor. Thus, it is not until after scaling that we can actually judge the quality of the data. The signal to noise ratio ***I/σ***(***I***), R_{sym} or even more précised indicator R_{pim}, data completeness and redundancy are used to determine the effective resolution of the data set.

2.2.10 Structure determination and refinement

The ultimate outcome in X-ray crystallography experiment is to obtain an interpretable electron-density map and build a model upon it. The electron density is calculated from the diffraction data by the equation:

$$\rho(xyz) = \frac{1}{v_c} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} |F(hkl)| e^{i\phi(hkl)} e^{-2\pi i(hx+ky+lz)}$$

where ***F***(*hkl*) is the structure factor, *hkl* is a unique set of Miller indices assigned to each individual reflection, and ν_c is the volume of crystallographic unit cell. In diffraction theory, ***F***(*hkl*) is treated as a complex number; therefore, it has a real part ($|\boldsymbol{F}(hkl)|$, also called amplitude) and an imaginary component ($\boldsymbol{\phi}(hkl)$, also called phase). That is to say, one can calculate the electron-density ρ(x,y,z) only if both the amplitude $|\boldsymbol{F}(hkl)|$ and the phase angle $\boldsymbol{\phi}(hkl)$ are known. Indeed, we can measure the intensities of the diffracted waves in an X-ray diffraction experiment, and then the intensities of reflections can be used to calculate the structure factor amplitude by

$$|F(hkl)| = \sqrt{I(hkl)}$$

However, we cannot experimentally measure the phase angle of each reflection. This is known as the so-called phase problem in X-ray crystallography. The common methods used to recover the lost phase information these days include: traditional isomorphous replacement, multiple-wavelength anomalous dispersion (MAD), single-wavelength anomalous dispersion (SAD), and molecular replacement. Since RIG-I has very low sequence similarity with known protein structure, SAD was used to determine the phase information. In SAD experiment, introducing a strong anomalous scatters into the protein crystals and collecting data using X-ray wavelength at or near the absorption edge of the anomalous scatterer can be used to resolve the phase ambiguity. For RIG-I Helicase-RD-dsRNA complex, selenomethionine-substituted protein crystals were prepared for SAD experiment in order to obtain the initial phase information. The program SHELXC/D was then used to locate the selenium atoms and the program Phaser or SHARP from CCP4 software suite was used to refine the initial heavy atom positions and calculate initial phases. The resulting electron density map was further improved by density modification techniques, such as solvent flattening. It improved the starting phase and led to better interpretable map. A partial model was built into the electron density map in COOT and the iterative process of model building and refinement was carried out in PHENIX to build the final model. Two R-factors, R_{work} and R_{free} were used to cross-validate model building and avoid over-fitting throughout the refinement process. Molecular replacement was used to solve all other RIG-I-Helicase-RD-RNA complex and ATP analog

data sets by using Phaser-maximum likelihood method implemented in the PHENIX software suite.

2.2.11 Small angle X-ray scattering and structural modeling

Small-angle X-ray scattering (SAXS) is a powerful method to study overall structure of protein or protein-nucleic acid complex in solution (Lipfert and Doniach, 2007; Svergun and Koch, 2002). Although X-ray crystallography can provide high-resolution structure, it usually presents a snapshot of the static structure in the crystal, which may be influenced by crystal packing force as well. Furthermore, sometimes it is very difficult to probe the structural changes induced by ligand binding in crystals, because of difficulty in obtaining the well-diffracting crystals. However, SAXS data can be used to determine the low-resolution macromolecular envelopes by *ab initio* modeling (Hura et al., 2009). This resolution is often sufficient to address key biological questions, such as conformational change and model building, especially in combination with X-ray crystallography data (Putnam et al., 2007). Therefore, this unique technique is becoming more and more valuable now days because it provides useful information about the overall structure of large complexes and their response to external stimuli like changes in pH, salt, or addition of ligands.

In a crystal diffraction experiment, if we place a crystal in an X-ray beam with a wavelength of ~0.5-2.5 Å, we will see light scattering from the crystal. The

scattering of X-ray by electrons from the repeating array of crystal lattices gives rise to interference effects. Constructive and destructive interference combine to yield an X-ray diffraction pattern in a regular fashion on the detector. We then use the diffraction intensities and phase angles to calculate the distribution of electron density within the protein and thereby derive the position of atoms in the protein relative to a common origin. For a SAXS experiment, if we put a protein solution in an X-ray beam, we see scattering pattern too. Since protein size is >> incident X-ray wavelength and protein in solution is randomly oriented, the X-ray scattering is radically isotropic and therefore all positional information is lost. However, if the scattering profile is measured at very low scattering angles (2θ), the scattering profile is the scattered intensity **I**(q) recorded solely as a function of the momentum transfer q, where q= $4\pi\sin\theta/\lambda$ ($\propto \theta$ for small θ). The scattering from the buffer alone must be subtracted from the scattering of the protein sample in order to obtain the final scattering dataset for the hydrated protein molecule. The reduction data consists of the intensity **I**(q_i), the momentum transfer q_i, and the standard error σ_i, where i= 1, 2, …, N and N is the number of points for which the scattering intensity is recorded. The scattering data can be then used to obtain the radius of gyration (Rg) determined from Guinier plot as well as other shape information from the pair-distribution function P(r). Rg is the root-mean-square of the distances of all regions to the center of mass of the particle weighted by their electron density and can be approximated according to $I(q) \approx I(0)\exp(-\frac{1}{3}Rg^2q^2)$, where the slope of curve yields Rg. P(r) represents the probability distribution of distances between scattering atoms within a particle

and can be determined based on the Fourier transform equation, $I(q) = 4\pi \int_0^{D\max} p(r) \frac{\sin qr}{qr} dr$, where D_{max} is the maximum distance in the particle. Usually, globular proteins have a P(r) function with a symmetric single peak (Guassian distribution), whereas elongated or extended proteins have a long tail at large r (Skewed distribution) and can have multiple peaks (Figure 3, left). In addition to obtaining the overall size and shape of a macromolecule, we also can analyze protein folding/unfolding and flexibility within in a protein by Kratky plot ($I(q)q^2$ vs q). Generally, Kratky plot for globular protein or compact structure appears as a bell-shape curve and becomes flat at high q values (Figure 3, right). Kratky plot for extended molecules, such as flexible structure, becomes more pronounced with an increase in $I(q)q^2$ at intermediate and high q. Unfolded protein, however, completely lose the bell-shape curve even at low q, and become a plateau at high q.

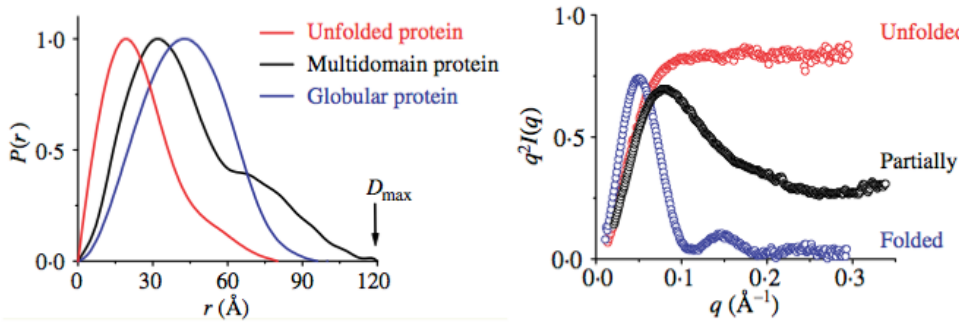


Figure 3. SAXS data analysis (adapted from John A. Tainer et al. Quarterly Reviews of Biophysics, 2007): P(r) function (left) and Kratky plot (right).

The structural model can be reconstructed *ab initio* from SAXS data as well. The basic idea for this method is to define a spherical search space that is big enough to represent the whole protein with radius $R_{sphere}=D_{max}/2$. Then the search volume is filled with dummy atoms with a size of $r << R_{sphere}$. These dummy atoms are assigned either as protein atoms or solvent atoms. Remember, only protein atoms will contribute to the scattering from the model. Now if we keep changing the positions and numbers of dummy atoms in the protein part, we will get a lot of different models. Once we compare the calculated scattering curve from those dummy atom models with the measured, experimental scattering data, we begin to narrow down the pool of test models. By *trial and error*, the structural model will be optimized in order to fit the best to the experimental scattering data. By this means, we can reconstruct low-resolution solution structures.

We employed SAXS to study the solution structure of RIG-I full-length protein in the absence or presence of RNA ligands to explore the conformational changes of RIG-I upon RNA binding. Two programs, DAMMIN and GASBOR, were used to generate the molecular envelope (Hura et al., 2009). Combining SAXS data with atomic resolution structures enabled us to generate more accurate model for RIG-I activation. In brief, SAXS data were collected at the CHESS beamline G1 using a Finger Lakes CCD X-ray detector system with a sample-to-detector distance 1450mm to make scattering vectors q range from 0.01 to 0.233 $Å^{-1}$, where $q= 4\pi\sin\theta/\lambda$ (2θ is the scattering angle and $\lambda= 1.296Å$). All samples were

exchanged into 50 mM HEPES 7.5, 50 mM NaCl, 5 mM DTT, 5% glycerol by size exclusion chromatography (Hiload 16/26 Superdex200, GE Healthcare) to minimize the discrepancies in background subtraction. Various programs in ATSAS software package were used to process and evaluate scattering data. Radius of gyration (Rg) was analyzed using the Guinier approximation with low angle data ($q<1.3/Rg$). The probability distribution of distances between scattering atoms within the macromolecule, P(r), and the maximum atom pair distance, Dmax, were determined from the scattering data using the GNOM algorithm. The program DAMMIF was used to calculate low-resolution *ab initio* shape from experimental SAXS profiles. Ten to twenty independent models were aligned, filtered, and averaged based on the occupancy using SUPCOMB and DAMAVER to reconstruct the final *ab initio* envelope, as judged by averaged normalized spatial discrepancies (NSD<1.0). The structure modeling was refined against the solution scattering data by rigid body docking using SASREF program. The orientation and position of individual domains in the structure models were further manually adjusted to minimize the discrepancies (χ^2) between the calculated scattering intensities and experimental scattering intensities computed using program CRYSOL (Bernado et al., 2007).

2.2.12 Thermal shift assay

Thermal shift assay was conducted with 10 μM of RIG-I Helicase-RD or full-length RIG-I with or without 12 μM of 14 base pair self-complementary dsRNA in 50 mM HEPES 7.5, 50 mM NaCl, 5mM MgCl2, 5 mM DTT, and a 5x dilution of

SYPRO Orange dye (Invitrogen). The fluorescence signal as a function of temperature was recorded using a Real Time PCR machine (Applied Biosystems). The temperature gradient is performed in the range of 25–80°C with a ramp of 0.2°C over the course of 60 minutes. Control assays were carried out with buffer in the presence or absence of RNA. Data were analyzed with the Excel-based worksheet DSF analysis, and Boltzmann sigmoid function was used to fit the fluorescence data to obtain the midpoint temperature for the thermal protein unfolding transition (*T*m) using the curve-fitting software Prism.

2.2.13 Hydrogen/deuterium exchange mass spectrometry

Many analytical methods, such as thermal shift assay, circular dichroism (CD), limited proteolysis, and analytical ultracentrifugation (AUC) have been used to study protein conformational changes upon binding ligands. However, all these methods provide only the global conformational changes. For example, CD tells us the changes of content of secondary structures within in protein, but it cannot determine where the conformational changes occur within in protein. Conversely, hydrogen/deuterium exchange (HDX) characterizes both global conformational changes and local conformational changes within in protein upon binding ligands. In hydrogen/deuterium exchange mass spectrometry (HDX-MS) experiment, protein is incubated with D_2O for different times. The backbone amide hydrogen atoms are capable of exchanging with hydrogen from D_2O in solution at the neutral pH. Solvent-exposed amide hydrogens have fast $^{1}H/^{2}H$ exchange rates, whereas the amide hydrogens buried in protein interior and/or involved in

hydrogen bonding have much slower deuterium uptake rates. Hydrogens bonded to carbon usually do not exchange with deuterium, while hydrogens in the sides exchange too quickly to be easily measured by mass spectrometry. The exchange reaction is quenched with cold acidic buffer (pH2.5). This step serves two purposes: 1) to slow down undesirable back exchange from deuterium to hydrogen and lock on the deuterium atoms that have already incorporated. 2) to facilitate acid protease digestion by mildly denature the protein in very low pH environment. After quench, the sample is then digested with pepsin and analyzed by liquid chromatography (LC)-mass spectrometry. Peptides that readily undergo mass shift represent greater deuterium exchange rates. This rapid $^{1}H/^{2}H$ exchange rate, in turn, suggests high solvent exposure. However, slow $^{1}H/^{2}H$ exchange rate doe not necessarily imply less solvent exposure and could be instead caused by high hydrogen bonding. Nonetheless, all those changes in deuterium uptake indicate there might be some global or local conformational changes taking place, either significant or subtle.

Hydrogen/deuterium exchange mass spectrometry experiments were performed as fellows. Briefly, 1.0 µl of RIG-I FL and Helicase-RD with or without RNA ligand (8 mg ml^{-1}) was diluted with 5.0 µl of D_2O buffer (50 mM HEPES, pH 7.5, 50 mM NaCl, 5 mM DTT, 5% glycerol) and incubated for 10, 100 and 1,000 s on ice. At the indicated times, the resulting mixtures were quenched by addition of 15 µl of ice-cold quench solution (100 mM TCEP, pH 2.4), immediately frozen on dry ice, and stored at −80 °C. Non-deuterated control was prepared in H_2O buffer (50 mM

HEPES, pH 7.5, 50 mM NaCl, 5 mM DTT, 5% glycerol) and then quenched and analyzed in the same way as deuterated samples. To correct back exchange during pepsin digestion and column separation, a completely deuteriumed RIG-I FL or Helicase-RD was prepared by 24 hours incubation of 1.0 µl of RIG-I protein with 20 µl of D_2O buffer containing 100 mM TCEP, pH 2.4. Frozen samples were thawed on ice and immediately manually injected into an immobilized pepsin column (66 µl bed volume) at a flow rate of 100 µl min^{-1} at 0°C, followed by passed through 200 µl of 0.05% trifluoroacetic acid at the same flow rate. After pepsin digestion, peptide fragments were separated by an online C18 HPLC column (Supelco, Sigma-Aldrich) using a linear acetonitrile gradient of 2–50% over 20 min at 200 µl min^{-1}. The eluate was then subjected on an electrospray-linear ion-trap mass spectrometer (LTQ, Thermo Scientific) for acquiring over a mass range from 300-2000 *m/z*. Sequest software was then used to search the sequence of the peptide ions. Mass analysis of the peptide centroids and peptide identification were carried out using HDExaminer (Sierra Analytics Inc.).

3. Results

3.1. RIG-I and HCV PAMP RNA interaction

3.1.1 Expression and purification of recombinant full-length human RIG-I

The wild-type cDNA encoding full-length human RIG-I have been successfully subcloned into three versatile prokaryotic expression vectors: pGEX-6p-1 (GE healthcare), pET28a-SKB2-His (a modified pET28a vector containing PreScission cleavage site instead of Thrombin recognition site), pET28a-His-smt3 vector (a modified pET28a vector producing N-terminal His_6-SUMO fusion protein). Recombinant full-length RIG-I was then overexpressed in *Escherichia coli* strain Rosetta 2 (DE3) (Novagen), and all three expression vectors were then compared based on protein expression levels. The optimization of overexpression and purification showed that SUMO fusion tag provided exceptionally high yields (~15mg/L; compare to ~0.5mg/L for GST fusion protein and ~0.2 mg/L for conventional expression using N-His_6 tag). It is also noteworthy that in an effort to increase protein expression and solubility, we extensively examined expression combinations, such as inducing at lower temperature, reducing IPTG concentration, inducing for less time, inducing earlier in growth, changing the growth medium, using glucose to control basal expression, expressing in different cell strains, trying heat shock growth or cold shock growth, checking different lysis buffer at different pH. It turned out that two important factors are essential for this high overproduction of full-length SUMO-fusion RIG-I in E. coli expression system: 4°C cold shock for at least 3 hours before induction at low temperature (18°C), and a lower IPTG concentration

(0.05-0.1 mM). Once the optimal conditions for overexpression were established with pET28a-His-smt3 vector, a combination of chromatography, including immobilized metal ion affinity chromatography (IMAC), hydroxyapatite column, heparin affinity column, blue sepharose column, ion-exchange column and gel-filtration chromatography were employed to determine the best purification scheme.

Briefly, full-length RIG-I protein was first purified by immobilized metal ion affinity chromatography (Figure 4A), followed by adding Ulp1 protease to remove SUMO tag (Figure 4B). Subsequently, recombinant RIG-I full-length protein was further purified to homogeneity through hydroxyapatite column, heparin affinity column, and gel filtration chromatography on 26/60 Sephacryl S-200 HR (GE, Healthcare) (Figure 4C-D). Among them, hydroxyapatite column is the most critical purification step for removing nucleic acid contaminations, whereas heparin column is important for getting rid of inactive or less active RIG-I protein. Those experiments demonstrated that E.coli expression system can be used to obtain large amounts of soluble RIG-I full-length protein in an easy-to-purify form, compared to baculovirus expression system described in literature (Saito et al., 2007).

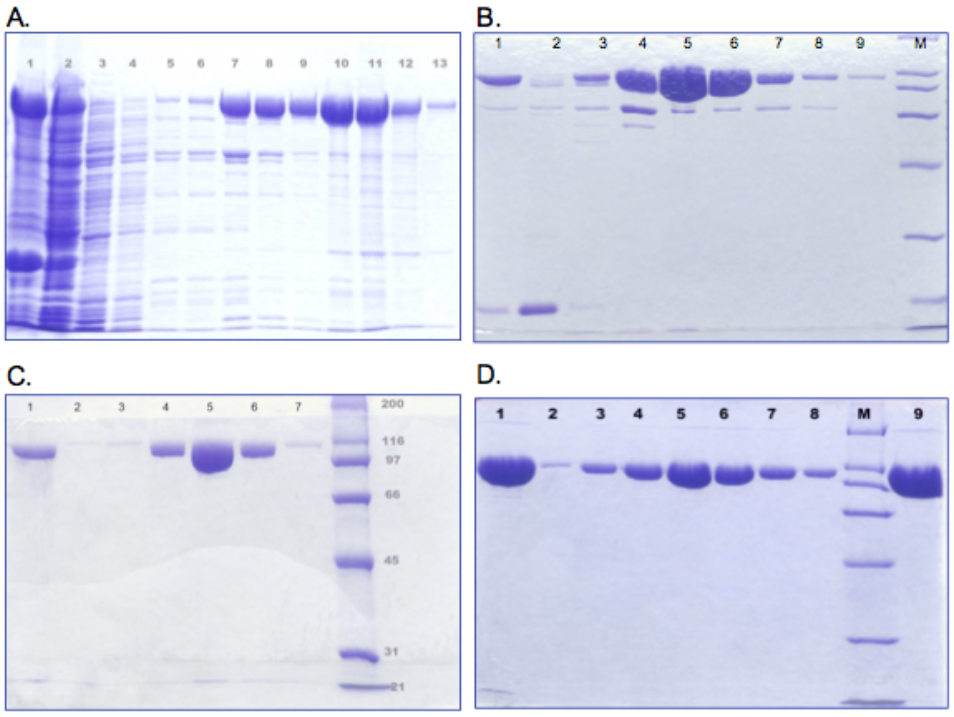


Figure 4. SDS-PAGE of RIG-I FL purification. (A) SDS-PAGE of elutes from immobilized metal ion affinity. Lane 1: total cell extract. Lane 2: supernatant of cell extract. Lane 3: elute using 10mM imidazole. Lane 4: elute using 20 mM imidazole. Lane 5: 30mM imidazole. Lane 6: 50mM imidazole. Lane 7-9: 100mM imidazole. Lane 10-11: 200 mM imidazole. Lane 12-13: 250 mM imidazole. (B) SDS-PAGE from hydroxyapatite chromatography. Lane 1: RIG-I protein after removing SUMO tag. Lane2: flow-through. Lane3: minor peak on hydroxyapatite chromatography. Lane 4-9: major peak fractions. (C) SDS-PAGE analysis of protein sample from heparin column. Lane 1: purified RIG-I sample from hydroxyapatite column. Lane 2-3: flow through, first peak fractions on heparin chromatography. Lane 4-7: second peak fractions. (D) SDS-PAGE analysis of the fraction from gel filtration chromatography.

3.1.2 Characterization of recombinant full-length human RIG-I

We have performed a series of biochemical and biophysical experiments to evaluate the purity, stability, homogeneity and monodispersity of recombinant full-length human RIG-I (Figure 5). Non-reducing SDS-PAGE of the RIG-I FL protein showed a similar migration to the reduced sample (Figure 5A), indicating that there are no inter-molecular and intra-molecular disulfide bond formation in RIG-I FL protein. Native PAGE experiment was also carried out to further confirm its purity, homogeneity and stability (Figure 5B). Isoelectric focusing (IEF) showed that the full-length protein has a net charge of zero at pH6.0 (Figure 5C). Additionally, UV scan (340~220 nm) provided a symmetric curve with a maximum at 280 nm and no shoulder at 320 nm (Figure 5E), implying that purified RIG-I full-length protein is free of nucleic acid contaminations with no obvious aggregate in solution. Moreover, the integrity of the full-length protein was verified by MALDI-TOF mass spectrometry (Figure 5F). The spectra showed three major peaks. One peak at m/z = 106419.4 agrees with the calculated molecular weight for full-length RIG-I (106,555.5 Da). The second peak at m/z = 53,233.3 and third peak at m/z = 35,484.03 corresponds to the doubly charged and triply charged species, respectively. In order to determine the oligomeric state of RIG-I under the native conditions, purified full-length human RIG-I protein was analyzed by the combination of analytical size-exclusion column and multi-angle laser light scattering. The sample showed a monomodal particle size distribution with a molar mass of 108,500 Da, corresponding to the molecular

weight of the monomer (106,555.5 Da) (Data not shown). Dynamic light scattering further showed the >95% monodispersity of RIG-I (Figure 5G).

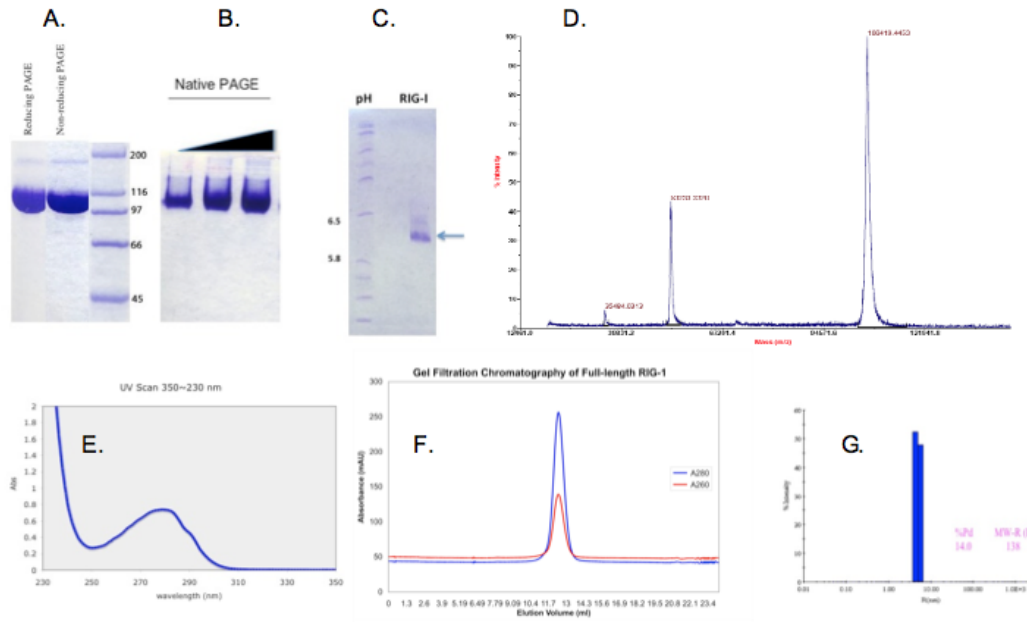


Figure 5. RIG-I FL characterization. (A) SDS reducing and non-reducing PAGE. (B) Native PAGE analysis of purified RIG-I. (C) Isoelectric focusing electrophoresis. (D) Mass spectrometry analysis of purified RIG-I to confirm the homogenity. (E) UV Scan 350-230 nm. (F) Analytical size-exclusion chromatography on Superdex 200 to test oligomeric state in solution. (G) Dynamic light scattering analysis to check the monodispersity of purified RIG-I.

3.1.3 dsRNA-dependent ATPase activity by RIG-I

RNA helicases are usally considered to be enzymes that catalyze the separation of double stranded RNA in an ATP-dependent manner (Singleton et al., 2007). They have been widely found in viruses, bacteria, and eukaryotes. RNA helicases play important roles in the rearrangement, disruption or modification of RNA structure and /or RNA-protein interactions. Most RNA helicases belong to Superfamily 2 (SF2), and have been classifed into the DEAD, DEAH, DExD and DExH subfamilies based on Walker B motif. (Singleton et al., 2007). It is generally believed that DExD/H RNA has nucleic acid-stimulated ATPase (ATP hydrolysis) activity. However, only a small number of DExD/H RNA helicases have exhibited ATP-dependent RNA helicase activity (duplex unwinding) in vitro. Based on the sequence alignment, RIG-I, MDA-5, and LGP2 belong to SF2 helicases and share almost all highly conserved signature motifs that mediate ATP and RNA binding (See amino acid sequence alignment in Appendix).

In collaboration with Dr. Smita S. Patel at UMDNJ, we investigated those activities in vitro using the highly purified RIG-I protein. In agreement with published data, our preliminary results showed that recombinant full-length RIG-I has no intrinsic ATP hydrolysis activity, but exhibits RNA-dependent ATPase activity (data not shown).

The fact that RIG-I-Like receptors harbor necessary DExH/D box helicase motifs suggested that these receptors might have helicase activity. Indeed, previous

studies had shown that RIG-I could use the energy derived form ATP hydrolysis to unwind dsRNA (Cui et al., 2008; Takahasi et al., 2008). Since the protein purity is crucial for detecting duplex unwinding activity and ruling out any false positive results due to protein impurities and contaminations, we next examined the helicase activity in vitro using our highly purified human RIG-I full-length protein. Surprisingly, in our hands, RIG-I did not show any helicase activity based on standard oligonucleotide displacement assays. RIG-I protein used in previous helicase tests might not be pure enough or RIG-I needs the interacting partners to exhibit duplex unwinding activity. Nontherless, our studies suggests that ATP hydrolysis by RIG-I is not involved in RNA unwinding, consistent with the observation from single molecule studies on RIG-I translocase (Myong et al., 2009).

3.1.4 Domain mapping by limited proteolysis

Limited proteolysis is a very useful technique for elucidating domain boundaries (Cohen and Chait, 2001; Fontana et al., 2004). The success of this technique resides in the fact that limited proteolysis of a globular protein often occurs at the “hinge” regions, exposed loops, or linkers between globular domains, which usually are flexible or disordered in the protein. To accurately define the domain boundaries of full-length RIG-I, the purified RIG-I full-length protein was subjected to limited proteolysis. Mass spectrometry and N-terminal sequencing was used to identify the corresponding proteolytic fragments (Gao et al., 2005). Trypsin proteolysis revealed a large N-terminal domain (~78KD), a linker region

(~19.7KD), and a small C-terminal domain (8.7 KD) (Figure 6A). In contrast to trypsin-limited proteolysis, subtilisin-partial proteolysis produced five stable fragments (Figure 6B). These observations demonstrated that full-length RIG-I is a multi-domain protein and that the region nearby Lys^{679}-Phe^{685} and Lys^{851} site are most likely exposed on the surface (Figure 6C).

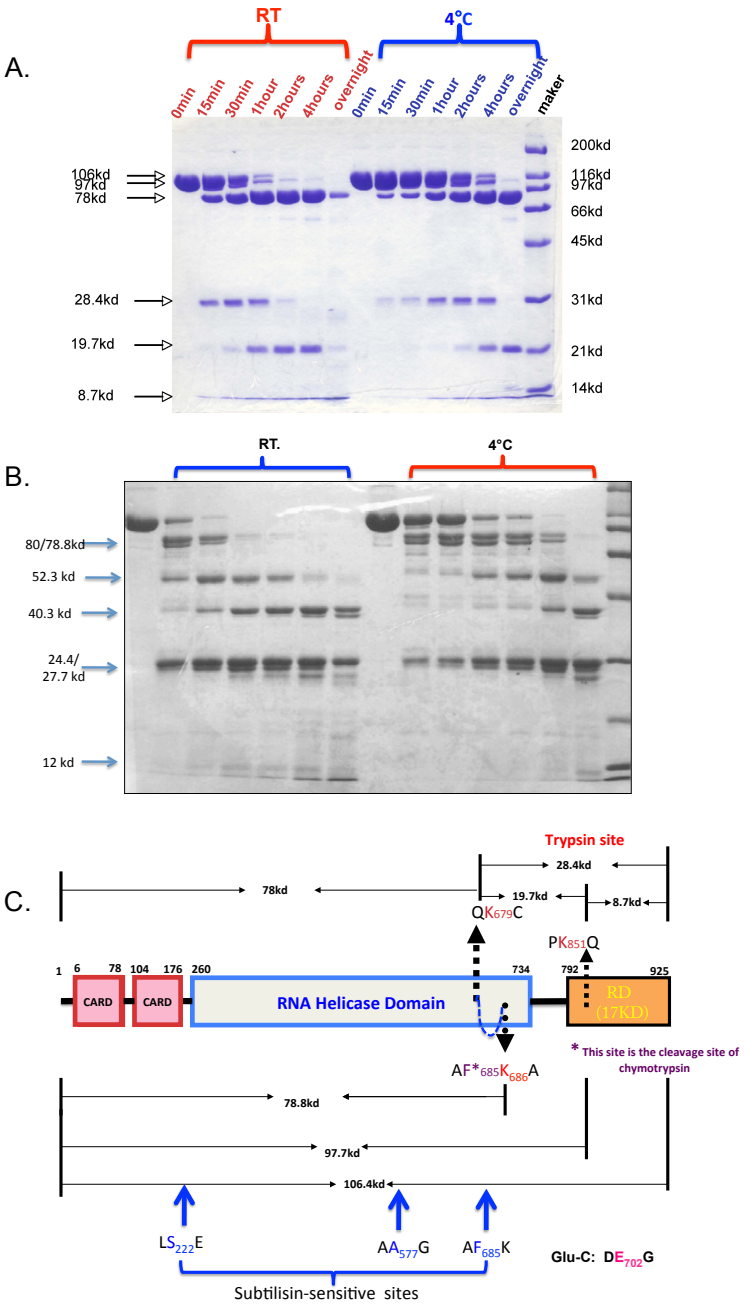


Figure 6. Domain mapping by limited proteolysis.

(A) RIG-I was treated with trypsin (1:300 mass ratio) at RT. (left) and 4 °C (right). (B) Subtilisin digestion. 27kd upper bands (685-925aa RD domain) always appear on the gel first, and then 24.4kd (1-222aa, CARDs domain). (C) Domain mapping of each proteolytic fragments. Protease-sensitive sites identified in this study are indicated.

3.1.5 investigation of RIG-I-HCV PAMP interaction by limited proteolysis

In addition to probing for domain boundaries, limited protease digestion can be applied to obtain evidence of protein conformational changes induced by ligand binding, since ligands may provide steric protection for the binding region that could not be afforded by the free-protein (Lorsch and Herschlag, 1998a, b; Plyte and Kneale, 2001). To further explore ligand-induced conformational changes within full-length RIG-I, limited protease digestion experiments were accordingly performed.

To our surprise, ATP analogs have little, if any, effect on the cleavage of RIG-I (Figure 7A). However, HCV PAMP RNA conferred strong trypsin resistance to such an extent that no cleavage occurred at Lys^{851} site even after 1 hour of incubation with trypsin. Moreover, poly I:C, a synthetic double-stranded RNA which mimics viral double-stranded RNA, offered the protection at both cleavage sites (Lys^{679} and Lys^{851}), especially the Lys^{851} site in the presence of ATP (Figure 7B). Similar results were also observed in subtilisin digestion with polyI:C, where the subtilisin-sensitive sites Ala^{577} and Phe^{685} were strongly protected, but Ser^{222} was more accessible, indicating CARD domain was released from constraint upon binding RNA (Figure 7C).

Taken together, these observations provide evidence for conformational changes within RIG-I upon binding the ligands and implied that (i) RIG-I probably undergoes a significant conformational change upon RNA binding rather than

ATP binding. (ii) In the case of poly I:C, ATP binding is critical for the further, but minor conformational change of RIG-I, which would be caused by the ATP hydrolysis. That is to say, there are at least two distinct events on the RIG-I conformational changes: RNA binding cause the major conformational changes of RIG-I and the ATP binding further gives rise to the second minor conformational change.

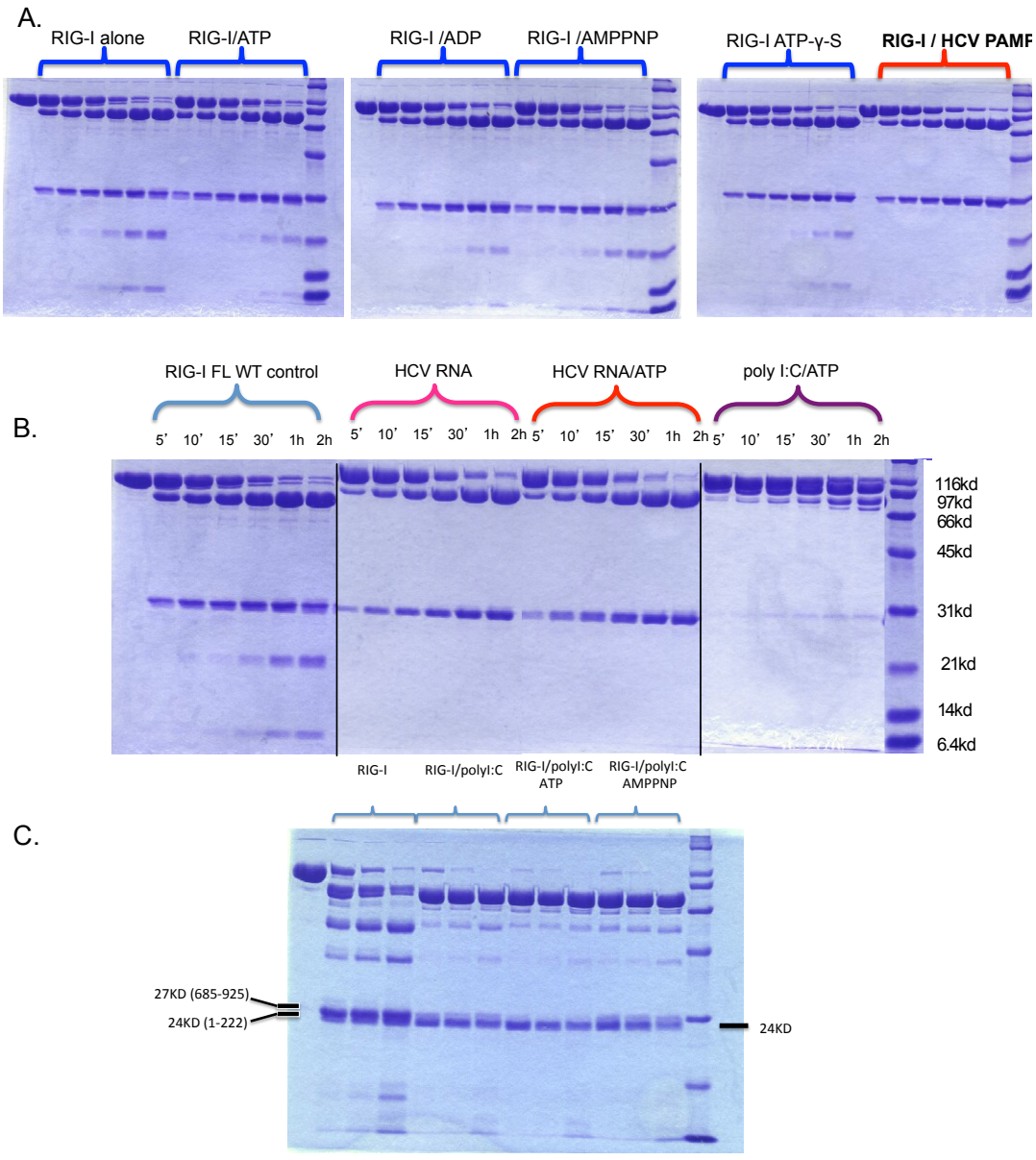


Figure 7. Investigation of RIG-I-RNA interactions by limited proteolysis

(A) Trypsin digestion of RIG-I FL in the absence or presence of ATP analogs. (B) Comparision of lmited trypsin proteolysis pattern of RIG-I in complex with HCV or poly I:C. (C) Limited substilisin proteolysis of RIG-I in complex with HCV or poly I:C. The upper fat band corresponds to RIG-I (222-925) based on N terminal sequencing.

3.1.6 Native gel electrophoresis of RIG-I with HCV PAMP

The host has developed mechanisms to detect specific viral PAMPs and initiate antiviral responses. Hepatitis C virus (HCV) is a positive-sense single-stranded RNA virus that replicates in the liver, and infects 200 million people worldwide. In collaborated with Dr. Takeshi Saito and Dr. Michael Gale at University of Washington School of Medicine, we were able to identify the poly-U/UC motif within HCV genome 3' non-translated region (NTR) and its replication intermediate as a PAMP substrate of RIG-I recognition (Figure 8). Highly purified recombinant human full-length RIG-I protein was used to verify the HCV PAMP RNA *in vitro* by gel shift assay. In brief, in vitro-transcribed poly-U/UC or X region RNA (6 pmol) were incubated with increasing concentration of purified RIG-I (0, 10, 20, 40, or 60 pmol). Native gel electrophoresis and subsequent SYBR green staining were then used to assess the complex formation. As shown in Figure 8C, the poly-U/UC, but not X region of HCV-3'NTR, formed a stable complex with purified RIG-I. Furthermore, gel shift analysis of complex formation between 25 pmol of purified RIG-I CARDs (residues 1–228, control) or full-length RIG-I (Durr et al.) and 10 pmol of poly-U/UC or X region RNA in the presence or absence of 5'-triphosphate moiety, revealed that 5′ terminal triphosphate on poly-U/UC motif was essential for mediating stable RIG-I-RNA interaction (Figure 8D). These results, together with RIG-I signaling and other in vivo analyses conducted by Dr. Takeshi's, defined the 100-nucleotide poly-U/UC region of the HCV genome and replication intermediate RNA as the HCV PAMP motif and potential substrate of RIG-I signalling.

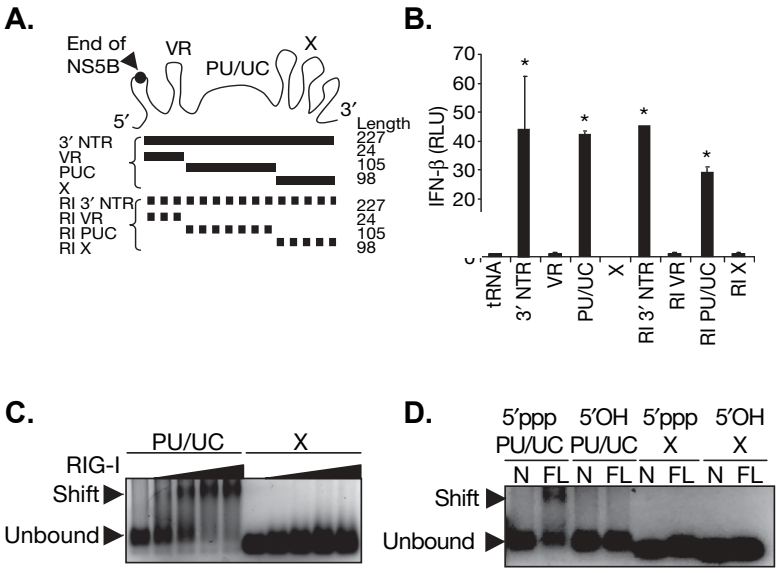


Figure 8. Identification of HCV PAMP RNA.

(A) The HCV 3'-NTR motifs and respective RNA constructs. RI and broken lines denote replication intermediate. PUC, PU/UC; VR, variable region. (B) IFN-β promoter activation, shown here as mean relative luciferase units (RLU), triggered by 1 mg of the indicated RNA species in transfected Huh7 cells. (C) RNA binding/gel-shift analysis of purified RIG-I with poly-U/UC or X region RNA (6 pmol) reacted with 0, 10, 20, 40, or 60 pmol of RIG-I protein. All RNAs contain 5'ppp. (D) Gel-shift analysis of complex formation between 25 pmol of purified RIG CARDs (residues 1–228, control) or full-length RIG-I (FL) and 10 pmol of poly-U/UC (PU/UC) or X region RNA containing 5'ppp or 5'OH as indicated. Arrows denote position of unbound RNA and RNA–RIG-I complexes (Saito et al., 2007).

3.2. Structure of RIG-I Helicase-RD-ADP•BeF_3-dsRNA ternary complexes

3.2.1 Expression, purification and crystallization trials of RIG-I constructs

Several truncation constructs listed in Figure 9A were designed based on previous functional studies and bioinformatics analysis such as secondary structure prediction, disorder prediction, globularity prediction by GlobPlot method and Hydrophobic Cluster Analysis method, as well as domain prediction by Pfam/SMART. Based on our limited proteolysis and structure-based sequence alignment with other helicases (eIF4A, vasa, mjDEAD, Hef), we also made eight deletion constructs between helicase motif IVa and V. All the truncations and deletions were cloned and expressed using the similar protocol as recombinant RIG-I full-length protein with some minor modifications, and were then purified to homogeneity through a series of chromatography steps including hydroxyapatite column, heparin affinity column followed by gel filtration chromatography. The observed molecular weights of all RIG-I constructs were consistent with their predicted molecular weights based on SDS-PAGE (Figure 9B).

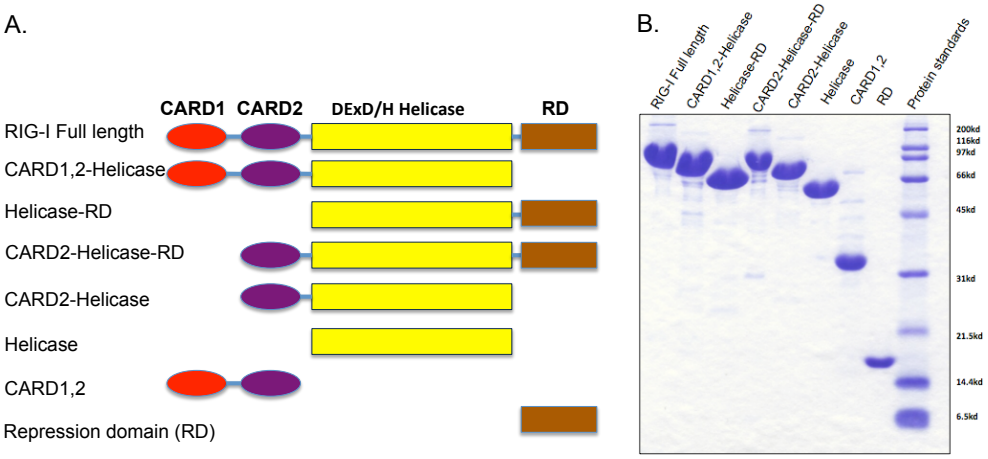


Figure 9. Constructs used in our studies.

(A) Schematic representations of the constructs of RIG-I.

(B) SDS-PAGE of the purified protein from RIG-I constructs.

Preliminary crystallization screens were set up by sitting-drop vapour-diffusion method in Intelli 96-3 low-profile plates (Art Robbins Instruments) at 293 K using a Gryphon crystallization robot (Art Robbins Instruments). Initial crystallization screening of those apo proteins did not yield any crystals or promising microcrystalline precipitates. Most drops showed heavy precipitate (amorphous precipitation) or phase separation. In addition, alternative strategies including limited in situ proteolysis for crystallization (Dong et al., 2007), soaking with ATP or non-hydrolysable analogues (e.g.. AMP-PNP, ADP, and ATP-gamma-S), crystallization of fusion protein with tag, and crystallization of homologous protein from different organisms (eg. mouse RIG-I) were also investigated; nevertheless, those attempts to obtain crystals have been unsuccessful except one construct, RIG-I (96-794). However, this crystal is very fragile and diffracted poorly (7 Å). Further optimization of crystallization and cryo-solution is under the way.

Since RNA-RIG-I interaction plays an essential role in RIG-I signaling pathway, it will be more important to investigate protein-RNA complex rather than protein alone. We hypothesized that the stable RIG-I-RNA complex formation would facilitate crystal growth, given that RNA binding stabilizes RIG-I by reducing the conformation flexibility based on our limited proteolysis experiments. In this thesis, I will focus on the structural studies of RIG-I in complex with RNA duplex. First of all, we needed to find out the domain contributions to RNA binding and ATP binding by gel shift assay and ATPase assay, respectively. We then used analytical gel filtration (Superdex 200 10/300GL) to monitor complex formation

and determine the ratio of protein:RNA to produce stoichiometric complexes (Ke and Doudna, 2004). Given that the length of RNA is a critical parameter for successful crystallization of protein-RNA complex (Hoggan et al., 2003; Jordan et al., 1985), RNA lengths starting from the minimal length (8bp) with increments of 2 base pairs up to 14-26 bps longer than the minimal sequence were further evaluated in order to find the most suitable RNA ligand for co-crystallization with RIG-I. Additives stabilizing crystal packing and decreasing mosaicity in RNA-protein crystals were examined as well using Hampton Research additive screen kit.

3.2.2 Electrophoretic mobility shift assay of RIG-I constructs

In collaboration with Dr. Smita S. Patel at UMDNJ, we have defined dsRNA with blunt ends as a potential ligand for RIG-I binding in vitro. To investigate the contribution of the individual domains to RNA binding, electrophoretic mobility shift assay (EMSA, or gel shift assay) was employed to assess the RNA binding activity and measure the apparent equilibrium dissociation constant ($Kd\text{-}_{apprent}$). We choose the length of 18 base pairs based on our previous work to evaluate the RNA binding activity of individual domains of RIG-I. As Figure 10 shows, at lower concentrations of protein only one shifted band is visible, whereas a second band of higher molecular weight appears at higher concentrations of RIG-I full-length protein and RIG-I (96-925). Interestingly, RIG-I Helicase-RD also shows two major bands corresponding to the monomeric and dimeric forms; however, the complexes become smeared out through monomer to dimer and

some RNA even remains trapped in the gel wells at the highest protein concentration. Helicase domain alone (232-794) still can binds the dsRNA with very low affinity, while RD domain binds dsRNA very tightly. DeltaRD (1-794), RIG-I (96-794), and CARDs domain lose completely the RNA binding activity in our gel shift assay. Taken together, our gel shift results indicate that: 1) more than one molecule of RIG-I can bind to one molecule of dsRNA; 2) CARDs interfere with dsRNA binding and is important for stabilizing the dimerization form of RIG-I; 3). RD is most important domain within RIG-I for dsRNA binding, but helicase domain is still required for achieving much higher binding capacity with dsRNA ligand.

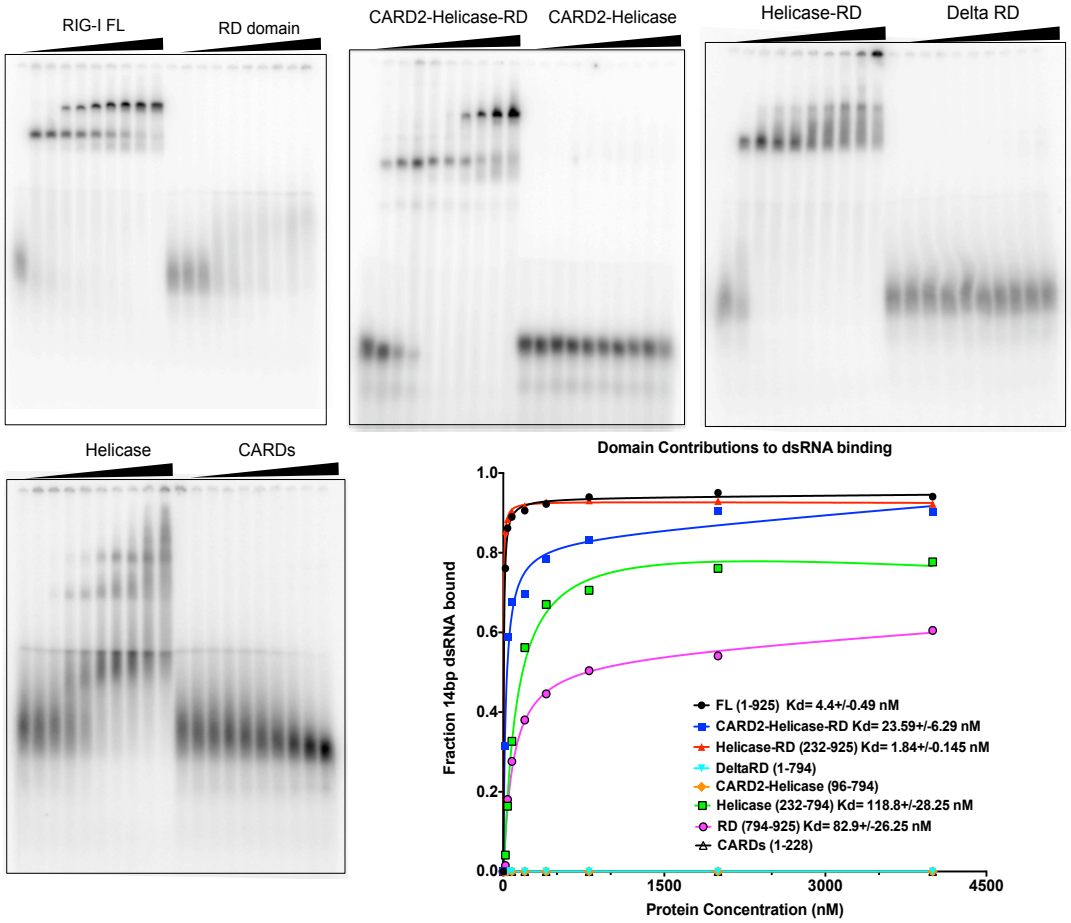


Figure 10. Electrophoretic mobility shift assay of RIG-I constructs. One-site binding model was used to analyze the data and determine apparent equilibrium dissociation constant (Kd).

Fluorescence anisotropy was further employed in Dr. Smita S. Patel's lab to accurately measure the equilibrium dissociation constant (Kd) and stoichiometry of the full-length RIG-I, helicase (residues 232-794), RD (residues 795-925), and helicase-RD (residues 232-925) for dsRNA (Table not shown). Titration of the 5'-fluorescein labeled 14 base pair dsRNA with increasing protein concentrations demonstrated that the helicase-RD binds the blunt-ended dsRNA stoichiometrically and with a very high affinity (Kd = 0.05 ± 0.02 nM). In contrast, the individual helicase domain and RD bind with an 8600-fold (Kd = 430 ± 121 nM) and a 50-fold (Kd = 2.6 ± 1.2 nM) weaker affinity, respectively, suggesting synergy between the RD and helicase domain in binding to dsRNA. Additionally, There is no discernable difference in RNA binding by helicase-RD upon the addition of the ATP analog, ADP•BeF_3.

In an attempt to further explore the specificity of the dsRNA towards RIG-I, we compared RIG-I full-length protein and RIG-I Helicase-RD in a competition experiment. Radioactive-labelled 18 bp dsRNA (0.1 nM) was mixed with RIG-I proteins (1 μM) in the presence of varying amounts of unlabelled competitor 18bp dsRNA (also referred to as cold RNA). There are two ways for performing this competition experiment, either pre-mixing hot RNA and cold RNA, or incubating hot RNA with proteins first for 5 min and then add cold RNA as indicated (Figure 11). As noted, for RIG-I Helicase-RD, premixing hot and cold RNA first resulted in a complete different gel shift pattern from adding the cold RNA after mixing hot RNA and proteins. RIG-I Helicase-RD still binds the hot RNA tightly even when

titrated with high concentration of cold RNA, while RIG-I full-length protein binds less and less hot RNA with increasing concentration of cold RNA (Figure 11A). This suggests that the RIG-I proteins can specifically recognize the double-stranded RNA motif, albeit RIG-I Helicase-RD bearing much slower K_{off} rate than full-length protein when bound to dsRNA. It is also noteworthy that increasing amounts of competitor RNAs resulted in the formation of a single faster migrating species that correspond to monomeric state of RIG-I-dsRNA complex, and readily reduced the formation of dimerization of RIG-I that most likely corresponds to two molecules of RIG-I bound to one molecule of RNA. Increasing the length of RNA to 30bp still shows two migration bands, which further suggests that RIG-I preferably binds to the end of dsRNA (data not shown).

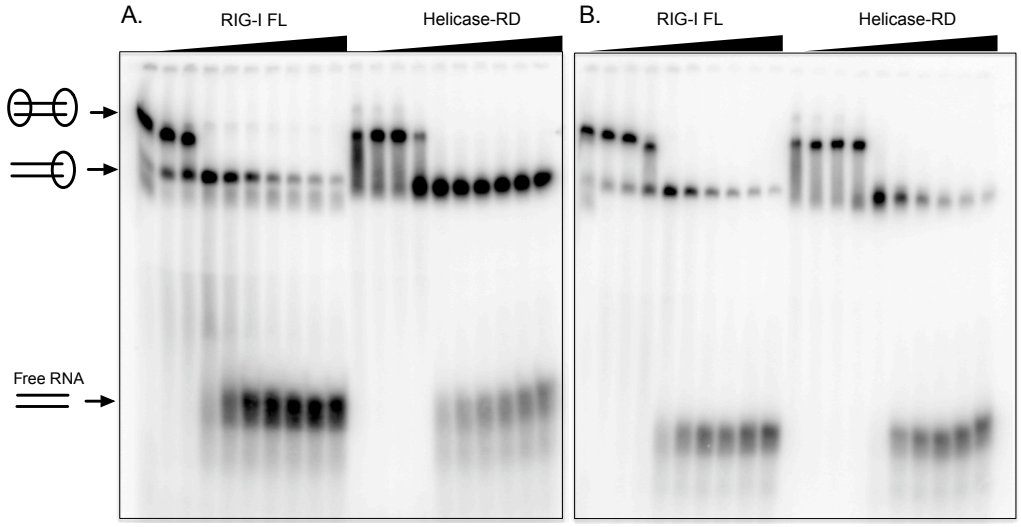


Figure 11. Competition experiment confirmed that RIG-I specifically binds dsRNA. (A) Hot 18bp dsRNA amount is constant. Mix radiolabeled RNA with protein first for 5 min and then add the increasing cold RNA (0, 1, 2, 5, 10, 20, 40, 80, 100 and 200 pmol in 25μl reaction). (B) Pre-mixing hot RNA and cold RNA, and then add cold RNA as indicated.

In addition, we tested different length (6bp, 10bp, 14bp, 18bp) of blunt-ended dsRNA by gel shift assay (Figure 12). Our gel retardation result showed that 14bp and 18bp bind slightly tighter than 10bp dsRNA, whereas 6bp dsRNA gives very low binding affinity. RIG-I Helicase-RD still forms a stable monomeric complex with 10-14 dsRNA with higher affinity (Figure 12, upper right panel and bottom right panel), while RIG-I FL begins to form dimeric complex with 14bp dsRNA at the saturating concentration (Figure 12, upper left panel and bottom left panel). When the length of dsRNA comes up to 18bp, RIG-I FL promptly becomes dimer even at low protein concentration, indicating that RNA length would be critical for RIG-I dimerization. Intriguingly, Helicase-RD forms smeared species between dimer and monomer bands. There are two possible explanations for this phenomenon: 1) CARDs are important for RIG-I dimerization, i.e., RIG-I devoid of CARDs cannot form stable dimer. 2) RIG-I Helicase-RD can bind the dsRNA ends and the stem, while RIG-I FL can only recognize the dsRNA ends. Therefore, Helicase-RD could form multimers in complex with long dsRNA, whereas RIG-I FL could only form a dimer with long dsRNA. This finding further supports a model in which the CARDs exert an inhibitory effect on RNA binding ability of the helicase domain, consistent with what is observed in gel shift assay of RIG-I (96-925) and RIG-I CARD2-Helicase (96-794) (Figure 10).

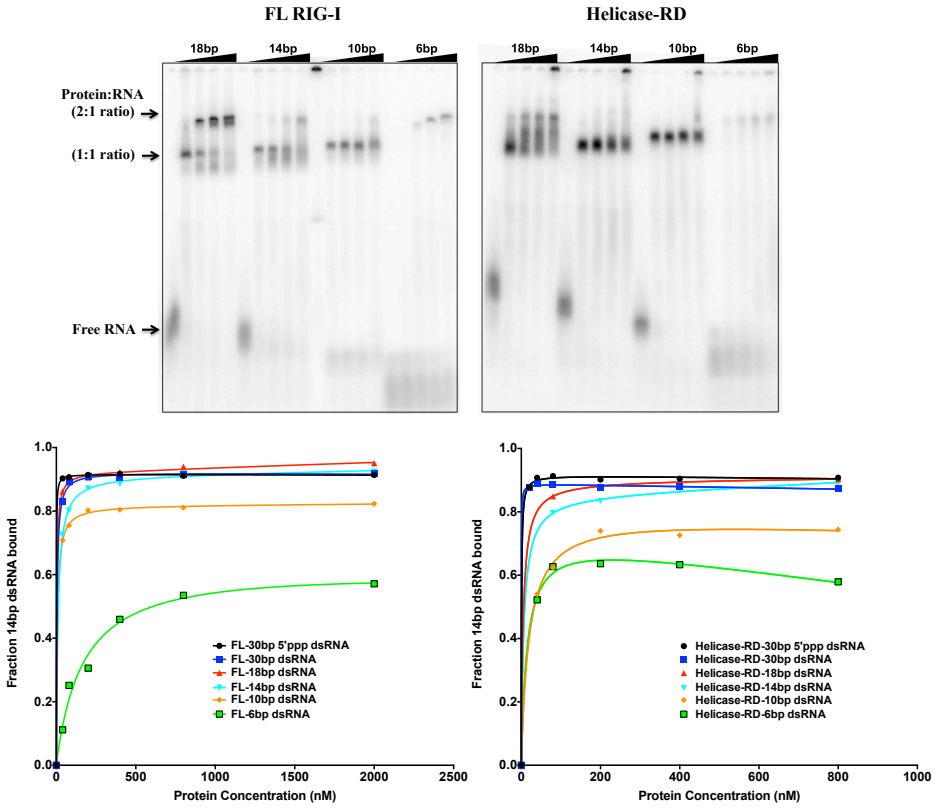


Figure 12. Gel shift assay with different length of dsRNA. ^{32}P-labeled RNA fragment was incubated with increased amount of RIG-I FL protein. Native-PAGE was then used to separate the RIG-I-RNA complex from free RNA. One-site binding model was used to analyze the data and determine apparent equilibrium dissociation constant (Kd).

3.2.3 ATPase activity of RIG-I constructs

Previous functional studies showed that mutation in the ATP binding site (K270A mutation in Walker A motif) of RIG-I fails to detect Newcastle disease virus and results in a dominant negative phenotype (Yoneyama et al., 2004), suggestive of an important role of ATPase activity in mediating antiviral signaling. Our preliminary studies demonstrated that RIG-I full-length protein exhibits a robust RNA-dependent ATPase activity in vitro. Since the different domains in RIG-I contribute differently to RNA binding, it was also of interest to dissect the relative contributions of the isolated individual domains to ATPase activity. As Figure 13 shows, the RIG-I full-length protein and RIG-I (96-925) have comparable ATP hydrolysis rates (~25 s^{-1} vs 32 s^{-1}) in the presence of the 14bp dsRNAs. Strikingly, Helicase-RD endows elevated ATPase activity [~65 M/(M × s)], while the helicase domain alone shows very weak ATPase activity [~0.3 M/(M × s)]. The ATPase activities of RIG-I deltaRD (1-794), RIG-I CARD2-Helicase (96-794) are undetectable, even though the intact helicase domain exists for those constructs. Obviously, CARDs and RD do not have ATPase activity in that the ATP binding motifs (Q motif, Walker A and B motifs) are completely missing in those constructs.

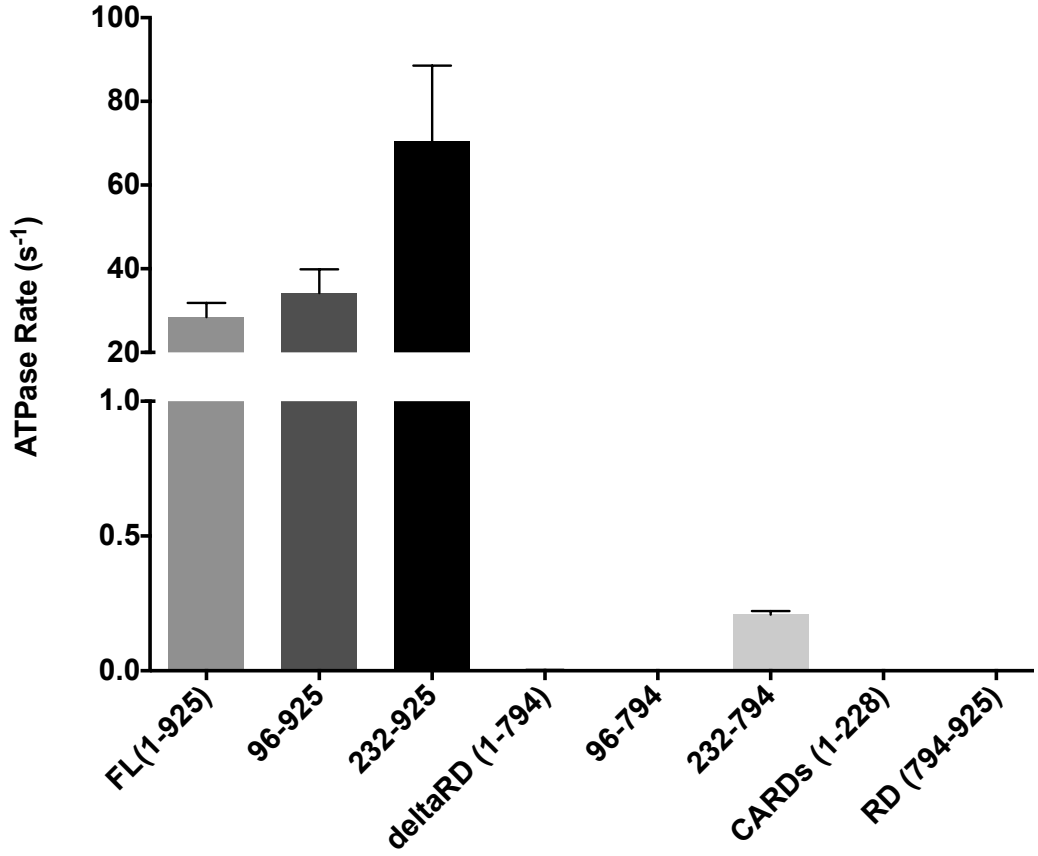


Figure 13. The relative contributions of the isolated individual domains to ATPase activity. 14bp dsRNA was used in this assay. RIG-I full-length protein and RIG-I (96-925) have comparable ATP hydrolysis rates, while Helicase-RD endows elevated ATPase activity. Helicase domain alone shows very weak ATPase activity. However, deltaRD (1-794), RIG-I CARD2-Helicase (96-794) does not show any detectable ATPase.

3.2.4 Crystallization, data collection

We tested the complex formation for RIG-I FL and Helicase-RD in complex with different length RNA ranging from 8bp to 28bp with an interval of 2bp increment. Consistent with what we found from fluorescence anisotropy and gel shift assay, size-exclusion chromatography (Zou et al.) demonstrated that RIG-I helicase-RD forms a 1:1 complex with the 8-14 base pair dsRNA (Figure 14). The stoichiometry of helicase-RD is different from the isolated RD, which binds to both ends of a 14 base pair dsRNA in a ratio of 2:1. The presence of a 5' ppp-dsRNA does not alter the stoichiometry of helicase-RD. However, RIG-I helicase-RD begins to form a 2:1 complex with the 18-28bp base pair dsRNA, whereas RIG-I FL rapidly dimerizes with excess of 14bp-28bp bp dsRNA. Subsequent crystallization trials of all those complexes showed that RIG-I Helicase-RD can readily crystallize with only 10-14bp dsRNA. Crystallization optimization further demonstrated that 14bp dsRNA gave rise to the best single crystals suitable for X-ray diffraction.

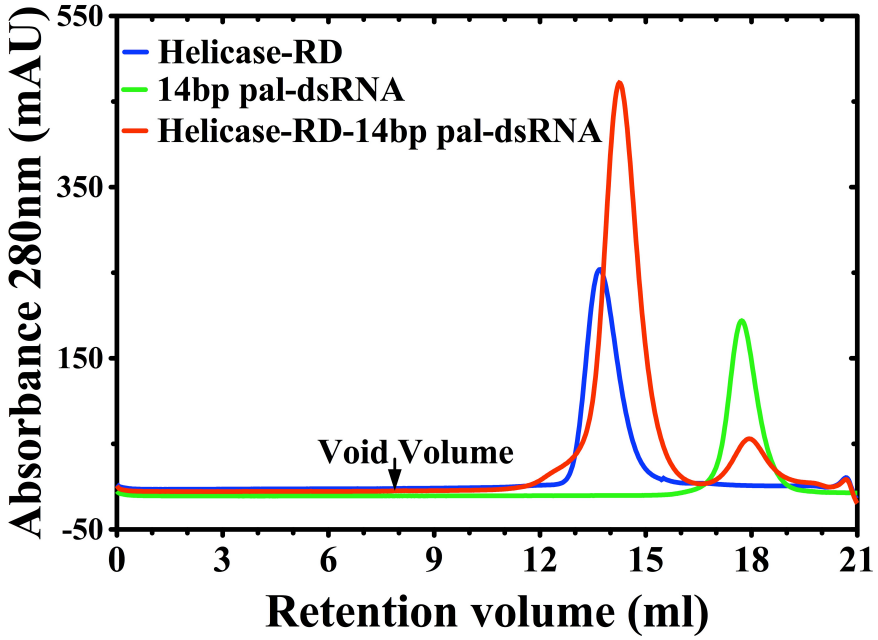
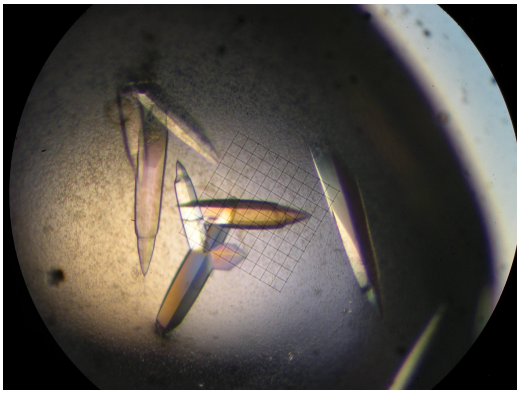


Figure 14. Stoichiometry of RIG-I helicase-RD bound to dsRNA. Chromatogram of size exclusion chromatography showing RIG-I helicase-RD (blue), 14 base pair dsRNA (green) and helicase-RD in complex to dsRNA (red). Helicase-RD•RNA complex forms a monomer (1:1 ratio), and elutes at a larger volume (14.3ml) compared to the free helicase-RD (13.7ml).

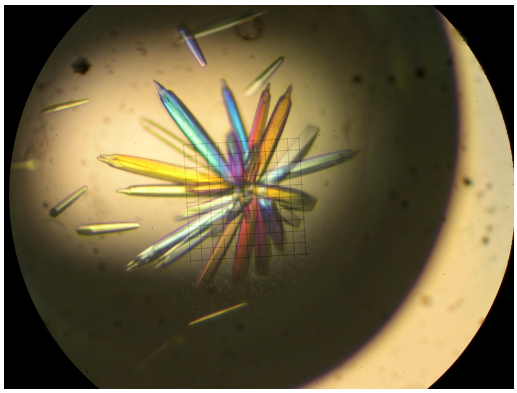
To be more specific, crystals of native and SeMet-substituted RIG-I Helicase-RD–14-bp dsRNA complex were grown by hanging drop vapor diffusion method at 20 °C (Figure 15 A-B). Aliquots (2.5 µl) of 15 mg ml^{-1} of protein-RNA complex in 50 mM HEPES 7.5, 50 mM NaCl, 5 mM DTT, 5mM $MgCl_2$ were mixed with 2.5 µl of reservoir solution containing 23% (w/v) PEG 3350, 0.25 M KNO_3, 100 mM MOPS (pH 7.8), 4% (v/v) 2,2,2-Trifluoroethanol. Crystals appeared after 2–3 days, and they grew to a maximum size of 0.15 × 0.15 × 0.5 mm over the course of 8 days. 5-iodo-uridine derivative RNA-Protein crystals appeared at the same condition.

RIG-I Helicase-RD-ADP•BeF_3-dsRNA ternary complexes were reconstituted by incubating 0.17 mM RIG-I Helicase-RD with 0.17 mM 14bp dsRNA, 2 mM ADP, 2 mM $BeCl_3$ and 5 mM NaF on ice for 30 min prior to crystallization. The rod-like single crystals were obtained in similar conditions as described above (Figure 15 C), except 0.25 M NaSCN was used instead of 0.25 M KNO_3. These crystals diffracted the best and the data were used to further refine to get the best and satisfactory model.

A.



B.



C.

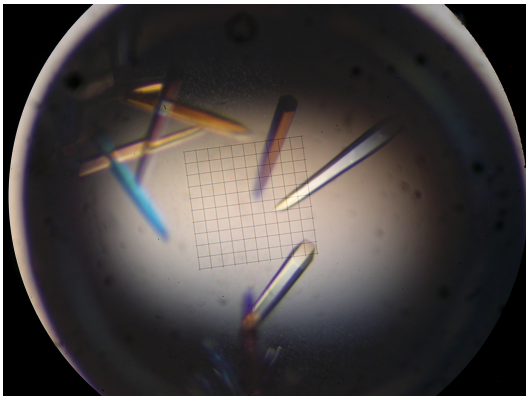


Figure 15. Crystals of RIG-I-RNA duplex

(A) Crystals of RIG-I Helicase-RD-–14-bp dsRNA.

(B) Crystals of SeMet-substituted RIG-I Helicase-RD-–14-bp dsRNA.

(C) Crystals of RIG-I Helicase-RD-ADP•BeF_3-dsRNA ternary complex.

For cryogenic data collection, crystals were gently transferred to crystallization solutions containing 4% (v/v) 2,2,2-Trifluoroethanol, 5% (v/v) (2R,3R)-(−)-2,3-butanediol as cryoprotectant, and then flash-frozen at 100 K. Reservoir solutions supplemented with different concentrations of glycerol, Ethylene glycol various, PEGs, DMSO, oils, ethanol, sucrose and glucose were also tested for cryoprotection of the crystals, but none of them led to comparable diffraction behaviour as 2,3-butanediol. Native and SeMet single-wavelength anomalous diffraction (SAD) datasets were collected on an ADSC Q315 CCD at the X29A beamline of the National Synchrotron Light Source (NSLS). For RIG-I Helicase-RD-ADP•BeF_3-dsRNA ternary complex crystals, diffraction data were recorded at the LRL-CAT 31-ID beamline of the Advanced Photon Source (APS). Data from iodouridine derivative crystals were collected at the Cornell High Energy Synchrotron Source (CHESS) F1 beamline. All diffraction data were processed using Mosflm and scaled in SCALA(1994). Assuming the presence of one molecule in the asymmetric unit gave a Matthews coefficient (V_M) of 2.74 $Å^3$ Da^{-1}, corresponding to a solvent content of 55%. Data-collection and processing statistics are summarized in Appendix table.

3.2.5 Structure determination and refinement

The structure of RIG-I Helicase-RD–14-bp dsRNA complex was solved by a single anomalous dispersion phasing method (Dauter et al., 2002) using SeMet data in combination with molecular replacement, using the RD structure (2QFB) as a partial model. Using SeMet data between 30 to 3.7 Å, SHELXD/HKL2MAP

(Schneider and Sheldrick, 2002) detected a total of 12 out of 14 possible selenium sites in the asymmetric unit. Initial phases were then calculated with Phaser (McCoy et al., 2007), and the density modification was carried out by DM. Phase extension to 3.2 Å by SOLOMON (Abrahams and Leslie, 1996) produced an electron-density map into which most of the protein and RNA residues could be built unambiguously. The model was built with COOT (Emsley and Cowtan, 2004) and refined in Refmac (Murshudov et al., 1997) and PHENIX (Adams et al.). Subsequent iterative rounds of manual model building and refinement produced a model with an R_{free} value of 37%. This model was further refined against RIG-I Helicase-RD-ADP•BeF_3-dsRNA ternary complex data set between 30 and 2.9 Å. Multiple rounds of individual B-factor refinement and translation-libration-screw (TLS) (Painter and Merritt, 2006) refinement were performed using PHENIX (Adams et al.) to improved the structure model, resulting in the final R_{work} and R_{free} values of 20.8% and 28.2%, respectively. The final models comprise 12 bp dsRNA and RIG-I residues 240–923, with the N termini (residues232–239) and loop regions (residues 496-503, 523-526) being completely disordered. Model validation was performed by using MolProbity (Davis et al., 2004), and no outliers were found in the Ramachandran plot.

The structure of RIG-I bound to 14 bp dsRNA with a 5-iodo-uridine was obtained by molecular replacement using the program Phaser (McCoy et al., 2007), which enabled us to determine the register of the palindrome RNA (Figure 16).

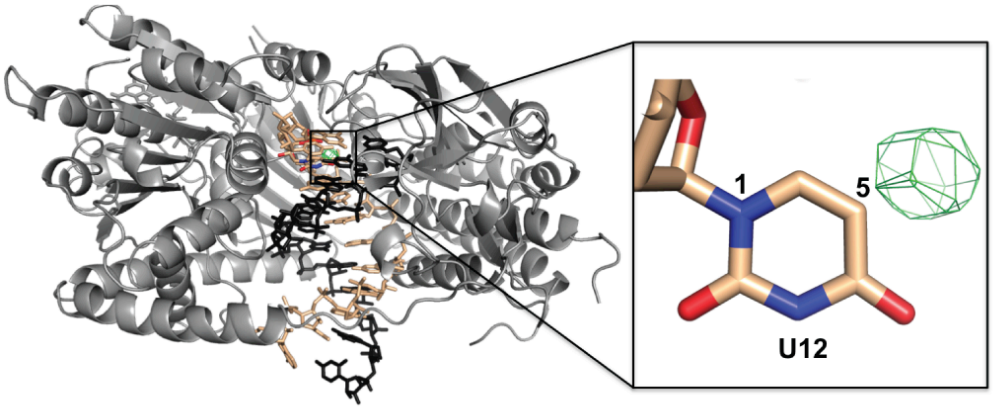


Figure 16. Confirmation of RNA sequence using 5-Iodo-Uridine substituted RNA. The sequence and orientation of the dsRNA in the structure was confirmed using data collected from a crystal substituted with 5-Iodo-uridine at position 12 of the dsRNA. Crystals of RIG-I helicase-RD, ADP•BeF_3 and 14 base pair dsRNA substituted with 5-Iodo-uridine at position 12 were grown under identical conditions as the native complex. 2Fo-Fc electron density maps at 3.1 Å resolution calculated using the protein alone contoured at 5σ is shown in green. The peak corresponds to an iodine atom at the 5 position of a uridine base.

3.2.6 Structure analysis of the ternary complex

3.2.6.1 Structure overview of RIG-I bound to dsRNA

Crystals of RIG-I helicase-RD in complex with ADP•BeF_3 and 14 base pair palindromic dsRNA diffracted to 2.9 Å resolution and the structure (PDB CODE: 3TMI) was determined by single-wavelength anomalous dispersion (SAD). The stoichiometry of the protein:RNA complex is 1:1, which is consistent with the results of biochemical experiments (Figure 14). The overall architecture of RIG-I helicase-RD reveals four domains, three from the helicase (domains 1, 2 and 3) and the RD, arranged in a ring structure on one end of the dsRNA. The polypeptide chain progresses from the first RecA-like helicase domain (domain 1) to an alpha helical domain (domain 3) followed by a second RecA-like helicase domain (domain 2) and lastly the RD. The linker connecting domain 1 to domain 3 (residues 446-469) (colored teal) forms a beta strand with the parallel beta sheet in domain 2. Domain 2 is connected to the RD via two alpha helices (colored orange) that form a prominent V-shaped structure. These helices make extensive interactions with both domain 1 and domain 2 with a buried surface over 1,500 Å2. Upon exiting the second alpha helix, an extended proline-rich (796KPKPVPD) loop makes the final connection to the C-terminal RD. Interestingly, one of the RIG-I phosphorylation sites is found at T770, which maps to the junction of the two V-shaped helices. The phosphorylation of T770 can repress RIG-I signaling, suggesting an important role for the linker in RIG-I regulation.

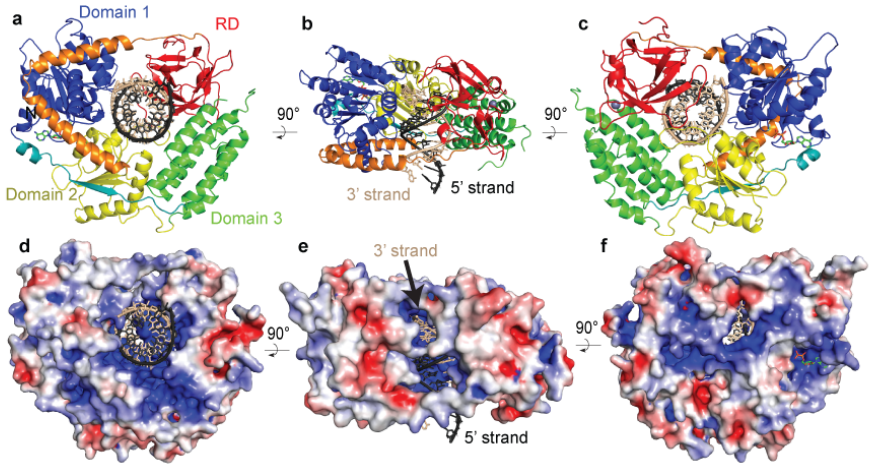


Figure 17. Structural overview of RIG-I helicase-RD. Schematic representation of the RIG-I helicase-RD, highlighting the RecA-like domain 1 (blue), the alpha-helical domain 3 (green), RecA-like helicase domain 2 (yellow), and RD (Fredericksen et al.). The linker connecting Domain 1 with Domain 3 is colored teal, while the V-shape linker between Domain 2 and RD is colored orange. The ADP•BeF_3 and dsRNA are shown in stick representation with the 5' and 3' strands of the RNA colored black and beige, respectively. A grey sphere denotes the position of the zinc ion in RD. The 3' and 5' strands are colored beige and black, respectively. (**d, e** and **f**) Surface of RIG-I helicase-RD colored for electrostatic potential at ±5 kT/e; blue (basic), white (neutral), and red (acidic). The views in panels **a**, **b**, and **c** are identical to **d**, **e** and **f**, respectively.

The helicase-RD molecule binds to one end of the dsRNA with the four domains encircling the dsRNA, which maintains an A-form helical conformation. The sequence and orientation of the dsRNA in the structure was confirmed using data collected from a crystal substituted with 5-Iodo-uridine at position 12 of the dsRNA (Figure 16). Looking down the long axis of the dsRNA from the blunt end where the helicase-RD binds (Figure 17a-c.) shows the following clockwise arrangement of the four domains: RecA-like helicase domain 1, RecA-like helicase domain 2, the alpha-helical domain 3, and the RD. The helicase domains form a crescent, and the RD closes it into a circle by interfacing with all three helicase domains.

Molecular surface analysis shows that one end of the dsRNA is capped by the entire helicase-RD molecule (Figure 17d-f); however, the opposite end of the dsRNA is quite exposed with the last two base pairs of the RNA being disordered. RIG-I helicase-RD has a small, highly basic channel that exposes the 3’ end. (Note: this will be referred to as the 3’-strand and is colored beige in the figures, while the opposite strand will be designated as the 5’-strand and colored black.) This opening may provide the flexibility to recognize dsRNA with 3' nucleotide overhang or 3' monophosphate, which is the product of RNase L digestion. On the other hand, the end of the 5’ strand abuts the RD domain and is not exposed to the solvent; thus 5' end extensions cannot be accommodated in this structure. Additionally there are several visible channels along the long axis

of the dsRNA, which may allow the domains to flex and accommodate bulges or noncanonical base pairs in the dsRNA (Figure 17).

3.2.6.2 Detailed interactions of RIG-I helicase-RD with dsRNA

All four RIG-I domains participate in binding the dsRNA, burying 1,500 Å2 of surface area to encircle about eight base pair of dsRNA (Figure 18). Both helicase and RD make contact with each strand. The majority of the contacts are with the sugar phosphate backbone of the RNA with very few base specific interactions. The RD binds to the dsRNA at the blunt end, making many contacts with the terminal base pair. F853 stacks over the terminal C1-G14 base pair, and the H830 and S854 form hydrogen bonds with the ribose 2'-hydroxyls of C1 (5'-strand) and G14 (3'-strand), respectively. Phosphorylation of S854 and S855 has been shown to negatively regulate RIG-I and, given the proximity of these residues to the blunt end of the RNA, would be predicted to adversely impact RNA binding. The RD/RNA contacts in this structure are identical to the previously determined structures of RD bound to dsRNA (Lu et al., 2011)(Lu et al., 2011)(Lu et al., 2011)(Durr et al., 2005; Lu et al., 2010; Wang et al., 2010). Residues H847, K861, and K858 of RD, which were reported to interact with the 5'-ppp, are not making any new interactions in the helicase-RD structure with 5'-OH dsRNA and ADP•BeF_3. The only contact that RD makes with the 3' strand other than at the blunt end is between S906 of the U7 phosphate backbone.

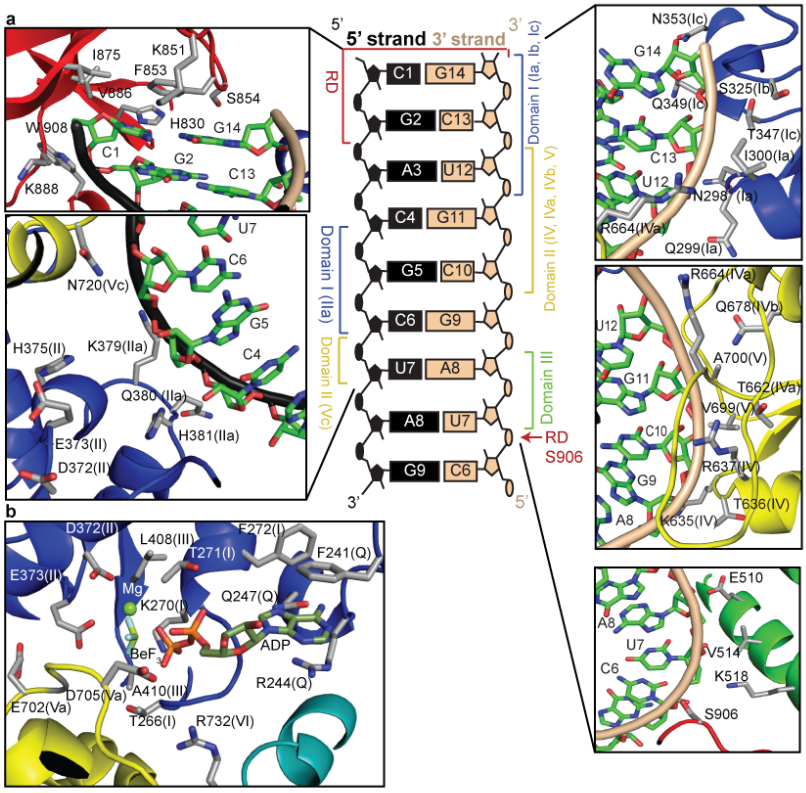


Figure 18. Interactions of RIG-I helicase-RD with dsRNA and ADP•BeF_3. A schematic representation showing the interactions between RIG-I domains and helicase motifs (given in parentheses) with dsRNA is located in the center. Detailed contacts are shown in the surrounding panels. Stick representation detailing the RIG-I helicase motifs interactions with ADP•BeF_3 and Mg^{2+} is shown in the lower left panel.

Helicase core domains 1 and 2 contain characteristic motifs critical for RNA binding and ATP hydrolysis. The helicase core binds to the terminal five nucleotides of the 3'-strand (G14-C10) and the middle of the 5'-strand at positions C4-U7 (Figure 18a). The helicase motifs Ia, Ib, and Ic found within domain 1 interact with the terminal three nucleotides of the 3'-strand. Residues in motif 1b (S325 and G326) and motif Ic (T347, Q349 and N353) make contact with the ribose-phosphate backbone of the terminal base of the 3'- strand (G14), while the backbone of N298 and I300, and the side chain of Q299 in motif 1a interact with C13 and U12. Motifs IV, IVa, IVb, and V in domain 2 continue these RNA interactions with the 3'-strand from U12 to C10. Residues K635, T636, and R637 in motif IV interact with the phosphate backbone of G11-C10. T662, G663, and R664 in motif IVa and Q678 in motif IVb contact the backbone of C13-G11. V699 in motif V interacts with the RNA backbone at C10.

All RIG-I-like family helicases contain a large insertion (Domain 3) between the core helicase domains, whose role has been unclear. Structure of Hef helicase in the absence of DNA showed domain 3 to be mainly alpha-helical, and our structure of the RIG-I shows a similar fold of domain 3 (Figure 19). An overlay of Hef with the RIG-I helicase-RD with dsRNA shows a rotation of domain 3 due to interactions with the dsRNA. An alpha helix within domain 3 (residues 506-522) of RIG-I is interacting with the 3'-strand and runs almost perpendicular to the minor groove of the dsRNA but does not contact the 5'-strand (Figure 18a).

Residues E510, V514 and K518 of domain 3 extend the interactions of the helicase to positions 7 and 8 of the 3’-strand.

Although the helicase extensively contact the 3'-strand, all three helicase domains are in close proximity to the 5'-strand, and there are specific contacts with the 5'-strand. Specifically, C4 and G5 of the 5'-strand contact a previously uncharacterized motif IIa containing the sequence 380QHPY (Y is highly conserved) in the helicase domain 1. This motif IIa immediately follows the 372DECH helicase motif II that coordinates the Mg^{2+} and stabilizes the ATP analog (Figure 18a-b). Similarly, N720 in motif Vc of domain 2 contacts the backbone at C6 and U7 of the 5’ strand. These contacts are newly identified due to a lack of structures for RNA helicases bound to dsRNA. These 5’ strand contacts may represent a general feature of helicases binding to dsRNA or dsDNA since a region similar to motif IIa was identified in the Swi/Snf2 family helicases and in the structure of a Rad54 homolog making contacts with dsDNA (Durr et al., 2005).

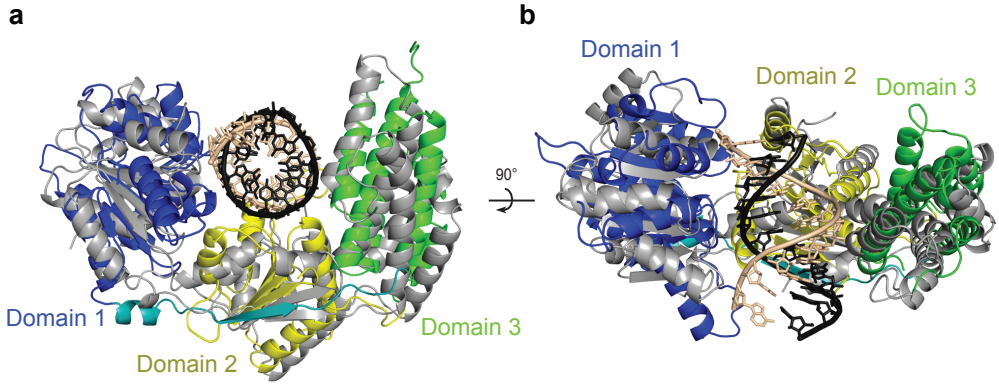


Figure 19. Structural comparison of Hef helicase in the absence of nucleic acid and RIG-I helicase in the presence of dsRNA. Overlay of *Pyrococcus furiosus* Hef helicase domain (PDB code 1WP9) with just the helicase domains of RIG-I and dsRNA. The RIG-I domains are colored similar to Fig. 1. The view in panel **b** is rotated 90° about a horizontal axis from panel **a**.

3.2.6.3 Comparison of RIG-I helicase-RD with previous structures

Although some of RIG-I helicase-RD interactions with dsRNA are similar to those observed in the structure of the RD with dsRNA (Lu et al., 2010; Wang et al., 2010) and other SF2 helicases, such as hepatitis C virus (HCV) NS3 and VASA bound to ssRNA or ssDNA, there are several significant distinctions. An overlay of the RIG-I helicase with the HCV NS3h bound to ssDNA26 (Figure 20a) shows an excellent match between the core helicase domains and makes a few predictions: 1) The overlay shows that domain 3 in NS3h is positioned where the RD is in RIG-I, which might indicate that domain 3 in NS3h interacts with the 5'-strand. 2) The ssDNA in NS3h overlays with the 3'-strand of the dsRNA bound to the RIG-I helicase, suggesting that the principal motor contacts of the RIG-I helicase are with the 3'-strand (Figure 20b). 3) A conserved Phe-loop implicated in RNA unwinding in NS3h bisects bases of the dsRNA at a position where the unwinding junction would be in NS3h29 (Figure 20b-c). In addition to predicting a role for this Phe- loop in strand separation, the absence of such a motif in RIG-I could explain why we and others (Myong et al., 2009) have failed to detect unwinding activity in RIG-I, although there are reports that RIG-I unwinds short dsRNA (Cui et al., 2008; Takahasi et al., 2008).

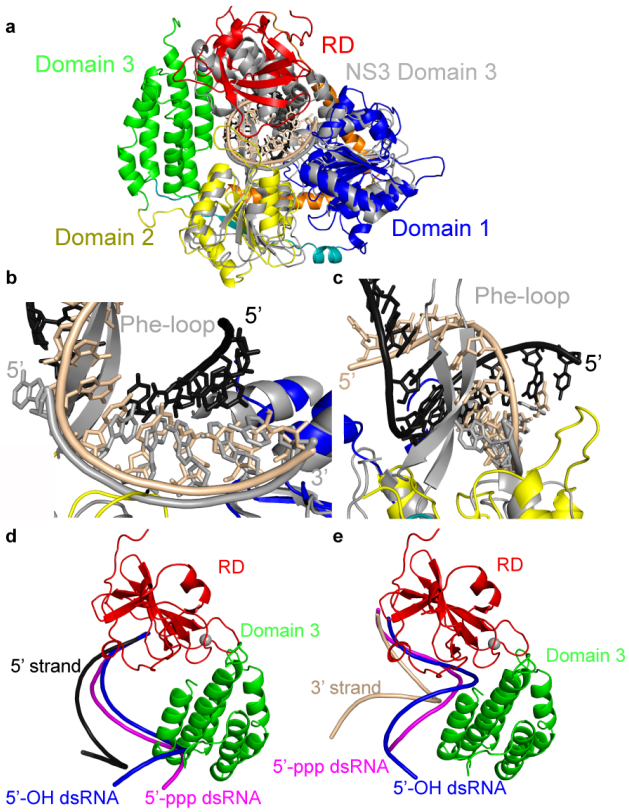


Figure 20. Comparison of RIG-I helicase RD with HCV NS3h and RD bound to 5'-OH and 5'-ppp dsRNA. (**a**, **b**, and **c**) Ribbons diagram showing the superposition of RIG-I helicase-RD•dsRNA•ADP•BeF_3 structure and NS3h bound to ssDNA (PDB code 3KQH) (grey). (**b**) Superposition of RIG-I helicase-RD with NS3h demonstrates that the ssDNA bound to NS3h overlays with the 3'-strand (beige) of the dsRNA bound to the helicase-RD. (**c**) The location of the Phe-loop of NS3h relative to the dsRNA of the RIG-I helicase-RD•dsRNA•ADP•BeF_3 structure. (**d** and **e**) Superposition of the 5'-OH (blue; PDB code 3OG8) and 5'-ppp dsRNA (magenta; PDB code 3LRR) based on the location of RD. For clarity the 5' strands (**d**) and 3' strands (**e**) are shown separately.

The structure of the RD bound to dsRNA with and without 5'-ppp is in agreement with the RD in the helicase-RD structure with a root mean square deviation of ~0.6Å for backbone atoms. However, the trajectory of the RNA helix is different. Superposition of the three structures demonstrates that the conformation of the dsRNA, with and without 5'-ppp in the RD alone structure, clashes with domain 3 (Figure 20d-e). In addition, contacts between the RD and the helicase domains have rotated residues R811, K849, K851, and H871, which were previously reported to contact the 3' strand, such that they no longer interact with the RNA.

3.2.6.4 ATP binding and coupling of ATPase to dsRNA binding

The ADP•BeF_3 molecule is bound at the interface of the core domains 1 and 2 and brings together amino acids in the helicase conserved motifs Q, I, II, III, V, and VI that are known to be involved in ATP binding/hydrolysis (Figure 18b). The Q motif recognizes the adenine base while the rest of the helicase motifs are involved in binding the triphosphate moiety and the Mg^{2+}. The side chains of F241 and the highly conserved R244 in the Q motif stack on either side of the adenine base. Q247 makes adenine specific contacts at N6 and N7, explaining the ATP dependence for RIG-I activity. The 269GKT helicase motif I contacts the phosphates and the BeF3 mimicking the γ-phosphate of ATP, as observed in other SF2 helicases. The D372 and E373 in the DECH motif II coordinate the Mg^{2+}. The helicase motifs V and VI that contact the ribose and phosphates of the ATP in other helicases are close to ADP•BeF_3 in our structure, but are not within hydrogen-bonding distances. Since the ADP•BeF_3 molecule represents the

ground state, these residues may engage during other steps of the catalytic cycle.

The ATPase activity of RIG-I is completely dependent on the presence of RNA and stimulated by dsRNA. The dsRNA-specific ATPase activity of RIG-I can be explained by motif IIa that contacts the 5'-strand. As indicated above, motif IIa resides in helicase domain 1 and it is present in a loop immediately following motif II, which coordinates the Mg^{2+} essential for ATP hydrolysis (Figure 18a-b). This specific arrangement of motif IIa makes it an effective dsRNA sensor to allosterically communicate RNA binding to ATPase stimulation.

3.2.7 Conformation changes of RIG-I upon binding dsRNA

3.2.7.1 Evidence from biochemical studies for conformational changes of RIG-I upon RNA binding

RIG-I exists in an autoinhibited state in the absence of RNA with little ATPase activity and activates upon binding viral RNAs. It has been proposed previously that the CARDs are inaccessible for signaling in the autoinhibited state, which is an important mechanism to safeguard against aberrant signaling. Limited proteolysis, and differential scanning fluorimetry were then employed to investigate the conformational changes of full-length and helicase-RD of RIG-I upon RNA binding. As Figure 21a showed, limited trypsin digestion led to a different cleavage pattern and a slower cleavage of RIG-I proteins at K679 and K685 in the presence of dsRNA. In particular, both full-length and helicase-RD

RIG-I show substantial protection from trypsin digestion at K851. Since K679, K685 and K851 do not have the direct contact with RNA, RIG-I must have undergone a conformational change so that those sites are not readily accessible to trypsin cleavage. To further test this hypothesis, thermal stability shift assay was conducted to examine RIG-I stabilization upon binding dsRNA (Figure 21b). dsRNA binding significantly enhances the apparent melting temperature (3.5±0.30°C for full-length RIG-I and 8.5±0.35°C for helicase-RD). Conversely, dsRNA and ATP analog act in synergy towards increase even greater Tm (4.7±0.15°C for full-length RIG-I and 10.6±0.10°C for helicase-RD). Given that an increase in stability indicates a reduced conformational flexibility, our results from thermal shift assay provide further insight into RIG-I conformational changes upon RNA and ATP binding. Considering the structure of ternary complex showed Helicase-RD surround the dsRNA, it is most likely that the compaction upon binding RNA confer the increase of RIG-I thermal stability. This compaction of helicase-RD upon dsRNA binding is also observed in gel filtration, where the helicase-RD complex with dsRNA elutes at a larger volume (14.3ml) compared to the free helicase-RD (13.7ml) (Figure 14).

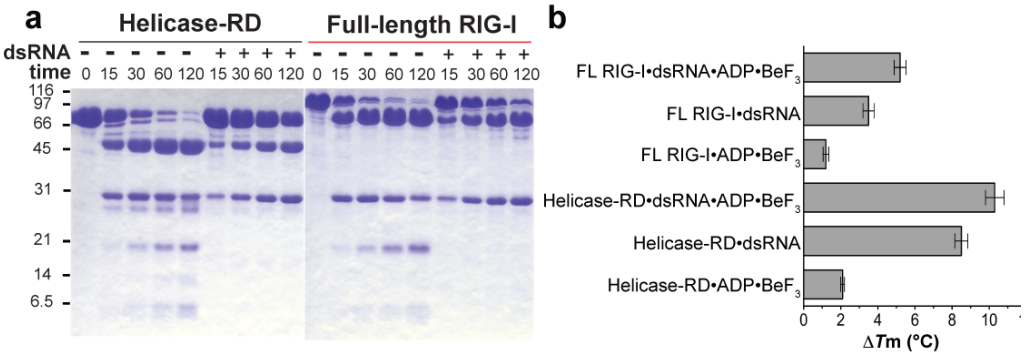


Figure 21. RIG-I-dsRNA interactions by limited proteolysis and thermal shift. (**a**) SDS-PAGE analysis of a time course (minutes) of limited trypsin digestion of Helicase-RD or full-length RIG-I in the absence or presence of 14 base pair pal-dsRNA. (**b**) DSF of RIG-I Helicase-RD or full-length RIG-I in the presence of 14 base pair pal-dsRNA and/or ADP•BeF_3 with respect to protein alone. The bar graph displays the mean melting temperature difference (ΔT_m) and the error bars represent the standard deviation from three independent measurements.

3.2.7.2 Evidence of conformational changes by SAXS studies

Small-angle X-ray scattering (SAXS) is a structural characterization technique that measures the size and shape of biological macromolecules in solution. It also can provide an experimental approach to obtain low-resolution structural information of biological materials in near-physiological conditions. Therefore, we further employed SAXS to studying the conformation changes between apo-form and dsRNA-bound RIG-I proteins in solution. Our SAXS data indicates that the radius of gyration (Rg) for helicase-RD decreases by 10 Å upon addition of dsRNA, indicative of a significant compaction of the protein. Full-length RIG-I produced a similar, albeit less dramatic, change of 2.3 Å in the Rg after the addition of dsRNA.

The SAXS data indicate that both full-length RIG-I and helicase-RD in the absence of dsRNA are in a conformation different from the dsRNA bound states; however, we were unable to produce acceptable models for the proteins without RNA. Kratky plot (s^2I as a function of s) of SAXS data can qualitatively evaluate the folded states of a protein. P(r) functions clearly show RIG-I becomes more compacted upon binding dsRNA (Figure 22a-b). For globular domains, the Kratky plot yields a parabolic peak, while for unstructured proteins the plots lack the peak and become linear with increasing s. Globular domains connected by flexible linkers yield intermediate plots. In the presence of RNA, both helicase-RD and full-length RIG-I yielded symmetrical parabolic Kratky plots (Figure 22c-d), consistent with these complexes being folded and globular. However, Kratky

plots of the proteins in the absence of dsRNA displayed a loss of parabolic shape and a decrease in peak amplitude, suggesting that the proteins do not maintain the structure and are more flexible in the absence of dsRNA.

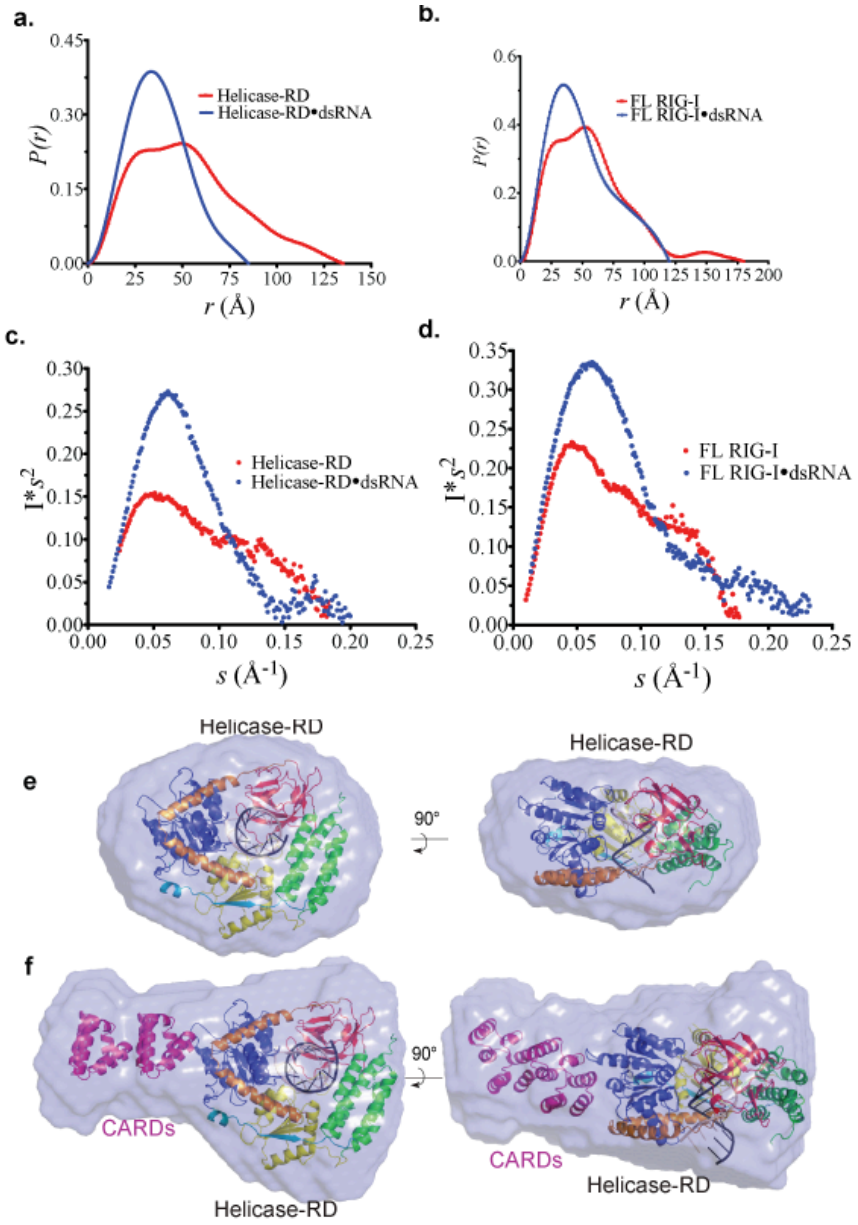


Figure 22. Conformatinal changes of RIG-I upon RNA by SAXS. (a-b) Distance distribution functions P(r) plot and (c-d) Kratky plot of scattering of Helicase-RD and full-length RIG-I in the presence and absence dsRNA. (e) Ab initio envelope of helicase-RD and dsRNA overlaid with the crystal structure of helicase-RD•dsRNA. (f) Ab initio envelope of full-length RIG-I and dsRNA overlaid with the crystal structure of helicase-RD•dsRNA with two copies of homologous CARDs added (PDB code 2VGQ).

Calculation of a series of ab initio models for full-length and helicase-RD in the presence of dsRNA followed by rigid body refinement yielded models with acceptable normalized spatial discrepancy (NSD) and χ^2 values. The SAXS envelope of helicase-RD with dsRNA is in excellent agreement with our X-ray crystallographic structure (Figure 22). Using a combination of SAXS data, the helicase-RD structure, and a homologous structure of a CARD, a model of full-length RIG-I bound to dsRNA could be established (Figure 22e-f). The helicase-RD can be accommodated on one side of the full- length RIG-I SAXS envelope while the two CARDs would protrude from the helicase domain 1. In this orientation it is likely that they are available for interaction with CARDs of the downstream signaling factors like IPS-1. Interestingly, the CARDs are positioned adjacent to the V-shaped linker, connecting domain 2 with the RD and the T770 phosphorylation site. The close proximity of the V-shaped linker and the second CARD, suggests a possible mechanism of RIG-I where the two alpha helices may serve as a hinge for RD movement from a repressed state to an activated state upon RNA binding. The orientation of the CARDs in this model is perpendicular to the axis of the dsRNA, allowing for multiple RIG-I molecules to bind to a single RNA.

Combining the observations from limited proteolysis, gel filtration profile, and thermal shift assay, we therefore predict that RIG-I consists of globular, folded domains connected with flexible linkers that become ordered and more compact upon RNA binding. Further structural and biophysical studies are needed to

better characterize the structure of RIG-I full length in the absence or presence of dsRNA.

3.3 Structure of RIG-I Helicase-RD in complex with different RNA ligands

3.3.1 Crystallization, data collection and structure determination

The chase for RIG-I ligand has been subject to intense research since the discovery of this pivotal viral RNA sensor (Baum and Garcia-Sastre, 2011; Kato et al., 2011; Onoguchi et al., 2011; Schlee et al., 2009a; Schlee and Hartmann, 2010). Recent work has found that RIG-I detects blunt-ended dsRNA with or without a 5'-ppp (Marques et al., 2006), single-stranded (ss) RNA marked by 5'-ppp (Hornung et al., 2006; Pichlmair et al., 2006), and the panhandle-like RNAs of negative strand virus such as influenza. To further identify the molecular determinants for RNA recognition, and to understand how RIG-I distinguishes viral RNA from self cellular RNA, we recently determined the crystal structures of RIG-I Helicase-RD bound to panhandle-like short hairpin RNAs in the presence and absence of 5’-triphosphorylated modification, bound to dsRNA with and without 5’ppp, and bound to chimaeric RNA-DNA duplex at high resolution.

5’-ppp hairpin RNA (5'-pppGAAUAUAAUAGUGAUAUUAUAUUC-3') is chemically synthesized from Biosynthesis Company and it contains a stem of 10 base pairs, a GUGA tetra loop, and 5’-triphosphate moiety. Blunt-end hairpin RNA has the same nucleotide sequence as 5’ppp hairpin RNA. Crystals of RIG-I Helicase-RD–5’ppp hairpin RNA complex were grown from 23% (w/v) PEG 3350, 0.25 M KSCN, 100 mM MOPS (pH 7.8). Crystals of RIG-I Helicase-RD–hairpin RNA ternary complex were obtained in similar conditions as described above, except 0.3 M NaSCN was used instead of 0.25 M KSCN. Crystals of RIG-I

Helicase-RD–5'ppp 14bp-dsRNA complex were grown in 24% (w/v) PEG 3350, 0.2 M KSCN, 100 mM MOPS (pH 7.8), 4% (v/v) 2,2,2-Trifluoroethanol. Crystals of RIG-I Helicase-RD–14bp-dsRNA complex (CGACGCUAGCGUCG) were grown in 24% (w/v) PEG 3350, 0.2 M KNO_3, 100 mM MOPS (pH 7.8), 4% (v/v) 2,2,2-Trifluoroethanol. Crystals of RIG-I Helicase-RD–14bp-chimaeric RNA-DNA complex (rCrGrArCrGrCrUAGCGTCG) were grown in 22% (w/v) PEG 3350, 0.25 M KSCN, 100 mM MOPS (pH 7.8).

For cryoprotection, all crystals were gently transferred to crystallization solutions plus 5% (v/v) (2R,3R)-(−)-2,3-butanediol, and then flash-frozen in liquid N_2. X-ray diffraction data were collected either at the X29A beamline of the National Synchrotron Light Source (NSLS), or at the Cornell High Energy Synchrotron Source (CHESS) F1 beamline. Diffraction intensities were indexed, integrated, and scaled with Mosflm and SCALA. Calculation of the Matthews coefficient suggested the presence of six molecules per asymmetric unit (V_M = 2.97 $Å^3$ Da^{-1}) with a solvent content of 61.9% in crystals of Helicase-RD–hairpin RNA with or without 5'triphosphate. However, Matthews coefficient calculation of crystals of Helicase-RD–dsRNA with or without 5'triphosphate indicated one molecule per asymmetric unit. Initial space group were determined by Pointless and further confirmed by decreases in both R_{work} and R_{free} during refinement. Molecular replacement was performed with Phaser using the previous determined structure of RIG-I Helicase-RD-ADP•BeF_3-dsRNA ternary complex (PDB: 3TMI), and structural model was then refined in PHENIX software package. Iterative rounds

of Model building in COOT were guided by the sigma A-weighted 2Fo-Fc electron density maps and further by inspection of omit maps calculated in PHENIX to prevent model bias. For hairpin RNA and 5'ppp hairpin RNA datasets, strict non-crystallographic symmetry (NCS) restraints were used during the early rounds of refinement, and were later relaxed in final rounds of refinement. Water molecules were added stepwise in final stages of refinement and then verified by occupancy and B-factor. MolProbity was used to perform validation of protein and RNA models. Statistics of the data processing and structure refinement are summarized in Appendix Table.

3.3.2 Structure overview and analyses

The overall structures of all the complexes are similar to the previous determined structure of RIG-I Helicase-RD-ADP•BeF_3-dsRNA ternary complex. They all share the similar closed conformation features. RIG-I helicase domains compleltely surrounds the RNA duplexes with RD domain capping the 5'-ends of duplex RNA through an extensive protein-RNA interactions and interdomain interactions (Figure 23A.) Thermal parameter distribution colored by the crystallographic temperature factor (B-factor) shows RIG-I has very high dynamtics even in the presence of RNA duplex (Figure 23B). B-factor describes by how much individual atoms or groups of atoms oscillate around their mean position in a protein structure and usally can be used to measure internal flexibility of a protein. The lowest B-value is observed helicase RecA domain 1 and most part of RecA domain 2 (dark blue). RIG-I helicase domain 3 and RD

domain, however, shows a much higher B-factor value, implying that those two domains are more dynamic and flexible during the recognition of RNA duplexes. The largest B-factor (colored in red) is observed in the loop region (665-690) within RecA domain 2 in 5'ppp duplex structures, where the electron density clearly shows more disorder than elsewhere in the structure. However, this region is actually well ordered in the crystal structures of duplexes without 5'ppp, strongly suggesting that this loop region might plays an important role in sensing 5'-triphosphate moiety. Superposition of RNA duplexes in all complex structures reveals that the orientation of the duplexes relative to RIG-I helicase-RD are almost the same (Figure 23C), and that the phosphate backbones of 9 base pairs of duplexes from the 5' end overlay pretty nicely, except 3' ends (Figure 23D). 3'-ends of duplex RNA is exposed to solvent and exhibits greater flexibility than 5' ends. Nevertheless, all RNA duplexes maintain the canonical A-form conformation, which is critical for RIG-I recognition.

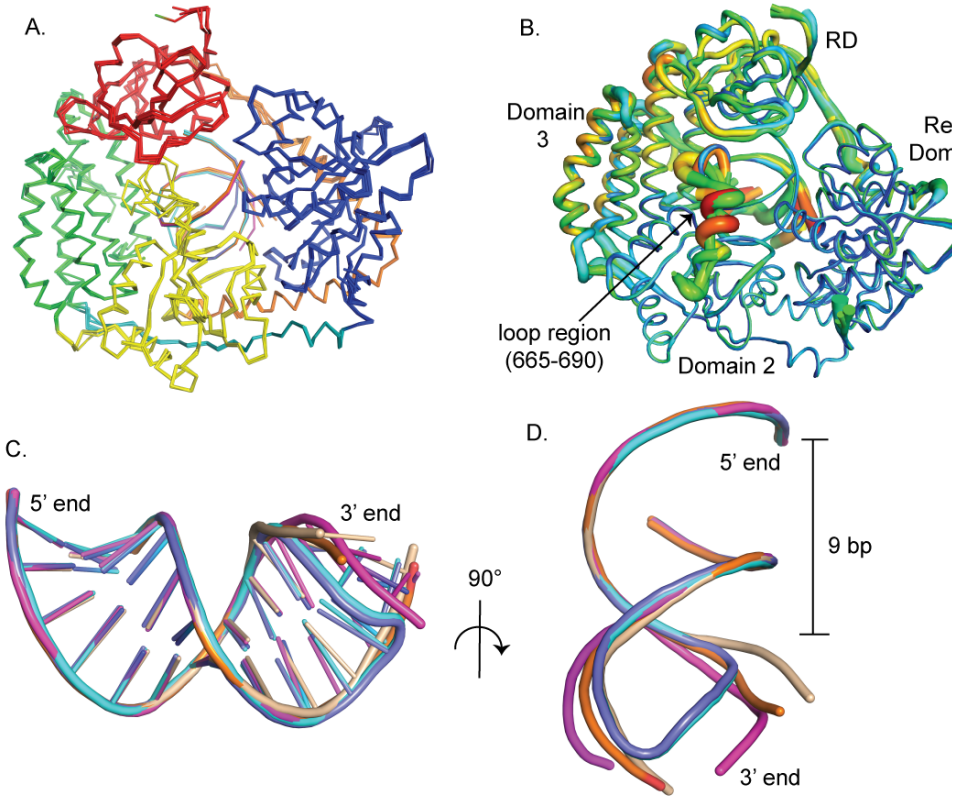


Figure 23. Structure overview and analyses of RIG-I in complex with different RNA duplex. (A). Overlay of all RIG-I Helicase-RD-RNA duplex structures. (B) Structure representation by B-fator showing the flexibility of complex. (C) Superposition of RNA duplexes in all complex structures. (D) Phosphate backbones of 9 base pairs of duplexes from the 5' end overlay pretty nicely, except 3' ends.

Consistent with what we found from diagrams of the B-factor distribution, overlay of RIG-I Helicase-RD-hairpin RNA structure with the structure of 5’ ppp-hairpin reveals a striking feature of loop region (663-672), which exhibits two conformations during the recognition of 5’ triphosphated end (Figure 24A). In the absence of 5’ triphosphate, this loop region makes hydrogen-bonding contacts with 5’ base (G1) through N668 and T667 (Figure 24B). However, this flexible loop does not make any contact with 3’ base (C24). Interestingly, an oxyanion hole is formed by K849 from RD with interaction with the negative charged oxygens on R664, Q669, N670 and T671 from the loop (Figure 24B). This oxyanion hole might be critical for stabilizition of the loop in a conformation that would suffer from steric hindrance by forming high-energy tetrahedral oxyanion intermediate or transition state through hydrogen bonding. It is also worth noting that K849 is completely conserved in RIG-I species, but not conserved at all in MDA5 and LGP2 receptors (Leu and Ile, respectively).

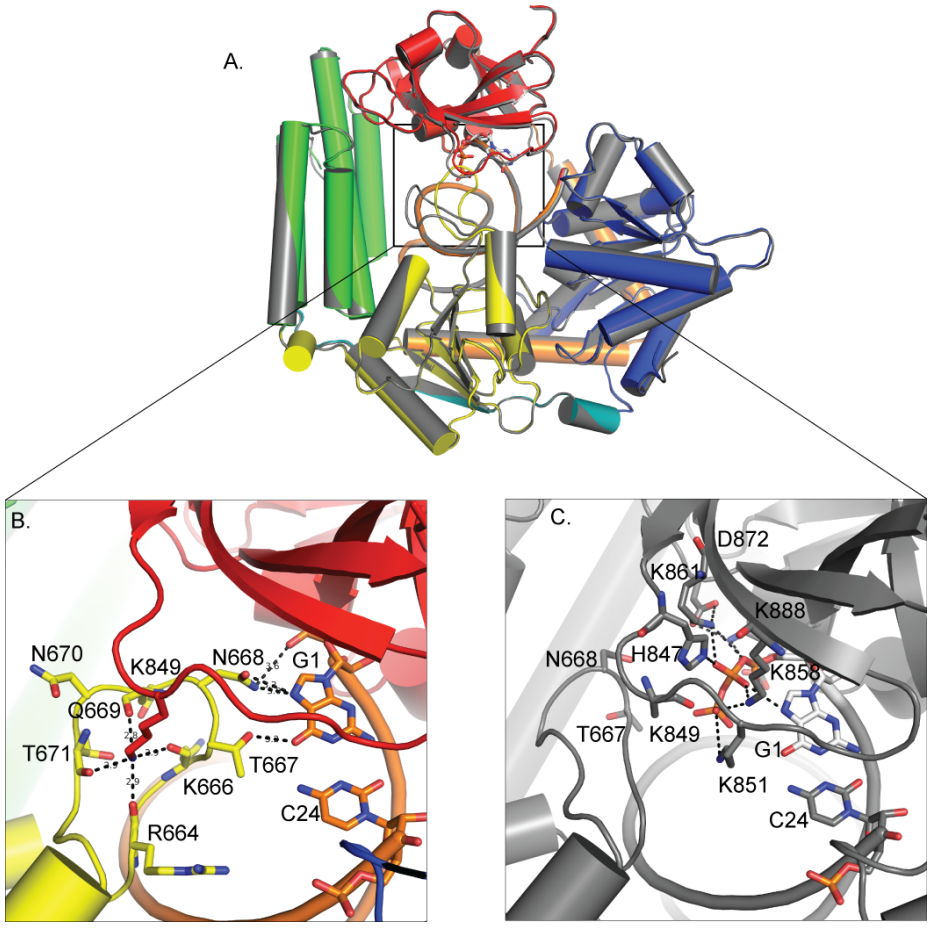


Figure 24. Loop movement in the absence or present 5'ppp. (A) Overlay of RIG-I Helicase-RD-hairpin RNA structure with the structure of 5' ppp-hairpin RNA reveals a motion of loop region (663-672). (B) Details of elaborate network of loop region with 5' end of blunt-end hairpin and K849 from RD domain. (C) Detailed interaction of 5' triphosphate and RD domain.

In the presence of 5’ triphosphate, this loop region moves away from 5’-triphosphates, and no longer makes contacts with 5’ base (Figure 24C). Intriguingly, the RNA 5’ triphosphate group is specifically recognized by RIG-I RD through extensive elctrostatic interactions. No intereaction has been obsereved between triphosphate and helicase domains in both 5’ppp hairpin RNA duplex and 5’ppp dsRNA structures. Specifically, K861 and K888 interact with the α-phosphate, and H847, K861, K858 and the G1 base make contacts with the β-phosphate. Among them, D872 forms a salt bridge with K861 and K888, which properly position K861 and K888 in the right place to faciliate the interaction with α- and β- phosphates. Additonaly, K851 forms hydrogen bonding (3.2 Å distance) with γ-phosphate, and make direct contacts with the 5’ G1 base. Sequence alignment of of the RD regions of RIG-I, MDA5, and LGP2 from different species clearly shows that this resides (K851) are well conserved. Interestingly, we could not see the clear electron density for the side chain of K849, and it could not make direct contacts with the flexible loop region. This might cause an even high flexibility for the loop region because K849 could not stablize the loop anymore once it moves away from the 5’ ends of duplex. This flexibility was further confirmed in our 5’ppp dsRNA structure, where almost no electron density is visible for this loop region.

Taken together, our structurals studies demonstrated that the loop region (663-672) acts as a “switch region”, which interacts with blunt end and 5’-triphosphate in two distinct conformations. It is most likely that the occupancy of triphosphate

in 5’ end of RNA duplex pushes the loop region away. Given that this loop region also makes contacts with the minor grove of RNA duplex in recent MDA5 helicase-RD-dsRNA structure, we speculate that this flexible loop and RD act together to contribute to the specific recognition of 5’ ends of PAMP RNA.

In additon, our recent structures established a novel helicase motif (Q507-E510-Q511 in helicase domain 3) that makes hydrogen-bonding interactions with the minor groove of dsRNA duplex (Figure 25F.) In RIG-I Helicase-RD-14bp dsRNA structure, Q507 and Q511 interact with the base G9 in 5’ strand, while E510 and Q507 contact with 2’OH of the A8 in 3’ strand, A8 in 5’ strand, respectively (Figure 25A). This interaction network is pronounced in RIG-I Helicase-RD-14bp 5’ppp dsRNA structure, where Q507 make hydrogen bonding with both the bases of A8 and G9 in 5’ strand (Figure 25B). This elaborate network is also observed in both hairpin RNA and 5’ppp hairpin RNA structures (Figure 25C-D). Hairpin RNA (5'-GAAUAUAAUAGUGAUAUUAUAUUC-3') and 5’ppp hairpin RNA (5'-pppGAAUAUAAUAGUGAUAUUAUAUUC-3') are composed of 10bp dsRNA duplex and a tetraloop. It is worth noting that the position of U18 in hairpin RNA corresponds to that of A8 in 3’ strand of dsRNA. We also determined a structure of RIG-I Helicase-RD in complex with an equivalent 14-bp chimaeric RNA-DNA duplex at 2.9 Å. The structure of complex shows the chiaeric RNA-DNA duplex makes almost identical contacts with RIG-I Helicase-RD as 14bp self-complimentary dsRNA. Interestingly, we did not get the crystals for RIG-I Helicase-RD in complex with an equivalent 14-bp chimaeric DNA-RNA duplex,

even though we still see the stable complex formation from gel filtration, which elutes at the same volume as RNA-DNA. This clearly indicates the importance of 5’ strand as a nature of RNA in crystallization. As Figure 25E shows, Q507 and Q511 still make extensive interactions with the base dG9 and dA8 in the 5’ strand, while Q507 and E510 no longer make hydrogen bonding with 2’OH of the ribose in dA8 due to the lack of 2’OH in dA8. Suprisingly, Q511 interact with both U7 in the 3’ strand and dG9 in 5’ strand, which has never seen in other RIG-I-duplex structures. One possible explanation for this observation is that Q511 make contacts with both strands to compensate for the loss of hydrogen bond contacts between E510 and A8 from the both strands.

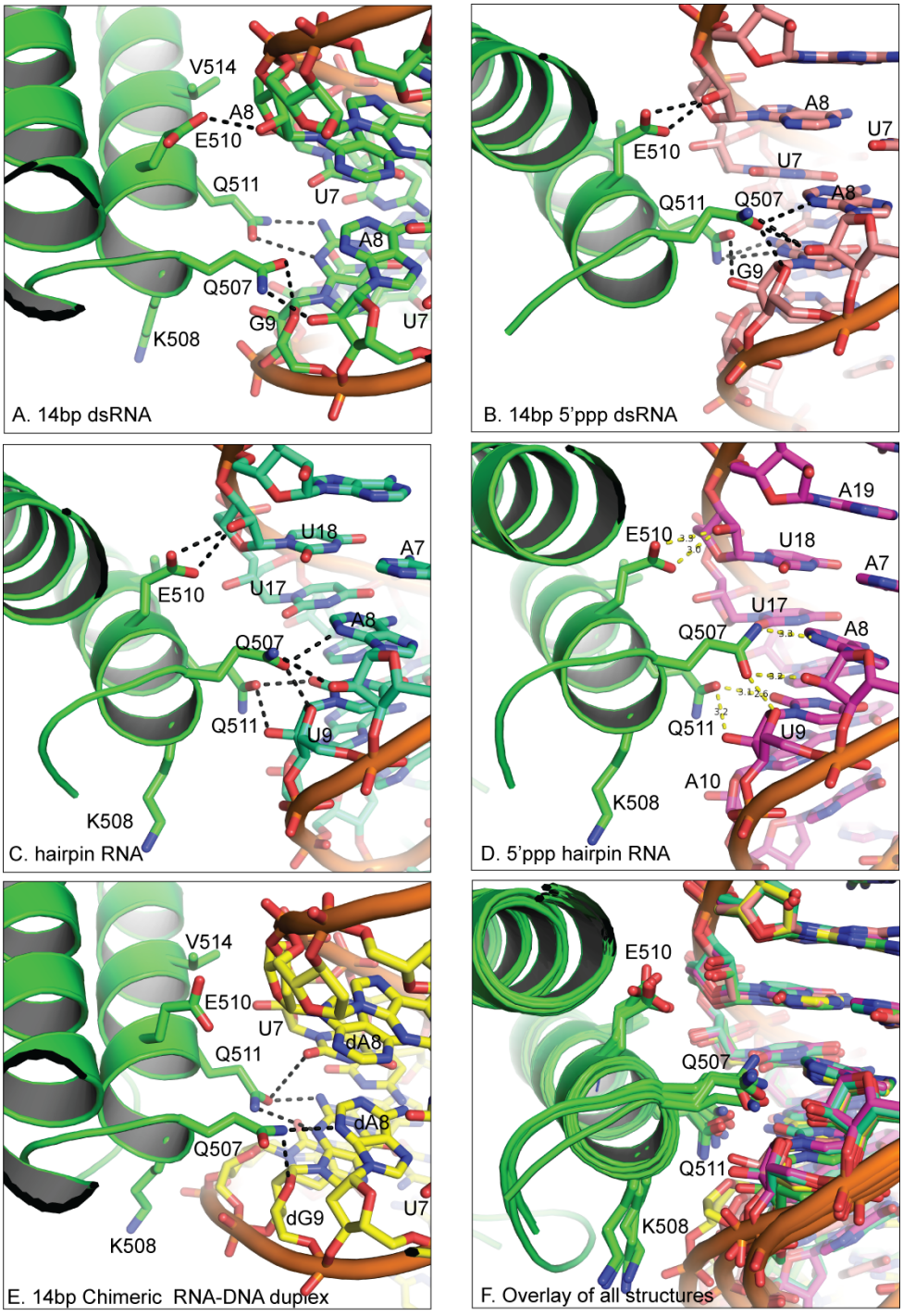


Figure 25. Helicase domain 3 (Q507-E510-Q511) makes extensive hydrogen-bonding interactions with the minor groove of dsRNA duplex.

3.3.3 Local conformational changes of RIG-I by HDX

To further investigte the conformational changes of RIG-I upon binding RNA duplex, we carried out HDX experiment with RIG-I FL and Helicase-RD in the presence or absence of 5'ppp hairpin RNA. The color shown in the heating map corresponds to the computed deuteration percentage of that region in RIG-I protein (see the color key in the upper right of the Figure 26). Hotter color represents more exchange, thus much exposed to solvent. Thus, this deuterium uptake information can be used for assessing conformational change in localized region of RIG-I upon binding RNA. The HDX profiles show that in the absence of 5'ppp hairpin RNA, RIG-I CARDs has low deuterium exchange rate (blue color), indicating that this domain is less exposed and flexible, either due to the protection from other domains or due to the presence of intramolecular hydrogen bonding (a sign of more alpha helices present in structure) (Figure 26A). However, the linker region (residues 180 to 245) between CARDs domain and Helicase domain, the internal region from residues 665 to 735 in helicase domain 2, and the Pro-rich region in the V-shape linker (residues 795-810) have much higher exchange rate (red color), strongly indicating that those linker regions are highly exposed to solution and therefore bears more flexibility. As shown in Figure 26B, higher deuterium uptake was observed in the N-terminal CARDs domain of RIG-I upon binding RNA duplex, indicating that CARDs are more exposed to the solvent. We also see the slower deuterium exchange rates in the helicase domain and RD region upon binding hairpin RNA, consistence with our observations that helicase-RD become more compared in the presence of RNA.

This is another evidence showing that RIG-I undergos conformaitonal changes upon binding RNA ligands.

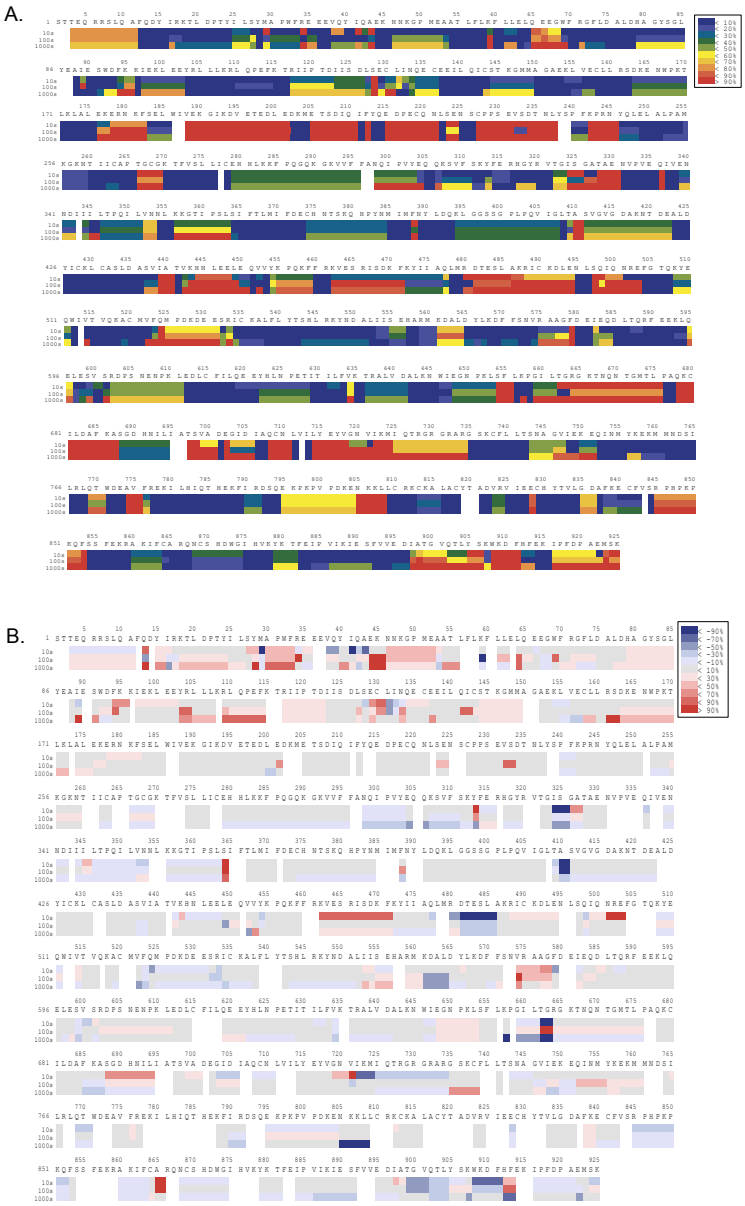


Figure 26. Local conformational changes of RIG-I by HDX. (A) HDX heating map profile of RIG-I FL alone. RIG-I CARDs has low deuterium exchange rate. (B) Comparision of RIG-I in the absence and presence of 5'ppp hairpin RNA duplex. Higher deuterium uptake was observed in the N-terminal CARDs domain of RIG-I upon binding RNA duplex.

4. Discussion

4.1 Role of the helicase/ATPase

Our structure shows a major role of the helicase domain of RIG-I in dsRNA binding; however, the role of ATPase/helicase function remains enigmatic. RIG-I has been reported to use its ATPase to translocate on dsRNA without RNA melting. Our structure showing an A- form dsRNA duplex bound to the helicase-RD is consistent with the dsRNA translocation model. The overlay of the helicase domain of RIG-I with the HCV NS3h suggests that RIG-I would use the 3'-strand for its motor contacts. Thus, translocation of RIG-I on dsRNA can be mechanistically similar to moving on single-stranded nucleic acid32. The ATPase can bring about periodic 'ratchet' type opening-closing of the two RecA-like helicase domains, which will allow RIG-I to track the 3'-strand and move on dsRNA. Translocation activity may facilitate RIG-I to displace RNA binding proteins and/or to find the end of the RNA. The translocation /ATPase can check to determine whether RIG-I is bound to self or non-self RNA. The self RNAs do not contain a blunt end or 5'-ppp, and thus RIG-I will continue to translocate off the RNA end and dissociate. On the other hand, RIG-I with its tight interactions with blunt-end and 5'-ppp will remain bound to the non-self RNA to eventually initiate signaling. We propose that our structure of the helicase-RD represents the end-point of such a search, where RIG-I has found the blunt end and the helicase-RD is engaged with the terminal base pair. We propose that such a conformation of the helicase-RD in the full-length RIG-I would represent the activated state of RIG-I.

4.2 Role of RD in RNA binding and ATPase

The crystal structure of the RIG-I helicase-RD shows that the RD interacts mainly with the blunt-end bp and a few nt near the 5'-end; however, the helicase interacts extensively with 3'-strand and 5'-strand. Despite its major contribution to dsRNA binding in the helicase-RD, the helicase alone shows substantially weak dsRNA binding affinity in the absence of RD, whereas RD alone binds the blunt-end dsRNA still relatively tightly compared with RIG-I FL. It appears that the RD is important in organizing the helicase domain to stably interact with the dsRNA, and there is synergy between the helicase and RD that promotes blunt-end dsRNA recognition. Why helicase alone shows weak ATPase, however helicase-RD shows robust ATPase activity? The most plausible explanation is that RD binds RNA, and then brings RNA close to the helicase domain. This causes the conformational rearrangement in helicase domain, making the ATP binding pocket more acessessible to ATP binding and hence activates ATP hydrolysis.

4.3 RIG-I activation model

The RIG-I also contains an autoinhibition mechanism to safeguard against aberrant signaling. In the absence of viral RNA, RIG-I exists in an autoinhibited state, which is proposed to be a 'closed' state in which the CARDs are interacting with the RD and not exposed for signaling. In the presence of RNA, RIG-I changes into an 'open' state where the RD is engaged with the RNA and the CARDs are exposed for active signaling. Our SAXS measurements are consistent with a conformational change in RIG-I upon dsRNA binding.

However, our structural data indicate that RIG-I has an open and flexible conformation in the absence of RNA and it changes into a closed conformation in the presence of dsRNA. The closed conformation captured in our helicase-RD crystal structure shows how dsRNA organizes the RIG-I domains into a ring around dsRNA. This helicase-RD crystal structure could be modeled into the molecular envelop of the full-length RIG-I bound to dsRNA, which reveals the conformation of the two CARDs. Our model of full-length RIG-I bound to dsRNA shows a compact helicase-RD around dsRNA and the two CARDs emerging in succession from the helicase-RD in a direction perpendicular to the long axis of the dsRNA. This conformation of the CARDs appears to be exposed and in a position to interact with other CARDs to initiate signaling events.

4.4 Speculations on how RIG-I binds to 5'ppp ssRNA

It has been reported that polyU-rich ssRNA with 5'ppp serve as RIG-I ligands for signaling. The RD domain binds to the 5'ppp ssRNA; however, it is difficult to imagine how the helicase would continue these interactions with the ssRNA in the conformation observed in our structure because its polarity of RNA binding would be incorrect. There are three possible solutions that can allow both helicase and RD to interact with the 5'ppp ssRNA: 1) The helicase-RD assumes a different conformational state to bind the 5'ppp ssRNA. 2) The RNA if long enough can loop around and bind the helicase domain in the orientation observed in our structure. 3) Two molecules of ssRNA bind to the helicase-RD,

one to the RD and another one to the helicase. In this way it would mimic the dsRNA structure to facilitate the recognition by RIG-I.

5. Conclusion and outlook

The innate immune system is the first line of host defense against pathogen invasion. The activation of the innate immune system relies on the recognition of pathogen-associated molecular patterns (PAMPs) by specific pattern-recognition receptors (PRRs). RIG-I is one of the crucial PPRs in the cytoplasm that induce innate immunity against HCV and other RNA viruses by selectively binding PAMP RNA containing 5' terminal triphosphate (5'ppp) or double-stranded RNA to signal IRF-3 activation, IFN-α/β expression, and induction of antiviral/immunomodulatory interferon-stimulated genes (ISGs). However, the detailed molecular mechanism for RIG-I activation remains elusive. In our studies, we overexpressed, purified and characterized full-length RIG-I protein. In addition, we identified a specific HCV RNA PAMP structure for RIG-I recognition. We then focused on determining the three-dimensional crystal structures of RIG-I-helicase-RD in complex with different RNA duplexes. Our structure presented here provided a detailed view of the role of the helicase in dsRNA recognition, synergy between RD and the helicase for RNA binding, organization of full-length RIG-I bound to dsRNA, and further evidence of a conformational change upon RNA binding. Furthermore, our structures provided a molecular basis of RIG-I for duplex RNA specific recognition with and without 5'-triphosphate. The structural and mechanistic details of RIG-I obtained from this study prompted us to propose a new, paradigm-changing model of PAMP recognition and RIG-I activation. Overall, my thesis work provides a greater mechanistic understanding

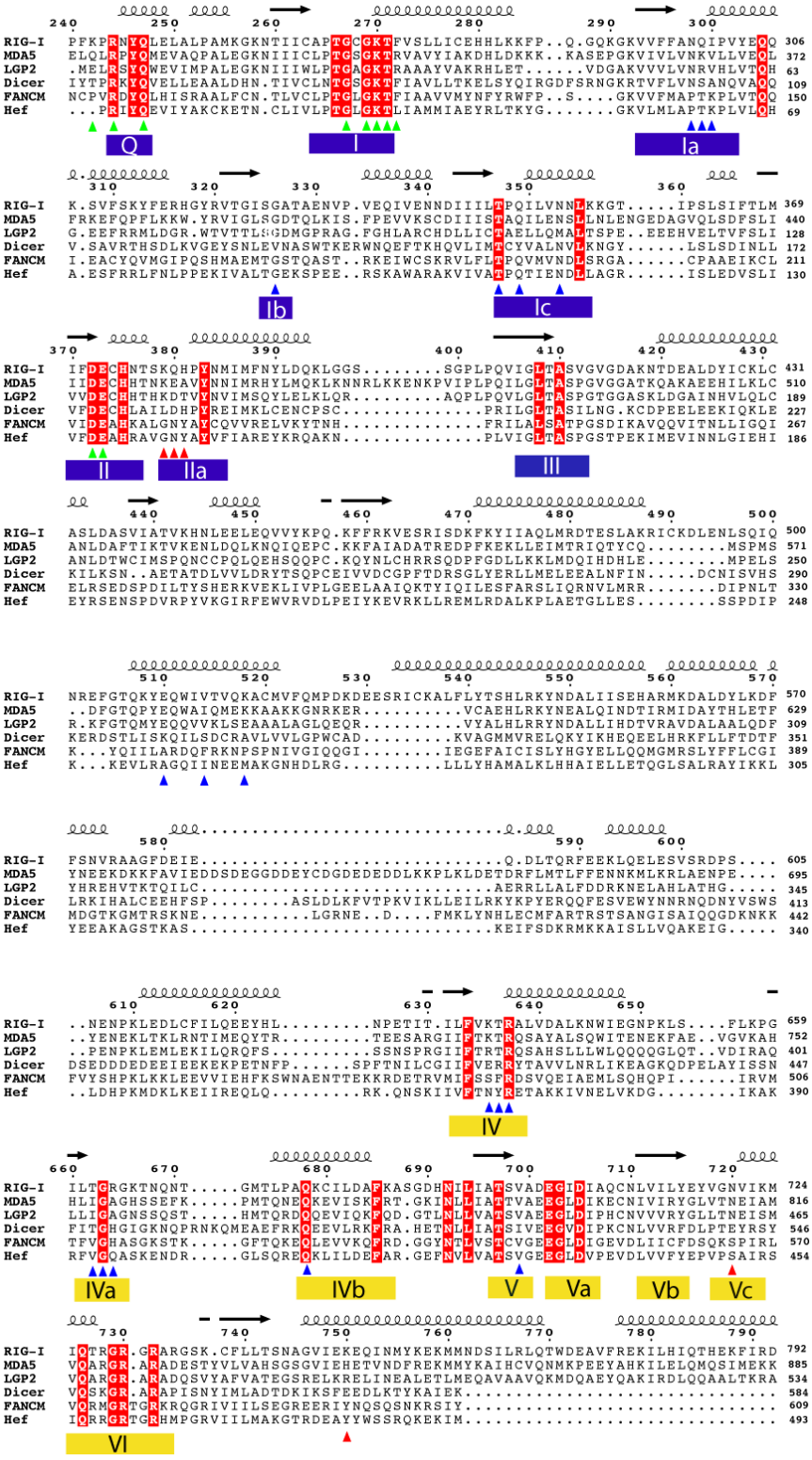
of innate immune activation to viral infection, and may have broader impact in other areas.

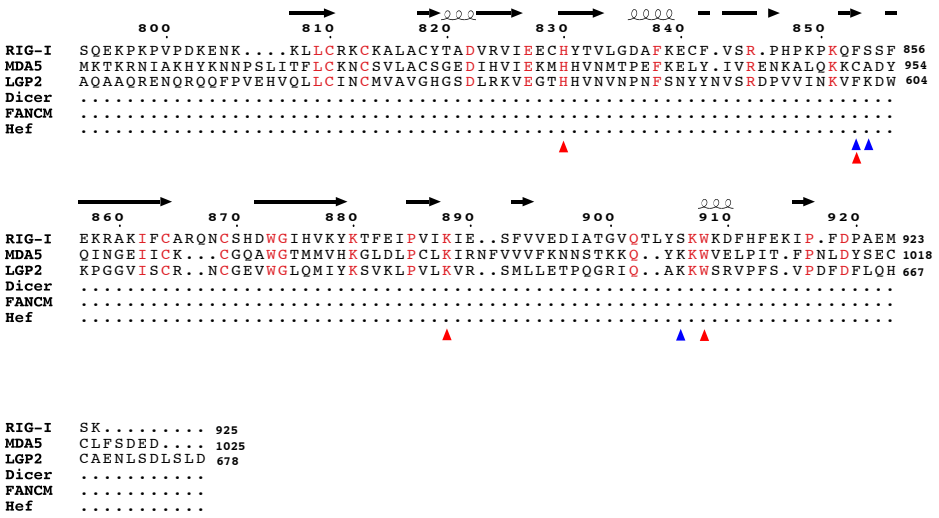
Despite these advances, our understanding of the molecular mechanism of RIG-I-mediated antiviral signaling is still limited and several outstanding questions remain unanswered. For example, the RIG-I Helicase-RD-dsRNA structures have shown a major role of the helicase domain in dsRNA binding (Jiang et al., 2011; Luo et al., 2011); however, there is still much to uncover what is underneath the hoods of ATP binding and hydrolysis in the course of RIG-I activation and signaling. RIG-I has been demonstrated to translocate rapidly and repeatedly on short dsRNA (20-50 bp) (Myong et al., 2009). Nevertheless, the biological relevance of RIG-I translocation activity still needs to be elucidated. For instance, how does the translocation activity correlate to RIG-I signaling? Another key question is the molecular mechanism between RIG-I, MDA5 and LGP2 for recognizing different viruses. RIG-I and MDA5 detects short dsRNA and long dsRNA, respectively. Likewise, the molecular mechanism underlying the sensing of nucleotide length has remained a mystery. In addition, so far all of the atomic structural data have been performed with short double-stranded RNAs. These dsRNAs do have extraordinarily high binding affinities to RIG-I, but they seem unlikely to stimulate sufficient IFN production *in vivo*. Therefore, an atomic structure of RLR bound to a real viral PAMP will undoubtedly shed new light on the precise molecular mechanism of viral RNA recognition and RIG-I activation in nature.

In addition, RIG-I is found to form multimeric complexes in vivo during viral infection (Saito et al., 2007), and it has also been shown to cooperatively form oligomers along long dsRNA in vitro (Binder et al., 2011). These observations raise the following questions: What does the role of the formation of higher-order RIG-I complexes play? Recent EM studies by Dr. Smita Patel's group clearly show RIG-I Helicase-RD can form multimers on the long dsRNA, whereas RIG-I FL cannot form a multimer (Personal communication). What is the underlying molecular basis for RIG-I dimerization or multimerization? Additionally, current model suggests polyubiquitinated RIG-I interacts with MVAS to relay the signals to downstream effectors. In this regard, it will be important to investigate the exact nature of interactions between RIG-I, MAVS, TRIM25 in the future. Regardless, the structural insights gained from RIG-I with and without RNA substrates will guide future studies aimed at finalizing our understanding of RLR activation and signaling in innate immunity. Furthermore, these studies will provide a framework for understanding the functions of other important DExD/H helicases, such as Dicer and FANCM, involved in RNA interference and DNA repair pathway, respectively.

6. Appendix

6.1 sequence alignment of RIG-I-like receptors





Sequence alignment of RIG-I-like receptors. Absolutely conserved residues in the helicase are shown as white text on a red background, while identical residues in the RD domain are shown as red text with a white background. Helicase motifs are boxed with colors according to domain localization as in structures. Roman numerals indicate the position of conserved motifs found in RNA helicases and RNA helicase-like proteins. Red triangle indicates the residues involved in 5' strand RNA binding, whereas the residues involved in binding to 3' stand RNA are denoted with a blue triangle. Green triangle indicates the conserved residues involved in ATP analog binding. The secondary structure and residues numbers of RIG-I are marked on the top of sequence alignment. Accession numbers for each protein are as follows: RIG-I (NP_055129.2), MDA5 (NP_071451.2), LGP2 (NP_077024.2), Dicer (NP_085124.2), FANCM (NP_065988.1), and Hef (NP_579744.1).

6.2 Summary of the X-ray Crystallographic Analyses

	24 mer blunt-end Haripin RNA	24 mer 5'ppp Haripin RNA
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* (Å)	111.5, 174.3, 308.3	111.2, 174.8, 309.3
α, *β*, *γ* (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	30.0-2.6 (2.74-2.6)	50-3.0 (3.37-3.2)
R_{merge}	0.156 (1.016)	0.279 (1.106)
R_{pim}	0.043 (0.335)	0.087 (0.353)
I/sI	13.0 (2.4)	10.0 (2.4)
Completeness (%)	90.6 (80.3)	99.5 (99.6)
Redundancy	13.0 (9.5)	11.1 (10.6)
Refinement		
Resolution (Å)	29.99-2.60(2.69-2.60)	49.44-3.1(3.21- 3.1)
No. reflections	166274	99666
R_{work}/R_{free}	0.21(0.30)/ 0.25(0.36)	0.20(0.30)/0.25(0.36)
No. atoms		
Protein	35025	34297
Ligand/ion	6	200/4
B-factors (Å^2)[a]		
Protein	48.80	68.20
Ligand/ion	47.10	57.20
R.m.s deviations		
Bond lengths (Å)	0.012	0.006
Bond angles (º)	1.42	1.09

	14bp dsRNA	14bpdsRNA -ADP-BeF_3
Data collection		
Space group	$P6_522$	$P6_522$
Cell dimensions		
a, *b*, *c* (Å)	174.8, 174.8, 110.1	175.4, 175.4, 109.2
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Resolution (Å)	57.2-2.95(3.1-2.95)	30.0-2.9(3.1-2.9)
R_{merge}	0.15(2.37)	0.18(2.83)
R_{pim}	0.024(0.384)	0.028(0.45)
$I/\sigma I$	16.9(2.0)	17.4(1.9)
Completeness (%)	99.9(100.0)	99.8(99.8)
Redundancy	40.5(38.2)	39.4(40.0)
Refinement		
Resolution (Å)	28.16 - 3.2 (3.004 - 2.9)	18.06-2.90
No. reflections	22444	22245
R_{work}/R_{free}	0.22(0.47)/0.27(0.45)	0.199/0.287
No. atoms		
Protein	5878	5196
Ligand/ion	19	544
B-factors ($Å^2$)[a]		
Protein	120.20	117.1
Ligand/ion	132.90	126.3
R.m.s deviations		
Bond lengths (Å)	0.002	0.014
Bond angles (º)	0.48	1.266

	14bp 5'ppp dsRNA	14bpp RNA-DNA chimeric RNA
Data collection		
Space group	$P6_522$	$P6_522$
Cell dimensions		
a, *b*, *c* (Å)	175.5, 175.3, 109.1	176.2, 176.2, 108.3
α, *β*, *γ* (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Resolution (Å)	51.3-3.2 (3.37-3.2)	46.12-2.9 (2.95-2.8)
R_{merge}	0.294 (2.744)	0.266 (3.398)
R_{pim}	0.046 (0.443)	0.041 (0.545)
I/s*I*	15.0 (2.1)	18.0 (1.5)
Completeness (%)	99.9 (99.9)	100.0 (100.0)
Redundancy	40.8 (38.4)	42.7 (38.1)
Refinement		
Resolution (Å)	46.2-3.2 (3.31-3.2)	34.99-2.9(2.952-2.85)
No. reflections	16816	24910
R_{work}/R_{free}	0.23(0.35)/0.28(0.40)	0.23(0.37)/ 0.29(0.41)
No. atoms		
Protein	5468	5779
Ligand/ion	30	1
B-factors (Å^2)[a]		
Protein	129.60	31.30
Ligand/ion	145.30	55.20
R.m.s deviations		
Bond lengths (Å)	0.008	0.015
Bond angles (º)	1.17	1.16

7. Reference

(1994). The CCP4 suite: programs for protein crystallography. Acta crystallographica Section D, Biological crystallography *50*, 760-763.

Ablasser, A., Bauernfeind, F., Hartmann, G., Latz, E., Fitzgerald, K.A., and Hornung, V. (2009). RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate. Nature immunology *10*, 1065-1072.

Abrahams, J.P., and Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr D Biol Crystallogr *52*, 30-42.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al*. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Akira, S., Uematsu, S., and Takeuchi, O. (2006). Pathogen recognition and innate immunity. Cell *124*, 783-801.

Andrejeva, J., Childs, K.S., Young, D.F., Carlos, T.S., Stock, N., Goodbourn, S., and Randall, R.E. (2004). The V proteins of paramyxoviruses bind the IFN-inducible RNA helicase, mda-5, and inhibit its activation of the IFN-beta promoter. Proc Natl Acad Sci U S A *101*, 17264-17269.

Baum, A., and Garcia-Sastre, A. (2011). Differential recognition of viral RNA by RIG-I. Virulence *2*, 166-169.

Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc *129*, 5656-5664.

Binder, M., Eberle, F., Seitz, S., Mucke, N., Huber, C.M., Kiani, N., Kaderali, L., Lohmann, V., Dalpke, A., and Bartenschlager, R. (2011). Molecular mechanism of signal perception and integration by the innate immune sensor retinoic acid-inducible gene-I (RIG-I). The Journal of biological chemistry *286*, 27278-27287.

Chiu, Y.H., Macmillan, J.B., and Chen, Z.J. (2009). RNA polymerase III detects cytosolic DNA and induces type I interferons through the RIG-I pathway. Cell *138*, 576-591.

Cohen, S.L., and Chait, B.T. (2001). Mass spectrometry as a tool for protein crystallography. Annu Rev Biophys Biomol Struct *30*, 67-85.

Cui, S., Eisenacher, K., Kirchhofer, A., Brzozka, K., Lammens, A., Lammens, K., Fujita, T., Conzelmann, K.K., Krug, A., and Hopfner, K.P. (2008). The C-terminal regulatory domain is the RNA 5'-triphosphate sensor of RIG-I. Molecular cell *29*, 169-179.

Cui, Y., Li, M., Walton, K.D., Sun, K., Hanover, J.A., Furth, P.A., and Hennighausen, L. (2001). The Stat3/5 locus encodes novel endoplasmic reticulum and helicase-like proteins that are preferentially expressed in normal and neoplastic mammary tissue. Genomics *78*, 129-134.

Dauter, Z., Dauter, M., and Dodson, E. (2002). Jolly SAD. Acta Crystallogr D Biol Crystallogr *58*, 494-506.

Davis, I.W., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res *32*, W615-619.

Dong, A., Xu, X., Edwards, A.M., Chang, C., Chruszcz, M., Cuff, M., Cymborowski, M., Di Leo, R., Egorova, O., Evdokimova, E., *et al*. (2007). In situ proteolysis for protein crystallization and structure determination. Nat Methods *4*, 1019-1021.

Durr, H., Flaus, A., Owen-Hughes, T., and Hopfner, K.P. (2006). Snf2 family ATPases and DExx box helicases: differences and unifying concepts from high-resolution crystal structures. Nucleic acids research *34*, 4160-4167.

Durr, H., Korner, C., Muller, M., Hickmann, V., and Hopfner, K.P. (2005). X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. Cell *121*, 363-373.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Fontana, A., de Laureto, P.P., Spolaore, B., Frare, E., Picotti, P., and Zambonin, M. (2004). Probing protein structure by limited proteolysis. Acta Biochim Pol *51*, 299-321.

Fredericksen, B.L., Keller, B.C., Fornek, J., Katze, M.G., and Gale, M., Jr. (2008). Establishment and maintenance of the innate antiviral response to West Nile Virus involves both RIG-I and MDA5 signaling through IPS-1. Journal of virology *82*, 609-616.

Gao, X., Bain, K., Bonanno, J.B., Buchanan, M., Henderson, D., Lorimer, D., Marsh, C., Reynes, J.A., Sauder, J.M., Schwinn, K., *et al*. (2005). High-throughput limited proteolysis/mass spectrometry for protein domain elucidation. J Struct Funct Genomics *6*, 129-134.

Hoggan, D.B., Chao, J.A., Prasad, G.S., Stout, C.D., and Williamson, J.R. (2003). Combinatorial crystallization of an RNA-protein complex. Acta Crystallogr D Biol Crystallogr *59*, 466-473.

Hornung, V., Ellegast, J., Kim, S., Brzozka, K., Jung, A., Kato, H., Poeck, H., Akira, S., Conzelmann, K.K., Schlee, M., *et al*. (2006). 5'-Triphosphate RNA is the ligand for RIG-I. Science *314*, 994-997.

Hura, G.L., Menon, A.L., Hammel, M., Rambo, R.P., Poole, F.L., 2nd, Tsutakawa, S.E., Jenney, F.E., Jr., Classen, S., Frankel, K.A., Hopkins, R.C., *et al*. (2009). Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). Nat Methods *6*, 606-612.

Imaizumi, T., Aratani, S., Nakajima, T., Carlson, M., Matsumiya, T., Tanji, K., Ookawa, K., Yoshida, H., Tsuchida, S., McIntyre, T.M., *et al*. (2002). Retinoic acid-inducible gene-I is induced in endothelial cells by LPS and regulates expression of COX-2. Biochem Biophys Res Commun *292*, 274-279.

Janeway, C.A., Jr., and Medzhitov, R. (2002). Innate immune recognition. Annual review of immunology *20*, 197-216.

Jiang, F., Ramanathan, A., Miller, M.T., Tang, G.Q., Gale, M., Jr., Patel, S.S., and Marcotrigiano, J. (2011). Structural basis of RNA recognition and activation by innate immune receptor RIG-I. Nature *479*, 423-427.

Johnson, C.L., and Gale, M., Jr. (2006). CARD games between virus and host get a new player. Trends Immunol *27*, 1-4.

Jordan, S.R., Whitcombe, T.V., Berg, J.M., and Pabo, C.O. (1985). Systematic variation in DNA length yields highly ordered repressor-operator cocrystals. Science *230*, 1383-1385.

Kang, D.C., Gopalkrishnan, R.V., Wu, Q., Jankowsky, E., Pyle, A.M., and 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*, 637-642.

Kato, H., Takahasi, K., and Fujita, T. (2011). RIG-I-like receptors: cytoplasmic sensors for non-self RNA. Immunological reviews *243*, 91-98.

Kato, H., Takeuchi, O., Sato, S., Yoneyama, M., Yamamoto, M., Matsui, K., Uematsu, S., Jung, A., Kawai, T., Ishii, K.J., *et al*. (2006). Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. Nature *441*, 101-105.

Katze, M.G., Fornek, J.L., Palermo, R.E., Walters, K.A., and Korth, M.J. (2008). Innate immune modulation by RNA viruses: emerging insights from functional genomics. Nature reviews Immunology *8*, 644-654.

Kawai, T., and Akira, S. (2006). Innate immune recognition of viral infection. Nat Immunol *7*, 131-137.

Kawai, T., and Akira, S. (2008). Toll-like receptor and RIG-I-like receptor signaling. Ann N Y Acad Sci *1143*, 1-20.

Kawai, T., Takahashi, K., Sato, S., Coban, C., Kumar, H., Kato, H., Ishii, K.J., Takeuchi, O., and Akira, S. (2005). IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. Nature immunology *6*, 981-988.

Ke, A., and Doudna, J.A. (2004). Crystallization of RNA and RNA-protein complexes. Methods *34*, 408-414.

Komuro, A., and Horvath, C.M. (2006). RNA- and virus-independent inhibition of antiviral signaling by RNA helicase LGP2. J Virol *80*, 12332-12342.

Kovacsovics, M., Martinon, F., Micheau, O., Bodmer, J.L., Hofmann, K., and Tschopp, J. (2002). Overexpression of Helicard, a CARD-containing helicase cleaved during apoptosis, accelerates DNA degradation. Curr Biol *12*, 838-843.

Kumar, H., Kawai, T., and Akira, S. (2009). Pathogen recognition in the innate immune response. Biochem J *420*, 1-16.

Lipfert, J., and Doniach, S. (2007). Small-angle X-ray scattering from RNA, proteins, and protein complexes. Annu Rev Biophys Biomol Struct *36*, 307-327.

Lorsch, J.R., and Herschlag, D. (1998a). The DEAD box protein eIF4A. 1. A minimal kinetic and thermodynamic framework reveals coupled binding of RNA and nucleotide. Biochemistry *37*, 2180-2193.

Lorsch, J.R., and Herschlag, D. (1998b). The DEAD box protein eIF4A. 2. A cycle of nucleotide and RNA-dependent conformational changes. Biochemistry *37*, 2194-2206.

Lu, C., Ranjith-Kumar, C.T., Hao, L., Kao, C.C., and Li, P. (2011). Crystal structure of RIG-I C-terminal domain bound to blunt-ended double-strand RNA without 5' triphosphate. Nucleic acids research *39*, 1565-1575.

Lu, C., Xu, H., Ranjith-Kumar, C.T., Brooks, M.T., Hou, T.Y., Hu, F., Herr, A.B., Strong, R.K., Kao, C.C., and Li, P. (2010). The structural basis of 5' triphosphate double-stranded RNA recognition by RIG-I C-terminal domain. Structure *18*, 1032-1043.

Luo, D., Ding, S.C., Vela, A., Kohlway, A., Lindenbach, B.D., and Pyle, A.M. (2011). Structural insights into RNA recognition by RIG-I. Cell *147*, 409-422.

Marques, J.T., Devosse, T., Wang, D., Zamanian-Daryoush, M., Serbinowski, P., Hartmann, R., Fujita, T., Behlke, M.A., and Williams, B.R. (2006). A structural basis for discriminating between self and nonself double-stranded RNAs in mammalian cells. Nature biotechnology *24*, 559-565.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

Meylan, E., Curran, J., Hofmann, K., Moradpour, D., Binder, M., Bartenschlager, R., and Tschopp, J. (2005). Cardif is an adaptor protein in the RIG-I antiviral pathway and is targeted by hepatitis C virus. Nature *437*, 1167-1172.

Meylan, E., Tschopp, J., and Karin, M. (2006). Intracellular pattern recognition receptors in the host response. Nature *442*, 39-44.

Murali, A., Li, X., Ranjith-Kumar, C.T., Bhardwaj, K., Holzenburg, A., Li, P., and Kao, C.C. (2008). Structure and function of LGP2, a DEX(D/H) helicase that regulates the innate immunity response. The Journal of biological chemistry *283*, 15825-15833.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr *53*, 240-255.

Myong, S., Cui, S., Cornish, P.V., Kirchhofer, A., Gack, M.U., Jung, J.U., Hopfner, K.P., and Ha, T. (2009). Cytosolic viral sensor RIG-I is a 5'-triphosphate-dependent translocase on double-stranded RNA. Science *323*, 1070-1074.

Onoguchi, K., Yoneyama, M., and Fujita, T. (2011). Retinoic acid-inducible gene-I-like receptors. Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research *31*, 27-31.

Painter, J., and Merritt, E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr *62*, 439-450.

Pichlmair, A., Schulz, O., Tan, C.P., Naslund, T.I., Liljestrom, P., Weber, F., and Reis e Sousa, C. (2006). RIG-I-mediated antiviral responses to single-stranded RNA bearing 5'-phosphates. Science *314*, 997-1001.

Plumet, S., Herschke, F., Bourhis, J.M., Valentin, H., Longhi, S., and Gerlier, D. (2007). Cytosolic 5'-triphosphate ended viral leader transcript of measles virus as activator of the RIG I-mediated interferon response. PLoS One *2*, e279.

Plyte, S.E., and Kneale, G.G. (2001). Limited proteolysis of protein-nucleic acid complexes. Methods Mol Biol *148*, 315-321.

Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys *40*, 191-285.

Rothenfusser, S., Goutagny, N., DiPerna, G., Gong, M., Monks, B.G., Schoenemeyer, A., Yamamoto, M., Akira, S., and Fitzgerald, K.A. (2005). The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I. J Immunol *175*, 5260-5268.

Saito, T., and Gale, M., Jr. (2007). Principles of intracellular viral recognition. Curr Opin Immunol *19*, 17-23.

Saito, T., Hirai, R., Loo, Y.M., Owen, D., Johnson, C.L., Sinha, S.C., Akira, S., Fujita, T., and Gale, M., Jr. (2007). Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2. Proceedings of the National Academy of Sciences of the United States of America *104*, 582-587.

Saito, T., Owen, D.M., Jiang, F., Marcotrigiano, J., and Gale, M., Jr. (2008). Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. Nature *454*, 523-527.

Satoh, T., Kato, H., Kumagai, Y., Yoneyama, M., Sato, S., Matsushita, K., Tsujimura, T., Fujita, T., Akira, S., and Takeuchi, O. (2010). LGP2 is a positive regulator of RIG-I- and MDA5-mediated antiviral responses. Proceedings of the National Academy of Sciences of the United States of America *107*, 1512-1517.

Schlee, M., Hartmann, E., Coch, C., Wimmenauer, V., Janke, M., Barchet, W., and Hartmann, G. (2009a). Approaching the RNA ligand for RIG-I? Immunological reviews *227*, 66-74.

Schlee, M., and Hartmann, G. (2010). The chase for the RIG-I ligand--recent advances. Molecular therapy : the journal of the American Society of Gene Therapy *18*, 1254-1262.

Schlee, M., Roth, A., Hornung, V., Hagmann, C.A., Wimmenauer, V., Barchet, W., Coch, C., Janke, M., Mihailovic, A., Wardle, G., *et al*. (2009b). Recognition of 5' triphosphate by RIG-I helicase requires short blunt double-stranded RNA as contained in panhandle of negative-strand virus. Immunity *31*, 25-34.

Schmidt, A., Schwerd, T., Hamm, W., Hellmuth, J.C., Cui, S., Wenzel, M., Hoffmann, F.S., Michallet, M.C., Besch, R., Hopfner, K.P., *et al*. (2009). 5'-triphosphate RNA requires base-paired structures to activate antiviral signaling via RIG-I. Proceedings of the National Academy of Sciences of the United States of America *106*, 12067-12072.

Schneider, T.R., and Sheldrick, G.M. (2002). Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr *58*, 1772-1779.

Seth, R.B., Sun, L., and Chen, Z.J. (2006). Antiviral innate immunity pathways. Cell Res *16*, 141-147.

Seth, R.B., Sun, L., Ea, C.K., and Chen, Z.J. (2005). Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. Cell *122*, 669-682.

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007). Structure and mechanism of helicases and nucleic acid translocases. Annu Rev Biochem *76*, 23-50.

Stetson, D.B., and Medzhitov, R. (2006). Type I interferons in host defense. Immunity *25*, 373-381.

Svergun, D.I., and Koch, M.H. (2002). Advances in structure analysis using small-angle scattering in solution. Curr Opin Struct Biol *12*, 654-660.

Takahasi, K., Yoneyama, M., Nishihori, T., Hirai, R., Kumeta, H., Narita, R., Gale, M., Jr., Inagaki, F., and Fujita, T. (2008). Nonself RNA-sensing mechanism of RIG-I helicase and activation of antiviral immune responses. Molecular cell *29*, 428-440.

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., and Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. J Mol Biol *229*, 105-124.

Vance, R.E., Isberg, R.R., and Portnoy, D.A. (2009). Patterns of pathogenesis: discrimination of pathogenic and nonpathogenic microbes by the innate immune system. Cell Host Microbe *6*, 10-21.

Wang, Y., Ludwig, J., Schuberth, C., Goldeck, M., Schlee, M., Li, H., Juranek, S., Sheng, G., Micura, R., Tuschl, T., *et al*. (2010). Structural and functional insights into 5'-ppp RNA pattern recognition by the innate immune receptor RIG-I. Nature structural & molecular biology *17*, 781-787.

Xu, L.G., Wang, Y.Y., Han, K.J., Li, L.Y., Zhai, Z., and Shu, H.B. (2005). VISA is an adapter protein required for virus-triggered IFN-beta signaling. Molecular cell *19*, 727-740.

Yoneyama, M., Kikuchi, M., Matsumoto, K., Imaizumi, T., Miyagishi, M., Taira, K., Foy, E., Loo, Y.M., Gale, M., Jr., Akira, S., *et al*. (2005). Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. J Immunol *175*, 2851-2858.

Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., Akira, S., and Fujita, T. (2004). The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. Nature immunology *5*, 730-737.

Zou, J., Chang, M., Nie, P., and Secombes, C.J. (2009). Origin and evolution of the RIG-I like RNA helicase gene family. BMC evolutionary biology *9*, 85.