A feasibility study on computational prediction on the death domain superfamily protein oligomers

Zhang, Jiawen

2020

Zhang, J. (2020). A feasibility study on computational prediction on the death domain superfamily protein oligomers. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/144233>

<https://doi.org/10.32657/10356/144233>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 18:27:47 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**A FEASIBILITY STUDY ON COMPUTATIONAL PREDICTION
ON THE DEATH DOMAIN SUPERFAMILY PROTEIN
OLIGOMERS**

JIAWEN ZHANG

SCHOOL OF BIOLOGICAL SCIENCES

2020

**A FEASIBILITY STUDY ON COMPUTATIONAL PREDICTION
ON THE DEATH DOMAIN SUPERFAMILY PROTEIN
OLIGOMERS**

JIAWEN ZHANG

SCHOOL OF BIOLOGICAL SCIENCES

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

2020

Statement of Originality

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

10-01-2020

.....
Date 10 Jan 2020

Zhang Jiawen

.....
JIAWEN ZHANG

Supervisor Declaration Statement

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

10-Jan-2020
.....
Date 10 Jan 2020

.....
BIN WU

Authorship Attribution Statement

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

10-01-2020

.....
Date 10 Jan 2020

Zhang Jiawen

.....
JIAWEN ZHANG

Acknowledgments

Firstly, I thank my family for their upbringing and supports.

Secondly, I would like to express my sincere gratitude to my supervisor, Nanyang Asst. Prof. Wu Bin for his guidance and support throughout my Ph.D. journey. His motivation, patience and immense knowledge impress me and encourage me to conduct scientific research from experimental design to a paper publication. I thank my thesis advisory committee, my co-supervisor Assoc. Prof. Mu Yuguang, Assoc Prof. Lu Lanyuan i for their guidance, prolific discussions during my Ph.D and Assoc.Prof Gao Yonggui for being my QE examiner.

Finally, I thank all the current and previous lab members for their help and support during my Ph.D. life and I enjoy a lot to work with them in such a warm and happy research environment. I thank for her excellent technical contribution to the work. I thank my fellow Ph.D. students Mr. Chenrui Xu, Mr. Zhaozhi Boo, Ms. Phuong Thao Huynh and Ms. Qin Gong for their contribution and moral support during my Ph.D. life. And I also thank Ms. Ziqi Long, Ms. Yaming Zhang, Ms. Audrey Chiew, and Ms. Chai Yoke Tin, Dr. Kelvin Chong and Dr. Renliang Yang for their support and kindness they have given me during the time of Ph.D. Specially, thanks for Ms. Qin Gong for contributing her NLRP1-density(Figure 3.3.3B) during the data processing as a validation candidate for my computational prediction.

Table of Contents

SUPERVISOR DECLARATION STATEMENT.....	4
ACKNOWLEDGMENTS.....	6
CHAPTER 1: BACKGROUND	17
CHAPTER 1.1: INNATE IMMUNITY	17
<i>Anatomical barriers</i>	<i>19</i>
<i>Inflammation</i>	<i>19</i>
<i>Complement system</i>	<i>19</i>
<i>Leukocytes</i>	<i>21</i>
CHAPTER 1.2: PATTERN RECOGNITION RECEPTORS (PRRs) AND THEIR HISTORY	22
<i>RIG-I-like receptors</i>	<i>24</i>
<i>Toll-like receptors</i>	<i>29</i>
<i>NOD-like receptors</i>	<i>32</i>
<i>AIM2-like receptors</i>	<i>37</i>
CHAPTER 1.3: CRYO ELECTRON MICROSCOPY AND HELICAL RECONSTRUCTION	38
Cryogenic electron microscopy.....	38
<i>Helical reconstruction</i>	<i>40</i>
CHAPTER 1.4: DEATH DOMAIN CONTAINING PROTEINS.....	42
<i>MyD88 and IRAK2/4 complex (Myddosome) formation by death domain interaction</i>	<i>48</i>
<i>RIG-I and MAVS interaction with homotypic CARD-CARD interaction.....</i>	<i>51</i>
<i>NOD1/2 and RIP2 induced signalling pathway.....</i>	<i>54</i>
<i>Higher order assemblies in signalling cascades.....</i>	<i>59</i>
CHAPTER 1.5: COMPUTATIONAL SIMULATION AND STRUCTURE PREDICTION METHODS	62
<i>Rosetta de novo structure-prediction</i>	<i>63</i>
<i>Molecular dynamics simulation and force fields</i>	<i>68</i>
<i>GalaxyWEB</i>	<i>73</i>
<i>Protein structure prediction using I-TASSER</i>	<i>75</i>
<i>Summary of current computational tools.....</i>	<i>77</i>
CHAPTER 1.5: AIM OF THE THESIS	79
CHAPTER 2: MATERIALS AND METHODS	82
CHAPTER 2.1: HELICAL PARAMETERS IN CARD-CARD FILAMENTS	82
<i>Monomer models</i>	<i>82</i>
<i>Helical filament model building</i>	<i>82</i>
<i>Rosetta prediction</i>	<i>84</i>
<i>Molecular dynamics refinement</i>	<i>85</i>
CHAPTER 2.2: STRUCTURAL STUDY OF CARD8 CARD	86
<i>Plasmid construction</i>	<i>86</i>
<i>Protein purification.....</i>	<i>86</i>
<i>Oligomerization assay</i>	<i>88</i>
<i>Cryo-electron microscopy</i>	<i>88</i>
<i>Data processing</i>	<i>89</i>

<i>Density and model build</i>	<i>90</i>
CHAPTER 3. HOMOLOGICAL FILAMENT STRUCTURE AND HETEROGENOUS PAIR INTERACTION PREDICTION OF CARD DOMAINS	92
CHAPTER 3.1 OVERVIEW OF THE CARD DOMAIN IN INNATE IMMUNITY	92
CHAPTER 3.2 COMPUTATIONAL CARD HOMOGENOUS FILAMENT STRUCTURE PREDICTION	98
<i>Molecular dynamics.....</i>	<i>104</i>
CHAPTER 3.3 CARD FILAMENT HELICAL SYMMETRY FIRST-LAYER PREDICTIONS	106
<i>CARDs with solved filament structures in PDB database.....</i>	<i>106</i>
<i>CARDs without solved filament structures in PDB database</i>	<i>108</i>
<i>Prediction of unknown NLRP1-CARD and CARD8-CARD filament structures.....</i>	<i>114</i>
<i>NLRP1-CARD prediction</i>	<i>115</i>
<i>CARD8-CARD prediction.....</i>	<i>116</i>
CHAPTER 3.4 HETEROGENOUS (PAIRWISE INTERACTION) STUDY ELUCIDATES THE “ACCESS CODE” BETWEEN CARD-CARD INTERACTIONS	118
<i>Validation of NOD1-/NOD2-CARD interaction with RIP2-CARD</i>	<i>118</i>
<i>Prediction of MDA5-2CARDs interaction with MAVS-CARD.....</i>	<i>119</i>
CHAPTER 3.5 DISCUSSION	122
CHAPTER 4: STRUCTURAL CHARACTERIZATION OF CARD8-CARD IN INNATE IMMUNITY	126
CHAPTER 4.1 CARD8-CARD STRUCTURAL ANALYSIS	126
<i>Overview of the CARD8 in innate immunity</i>	<i>126</i>
<i>Architecture design of CARD8-CARD construct</i>	<i>127</i>
<i>Data processing for CARD8-CARD</i>	<i>129</i>
<i>CARD8-CARD filament structure analysis</i>	<i>131</i>
CHAPTER 4.2 DISCUSSION	137
CHAPTER 5. CONCLUSIONS AND FUTURE DIRECTIONS	139
CHAPTER 5.1 CONCLUSIONS	139
CHAPTER 5.2 FUTURE DIRECTIONS	141
REFERENCES	143
APPENDIX.....	155

List of Figures

Figure 1.1.1 Comparison between innate immune response after infection and adaptive immune response	17
Figure 1.1.2 The immune recognition and receptors in innate and adaptive immunity. ..	18
Figure 1.1.3 Reaction Cascade of the Complement System: Classical, Alternative and Lectin Pathway, Amplification Loop, Terminal Pathway, and Membrane Attack complex.....	21
Figure 1.2.1 Domain structure of RLRs	26
Figure 1.2.2 Model of Signal Activation by Full-Length RIG-I and MAVS	28
Figure 1.2.3 TLRs recognize microbes by binding to PAMPs	30
Figure 1.3.1 The cryo-EM structure of the tail of T4 bacteriophage.	40
Figure 1.4.1 Structure comparison of PYRIN, DED, CARD and DD domains	42
Figure 1.4.2 Structure-based amino acid sequence alignment of representatives of the four different subfamilies of the death-domain superfamily listed in Figure 1.4.1	43
Figure 1.4.3 A Death Domain (DD) mediates three interaction types through six interaction patches.	44
Figure 1.4.4 Structure of PIDDosome consisting of PIDD and RAIDD DDs.	46
Figure 1.4.5 Structure of Myddosome, formed by DDs from Myd88, IRAK2 and IRAK4.	48
Figure 1.4.6 Helical assemblies of RIG-I and MAVS DDs.....	50
Figure 1.4.7 Illustrations of protein domain structure of human NOD1, NOD2, and RIP2. The domains are highlighted in colors.....	54
Figure 1.4.8 The details of Ia:Ib, IIa:IIb, and IIIa:IIIb interactions in RIP2–CARD filament	56
Figure 1.4.9 Model of sequential assembly in TLR/IL1-R signalling	58
Figure 1.4.10 Schematic diagram for mechanism of PrgJ-NAIP2–nucleated polymerization of NLRC4, followed by caspase-1 activation.....	59
Figure 3.1.1 Overview of the helical symmetry and current CARD filaments architectures.	93
Figure 3.2.1 Schematic of interacting surface and architecture of CARD-CARD filaments and computational design protocol.....	95
Figure 3.2.2 Helical symmetry parameter score screen return map (grid search) of MAVS-CARD	98
Figure 3.2.3 MAVS-CARD Molecular dynamics simulation energy, RMSD, and superimposition	100
Figure 3.3.1 All score/relax maps for a screen of applying helical symmetry-rise from 4Å to 5.9Å, rotation angle from 90° to 109° for all other six CARD filaments with solved cryo-EM filament structures in the PDB database.....	107

Figure 3.3.2 All score/relax map return for a screen of applying helical symmetry-rise from 4Å to 5.9Å, rotation angle from 90° to 109° for all other CARD filaments without solved cryo-EM filament structures in PDB database(unit:REU). The CARDS with predicted monomer structures are labelled with *	113
Figure 3.3.3 NLRP1-CARD prediction score and validation structures	115
Figure 3.3.4 CARD8-CARD prediction score and validation structure	117
Figure 3.4.1 Schematic illustration and score of NOD1: RIP2, NOD2: RIP2, RIG-I: MAVS, MDA5: MAVS pairwise interactions	121
Figure 4.1.1 The architecture of CARD8 and comparison of cryo-EM image of CARD8-CARD-SNAP and CARD8-CARD	128
Figure 4.1.2 Cryo-EM structure determination of CARD8-CARD filaments	130
Figure 4.1.3 CARD8-density and sidechain fitting	131
Figure 4.1.4 Cryo-EM structures of CARD8-CARD filaments	132
Figure 4.1.5 CARD8-CARD filament structure analysis	134
Figure 4.1.6 Sequence alignment and Type IIa-b interface for inflammasome CARD assemblies	136

List of Tables

Table1.2.1 pattern-recognition receptor families.....	23
Table1.5.1 Summary of terms in the REF15 energy function for proteins.....	65
Table 1.5.2 The summary and comparison of current computational tools advantage and disadvantage	79
Table3.1.1 Summary of human solved and unsolved CARD structures	94
Table3.1.2 Summary of solved human CARDS filament structures parameters: rise, rotation angle, and diameter of filaments, PDB ID, and the year of the structure being solved and deposited.....	95
Supplement Table S1 1 Data collection, map and model refinement, validation of CARD8 filament structures	155

List of Abbreviations

AD	acidic transactivation domain
AIM2	Absent in melanoma 2
ALRs	absent-in-melanoma (AIM)-like receptors
ASC	apoptosis-associated speck-like protein containing a CARD
ATP	adenosine triphosphate
BIR	baculovirus inhibitor of apoptosis repeat
C3	Complement component 3
CARD	caspase activation recruitment domain
CCD	charge-coupled device
CLRs	C-type lectin receptors
CIITA	Class II transactivator
CTD	C-terminal domain
DAMPs	danger associated molecular patterns
DD	death domain
DED	death effector domain
FIIND	function to find domain
GPI	Glycosylphosphatidylinositol
HIN	Hematopoietic expression, interferon-inducible nature, and nuclear localization
iE-DAP	γ -D-glutamyl-meso-diaminopimelic acid
IFI16	Gamma-interferon-inducible protein 16
IFNs	type I interferons
IL-1	interleukin 1
IRAK	IL-1R-associated kinases
LGP2	laboratory of genetics and physiology 2
LP	lipoproteins
LPS	bacterial lipopolysaccharide
LRRs	leucine-rich repeats
LTA	lipoteichoic acid
MALT1	mucosaassociated lymphoid tissue lymphoma translocation protein 1
MAPKs	mitogen-activated protein kinases
MAVS	mitochondrial antiviral-signaling protein
MBL	mannose-binding lectin
MD	molecular dynamics
MDA5	melanoma differentiation associated factor 5
MDP	muramyl dipeptide
Myd88	myeloid differentiation primary-response protein 88
NACHT	predicted nucleoside-triphosphatase (NTPase) domain
NAIP	NLR family apoptosis inhibitory proteins

NCS	non-crystallographic symmetry
NF-κB	nuclear factor- κ B
NLRC	Nod-like receptor subfamily C
NLRP	Nod-like receptor subfamily P
NLRs	NOD-like receptors
NMR	nuclear magnetic resonance
NOD	nucleotide oligomerization domain
PAMPs	pathogen-associated molecular patterns
PCA	Principal component analysis
PGN	peptidoglycan
PRRs	pattern recognition receptors
PYD	pyrin domain
RHIM	RIP homotypic interaction motif
RIP2	receptor-interacting serine/threonine-protein kinase2
RLRs	retinoic acid-inducible (RIG)-I-like receptors
SEM	screening electron microscope
TAK1	transforming growth factor- β -activated kinase
TEM	transmission electron microscope
TIR	Toll/IL-1R
TIRAP	TIR domain containing adaptor protein
TLRs	Toll-like receptors
TRAF6	tumor-necrosis factor-receptor-associated factor 6
TRIF	TIR-domain-containing adapter-inducing interferon- β
TRIM	TRIF related adaptor molecule
WHD	winged-helix domain

Summary

The immune system can be classified as the innate immune system and the adaptive immune system. Innate immunity refers to nonspecific immune defense mechanisms to destroy external invaders and comes into play immediately or within hours of an antigen's exposure to the body. In innate immunity, pathogen recognition receptors (PRRs) play an essential role in detecting pathogenic infections and propagating such immune signals to downstream effectors. PRRs are germline-encoded sensors that recognize pathogen-associated molecular patterns (PAMPs) as well as damage-associated molecular patterns (DAMPs). Currently, there are several subfamilies of PRRs, including C-type lectin receptors (CLRs), RIG-I like receptors (RLRs), NOD-like receptors (NLRs), AIM-like receptors (ALRs) and Toll-like receptors (TLRs), etc. Dysregulation of these PRR- related signaling pathways may lead to defective pathogen detection, and in other cases, autoimmune diseases.

How those proteins in the immune system quickly recognize the invaders and collaborate with allies is the cutting-edge point of studying immunology and can benefit both doctors and patients to provide new avenues to modulate innate immunity for therapeutic purposes. With the development of computational power, the central processing units (CPUs) and graphics processing units (GPUs) advancement enables scientists to study the protein dynamics, including protein folding and protein movements by computer in larger timescales. Using computational approaches to predict structures from primary sequence has already been widely adopted and has been proved to be a useful tool for structural biologists.

Caspase activation and recruitment (CARD) domain, as one of the Death Domain superfamily members, adopts a typical death fold motif, consisting of six antiparallel α -helix and arranged in an unusual Greek Key topology. Based on existing solved cascade-like structures of the CARDs, we hypothesized that this filamentous module could apply to all CARD domains and that the CARD domain filament-like oligomer complex is the active module in innate immunity to conduct signaling and immune responses.

Our computational approach design adopts the classical structural prediction methods and combines it with our own scripts to predict the CARD filament structures in a computational manner. Traditional structural software packages that were used to solve structures from X-ray crystallography or Cryo-electron microscope (Cryo-EM) data, such as Chimera and Coot, were applied to build a series of models with scripts running automatically. In our approach, both monomer structures from PDB databased or sequences in FASTA format can be accepted as input. Rosetta de novo Monte-Carlo structure-prediction was adopted to predict the stability of CARD filament structures models for selecting helical symmetry parameters (rise and rotation angle). Molecular dynamics (MD) methods were also adopted to refine the target models before the final predicted filament model.

In general, our computational approach provides a prediction of CARDs filaments' helical pattern for structural biologists and paves the way for model-driven quantitative structural analysis with diverse applications in CARD and other Death Domain family studies. Our hypothesis is that filamentous modules could be present in all CARD domain and that CARD domain filament-like oligomer complex is the active module in innate immunity to conduct signalling and lead to

immune response. Two candidates forming filamentous oligomers, NLRP1-CARD and CARD8-CARD, were successfully validated by using biochemistry and structural studies of their filament structures.

In order to investigate the signaling transduction process and validate our prediction and hypothesis, we conducted the structural study of the CARD8-CARD oligomeric complex by cryo-EM. The CARD domain of CARD8 (CARD8-CARD) oligomerizes to form the filamentous structure in responses to upstream CARDS and activate Caspase1 to conduct the inflammatory response. We successfully resolved the structure of the filamentous CARD8-CARD complex with a resolution of 3.7Å(central segment). With the structure at near-atomic resolution, we managed to demonstrate the mechanism of CARD8-CARD crosstalk to upstream or downstream CARDS.

Chapter 1: Background

Chapter 1.1: Innate immunity

Innate immunity refers to non-specific immune defense mechanisms, and it is one of the two main immune response in animals, plants, fungi, insects, and other primitive multi-cellular organisms. (Figure 1.1.1) (Figure 1.1.2) Other than the adaptive immune system, innate immunity is a defense against external infectious organisms and other invaders that come into play within hours of an antigen's exposure with no specific for pathogen species.

Figure 1.1.1 Comparison of the timeline between innate immune response and adaptive immune response after infection. Innate immune responses typically arise and reach its peak within hours when infection detected and the adaptive immune responses usually take several days.

There are a lot of different components that function in various ways to conduct all the functions in innate immunity. There are significant functions of innate immunity, which include:

- 1) To perform its function and act as a physical and chemical barrier to all possible agents that potentially will cause infection.
- 2) Producing chemical factors, including those specialized chemical mediators, which are also called cytokines, and recruiting immune cells to the sites of the infection.
- 3) Identification and removal of foreign substances in all organs, blood, and lymph, as well as tissues by specialized white blood cells.
- 4) Activation of complement cascade to recognize foreign bacteria, and then activate specific cells, that promotes release of the antibody complexes and clearance of dead cells inside the body.

Figure 1.1.2 The mechanisms of immune recognition/response in innate and adaptive immunity. The innate immune system in the early responses to pathogens and the adaptive immune system in late responses. The mechanisms of innate immunity provide the initial defence against infections. Some prevent infections (e.g. epithelial barriers) and others eliminate microbes (e.g. phagocytes, natural killer (NK) cells, and the complement system). The adaptive immune responses are mediated by lymphocytes and their products. Antibodies produced by B lymphocytes can block the infections and help eliminate microbes, while T lymphocytes eliminate intracellular microbes.

Anatomical barriers

Physical barriers are one kind of anatomical obstacle. Besides physical barriers, chemical and biological barriers are the other two types of anatomical barriers. Most infectious agents are not able to permeate the first line physical barrier- the epithelial surface.

Inflammation

Inflammation is part of the innate immune response of body tissues and is one of the first responses of the immune system to all kinds of infection or irritation. It is a protective response that involves immune cells, blood vessels, and a lot of molecular mediators in response to the outside antigens and other harmful stimuli such as pathogens and damaged cells. It is usually stimulated by chemical factors (cytokines) released by the injured cells and will try to establish a new physical barrier that prevents the spread of infections, and at the same time, to enhance the healing of damaged tissues and to remove external invaders.

Complement system

The complement system, a biochemical cascade, helps antibodies and labels them for destruction, clears pathogens, and triggers the recruitment of inflammatory factors. It is composed of multiple plasma proteins that are synthesized in the liver which can be divided into mainly three different classes, classical, alternative, and lectin. In all three pathways, complement component 3(C3)-convertase cleaves and activates C3, creating C3a and C3b, and causes a cascade of further cleavage and activation events. C3b binds to the surface of pathogens, leading to greater internalization by phagocytic cells by opsonization.

The classical complement system begins when the antibody recognizes and binds the pathogen and activate complement component 1(C1)-complex. C1 then acts as a protease and cleaves complement component 2(C2) and complement component 4(C4), producing C4a, C4b, C2a, and C2b. C4b and C2a bind to form the classical pathway C3-convertase (C4b2a complex), which promotes cleavage of C3 into C3a and C3b.

The alternative complement system starts spontaneously as the C3 hydrolysis breakdown of the internal thioester bond. This hydrolyzed C3 can bind and activate Factor B and cleave another C3 molecule into C3a and C3b. Once the alternative C3 convertase enzyme is formed on a pathogen or cell surface, it may bind covalently another C3b, to form the complement components 5 (C5)-convertase. This enzyme then cleaves C5 to C5a, a potent anaphylatoxin, and C5b. The C5b then recruits and assembles complement components 6,7,8 and 9 (C5, C6, C7, C8 and C9) molecules to assemble the membrane attack complex.

The lectin pathway starts with mannose-binding lectin (MBL), opsonin or ficolin binding to certain other sugars. MBL recognizes mannose residues on pathogen surface, which activates the MBL-associated serine proteases and then split C4 molecule into C4a and C4b and C2 into C2a and C2b. Then C4b and C2a bind together to form the C3 convertase just as the classical pathway.

Figure 1.1.3 The Classical and Alternative Complement Pathways: The classical and alternative complement pathways start off differently, but end in the same cascade of complement proteins that combine to form a membrane attack complex.

Leukocytes

Leukocytes are also known as white blood cells (WBCs). Leukocytes are a particular cell type and differ from other cells in the body because they exist throughout the whole body and are not particularly associated with a specific organ or specific tissue. The leukocytes are able to circulate around the entire body and interact with the potential invaders.

Chapter 1.2: Pattern recognition receptors (PRRs) and their history

Germ-line encoded pattern recognition receptors (PRRs) of innate immunity, which include Toll-like receptors (TLRs), RIG-I-like receptors (RLRs), NOD-like receptors (NLRs), Absent in melanoma 2 (AIM2) -like receptors, and C-type lectin receptors, are responsible for sensing the presence of infection and microorganisms (Gordon 2002, Palm and Medzhitov 2009, Takeuchi and Akira 2010, Brubaker, Bonham et al. 2015). They recognize conserved structures among all microbial species, which are called PAMPs. Recent research evidence indicates that PRRs identify endogenous molecules released from damaged cells as well, DAMPs (Beutler, Castonguay et al. 2006).

The very first innate immunity signaling receptor was discovered in 1996 from *Drosophila melanogaster*. It was demonstrated that the production of antimicrobial molecules depends on the activation of a signal cascade mediated by TLRs (Lemaitre, Nicolas et al. 1996). Subsequent studies identified a human homolog of Toll-like protein and demonstrated the first innate immune response that Toll-like proteins could sense bacterial lipopolysaccharide (LPS) and induce the host defense (Medzhitov, Preston-Hurlburt et al. 1997, Akira and Takeda 2004). After that, as the genome-sequencing technology improved rapidly, many genes associated with PRRs have been identified and classified into five groups as shown in Table 1.2.1 (Meylan, Tschopp et al. 2006).

Delving into the structural view of how innate immune receptors transduce the signals, there are studies that show that innate immune receptors act through forming higher-order complexes to transduce the ligand-binding signals. These higher-order supramolecular signaling assemblies are composed of intracellular adaptors, signaling enzymes and their substrates (Qiao and Wu 2015).

Table 1.2.1 pattern-recognition receptor families, adapted from (Meylan, Tschopp et al. 2006)

Family	Members	Shared domains	Receptor locations
TLR	1-10 in humans, 1-9 and 11-13 mice	leucine-rich repeat (LRR), Toll/interleukin-1 receptor (TIR)	Cell surface, endosomal compartments
CLR	Dectin-1, Dectin-2	C-type lectin	Cell surface
NLR	Nucleotide-binding oligomerization domain-containing protein 1 (NOD1, NLRC1), Nucleotide-binding oligomerization domain-containing protein 2 (NOD2, NLRC2), NOD-like receptor subfamily C (NLRC)3-5, NOD-like receptor subfamily P (NLRP)1-9 and 11-14, NLR family apoptosis inhibitory proteins (NAIP) 1, -2, -5, -6,	Nucleotide binding, LRR	Cytoplasm, plasma, and endosomal membrane associated
RLR	Retinoic acid-inducible gene-I (RIG-I), melanoma differentiation associated factor 5 (MDA5), laboratory of genetics and physiology 2 (LGP2)	DEXD/H helicase	Cytoplasm
ALR	AIM2, Gamma-interferon-inducible protein 16 (IFI16)	PYRIN, Hematopoietic expression, interferon-inducible nature, and nuclear localization (HIN)-200	Cytoplasm, nucleus (IFI16)

One of the main functions of PRRs is to identify the conserved components or patterns of foreign pathogens called PAMPs. PRRs also function to recognize the endogenous molecules (nuclear or cytosolic proteins) that are released from those damaged cells or exposed to its surface, which are

generally referred as DAMPs (Medzhitov and Janeway 2000, Medzhitov, Mackay et al. 2000, Bianchi 2007). PRRs can recognize both PAMPs and DAMPs and then lead to the activation of the innate immunity signaling pathways, which promotes the inflammatory response to eliminate the external microbes(Bianchi 2007). These PRRs are involved in sensing not only the outside invaders-pathogens but also those endogenous non-microbial “damage” and stress messages(Medzhitov and Janeway 2000). Both these two types of signals will lead to the activation of inflammatory signaling pathways such as nuclear factor-kappa-light-chain-enhancer of activated B cells (NF- κ B) and mitogen-activated protein kinases (MAPKs)(Seger and Krebs 1995).

RIG-I-like receptors

RIG-I-like receptors (RLRs) are a type of intra-cellular PRRs in innate immunity that serve as a frontline defense of the body in most tissues against viral infections.

Back in 1997, RLRs were first identified to be associated with promyelocytic leukemia cells. Subsequent studies found that a porcine homolog of the retinoic acid-inducible gene I played important roles in both porcine reproductive and respiratory syndrome((Zhang, Wang et al. 2000). More and more studies on RIG-I genes revealed they are related to the induction of type I interferon and are involved in essential processes such as RNA interference, and are found in almost all organisms from viruses, bacteria to mammals(Yoneyama, Kikuchi et al. 2004).

RIG-I belongs to the DExD/H box RNA helicases subfamily and plays important roles in sensing viral RNA. It is responsible for conducting the innate immune response(Loo and Gale 2011, Kell

and Gale 2015). Because of their ability to bind to duplex RNA without hydrolyzing ATP, RLRs, including MDA5, LGP2, and RIG-I itself, are also classified into the category of double-stranded RNA(dsRNA)-dependent ATPases(Luo, Kohlway et al. 2014). These three RLRs are similarly organized. They all have a central RNA helicase core which consists of two RecA-like helicase domains that help recognition of dsRNA while a pincer-shaped zinc-binding C-terminal domain(CTD) serves as a linker to connect with a helicase domain (also known as a repressor domain). It is important to note that RIG-I and MDA5 contain N-terminal tandem CARDs, while LGP2 lacks the N-terminal CARDs. (Figure1.2.1)The tandem CARDs are believed to transduce signals(Saito, Hirai et al. 2006, Cui, Eisenacher et al. 2008). However, LGP2 still retains the ability to bind the dsRNA with a quite high affinity and can regulate the signals conducted by RIG-I and MDA5 at the same time(Yoneyama, Kikuchi et al. 2005, Murali, Li et al. 2008, Pippig, Hellmuth et al. 2009). 23% amino acid similarity has been shown between the N-terminal CARD domains of MDA5 and RIG-I, while 35% has been shown for the central helicase domain(Yoneyama, Kikuchi et al. 2005). Subsequent studies on the function of MDA5 and RIG-I demonstrated that they both play roles as a defense against viruses but can detect different types of viruses. RIG-I can identify 5' triphosphate group (5'ppp) with a blunt dsRNA end or single-stranded RNA (ssRNA) hairpins that are mostly found in encephalomyocarditis virus, Theiler's virus and Mengovirus, while MDA5 is believed to detect long-duplex RNAs in the genome of dsRNA viruses(Yoneyama and Fujita 2009).

RIG-I like receptor (RLRs)

Figure 1.2.1 Domain structure of RLRs. The respective domains are highlighted in color. The tandem CARDS having two domains, CARD1 and CARD2 at the N-terminus for signal transduction; central DExD/H box RNA helicase domain (Hel-1, Hel-2i and Hel-2) with RNA-dependent ATP hydrolysis activities, and C-terminal domain (CTD) for RNA binding. Only RIG-I and MDA5 contains CARD domain.

Activated RIG-I and MDA5 can be engaged in inflammation, characterized by the pro-inflammatory interleukin-1 (IL-1) family of cytokines such as interleukin-1 β (interleukin-1 β) and interleukin-18 (IL-18). RLRs can signal NF- κ B-dependent pro-interleukin-1 β via interacting with Mitochondrial antiviral-signaling protein (MAVS) and two other members of the CARD family- CARD9 and B cell lymphoma/leukemia 10 (BCL-10) in response to virus infection by dsRNA (Hornung, Ellegast et al. 2006, Franchi, Eigenbrod et al. 2014). And for LGP2, there are studies that LGP2 can act as an inhibitor to regulate the signaling pathway mediated by RIG-I or MDA5 (Rothenfusser, Goutagny et al. 2005). Studies showed LGP2-deficient mice were resistant to lethal vesicular stomatitis virus infection which indicates that LGP2 needs either RIG-I or MDA5 to sense the viral RNA to conduct signals downstream in innate immunity (Venkataraman, Valdes et al. 2007, Ouda, Onomoto et al. 2011).

In innate immunity, RLRs are responsible for mediating the transcriptional upregulation of the type I interferons (IFNs). In general, RIG-I and MDA5 have quite similar domain architecture and

share a similar mechanism from the point of view of innate immunity. They both contain two N-terminal CARD domains that trigger MAVS nucleation via a homotypic CARD-CARD interaction and activate the antiviral signaling pathway(Wu, Peisley et al. 2014, Wu and Hur 2015).

Upon studying the mechanism of how RIG-I induces MAVS on immune response, a model of the signal activation has been proposed (Figure 1.2.2). In this proposal, the RNA binding domain of RIG-I, which is also called helicase-CTD, forms filamentous oligomers when encountering the viral dsRNA. The filamentous oligomers are surrounding dsRNA with approximately the length of dsRNA, and require the presence of ATP(Peisley, Wu et al. 2013). Subsequently, the filament will bring together 2CARDS and will induce the tetramerization of 2CARDS and connecting to the core filament with a 50 aa flexible linker(Peisley, Wu et al. 2013). RIG-I 2CARDS tetramer serves as a helical template and directly stimulates MAVS filament formation without any co-factor. A MAVS-CARD filament, which is anchored to the mitochondrial outer membrane by a ~400 aa linker with multiple tumor-necrosis factor-receptor-associated factor (TRAF) binding sites, can bring together and recruit the TRAF molecules through distinct binding motifs to conduct the downstream IFN- α/β signaling pathways(Hou, Sun et al. 2011, Takamatsu, Onoguchi et al. 2013).

Figure 1.2.2 Model of Signal Activation by Full-Length RIG-I and MAVS, adapted from (Wu, Peisley et al. 2014). Upon encountering viral dsRNA with 5' ppp and blunt end, the RNA binding domain (helicase-CTD) of RIG-I forms filamentous oligomers along the length of dsRNA in an ATP-dependent manner (Peisley et al., 2013). This filament formation brings together 2CARDs into proximity and induces tetramerization of 2CARD while being tethered to the core filament via a 50 aa flexible linker (Peisley et al., 2013). Tetramerization of 2CARD may also be induced by K63-linked polyubiquitin chains (K63-Ubn), which stabilizes the 2CARD tetramer by bridging between adjacent 2CARDs and wrapping around the tetramer (Peisley et al., 2014). The 2CARD tetramer resembles the lock-washer (helical ring) structure, which serves as a helical template that nucleates the MAVS-CARD filament. MAVS-CARD is anchored to the mitochondrial outer membrane via a ~400 aa long linker, which contains the TRAF binding sites. Filament formation of MAVS-CARD would bring together multiple TRAF binding sites into proximity, which appears to be important for efficient recruitment of TRAF molecules and activation of further downstream signaling pathway (Hou et al., 2011; Takamatsu et al., 2013)

Toll-like receptors

TLRs were among the first ones to be identified and are the best-characterized class of PRRs. TLRs are expressed in innate immune cells as well as other non-immune cells, and they can mainly be classified into two subfamilies-surface TLRs and intracellular TLRs. TLR1, TLR2, TLR5, TLR6, and TLR10 are located on the cell surface while TLR3, TLR7, TLR8, TLR9, TLR11, TLR12, and TLR13 are located in the endosome. TLR4 can signal both on the cell surface and in the endosome(Kawai and Akira 2010, Celhar, Magalhaes et al. 2012), and to activate two signaling pathways – one pathway is that TLR4 is activated by TIR domain containing adaptor protein (TIRAP) and MyD88, and the second pathway is activated by TIR-domain-containing adapter-inducing interferon- β (TRIF) and TRIF related adaptor molecule (TRIM). Recent study shows these two signaling pathways are induced sequentially and that the TRAM-TRIF pathway is operational from early endosomes following endocytosis of TLR4(Kagan, Su et al. 2008). TLR4 induces a TIRAP–MyD88-dependent signaling pathway from the cell surface that leads to downstream activation of MAPKs, NF- κ B activation, and induction of proinflammatory cytokines. Upon subsequent endocytosis, TLR4 triggers a TRAM–TRIF-dependent signaling pathway that results in the induction of IFN1. Trapping TLR4 at the surface after LPS stimulation using the dynamin inhibitor dynasore will reduce signaling through the TRIF-dependent pathway, but not inhibit the MyD88-dependent pathway(Kagan, Su et al. 2008).

Cell surface TLRs mainly contribute to recognizing microbial envelope molecules like lipids, some lipoproteins, and proteins. Endosomal TLRs can recognize endosomal PAMPs nucleic acid derived from viruses as well as endogenous nucleic acids, unmethylated CpG motifs and lead to production of cytokines and interferons type I(IFN1). For example, a defective mutation of TLR4

mouse showed resistance to LPS-induced sepsis which in result proved TLR4 could recognize bacterial LPS (Poltorak, He et al. 1998). Research from the database reveals totally thirteen members (10 in human and 12 in mouse), and they all together can recognize a broad spectrum of pathogens(Medzhitov 2001). (Figure 1.2.3) Endosomal TLRs are primarily expressed in dendritic cells, and their activation can induce the secretion of IFN-1 and lead to upregulation of IFN-inducible genes, to recruit helper T cells, and promote B-cell activation and antibody production in adaptative immunity(Trivedi and Greidinger 2009). For example, TLR4 at the plasma membrane induces the TIRAP-MyD88 pathway and get endocytosed into endosomes to induce the TRAM-TRIF pathway(Kagan, Su et al. 2008).

Figure 1.2.3 Receptors that detect infection. TLRs recognize microbes by binding to PAMPs. TLR3,7,8 bind to RNA viruses; TLR9 binds to DNA viruses; TLR 1,2,6 bind to lipoproteins (LP) and TLR2 binds to lipoteichoic acid (LTA) in Gram-positive bacteria; TLR4 binds to LPS and TLR5 binds to Flagellin in Gram-negative bacteria; TLR9 binds to the DNA in Gram-negative bacteria. TLR2 binds to Zymosan beta-Glycan and TLR9 binds to DNA in fungi; TLR 2,4 bind to glycosylphosphatidylinositol (GPI) anchors and TLR9 binds to DNA in protists,

TLRs are transmembrane proteins, and each TLR contains an ectodomain with LRRs that can help PAMP recognition. The horseshoe-like structure of the ectodomain can interact with potential PAMPs or DAMPs as a homo or heterodimer (Kawasaki and Kawai 2014).

TLRs are located in the membrane of cells and are believed function as dimers. Some TLRs are pre-dimerized in apo state. For example, plasma membrane, membranes of endosomes and lysosomes. So after TLRs successfully detect the invading pathogens, TLRs can dimerize first and undergo a conformational change to and recruit the downstream signaling molecules, such as myeloid differentiation primary response protein 88 (Myd88), IL-1R-associated kinases (IRAKs), TAK 1-binding protein 1/2(TAB1/TAB2), transforming growth factor- β -activated kinase (TAK1), and TRAF6 (Akira and Takeda 2004). MyD88 forms a supermolecular complex with the IRAK kinase, which is called Myddosome through Death Domain-Death Domain(DD-DD) interaction(Lin, Lo et al. 2010). Recent evidence suggests some surface TLRs exist as dimers. For example, two TLR4 proteins can loosely associate with each other to form a dimer can become more stable when binding to a microbial molecule(Latty, Sakai et al. 2018).

NOD-like receptors

NLRs are cytoplasmic proteins that were first discovered in plants, and they serve their function to recognize and fight against microbial and parasitic pathogens in innate immunity(Jones and Dangl 2006). In vertebrates, NLRs are believed to have originated in teleost fishes. Members of NLRs family share a similar domain architecture. NLRs have an N-terminal effector binding domain, an intermediary nucleotide-binding domain (NBD) that is responsible for nucleotide recruitment and self-oligomerization, and a C-terminal LRR domain that detects conserved pathogenic patterns to modulate NLR activity. Based on the N-terminal domain, NLRs can be further divided into several subfamilies. Two of the most prominent of these subfamilies are the

NLRC (C for CARD) and NLRP (P for PYD) subfamilies. The NLRC proteins may contain one or more CARD at their N-terminus, while the NLRP proteins have a PYD instead.

NLRPs, as the most characterized subfamily of NLRs, consist of a total of 14 distinct family members. Other subfamilies include NLRB (B for BIR) and NLRX (X for unknown). Typically, NLRX is supported and approved to be classified as an individual subfamily by Human Genome Organization Gene Nomenclature Committee (Ting, Lovering et al. 2008) and NLRX1 is one of the examples located in the mitochondria (Inohara and Nunez 2001). Different N-terminal domains contribute its different functions.

The NLRC subfamily contains a variety of proteins including NLRC1 (NOD1), NLRC2 (NOD2), NLRC3 (NOD3), NLRC4, NLRC5, and NLRX1. They are classified by their common characteristics: one or two CARD domains in the N-terminus, which are associated with recruitment of caspase1 or other kinases to mediate other molecules in innate immune signaling pathways (Kim, Shin et al. 2016) (Figure 1.2.4).

Subfamily	Gene	Structure
NLRA	CIITA
NLRB	NAIP
NLRC	NOD1,NLRC4
	NOD2
	NLRC3, NLRC5, NLRX1
NLRP	NLRP1
	NLRP2-9,11-14
	NLRP10

Figure1.2.4 The NLR subfamilies(AD, acidic transactivation domain; NACHT, predicted nucleoside-triphosphatase (NTPase) domain; BIR, baculovirus inhibitor of apoptosis repeat; X, unidentified; FIIND, function to find domain)

For protein members in NLRCs such as NLRC3, NLRC5, and NLRX1, they contain an undefined N-terminal domain rather than a typical CARD domain. They are still classified into the NLRC subfamily due to its polygenetic relationship in sequence(Proell, Riedl et al. 2008, Ting, Lovering et al. 2008).

Among all NLRs, NOD1 and NOD2 are two of the very first characterized NLRs that can signal via NF- κ B and MAPK following stimulation in an inflammasome independent manner and mainly contributed to first line defending against invading microbes in the intestine and maintain the homeostasis in the gut(Motta, Soares et al. 2015) . NOD1 and NOD2 undergo conformational changes and self-oligomerize. Both of them then recruit their common downstream adaptor, which is called Receptor-interacting serine/threonine-protein kinase 2- RIP2 (also referred to as RICK, RIPK2)(Ogura, Bonen et al. 2001, Caruso, Warner et al. 2014). For example, NOD2 will self-

oligomerize upon MDP binding and recruit RIP2 via CARD-CARD interactions. The activation of RIP-2 will then lead to the downstream activation of NF- κ B and MAPK.

NLRC4 is a typical member of the NLRC family with an N-terminal CARD domain, and is responsible for inflammasome formation and activating innate immune responses (Zhao, Yang et al. 2011). Other NLRCs, such as NLRC3 and NLRC5, are reported function as an adaptor in adaptive immunity while NLRCX1 regulates viral infection-related immune responses in mitochondria. NLRC3 has been proved to inhibit T-cell activation and TLR4-dependent TRAF6 (Schneider, Teufel et al. 2012). NLRC5 is a negative regulator for MHC class I and its inflammatory responses (Benko, Magalhaes et al. 2010). (Figure 1.2.5)

Figure 1.2.5 Overview of NLR mediated signalling pathways, adapted from (Motta, Soares et al. 2015). Overview of NOD-like receptor cellular pathways. The NLR family can be divided into four broad functional categories: transduction signaling, autophagy, transcriptional activation, and inflammasome assembly. While NOD1 and NOD2 are known to activate the NF- κ B pathway, other NLR receptors (e.g., NLRP2 and NLRP4) are known to be negative regulators of this pathway. Inflammasomes can be assembled by various NLR receptors and triggered by multiple stimuli. NOD2 is also known to recognize bacteria at the cell entry site and to recruit. Class II transactivator (CIITA) and NLRC5 have been described as transactivators of major histocompatibility complexes.

The NLRPs subfamily is typically characterized by its N-terminal PYD. Compared to members in NLRC receptors, the numbers of NLRPs and its homologs can be quite different and vary a lot between different species. With further study of NLRPs, more functions of NLRP receptors have been reported. NLRPs, such as NLRP1, NLRP3, and NLRP6, can trigger immune responses by a

homotypic Pyrin-Pyrin interaction (PYD-PYD) upon mediating ASC(Apoptosis-associated speck-like protein containing a CARD)-PYD to conduct inflammatory responses, which will lead to the activation of caspase1 to process the downstream release of IL-1 β and IL-18(Zhou, Yazdi et al. 2011, Levinsohn, Newman et al. 2012, Wang, Zhu et al. 2015).

As indicated in Figure 1.2.4, NLRP10 has a unique domain architecture and does not have an LRR domain. There are reports that NLRP10 performs as a negative regulator for inflammasome formation and is responsible for conducting the adaptive immune response(Eisenbarth, Williams et al. 2012). As for other members of NLRPs, NLRP2 has been reported as essential to inhibit NF- κ B activity for early embryonic development in mouse(Fontalba, Martinez-Taboada et al. 2007). NLRP4 can negatively regulate autophagy with the binding of beclin1(Jounai, Kobiyama et al. 2012). NLRP5 and NLRP7 are part of immune signal pathways related to mitochondrial and lipopeptide recognition, respectively(Khare, Dorfleutner et al. 2012, Tsuda, Tozaki-Saitoh et al. 2012). NLRP11 can perform important roles in attenuating TLR signaling pathways by degrading TRAF6 in the presence of ubiquitin and ligase(Wu, Su et al. 2017).

AIM2-like receptors

Absent in melanoma 2(AIM2) like receptors (ALRs) are a relatively new class of PRRs. ALRs were first identified as IFN-inducible genes in mice(Kingsmore, Snoddy et al. 1989) and only until recent years were characterized as a new class due to the finding that AIM2 was able to directly bind with double-stranded DNA and recruit downstream adaptors to conduct inflammatory responses in the cytoplasm by activating ASC that is similar in function and role as NLRs (Burckstummer, Baumann et al. 2009).

Since then, a few more members of ALRs have been identified sharing a Pyrin signaling domain and a DNA-binding hematopoietic expression, interferon-inducible nature, and nuclear localization (HIN) domain (Albrecht, Choubey et al. 2005, Jin, Perry et al. 2012, Gray, Winship et al. 2016). A single ALR has been annotated in cows while four (AIM2, PYHIN1, IFI16, and MNDA) in humans and thirteen in mice have been annotated (Gray, Winship et al. 2016). However, the phylogenetic study shows no orthologs among mammalian ALRs except AIM2 (Brunette, Young et al. 2012).

Chapter 1.3: Cryo Electron microscopy and helical reconstruction

Cryogenic electron microscopy

Cryogenic electron microscopy is also known as cryo-EM. The principle of the cryo-EM is that sample is taken images by a transmission electron microscopy under the cryogenic conditions (Fernandez-Leiro and Scheres 2016). It has been more than 40 years for scientists working together to build up the whole cryo-electron microscopy system, which includes the detectors for microscopy, automation system, software and other stuff that ensure to achieve higher resolution (Milne, Borgnia et al. 2013).

In transmission electron microscopy, the radiation is electrons that emitted by the filament under a high vacuum and get accelerated to the microscopy column at the accelerating voltages of 80-300KV in most cases. Then the scattered electrons are focused by the microscopy's

electromagnetic lenses after passing through the specimen. Comparing to the optical microscopy which adopts visible photons serve as the source of radiation, contributes the largest difference in its resolving power because the microscopy resolution is directly influenced by the wavelength of the radiation source of microscopy: the shorter the wavelength is, the higher the resolution can be attained. While the possible wavelength of visible light lies within the range of 4000-7000Å ($1\text{\AA}=10^{-10}\text{m}$) is significantly larger than a typical electron microscopy that wavelength of electron radiation sources approximately 0.02Å can be achieved at 300KV(Milne, Borgnia et al. 2013). In this way, the wavelength of the electron radiation is no longer a limitation for imaging biological macromolecules in electron microscopy. And the possibility of using electron microscopy to imaging biological structures have been successfully demonstrated by Joachim Frank in 1975 who analyzed the fuzzy 2D images of bacteriorhodopsin and determine the 3D structures into a resolution at around 7Å(Frank 1975).

Cryo-electron microscopy is one of the two approaches that scientist raised to solve this problem. By keeping the sample in a cryogenic temperature, the sample is put in a grid that plunged into ethane to get frozen specimens maintained liquid nitrogen or liquid helium temperatures. Four decades ago, Fernandez-Moran advocates that low temperature can slow down and even prevent the evaporation of the water in high vacuum which is good enough to preserve the hydrated state of the sample in electron microscopy. And measurements on the decay of diffraction intensities established that cryo-electron microscopy can reduce the effects of radiation damage caused by electrons at room temperature(Fernandez-Moran 1960). So in this way, water is no longer harmful to samples to get a high-resolution images. Another big progress of cryo-electron microscopy that have been achieved is that Jacques Dubochet successfully vitrified water in the preparation of

samples in the early 1980s(Dubochet, Adrian et al. 1988). According to Dubochet's report, biological samples can be cooled rapidly so that water molecules vitrified to form an amorphous solid and to protect the sample structure. And with the development of the devices for cryo-specimen handling and other progress in software, the modern cryo-electron microscopy has become a simple and powerful method to attain high resolution observations of biological samples and can use images from microscopy to get high resolution models(Fernandez-Leiro and Scheres 2016).

Helical reconstruction for cryo-EM

The very first biological structure solved by 3D electron microscopy reconstruction is the tail of T4 bacteriophage with helical symmetry(De Rosier and Klug 1968)(Figure 1.3.5). Comparing to the single particle reconstruction, the parameters in helical reconstruction that need to be determined is fewer than single particle reconstruction that every orientation for individual particle projection image needs to be determined. In helical reconstruction, projections images are repeating and have fixed orientations(He and Scheres 2017). Only with the determination of the parameters describing helical symmetry (diameter, rise, rotation angle) and the orientation of the image, the noise of the background can be efficiently reduced and average out particles.

Figure 1.3.1 The cryo-EM structure of the tail of T4 bacteriophage, adapted from (De Rosier and Klug 1968). (A) Micrographs of a tail of T4 bacteriophage by negative stain. (B) Diffraction pattern of the tail object in image A. (C) the helical model of T4 bacteriophage tail.

Compared to asymmetrical molecules, helical objects are much easier to apply 3D reconstruction as each single projection contains sufficient information (He and Scheres 2017). Specifically, the orientations of single particles for 3D reconstruction are all independent, while asymmetrical subunits in helical assembly are fixed and have relative fixed orientations (He and Scheres 2017). With suitable helical symmetry, signal to noise ratio is largely improved for helical reconstructions by averaging a huge number of asymmetrical subunits. In addition, helical reconstruction are exempted from preferred orientation problem by lying flat on the specimen support with a 360° range of views (Egelman 2015).

To achieve helical real-space reconstruction, it was introduced in 1997 to cut helical filaments into short segments by a method of Fourier-Bessel reconstructions (Beroukhim and Unwin 1997). In

2000, iterative helical real-space reconstruction (IHRSR) algorithm was developed for a better achievement of segmentations (Egelman 2000). By this approach, filaments are segmented by overlapping boxes at equal intervals with suitable width. In this way, segments are applied to iterative algorithms in single-particle analysis for 3D helical reconstructions (Egelman 2000). When representative 2D images from all segments are determined, good particles are selected and aligned for helical 3D reconstruction by back-projection for several iteration (He and Scheres 2017). During this process, the vital parameters of helical rise and twist are determined and iterated for the best pairs (He and Scheres 2017). In this way, 3D density can be obtained for refinement.

Chapter 1.4: Death domain containing proteins

The Death Domain superfamily is the largest and well-studied domain superfamily which acts as a critical component of the innate immune system by conducting apoptotic and inflammatory signals via assembling of oligomeric molecular complexes (McEntyre and Gibson 2004, Park, Lo et al. 2007, Valmiki and Ramos 2009).

The Death Domain superfamily is composed of four subfamilies, which are Death Domain(DD), PYD, CARD as well as the Death Effector Domain (DED) subfamilies(Hofmann and Bucher 1997, Kohl and Grutter 2004). The assembling of oligomeric signaling complexes is formed by homotypic interactions between domains from each subfamily and only interact within the same subfamilies with no cross interactions between different subfamilies(Kwon, Yoon et al. 2012). These proteins are most evolutionarily conserved in mammals, drosophila, *C. elegans* and other multicellular organisms(Park, Lo et al. 2007). According to human genome analysis, there are 32

death domains, 7 death effector domains, 28 CARD domains, and 19 PYD domain proteins in the human genome.

Members of the death domain family share a common structural motif consisting of six antiparallel α -helices and are arranged in an unusual Greek Key topology, despite having a quite low sequence conservation, which causes the variability in length and orientation of helices yielding a diverse interaction surface and specific binding affinity within each subfamily. Four structurally most conserved proteins ASC-PYRIN, FADD DED, Caspase9 CARD, and TNFR-1 DD are representatives of four subfamilies of DD superfamily (Figure 1.4.1)(Figure 1.4.2). Because of that, they can self-associate or bind to each other forming oligomers and large protein complexes in a special helical symmetry via homotypic interaction within the death domain family (Ferrao and Wu 2012).

Figure 1.4.1 Structure comparison of PYRIN, DED, CARD and DD domains, adapted from (Liepinsh, Barbals et al. 2003). All four proteins are of human origin. (a) PYRIN domain from ASC (PDB ID: 1UCP). The six helices are numbered. (b) Death effector domain from FADD (PDB ID: 2GF5). (c) Procaspase-9 prodomain (Casp-9 CARD, PDB ID: 4RHW). (d) TNFR-1 death domain (PDB ID: 1ICH). The views in the left and right panel differ by a 90° rotation of the molecules about a horizontal axis. The structures were oriented for best match of helix 5.

	helix 1	helix 2	helix 3	
ASC PYRIN	MGRARDAILDALENL	-TAEELKKFKLK	LLSVPLREGYGRIPRGALLSM-	47
FADD DED	MDPFLVLLHSVSSSL	-SSSELTTELKGLCLGR	-----VGKRRKLERVQ	40
CASP-9 CARD	SMD EADRRLLRRCLRLVEEL	QVDQLWDVLLSREL	-----FRPHMIEDIQ	45
TNFR DD	HKPQSLDTDDPATLYAVENV	P-PLRWKEEFVKRLG	-----LSDHEIDRL	360
	helix 4	helix 5	helix 6	
ASC PYRIN	-----DALDLTDKLVSFYL	-ETYGAELTANVLRDM	GLOEMAGQLQAATHQ	91
FADD DED	-----SGLDLFSMLLEQNDLEPGHTE	LLRELLASLRR	HDLLRRVDDFE	93
CASP-9 CARD	RAGSGSRRDQARQLIIDLETR	---GSOALPLFISCL	DTGQDMLASFIRTNRQAG	96
TNFR DD	LQNGRCLREAOYSMLATWRRRT	TPREATLELGRVLRDM	MDLLGCLEDIEEALC	413

Figure 1.4.2 Structure-based amino acid sequence alignment of representatives of the four different subfamilies of the death-domain superfamily listed in Figure 1.4.1, adapted from (Liepinsh, Barbals et al. 2003). Boxes delineate the helix boundaries. Yellow bands identify structurally conserved residues contributing to the core of the domains.

These death domain proteins can interact with each other by three different types of interactions via the three different types of interfaces within their subfamily(Ferrao and Wu 2012). Their interfaces are classified into three categories as well. The single DD-DD interaction was first found in a solved crystal structure of Apaf-1 CARD that interacts with Caspase-9 CARD domain. This asymmetrical dimer is an asymmetrical structure, and the interaction between Apaf-1 and Caspase-9 is known as a typical type I interaction (via interfaces Ia/Ib)(Qin, Srinivasula et al. 1999) (Figure 1.4.3). Non-conserved residues on each CARD form two highly complementary charged surfaces. In this solved crystallography structure, helices H1 and H4 of Caspase-9 together form a positive charged type Ia interface that can interact with the other negative charged Ib surface formed by helices H2 and H3 in Apaf-1(Weber and Vincenz 2001). The stability of Apaf-1 and Caspase-9 heterodimer is mainly contributed by the interaction residues pair D27/E40 of Apaf-1 and R13/R52/R56 of Caspase-9(Ferrao and Wu 2012).

Figure 1.4.3 A Death Domain (DD) mediates three interaction types through six interaction patches, adapted from (Kersse, Verspurten et al. 2011). The three asymmetric interaction types: type I (a), type II (b) and type III (c) derived from the PIDDosome (PDB ID: 2OF5). (d) The six possible interaction patches on the DD of TNFR-1. The composite model is rotated in four stages of 90° around the Y-axis.

Shortly after that, scientists also observed the type II interaction in the crystallographic structure of P53-induced protein with a death domain (PIDD) homodimer. For the type II interaction (IIa/IIb), it was first found that Pelle and Tube death domain homodimer are totally different from the interaction observed in the previous Apaf-1/Caspase-9 crystallography structure (Ferraro and Wu 2012) (Figure 1.4.3c). The type IIa interface is located at the Pelle side death domain, formed by the C-terminal of helix 4 as well as the loop between helices 4 and 5. While the type IIb interface of Tube is composed of a groove that is contributed by the corners of helices 1 and helices 2 on one side and helices 6 on the other side (Xiao, Towb et al. 1999). This type of interaction has been further confirmed when the PIDDosome structure was solved (Park, Logette et al. 2007).

The Type III interaction was also first observed in the crystallographic structure of the PIDDosome that includes all three types of interactions, showing how these binary interactions form an oligomeric DD complex (Figure 1.4.4A). In the solved PIDDosome structure, a total of five PIDD DDs and seven RIP-associated ICH-1/CED-3 homologous protein with a death domain (RAIDD) DDs assemble together to form a ternary structure complex (Figure 1.4.4B). The last layer of the PIDDosomes is composed of five PIDD DDs, the middle layer is formed by five RAIDD DDs, with a top layer containing another two RAIDD DDs. The newly found type III interaction is formed by the IIIa interface of helices 3 from the first death domain and two loops- the first loop between helices 2 to 3 and the other loop between helices 3 and 4 on the second death domain (Park, Logette et al. 2007). In this solved PIDDosome structure, all three types of interactions have been observed. Type I and type III interaction exist between PIDD: PIDD, RAIDD: PIDD, and RAIDD: RAIDD. Type II interaction is observed between RAIDD: PIDD and RAIDD: RAIDD (Ferraro and Wu 2012).

Figure 1.4.4 Structure of PIDDosome consisting of PIDD and RAIDD DDs. (A) Planar schematic of PIDDosomes by three different interfaces. (B) Crystal structure of PIDDosome colored as in (A).

Just like the PIDDosome oligomeric complex, the type I, type II, type III interactions among all death domains remain highly conserved in forming all of its oligomeric signalling complexes, which is the mechanism of how DDs to oligomerize with macromolecules and to conduct signalling(Ferrao and Wu 2012). Although DDs structures are conserved, there are still a lot of variations in residues because of its low sequence conservation, and this contributes to versatile interaction surfaces and all kinds of different types of interactions(Ferrao and Wu 2012).

MyD88 and IRAK2/4 complex (Myddosome) formation by death domain interaction

DD hetero-oligomers have been found between Mdy88 and IRAK DDs forming a four-layer complex and the surfaces determine recognition through both networks of intermolecular hydrogen bonds and van der Waals interaction(Lin, Lo et al. 2010).

In the TLR signaling pathway, TIR domains of the TLR receptors recognize PAMPs and initiate the immune response by recruiting downstream TIR-containing proteins such as Myd88, TIRAP, and TRIF (Yamamoto, Sato et al. 2003).

Myd88 is composed of an N-terminal DD, a short intermediate domain (ID) , and a C-terminal TIR domain(Yamamoto, Sato et al. 2003). Myd88 interacts with Interleukin-1 receptor-associated kinase (IRAKs) by its death domain (DD). IRAKs are a protein family containing IRAK1, IRAK2, IRAK3 and IRAK4, which are characterized mainly by containing a death domain and a Ser/Thr kinase or kinase-like domain(Lin, Lo et al. 2010).

The solved crystallographic structure with a resolution up to 3.4Å showed that Myd88-IRAK4-IRAK2 DD is a four-layer complex called the Myddosome, which is formed by a total of fourteen death domains forming a filamentous structure with about 110Å height, 70 Å diameter, rotation angle of $98 \pm 2^\circ$ between consecutive DDs and helical axis of around 6Å. The last two layers on the bottom are six Myd88 DDs followed by four IRAK4 DDs as the middle layer and with four IRAK2 in the top. The Myd88 and IRAK4 DDs first form a binary Myddosome complex and then further recruit IRAK2 to form the ternary Myddosome complex. This forms a hetero filamentous oligomer with a single-stranded left-handed helical structure (Lin, Lo et al. 2010) (Figure 1.4.5B).

Although Myd88, IRAK4 and IRAK2 DDs contain a conserved DD motif with six α -helices, they still have some different structural features due to their low sequence similarity. Myd88 tends to have short $\alpha 3$ helices with an extraordinarily $\alpha 6$ from residue 99 to 117 while IRAK4 and IRAK2 do not (Lin, Lo et al. 2010). As for the loop regions, Myd88 tends to have the longest loops connecting $\alpha 1$ $\alpha 2$ helices and $\alpha 4$ $\alpha 5$ helices. At the same time, Myd88 also has the shortest loop lying between $\alpha 3$ and $\alpha 4$, while IRAK4 has the longest loop between $\alpha 2$ and $\alpha 3$ (Lin, Lo et al. 2010). The three types of interactions can also be observed in the Myddosome complex. Type I and type II mostly mediate interactions between the layers, whereas type III represents contacts between adjacent DD molecules in the helical spiral (Lin, Lo et al. 2010). Superposition of the three type I interaction pairs, MyD88–MyD88, MyD88–IRAK4 and IRAK4–IRAK2, showed that they are highly similar to each other (Lin, Lo et al. 2010) (Figure 1.4.5). The type I interaction is formed by helices 2 and helices 3 on one DD and helices 1 and helices 4 region on the other DDs (Lin, Lo et al. 2010). The type II interaction is formed from two opposite edges of the DDs, centered around the helices 4 and helices 5 loop on one DD and around the helices 1 and helices 2

loop on the other DDs (Lin, Lo et al. 2010). The type III interaction is formed between helices 3 of one DD with the edge of the second DD composed of the helices 1-2 loop and helices 3-4 loops (Lin, Lo et al. 2010). The type I and type II interactions that exists in Myddosome are more extensive in buried surface areas than the type III interaction. Based on Myddosome structural analysis of surface charge and shape complementarity, either the top or the bottom surface of Myd88 layers have better shape complementarity compared to IRAK death domains. The top surface of IRAK4 has good charge complementarity and shape complementarity with the bottom surface of IRAK2, resulting in specific recruitment of IRAK2. Because of the incompatibility between the top and bottom surfaces of IRAK2, only one layer of IRAK2 can assemble into the helical oligomer (Ferraro and Wu 2012).

Figure 1.4.5 Structure of Myddosome, formed by DDs from Myd88, IRAK2 and IRAK4. (A) Planar schematic of the Myddosome ternary complex. The Myddosome is a single stranded left-handed helical oligomer, consisting of six Myd88 DDs, four IRAK4 DDs and four IRAK2 DDs. Their interaction is supported by three interaction interfaces. (B) The crystal structure of Myddosome, coloured as in (A).

RIG-I and MAVS interaction with homotypic CARD-CARD interaction

The most recent report finds that on the top of RIG-I 2CARDs, the MAVS CARD forms a hetero-oligomeric complex first, then MAVS CARD follows it as a template that further extends to form homo-filament-oligomers in a certain helical symmetry(Wu, Peisley et al. 2014).

RIG-I and MDA5 are both from a conserved family of immune receptors called RIG-I like receptors that detect viral dsRNA at the early stage of infection. RIG-I like receptors are important in activating downstream signaling pathways to induce immune responses by activating the type I IFN genes and releasing IFNs (Yoneyama, Kikuchi et al. 2005). With the detection of the viral dsRNA, dsRNA will interrupt the interaction between the CARD domain of either RIG-I or MDA5 and its helicase domain. Subsequently the central DExD/H motif helicase domain and CTD of RIG-I (or MDA5), which together function as an RNA recognition unit, interact with the viral RNA in the same binding site to become activated. The activated RIG-I tandem CARD domain forms the hetero-oligomer complexes as a tetramer which then will bring the MAVS CARD into proximity and follow it as a template to form a filamentous structure(Wu and Hur 2015)(Figure 1.4.6).

Ia/b, IIa/b, IIIa/b, forming both intermolecular interactions (Ia: Ib, IIIa: IIIb) and intramolecular interactions (IIa: IIb) (Figure 1.4.6).

The MAVS-CARD follows the RIG-I hetero oligomer complex as the template, and then forms a filamentous oligomeric complex, with a helical symmetry quite similar to the RIG-I 2CARD tetrameric structure. In the solved structure of the MAVS-CARD filament, the MAVS-CARD is in the form of a helical symmetry of an axial rise of 5.13 Å and a rotation angle as 101.1 °(Wu, Peisley et al. 2014). Each MAVS-CARD interacts with six neighboring subunits to form homotypic CARD-CARD interactions with six distinct interfaces (Ia/b, IIa/b, IIIa/b) from different MAVS-CARDS. Thus, the MAVS-CARD filament is best described as a single-stranded left-handed helix, in which IIIa: IIIb interface constitutes the intra-strand interaction, whereas Ia: Ib and IIa: IIb interfaces form the inter-strand interactions.

The solved crystallographic structure of the RIG-I tandem CARD and MAVS-CARD oligomer is reported to be a single-stranded left-handed helical tetramer, which gives us a better understanding of how the transient interaction is important for signal transduction (Wu, Peisley et al. 2014) (Figure 1.4.6). It showed that MAVS-CARD is stacked on top of the second RIG-I CARD by IIa: IIb interactions, and four copies of RIG-I: MAVS-CARD fusion proteins form a helical tetramer, which is aligned with the parameters of RIG-I CARD structure (Wu, Peisley et al. 2014). Comparison of the MAVS-CARD:MAVS-CARD homotypic interaction and the interaction between RIG-I 2CARD and MAVS-CARD interactions shows that the residues located in the interfaces of Ia, IIa, IIIa of MAVS-CARD are the very likely the same residues that interact with Ib, IIb, IIIb interfaces of MAVS-CARD and RIG-I 2nd CARD. Sequence alignment demonstrated

that the residues in Ib, IIb, IIIb of MAVS CARD and RIG-I 2nd CARD are dissimilar, which suggests that the plasticity of MAVS CARD surface allows to interact with different surface residues (Wu, Peisley et al. 2014)

NOD1/2 and RIP2 induced signalling pathway

Nucleotide-binding oligomerization domain-containing protein 1 (NOD1) is the first member from the NLR family that has been identified by sequence homology searching, followed by Nucleotide-binding oligomerization domain-containing protein 2 (NOD2) which is closely related to NOD1 with an additional N-terminal caspase activation and recruitment domain (CARD) domain (Ogura, Bonen et al. 2001). Both NOD1 and NOD2 are composed of one or two CARDs at the N terminus, while followed by a typical NOD, and multiple leucine-rich repeats (LRRs). Both NOD1 and NOD2 are intracellular pattern recognition receptors domain, and there are several reports showing that NOD1 and NOD2 play a role in activating the NF- κ B family of transcriptional regulators in response to peptidoglycan (PGN) fragments (Motta, Soares et al. 2015). NOD1 and NOD2 are widely expressed in the intestine to maintain gut homeostasis to sense and fight bacterial infection (Kim, Shin et al. 2016). NOD1 can sense γ -D-glutamyl-meso-diaminopimelic acid (iE-DAP), which is an important molecule that mostly is seen on the surface of all Gram-negative bacteria and even some Gram-positive bacteria. NOD2 is reported to sense muramyl dipeptide (MDP), which is also widely distributed among Gram-positive and Gram-negative bacteria. And with all these respective stimuli, NOD1 and NOD2 transduce signaling downstream and become activated by self-oligomerization and recruiting their common downstream adaptor called RIP2 (Caruso, Warner et al. 2014).

The mechanism of how PGN fragments activate NOD1/2 is still not very clear, but several possible theories have been reported. One study suggests endo-lysosomes are the entry sites for a variety of pathogens, and therefore the sites at which the immune system first senses their presence. Internalized pathogens also activate NOD1 and NOD2 in the cytosol(Nakamura, Lill et al. 2014). With the existence of surface plasmon resonance (SPR) data, the direct interactions between MDP and NOD2, Tri-DAP and NOD1 are recognized(Laroui, Yan et al. 2011).

A previously solved crystallographic structure of ubiquitin-bound NOD1-CARD showed that NOD1-CARD consists of a classic CARD domain motif. NOD1-CARD has six α -helices around a hydrophobic core. And core residues are highly conserved in NOD2 Tandem CARD(Ver Heul, Gakhar et al. 2014). In the solved NOD1-CARD structure, helix2 to helix5 form an antiparallel bundle, while helices 6 and 1 are positioned on the top of helices 5 and 4(Ver Heul, Gakhar et al. 2014). However, the tandem CARD structure of NOD2 remains unknown, though NOD2 tandem CARD is believed to be highly structurally similar to NOD1-CARD.

With the solved crystallographic structure of full length NOD1/NOD2 without CARD, scientists have looked into the structure to have a better understanding of the NOD1/NOD2 mechanism. NOD1/NOD2 share a similar domain architecture. They both have a NOD domain consisting of a nucleotide-binding domain (NBD), a winged-helix domain (WHD), and helix domains (HD1 and HD2) (Figure 1.4.7). The inactivated NOD1/NOD2 is stabilized by ADP-mediated interaction of NBD and WH. LRRs occlude these domains to keep NLR proteins in a monomeric state. Bound ATP is the key to stabilizing the entire NACHT domain in a closed and inactivated form of NOD1/NOD2. LRRs is assigned to detect the invading pathogens, HD2 mediates the conformational change of the NBD, WH as well HD1 together to allow the ADP-ATP exchange, and consequently cause the self-oligomerization and mediate the downstream signalling. The

LRRs allow the NOD2 to lock the dormant form of NOD2 by linking the NBD and HD2 instead of leaving the NBD exposed alone (Maekawa, Ohto et al. 2016).

Figure 1.4.7 Illustrations of protein domain structure of human NOD1, NOD2, and RIP2. The domains are highlighted in colors.

In the presence of the stimulus, active NOD1/NOD2 will activate receptor-interacting serine/threonine-protein kinase2 (RIP2, RICK or RIPK2). Active RIP2 will recruit nearby I κ B kinase (IKK) complexes, an essential process to recruit and activate the serine/threonine kinase TAK1 (Caruso, Warner et al. 2014). TAK1 is believed to be a prerequisite for the activation of the IKK complex as well as to induce the MAPK pathway. The IKK complex will mediate the phosphorylation of I κ B α , which is the inhibitor of NF- κ B, and cause the polyubiquitination and activation of this inhibitor. Thus, the co-localization of the IKK complex and TAK1 leads to the phosphorylation of I κ B β and the degradation of I κ B α at the same time to allow the nuclear translocation of NF- κ B, leading to downstream expression of the target gene (Caruso, Warner et

al. 2014).

The activation of RIP2 by NOD1/NOD2 as the common downstream adaptor has been studied for a long time. However, how NOD1/NOD2 activates RIP2 still remains unknown. Recently a report was claiming that the CARD domain from NOD1/NOD2 and RIP2 are the key for conducting this signalling. When stimulated, NOD1/NOD2 becomes activated and can oligomerize and recruit their common downstream adaptor RIP2 by homotypic CARD-CARD interactions, which is a key protein for immediately downstream signalling pathway. This was proved that NF- κ B could be activated by overexpression of the NOD1/NOD2 and RIP2 proteins and their CARD domain constructs. After the deletion of intrinsic RIP2 in HEK293T cells, NOD1/NOD2 full length, as well as NOD1/NOD2 CARD domain alone, lost their function to activate NF- κ B (Gong, Long et al. 2018). RIP2 CARD is reported to be able to self-oligomerize and to form a filamentous structure to activate downstream signalling. RIP2 CARD filaments was a left-handed helical filament. The helical rise is 4.936Å, and the rotation angle is -101.4°. The RIP2 CARD filament is stabilized by all three types of interactions via three types of interfaces, which are type I, type II, and type III interfaces. And the key residues involved in three types (Ia/b, IIa/b, IIIa/b) of interactions are shown in Figure 1.4.8.

Reports suggest that NOD1/2 CARD alone can activate the oligomerization of seed monomer RIP2 CARD to form a helical oligomeric structure while oligomerized RIP2 CARD cannot activate the NOD1/2 CARD seed to oligomerize, suggesting that both NOD1/NOD2 activate RIP2 in a top-down interaction mode (Gong, Long et al. 2018).

Figure 1.4.8 The details of Ia:Ib, IIa:IIb, and IIIa:IIIb interactions in RIP2–CARD filament, adapted from(Gong, Long et al. 2018)

Transcriptional factors of cytokines and chemokines get activated by NF- κ B to release TNF- α , IL-1 β and IL-18 from cells to defend them from pathogens(Caruso, Warner et al. 2014). But there are also reports showing that NOD1/NOD2 senses PGN fragments and initiates autophagy, which does not require the participation of RIP2 protein or the activation of NF- κ B. Instead, NOD1/NOD2 interacts with ATG16L1 at the plasma membrane and initiates autophagy of bacteria(Travassos, Carneiro et al. 2010). And there are studies which demonstrated that NOD1/NOD2 mediates ER-stress-induced inflammation by participating in the production of pro-inflammatory cytokine IL-6 in mouse and human cells as well, which shows a novel link with innate immunity and ER stress-induced inflammation(Keestra-Gounder, Byndloss et al. 2016).

In the innate immune system, invading bacteria can be recognized by multiple PRRs, so a deficiency of NOD1/NOD2 may only have a modest effect on pathogen clearance. So when TLR signaling is absent or reduced, the activity of NOD1/NOD2 may become important(Kim, Park et al. 2008, Caruso, Warner et al. 2014). Recently, the theory of modulating NOD1/NOD2 signaling with small molecules has been popular, and a lot of efforts have been made to help pathogen clearance and modulate inflammation that potentially can be applied to diseases such as Crohn's

disease (CD) and early-onset disease (EOS) (Caruso, Warner et al. 2014). These molecules have potential to treat the diseases associated with deregulated NOD1/NOD2 signaling such as asthma, arthritis, leprosy, and periodontitis (Caruso, Warner et al. 2014).

Higher order assemblies in signalling cascades

Higher order signaling complexes, often referred as supramolecular organizing centers (SMOCs), is a universal feature and plays important roles in innate immunity (Qiao and Wu 2015). The classical view of signal transduction - involving a serial reaction in which ligands induce conformational changes in receptors followed by the activation of enzymes and the generation of second messengers. The concept for PRR SMOCs supports the formation of higher order signaling complexes as the new mode of signal transduction and amplification.

In Myddosome, a hierarchy in formation of the helical signaling tower was found and this specific helical assemblies were proved to be the standard mode for DD superfamily assemblies. As previously mentioned, IRAK2 DD can interact with the MyD88-IRAK4 complex but not interact with either MyD88 or IRAK4 alone (Lin, Lo et al. 2010). This indicates a new concept that IRAK4 recruitment to MyD88 is an event upstream to IRAK2 recruitment in the TLR/IL-1R signaling pathways. In addition, MyD88 DD can form oligomers under a high protein concentration while IRAK4 DD is a monomer in solution. This indicates the weak MyD88 oligomer would be stabilized first and acts as an initial platform for IRAK4 recruitment and then IRAK1 and IRAK2 recruitment (Figure 1.4.9A). In the Myddosome structure, IRAK4 DDs and IRAK2 DDs pack closely against each other. The C-terminus of IRAK4 DD points towards the IRAK2 DD layer,

while the C-terminus of IRAK2 DD points towards the IRAK4 DD layer. This arrangement would then bring the kinase domains of IRAK4 to the proximity of IRAK1 or IRAK2 to allow efficient phosphorylation by IRAK4 (Fig.1.4.9B).

Figure 1.4.9 Model of sequential assembly in TLR/IL1-R signalling, adapted from (Lin, Lo et al. 2010). (A) On ligand binding, TLR/ IL-1Rs, including the cytoplasmic TIR domains, are dimerized or oligomerized. This results in the recruitment of other TIR containing adaptor proteins. In this case, MyD88 is recruited to the receptor complex and the death domain is oligomerized. In the presence of IRAK4, the death domain of IRAK4 (I41) can be recruited to the oligomerized MyD88 DDs through the three interfaces (M3 , M4 , M6) and quickly forms the binary Myddosome complex. Downstream kinases (IRAK2 in this case, but IRAK1 as well) can then be recruited in a similar fashion and signalling is triggered. (B) A model of the TLR signalling complex that recruits the MyD88–IRAK4–IRAK2 complex with proteins drawn to scale. TLRs, cyan and green (PDB code 3FXI for the extracellular domain of TLR4 and PDB code 2J67 for the TIR domain of TLR10). MD2, yellow and magenta (PDB code 3FXI in complex TLR4). Orange, MyD88–IRAK4–IRAK2 complex. Red, IRAK4 kinase domain (PDB code 2NRU). Blue, IRAK2 kinase domain using that of IRAK4.

Another example in innate immunity to generate SMOCs to induce the oligomerization of CARDs is the NLR family apoptosis inhibitory protein 2 (NAIP2) bind conserved bacterial ligands, such

as the bacterial rod protein PrgJ, and recruit NLRC4 as the inflammasome adapter to activate innate immunity. In NAIP2-NLRC4 inflammasome, the sensing part is NAIP and the assembling part is NLRC4. NAIP2 binds to the molecular pattern of bacteria and become active then recruits NLRC4 to assemble into a multi-disk higher-order oligomer structure to serve as an inflammatory adapter to activate caspase1 filament assembling(Zhang, Chen et al. 2015).Here, NAIP2 initiate NLRC4 polymerization by pushing the WHD of NLRC4 for conformational changes, overcoming the NLRC4 auto-inhibition, then the newly activated NLRC4 triggers activation of another NLRC4 molecule by using its newly exposed nucleating surface to repeat recruitment until a complete disk is formed(Figure 1.4.10). The disk provides a platform of NLRC4-CARD cluster to induce oligomerization and recruit the caspase-1 to form filaments through CARD-CARD interactions.

Figure 1.4.10 Schematic diagram for mechanism of PrgJ-NAIP2–nucleated polymerization of NLRC4, followed by caspase-1 activation, adapted from (Zhang, Chen et al. 2015). NAIP resting state is similar to the NLRC4 inactive conformation. Once infected, a NAIP recognizes its specific bacterial ligand(PrgJ). The specific ligand drives the NAIP into the open, activated conformation. Then the ligand-bound NAIP uses its nucleating surface to interact with the adapter NLRC4. The interaction forces the WHD and its linked C-terminal region to change into the activated conformation. The exposed nucleating surface to repeat recruitment and activation of additional NLRC4 molecules, until a complete disk is formed or until the NLRC4 concentration falls below the dissociation constant of the interaction. NLRC4 clustering induces oligomerization of the CARD of NLRC4, enabling the recruitment of caspase-1 through CARD-CARD interactions and triggering caspase-1 activation.

There are also studies showing that RLR signaling adaptor MAVS forming helical filaments to activate the interferon pathway(Hou, Sun et al. 2011, Wu, Peisley et al. 2014) and BCL-10 assembling into similar helical filaments to activate mucosa associated lymphoid tissue lymphoma translocation protein 1(MALT1) for NF- κ B signaling downstream of several innate and adaptive immune receptors, have confirmed that death domains and their related domains might be able to form large helical filamentous signaling complexes. As previous mentioned, RIG-I, MAVS, and RIP2 all form similar filamentous structures as Myddosome. They all adopted helical symmetry and form filaments although their monomer orientations and interfaces, and helical parameters are different. A sensor protein (receptor) in a pathway get activated first in the presence of ligands subsequently oligomerizes. The oligomerized sensor then provides the platform to nucleate a downstream protein to polymerize into filaments. And this platform may ensure a small number of sensor proteins (receptor) can activate various downstream signaling proteins.

These findings indicate oligomerization may be a general principle in promoting the assembly of high order SMOCs upon stimulation to mediate innate immune signal transduction and prove that to generate SMOCs and the high order assemblies are of great significance of studying the biology of signaling cascades in innate immunity.

Chapter 1.5: Computational simulation and structure prediction methods

Computational biology involves the development of computational technology and application of data-analytical and theoretical methods, as well as adopting mathematical models and

computational simulation techniques to study biology. It is broadly defined and includes foundations in molecular biology, structural biology, genetics, genomics, biophysics, and statistics. As such, Local optimization searching tools/algorithms such as Rosetta, GalaxyWEB, I-TASSER have been widely used in structure prediction(Simons, Kooperberg et al. 1997, Zhang 2008, Ko, Park et al. 2012, Yang, Yan et al. 2015).

Rosetta de novo structure-prediction

Rosetta has a lot of applications, including structure prediction, protein design, docking, scoring and relax analysis. Now Rosetta is a reputable molecular structure prediction software that has been employed successfully in numerous protein design and other applications(Bowman and Pande 2009). Rosetta 3 was released as the cleaned-up version code from the early FORTRAN to C++, and Rosetta 3 allowed the extension of Rosetta into PyRosetta(Chaudhury, Lyskov et al. 2010), FoldIt, RosettaScripts(Fleishman, Leaver-Fay et al. 2011), and ROSIE(Moretti, Lyskov et al. 2018). Among them, PyRosetta is a package containing the Python-bindings of the Rosetta 3 and RosettaScripts is an XML-like language for specifying modeling tasks in the Rosetta framework providing access to protocol-level functionalities such as rigid-body docking and sequence redesign. Foldit is an online puzzle game in protein folding as part of the experimental research of Rosetta, and ROSIE is a web front-end for Rosetta software suite that can fully utilize all functionalities in Rosetta 3.

In Rosetta, the energy of a biomolecule is calculated by all atomic interactions in a globular protein made of L-amino acids in Rosetta all-atom force-field (RAAFF). Rosetta energies are expressed in a generic unit, called the Rosetta Energy Unit (REU). Score functions in Rosetta are weighted

sums of energy terms under a certain force field, some of which represent physical forces like electrostatics as well as van der Waals' interactions among different atoms, while others represent statistically created terms, such as the probability of discovering certain possible torsion angles in Ramachandran space(Leaver-Fay, O'Meara et al. 2013, O'Meara, Leaver-Fay et al. 2015).

The Rosetta scoring function 2015(*REF15*) calculates the accurate interactions between polar atoms that have large impact to predict and design the macromolecular structures. In order to model these polar atom interactions, hydrogen bonds(H-bonds) and H-bond networking, which obviously play an important role in stabilizing polar interactions like helping the molecules make conformations suitable for binding and catalysis, have been put into one of the procedures in Rosetta scoring function combined with a distance-dependent Coulomb model of electrostatics(Kortemme, Morozov et al. 2003).

The Rosetta scoring function 2015 (*REF15*) will output a table file containing all the possible energy factors that might contribute to the atomic interactions(Alford, Leaver-Fay et al. 2017). The total energy, called ΔE_{total} , is computed as a linear combination of energy terms E_i that are calculated as functions of geometric degrees of freedom (Θ) and chemical identities (aa) and scaled by a weight on each term (w_i), as shown in below:

$$\Delta E_{\text{total}} = \sum_i w_i E_i(\Theta_i, \text{aa}_i)$$

The Rosetta energy function includes energies of interactions between nonbonded atoms pairs (van der Waals interactions, electrostatic interactions between fully and partially charged atoms, isotropic solvation energy), empirical potentials to model hydrogen and disulfide bonds, statistical potentials(backbone ϕ and ψ angles, likelihood of placing a specific amino acid side chain given

an existing ϕ , ψ backbone conformation) to describe backbone and side-chain torsional preferences, and terms important for native structural feature recapitulation(Alford, Leaver-Fay et al. 2017).

The energy terms and weights used are summarized in Table 1.5.1.

Table 1.5.1 Summary of terms in the REF15 energy function for proteins.

Term	Description	Weight	Units
fa_atr	Attractive energy between two atoms on different residues separated by distance, d	1.0	kcal/mol
fa_rep	Repulsive energy between two atoms on different residues separated by distance, d	0.55	kcal/mol
fa_intra_rep	Repulsive energy between two atoms on the same residue, separated by distance, d	0.005	kcal/mol
fa_sol	Gaussian exclusion implicit solvation energy between protein atoms in different residues	1.0	kcal/mol
lk_ball_wtd	Orientation-dependent solvation of polar atoms assuming ideal water geometry	1.0	kcal/mol
fa_intra_sol	Gaussian exclusion implicit solvation energy between protein atoms in the same residue	1.0	kcal/mol

fa_elec	Energy of interaction between two non-bonded charged atoms separated by distance, d	1.0	kcal/mol
hbond_lr_bb	Energy of short-range hydrogen bonds	1.0	kcal/mol
hbond_sr_bb	Energy of long range hydrogen bonds	1.0	kcal/mol
hbond_bb_sc	Energy of backbone-side chain hydrogen bonds	1.0	kcal/mol
hbond_sc	Energy of side chain to side chain hydrogen bonds	1.0	kcal/mol
dslf_fa13	Energy of disulfide bridges	1.25	kcal/mol
rama_prepro	Probability of backbone ϕ, ψ angles given amino acid type	0.45 kcal/mol/kT	kT
p_aa_pp	Probability of amino acid identity given backbone ϕ, ψ angles	0.4 kcal/mol/kT	kT
fa_dun	Probability that a chosen rotamer is native-like given backbone ϕ, ψ angles	0.7 kcal/mol/kT	kT

omega	Backbone-dependent penalty for cis ω dihedrals that deviate from 0° and trans ω dihedrals that deviate from 180°	0.6 kcal/mol/AU	Arbitrary Units (AU)
pro_close	Penalty for an open proline ring and proline ω bonding energy	1.25 kcal/mol/AU	Arbitrary Units (AU)
yhh_planarity	Sinusoidal penalty for non-planar tyrosine χ^3 dihedral angle	0.625 kcal/mol/AU	Arbitrary Units (AU)
ref	Reference energies for amino acid types	1.0 kcal/mol/AU	Arbitrary Units (AU)

In this output table of scoring function, the total_score represents the total weighted score for the input structure. Normally, the positive total_score indicates the structure is unfavorable in the force field, which does not necessarily mean that the structure is unstable but means there might be some minor steric clashes existing.

To avoid confusion due to these positive-scores, and to thoroughly evaluate different conformations based on their Rosetta all-atom score, Rosetta provides a refined ‘relax’ protocol with the same score function that intended to eventually be scored with (Nivon, Moretti et al. 2013, Conway, Tyka et al. 2014). All in all, the Rosetta relax application is a Monte Carlo optimization method for simple all-atom refinement for structures in the Rosetta force-field.

Rosetta tries to solve the structure optimization using random or stochastic moves. Most Rosetta Monte Carlo methods use the Metropolis-Hastings algorithm, which involves iterated steps in which one makes a random move, considers the change in the energy of a pose as a result of the move, and then accepts or rejects that move based on a rule called the Metropolis Criterion.

One thing that needs to be mentioned is that Rosetta is not reliable to do extensive, especially global conformational refinement but only searches the local conformational space around the starting structure (Khatib, Cooper et al. 2011, Tyka, Keedy et al. 2011). This sometimes might cause some problems when the structure provided is exactly the right structure of your intention, and you need it to be kept unchanged during the energy acquiring step yet it is mostly safe to use when this protocol applies at a relatively large scale of sorting or selecting under the exact same circumstances. The Rosetta relax application generally takes structure inputs and returns full-atom relaxed structures as feedback. All in all, relax usually serves in the final step in a relatively complex protocol aiming to get the lowest energy structures that are of interest to the user under most circumstances.

Molecular dynamics simulation and force fields

Molecular dynamics (MD) simulations calculate the time-dependent behavior of molecular systems, providing details of the atomic information and process of biomolecular conformational change (Karplus and McCammon 2002). The ensemble average is represented by the time average of long simulation properties after reaching the equilibration as the prerequisite of running MD, which means the trajectories of MD simulations can estimate many thermodynamic properties of the MD (Molnar 1999). In a MD simulation, molecules driven by the forces acted on atoms follow the numerical integration of the Newton's equations of motion of interacting particles (atoms):

$F=ma$. Integration of the equations of motion such as verlet leapfrog algorithms then yields a trajectory that describes the positions, velocities and accelerations of the particles as they vary with time($t+ \Delta t$). The forces among atoms are typically determined by their potential energies which is calculated referring to a specific interatomic potentials or empirical all-atom force fields. To calculate a trajectory, one only needs the initial positions of the atoms, an initial distribution of velocities and the acceleration, which is determined by the gradient of the potential energy function. The Verlet algorithm assumes the positions, velocities and accelerations can be approximated by a Taylor series expansion and uses positions and accelerations at time t and the positions from time $t-\Delta t$ to calculate new positions at time $t+\Delta t$. The Verlet algorithm uses no explicit velocities.

$$r(t+ \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2$$

$$v(t+ \delta t) = v(t) + a(t)\delta t + \frac{1}{2}b(t)\delta t^2$$

$$a(t+ \delta t) = a(t) + b(t) \delta t + \dots$$

Where r is the position, v represents the velocity (the first derivative with respect to time), a is the acceleration (the second derivative with respect to time),

$$r(t+ \delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2$$

$$r(t-\delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2$$

Combine these two can get

$$r(t+ \delta t) = 2r(t) - r(t- \delta t) + a(t)\delta t^2$$

And the leap-frog algorithm:

$$r(t+ \delta t) = r(t) + v(t+\frac{1}{2}\delta t)\delta t$$

$$v(t+\frac{1}{2}\delta t) = v(t-\frac{1}{2}\delta t) + a(t)\delta t$$

The velocities are first calculated at time $t+1/2\delta t$ this way, the velocities leap over the positions, then the positions leap over the velocities.

MD was first proposed by Fermi E., Pasta J., Ulam S. in the 1950s (Fermi E. 1955). Then it has been further adapted to be applied for large biomolecular systems in the 1970s (Warshel and Levitt 1976, McCammon, Gelin et al. 1977).

The general term of the force field is written as below:

$$E_{mm} = E_{bonded} + E_{non-bonded}$$

E_{bonded} represents those direct interactions such as bonded, angle and dihedral terms, and $E_{non-bonded}$ mainly refers to those non-bonded interactions such as electrotonic interaction and van der Waals interactions.

As previously mentioned, all atomic interaction can be described by the specific force field in MD system, and in most cases, they are classified into bonded interactions and non-bonded interactions. Calculating the non-bonded interactions such as electrostatics terms with Coulomb's law and van der Waals interaction terms with Lennard Jones potential(Zhang, Zhang et al. 2004) is the most computationally expensive and time-consuming part when doing a MD simulation.

The force field is the collection of the parameters and functional forms of calculating all bonded and non-bonded terms. There are many commonly used atomistic molecular mechanics force fields including Amber(Cornell, Cieplak et al. 1995), Charmm (Brooks, Brooks et al. 2009), Gromos, and OPLS-AA(Jorgensen, Maxwell et al. 1996).In spite of all these diverse atomistic force fields available for MD simulation, most of them share a same potential energy equation:

$$\begin{aligned}
V(r) = & \sum_{bonds} k_b (b - b_0)^2 \\
& + \sum_{angles} k_\theta (\theta - \theta_0)^2 \\
& + \sum_{torsions} k_\varphi [\cos(n\varphi + \varphi_0) + 1] \\
& + \sum_{\substack{nonbond \\ pairs}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} \right]
\end{aligned}
\tag{1.1}$$

While Σ_{bonds} , Σ_{angles} , $\Sigma_{torsions}$ describes bonded bonds interactions with a force constant (k_b , k_θ and k_φ) and an equilibrium status (b_0 , θ_0 , and φ_0) respectively, the $\Sigma_{nonbond}$ pairs describes all the non-bonded interactions (electrostatics and van der Waals terms) over pairs of two separated particle i and j . Electrostatics adopts Coulomb's law to interact particle pairs via employing partial charge q_i and pair distance r_{ij} . At the same time, Van der Waals term incorporates attractive and repulsive interactions represented by a Lennard-Jones 6-12 potential. The non-bonded Lennard-Jones or Buckingham, and Coulomb interactions are calculated on the basis of a list of non-bonded atoms within a certain radius.

The force field's accuracy contributes mainly to reproducing similar molecular properties from MD simulation when compared with experimental methods. Since most biomolecules exist and function in solvent, applying an explicit solvent method on a large biomolecule is necessary, yet it will significantly increase the degrees of freedom of the whole system during the MD simulation and cause heavy use of computational resources. To save the resources consumption, scientists also developed implicit solvent methods to reduce certain degrees of freedom and to simulate the

biomolecule through theoretical estimation or statistical potentials(Lee and Feig 2003, Lee and Olson 2004, Decherchi, Masetti et al. 2015)

Nowadays, the MD has been used to achieve to simulate much more complex systems due to sophisticated computer power resources. As such, the MD applications have advanced from small peptides with a few hundred atoms to functionally relevant biomolecules in an explicit solvated environment, for example, RNA/DNA in water, membrane proteins in lipid bilayers (Lindahl and Sansom 2008), whole viral capsids (Freddolino, Arkhipov et al. 2006), as well as nucleosomes (Roccatano, Barthel et al. 2007, Sun, Li et al. 2017). Newton's equation of motion is applied to calculate the movements of atoms in the molecular system in MD over time. Once the properties of the system do not change over time, then the MD simulation is considered as converged(Haile 1978). However, it's rare to actually achieve the final convergence of one simulation , the ability of computer simulations is limited to short timescales and some slow conformational changes are poorly characterized(Faraldo-Gomez, Forrest et al. 2004). For short protein simulations the first few principal components will always be caused by random diffusion, since the time is too short to reach barriers in the potential and the final direction and amplitude of the motions cannot be estimated(Hess 2002). The accuracy of MD is critically dependent on the force field, which describes the forces and energies of interaction within the system, as well as the simulation sampling, which determines what biologically relevant timescales are accessible(Marzinek, Lakshminarayanan et al. 2016). For sampling, backbone atom conformational space distribution is often analysed using principal components analysis (PCA). Block analysis are used to calculate the percentage of total sampled conformations that are required to achieve convergence(Faraldo-Gomez, Forrest et al. 2004). In addition, radius of gyration, end-to-end distance, and solvent accessible surface area (SASA) can also be an assessment of the system if those had converged

within a low level during the simulation sampling. By calculating the covariance matrix of the Cartesian coordinates of a representative subset of atoms (typically the C α atoms) of the protein, PCA is a power tool to quantify the extent of two protein simulations or fragments. The covariance between two coordinates x_i and x_j is defined as below:

$$C_{ij} = \overline{(x_i - \bar{x}_i)(x_j - \bar{x}_j)}$$

Where the bar defines the ensemble average.

To determine the differences between two average structures in MD, the overall similarity between two structures in a certain window W , $\mathbf{X}_w$ and $\mathbf{X}_{w'}$ can be quantified by defining their RMSD:

$$\text{RMSD}(\bar{\mathbf{X}}_w, \bar{\mathbf{X}}_{w'}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\bar{\mathbf{x}}_{w,i} - \bar{\mathbf{x}}_{w',i})^2}$$

where $\bar{\mathbf{x}}_{w,i}$ is the 3-dimensional vector defining the average position of atom i in window w .

Alternative MD simulation approaches such as replica exchange MD (REMD) and simulated annealing (SA) have been developed to overcome limitations associated with sampling the conformational landscape of a given molecular system.

GalaxyWEB

GalaxyWEB predicts protein structures from the sequence by template-based modelling and refines loops or terminal regions by ab initio modelling (Ko, Park et al. 2012). GalaxyWEB was assessed to be one of the best performing template-based modelling servers to predict three-

dimensional protein structures. It can generate reliable core structures from multiple templates(Ko, Park et al. 2012).

The GalaxyWEB structure prediction performs well with secondary structure, especially for monomeric structures of small proteins. Templates for the prediction candidates are first selected based on a unique scoring system in GalaxyWEB. The HHsearch has been re-scored and placed more weights on the secondary structure score for more difficult targets. The re-ranking score is a weighted sum of the Z-score of the HHsearch sequence score, Z_{seq} , and that of the HHsearch secondary structure score, Z_{ss} ,

$$S = Z_{seq} + wZ_{ss},$$

where the weight w depends on the target difficulty estimated by the probability for the HHsearch top ranker, P , as

$$w = \begin{cases} 1.0 & (P \geq 90) \\ 1.5 & (80 \leq P < 90) \\ 2.0 & (60 \leq P < 80) \\ 2.6 & (P < 60). \end{cases}$$

Among the new top homologs, all templates are re-selected by removing structural outliers. The selection is based on the calculated for the aligned core regions between candidates and templates(Zhang and Skolnick 2005). Unaligned termini are deleted from the core regions, and terminus sequence alignments are attached afterwards. Subsequently conformational space annealing (CSA) global optimization and triaxial loop closure algorithm is extensively used to generate geometrically proper backbone structures for loops. After these steps, the initial model is generated. Unreliable local regions from the initial model are detected and optimized by CSA optimization of hybrid energy that consists of physics-based terms and knowledge-based terms at

the same time(Park, Ko et al. 2011, Park and Seok 2012). The energy function is expressed as follows:

$$E = E_{\text{bonded}} + E_{\text{soft-sphere}} + w_1 E_{\text{DFIRE}} + w_2 E_{\text{dDFIRE}} + w_3 E_{\text{neighbor}}$$

The first two terms(E_{bonded} $E_{\text{soft-sphere}}$) are Physics-based terms and the other three terms(E_{DFIRE} , E_{dDFIRE} , E_{neighbor}) are knowledge-based terms. E_{bonded} represents bonded energy including bond lengths, bond angles, torsion angles. $E_{\text{soft-sphere}}$ represents soft-sphere potential to prevent steric clashes. E_{DFIRE} is the DFIRE statistical potential, and E_{dDFIRE} represents the additional orientation-dependent terms of dDFIRE. The neighbor energy E_{neighbor} is an estimate of the solvation free energy as a function of solvent accessibility of each residue measured by the number of its neighboring residues.

The input will be a protein sequence in the FASTA format. Average run time for a 500 residues protein structure prediction job is around 7 hours. The output will be five top-ranked models listed on the server. Full sets of the models generated can also be downloaded as a tar file at the same time. The model quality is measured by two methods based on *RMSD* of unreliable loop region (ULR) C α atoms (calculated after best superimposition of overall structures) and mean global distance test-total score (GDT-TS).

Protein structure prediction using I-TASSER

I-TASSER (Iterative Threading Assembly Refinement, <http://zhanglab.ccmb.med.umich.edu/I-TASSER>) is an online server for protein structure prediction and structure-based function annotation (Yang and Zhang 2015, Yang and Zhang 2015). Based on iterative structural assembly

simulations and multiple threading alignments, I-TASSER can generate full-length atomic-level structural models. All biological functions of the protein, such as ligand-binding sites information, enzyme commission number, and gene ontology terms can be inferred from protein function databases that are already known and comparisons with both sequence and structure profiles of the protein can be made to get the most precise result.

Starting from the amino acid sequence, I-TASSER is designed to identify whether there exists any homologous structure templates from the RCSB PDB databank (Dutta, H et al. 2007) using LOMETS(Wu and Zhang 2007), a meta-threading algorithm that consists of multiple individual threading programs. The entire topology of the predicted model is constructed by reassembling the aligned fragments from LOMETS templates and other super-secondary structure segments, with structures of unaligned regions from Monte Carlo simulations(Zhang, Kolinski et al. 2003). The near-native models with the lowest-free-energy conformations are identified from a pool of protein structure decoys by a clustering algorithm called SPICKER(Zhang and Skolnick 2004) through the clustering of Monte Carlo trajectories.

I-TASSER server will generate predictions of up to five models, and each model returned together with a confidence score. And the confidence score of each model is defined as below:

$$C\text{-score} = \ln \left(\frac{M/M_{\text{tot}}}{\langle \text{RMSD} \rangle} * \frac{1}{N} \sum_{i=1}^N \frac{Z_i}{Z_{\text{cut},i}} \right)$$

where M/M_{tot} is the number of structure decoys in the SPICKER cluster divided by the total number of decoys generated during the I-TASSER simulations. $\langle \text{RMSD} \rangle$ is the average backbone *RMSD* of the decoys to the cluster centroid. $Z_i/Z_{\text{cut},i}$ is the normalized *Z*-score of the best template gene, rated by the *i*th LOMETS threading program. This confidence score is highly

correlated with the quality of the predicted model. The confidence score lies between -5 to 2. Especially when the confidence score is higher than -1.5, usually it means the predicted model has a correct fold(Zhang 2008).

Summary of current computational tools

The current computational tools, including Rosetta, I-TASSER, GalaxyWEB, which all can be used to predict structures of proteins, have their own advantages and disadvantages.

I-TASSER is an online resource for automated protein structure prediction and structure-based protein function annotation(Yang and Zhang 2015). I-TASSER, as previously mentioned, it is a hierarchical templated-based method for protein structure prediction. The model is constructed by reassembling the continuous-aligned fragment structures from the templates, while the unaligned parts are constructed on replica-exchange Monte Carlo simulations. After several rounds of structure reassembly, the low free-energy conformations of the models are refined by full-atomic MD simulation. The models' global accuracy was estimated by a confidence score and residue-specific local accuracy were assessed by on a non-redundant set of 635 proteins with their structure models generated by I-TASSER. One of the advantages is that I-TASSER prediction combines the Monte Carlo simulations with the MD simulation into the prediction at the same time. On the other hand, I-TASSER requires an extensive CPU resource, which is a disadvantage of I-TASSER, also takes relatively longer time to get the result back. One single structure prediction of up to 100 amino acid sequence can take more than 50 hours on the I-TASSER server. Another disadvantage

is that I-TASSER did not make any refinements for symmetric oligomer structure prediction. It does not have any options that can define symmetry of the model, so it is not suitable to predict oligomers with symmetry. The accuracy of prediction is also less reliable in terms of higher order proteins with complicated symmetry.

Rosetta approach, as previously mentioned, is to sample a diverse set of conformations using the Rosetta fragment assembly method followed by all atom refinement, and select final models based on clustering and Rosetta all-atom energy (Conway, Tyka et al. 2014). It adopts Monte-Carlo simulation and require less computer resources than I-TASSER and has an online server (ROSIE) with powerful computational resources that open to everyone, which can accelerate the prediction and fulfill the demand of prediction all over the world. Rosetta provides symmetric docking for predicting high-order oligomers. This is typically set up to be the x-axis of the system and translation is only allowed along x. A slide degree of freedom is randomly selected, and subunits are translated along this direction until contact is found with other subunits. The next translational degree of freedom is selected and a slide into contact is performed to optimize the rigid body orientation using Monte Carlo search using a low-resolution energy function. However, the helical symmetry is not within the scope of symmetric docking of the Rosetta. The Rosetta symmetric docking is only applicable to cyclical and dihedral symmetry with limited number of sub-units. It is not feasible to predict oligomeric complexes with a helical symmetry by using Rosetta symmetric docking function. GalaxyWEB predicts protein structures from the sequence by template-based modelling and refines loops or terminal regions by ab initio modelling. It also provides symmetric docking, but not applicable to predict helical symmetry filament structures.

Table 1.5.2 *The summary and comparison of current computational tools advantage and disadvantage (required resources, accuracy, and prediction of monomer, low-order oligomer, and high-order oligomer)*

Computational tools	Required resources	Accuracy	Monomer prediction	Low-order oligomer prediction	High-order oligomer prediction
I-TASSER	High	High	Yes	No	No
GalaxyWEB	Relatively low	Relatively low	Yes	Yes	No
ROSIE	Relatively low	High	Yes	Yes	No

Chapter 1.6: Aim of the thesis

Innate immunity generally refers to immune defense mechanisms to destroy external invaders that come into play immediately or within hours of an antigen's appearance in the body. How those proteins in the immune system quickly recognize the invaders and collaborate is important in the study of immunology and can benefit both doctors and patients to provide new avenues to modulate innate immunity for therapeutic purposes.

In innate immunity, the oligomer is the active module, while the monomer is not the key to the immune response itself. The death domain is composed of four different families (DD, CARD, DED, PYRIN), and they tend to form a cascade macromolecule modulation to conduct their

function. Biologists all around the world are dedicated to studying it in both biochemistry way and computational ways.

Current classical biochemistry approaches for these macromolecules' structures determination in signal pathways adopt solid nuclear magnetic resonance (NMR), X-ray crystallography and Cryo-EM. These three methods all require highly purified materials and are quite laborious at the same time.

Current computational approaches adopting template-based modeling, and local optimization searching refinement are suitable for single monomer or small simple-symmetric oligomers. However, these approaches are not able to predict the structures of large multi-component complexes with helical symmetry. To solve this issue and to explore a new feasible computational-based approach so that studying the structural basis of filamentous macromolecules in helical symmetry in innate immune signaling pathways is the main aim of my project and this thesis.

My computation study focused on how the CARD domain forms the homo-oligomers by evaluating the physical compatibility of structure, such as the residue-residue attractive/repulsive interaction in a particular force field. CARD helical filament structures are vital to illustrate the CARD-CARD interaction in innate immunity with three types of interaction. What we want to achieve here is to predict and design helical filaments solely based on known biophysical knowledge about the compatibility of the structure by refining the existing methods and model building. Existing computational methods such as Rosetta, I-TASSER, and GalaxyWEB, can predict the structures. All three servers are able to use templated based methods to predict from sequence to monomer structure yet there is a research gap for them to predict macromolecular complexes of the helical symmetry. For I-TASSER, it is to be used to predict from sequence only

while no oligomeric structures prediction can be performed. GalaxyWEB and Rosetta are able to do symmetric docking to predict symmetric oligomers but is not suitable to be applied to predict the helical symmetry filaments.

Evolutionarily speaking, amino acid residues contain information to determine the protein structure. Death Domain may be a great candidate to study this due to its low sequence conservation and structure conservation. It is necessary to compare dozens of Death Domain members all together at the same time to obtain clues of rules of homotypic helical oligomerization.

My cryo-EM study focused on the illustrating the structural basis for CARD8 CARD-CARD interaction and CARD8-driven inflammasome assembly. CARD8 is the first known human inflammasome sensor that can directly activate pro-caspase-1 without ASC. By resolving the near-atom resolution (3.7Å) cryo-EM structures of human CARD8-CARD filaments, my study demonstrate that CARD8-CARD filaments have distinct topologies to mediate specific binding to pro-caspase-1. And this structural study enabled us to compare the structures of all known human inflammasome CARD complexes and discover more about the principle for inflammasome CARD assembly. The solved CARD8-CARD filaments structure can serve as a successful candidate that cross validate our computational prediction of helical parameters of CARD8 filaments, providing a potential application for general assembly principles for other CARD assemblies in innate immunity.

Chapter 2: Materials and methods

Chapter 2.1: Helical parameters in CARD-CARD filaments

Monomer models

All monomer CARD structures were either obtained from the solved structures published in the RCSB PDB database (protein data bank) or from the I-TASSER server. For CARDS with available monomer structures in the PDB database, we input these into our simulations. For those CARDS without available monomer structures on the PDB server, we chose to use the I-TASSER server to predict the monomer structures. The sequences for functional CARD domain regions were obtained from Uniprot database(www.uniprot.org). The certain sequence for each CARD in FASTA format was used in I-TASSER server to predict models for individual CARD. And for each CARD, the one with the highest I-TASSER confidence score was selected.

Helical filament model building

All helical filaments models were built based on homemade scripts. Chimera commands allow building models in a certain symmetry, including helical symmetry. The *sym* command generates symmetry-related copies of a molecule model. The molecule model to copy was specified by model number. Helical symmetry was defined by rise, angle, copy number, and the coordinates of the center. As such, one helical filament structure was built with the input monomer structure, and the parameters defining the helical symmetry. An example of generating the helical symmetry was “h,5, 90, 8”, which generated a multicopy molecule with a 5Å rise, 90° rotation angle helical

filaments with eight units copies of individual monomer.

In order to produce a large amount of helically symmetric filament structures with different helical parameters, scripts named chimera.sh that automatically utilize the chimera command line were made. Several description files were generated, which are “execute-chim.sh”, “myscript.com”. “Myscript.com” files were generated to describe the parameters of helical symmetry of filament molecules and “execute-chim.sh” were generated to execute every single chimera command file and combine molecules units into one PDB file. In all scripts were made, z-axis was adopted as the default axis to extend the filaments. The initial orientation of monomers was adapted from superimposing the monomer orientations of the solved filament structures in PDB data bank. The key scripts adopted were listed as below:

```
echo "sym #0 group h, $rise, $angle, 8, -3 center 0,0,0 >>myscripts-$rise-$angle.com
```

While #0 represents the unit number of the single-molecule(monomer structure) that is active to be executed; 8 represents the total unit numbers of molecules that combined into one filament, the example above means eight units of molecules were combined to final filaments; -3 represents the offset of copies of molecules in filaments. The integer offset (default 0) allows extending the helix in both directions. The example above would give n = 6 copies total, with two copies molecules in the negative axis direction, one at the identity position, and three in the positive axis direction; 0,0,0 represents the coordinates of center of the filaments. The example above would position the monomer molecules around the coordinate 0,0,0 and extend the filament along the axis. All parameters can be adjusted accordingly.

Other scripts and details are listed in the *Appendix*.

Rosetta prediction

The Rosetta scoring function was used to calculate the energy to evaluate the stability and the rationality of helical filament structures. An optimized energy function or score function called REF15 was adopted by Rosetta to calculate of all atomic interactions in a globular protein sample(O'Meara, Leaver-Fay et al. 2015). Some of the weighted sums of energy terms include physical forces like electrostatics and van der Waal's interactions, while some others represent statistical terms such as the probability of the potential torsion angles in Ramachandran space (Alford, Leaver-Fay et al. 2017). The total energy, called ΔE_{total} , is computed as a linear combination of energy terms E_i that are calculated as functions of geometric degrees of freedom (Θ) and chemical identities (aa) and scaled by a weight on each term (w_i), as shown in below:

$$\Delta E_{\text{total}} = \sum_i w_i E_i(\Theta_i, \text{aa}_i)$$

The Rosetta “relax” function was used for relaxing the helical filament structures. The “relax” function can find the lowest-scoring variant, running both the packer and the minimizer alternately. The relax function first minimizes the filament structures and performs a Monte-Carlo simulation in local space, and then gives a result of the evaluation of the helical filament structure, which is similar to performing a solely scoring function.

Multiple models were evaluated by scoring/relax function in a parallel mode on a workstation server. Each model underwent three repeats and the each run took around 10- 20 minutes.

Scripts and details are listed in the *Appendix*.

Molecular dynamics refinement

When setting up the MD simulation systems, all input coordinates for the CARD domain contained filament structures were based on homemade scripts and respective monomer structures. Gromacs 4.5.5 was adopted to run all MD simulations. The AMBER99SB-ILDN protein, nucleic AMBER94 forcefield, was used to describe all the protein parameters. All the complexes were explicitly solvated by the of the Gromacs software package in a cubic water box with at least 10Å distance around the entire complex and electrically neutralized using K⁺ and Cl⁻ ions at the concentration 0.15M. And the final size of the systems was about 500,000 atoms. Each system was minimized under both NVT and NPT with V-rescale thermostat for temperature control and barostat for pressure control for 1000 steps at 1atm and 298K using a time-step of 2 fs, and a cut-off of 10Å for Lennard-Jones interactions, Lincs constraints for all bonds, and PME for long-range electrostatics. V-rescale temperature coupling and Parrinello-Rahman pressure coupling have been applied to 3 replicas for 1ns.

The *RMSD* of certain atoms in a molecule with respect to a reference structure was calculated with the program “gmx_rms” by least-square fitting the structure to the reference structure ($t_2=0$) and subsequently calculating the *RMSD*.

$$RMSD(t_1, t_2) = \left[\frac{1}{M} \sum_{i=1}^N m_i \|\mathbf{r}_i(t_1) - \mathbf{r}_i(t_2)\|^2 \right]^{\frac{1}{2}}$$

Where $\sum_{i=1}^N m_i$ and $\mathbf{r}_i(t)$ is the position of atom i at time t . The *RMSD* can be calculated of the backbone, or of the sidechain, or of the whole protein.

Chapter 2.2: Structural study of CARD8-CARD

Plasmid construction

Caspase recruitment domain-containing protein 8(CARD8) is a protein participates in regulatory mechanism that coordinates immune response and pyroptosis via CARD domain and activation of proinflammatory caspases. For recombinant CARD8 protein sample preparation, human CARD8-CARD(residue 451-537, full length 1-537) with or without C-terminal SNAP tag (NEB) was cloned between XmaI and EcoRI restriction sites into pET47b (Novagen) plasmid. SNAP tag was added for enhancement for solubility of protein and can be further specifically labelled with a suitable ligand, such as Alexa fluorophore conjugated Brilliant green(BG). To generate the CARD-SNAP recombinant proteins, SNAP-tag has been inserted into pET47b (Novagen) between XmaI and EcoRI restriction sites. A FLAG tag was added to N-terminus. And an HRV 3c protease cleavage sequence was added in between the CARD8-CARD domain and SNAP-tag, which was used for SNAP-tag cleavage for further construct optimization to collect good-quality CARD8-CARD-only filaments micrographs under cryo-electron microscopy because that SNAP tag would be on the outer surface of the filament and they might interact with the filament in a non-specific and non-repetitive manner, interfering with the symmetry and helical repetition of the filament. For all constructs, the sequences were checked with Bio-Basic/1st Base sequencing service. The sequencing result and the primer designed were listed in *Appendix*.

Protein purification

For the expression of the His-tagged recombinant proteins, we used BL21(DE3) (NEB) *E. coli* for the constructed plasmids to transform into and grew it at 37 °C for 16–18h. Then we transferred the so-called starting culture (usually 50ml) into a larger volume of lysogeny broth (LB) to be

grown. When the optical density (OD) reached around 0.6-0.8, then we lowered the temperature of the shaker down to 16 °C and induced the *E. coli* bacteria with 0.5mM Isopropyl β - d-1-thiogalactopyranoside (IPTG) and grew the cells overnight (around 16-18h) and harvested the bacteria on the next day. The bacteria were harvested and spun down at $4600 \times g$ (Avanti JXN series, Beckman Coulter) at 4°C for 25 mins. Then we discarded the supernatant and resuspended the pellet with lysis buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole, and 10% glycerol). Resuspended cells were then lysed by LM20 Microfluidizer and then centrifuged again at a speed of $30,000 \times g$ (Avanti JXN series, Beckman Coulter) for 30 mins at 4 °C to separate supernatant and pellet. The solubility and purity of protein was initially checked by running 15% SDS PAGE gels. Then we collected the supernatant and the supernatant was passed through a pre-equilibrated column with nickel-nitrilotriacetic acid (Ni-NTA) agarose beads (ThermoFisher Scientific) by gravity. After the supernatant containing our protein passed the beads, we washed the column with the lysis buffer containing 20 mM imidazole to remove all nonspecific binding proteins in the beads. His-tagged proteins were then eluted with the elution buffer that contained a gradient volume of imidazole (100mM, 150mM, 200mM, 250mM,300mM). And we further purified our recombinant proteins by using buffer A (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 1 mM DTT) in size-exclusion chromatography (Superdex 200 increase, GE Healthcare) and collected the certain elution fractions of protein. At this step, the purified proteins were the oligomeric pre-activated seed fraction. Purified soluble filament seed was buffer exchanged into 10 ml of 6 M guanidinium, and gradually dialyzed against 3, 2, 1.5, 1, 0.8, and 0.6 M guanidinium containing buffer A in the cold room, and eventually in buffer A. The folding status was further checked by Bis-Tris Native-PAGE gel (Life) with fluorescent dye BG-Alexa Fluor 647 (NEB). One microliter of 100 μ M BG-Alexa dye stocks (dissolved in water containing 5% DMSO) were

added into every 100 μ l of pre-labeling protein stock at a concentration of 20 μ M. The native PAGE gel was run at a voltage of 200 V for 60 min and visualized by Typhoon FLA7000(GE Healthcare) with 100- μ m pixel size at channel 647 using PTM 600. After this step, we had the purified CARD8-CARD proteins in a monomeric form. The SDS PAGE and Native PAGE result were in the *Appendix*.

Oligomerization assay

CARD8-CARD monomers were concentrated to 10mg/ml to form the filament. Some constructs may need higher concentration to reach the oligomeric form, for example, CARD8-CARD-SNAP, may need to gain a concentration more than 20mg/ml until the protein sample solution forming cloudy texture was observed, and homogeneous long filaments formed spontaneously. After getting the thick filament, in-house-prepared maltose binding protein(MBP)-3C-protease then was added to digest the thick filament containing SNAP-tag, and the digestion was performed at 4° overnight. Then we further purified the core filament, the cleaved-off SNAP tags, and the proteases. Thick CARD8-CARD-SNAP filaments and 3C-protease-digested core CARD8-CARD thin and long filament were further purified using size-exclusion column Superose 6 (GE Healthcare). To form long CARD8-CARD filaments, N-terminal His-tagged CARD domain-only constructs were expressed and purified according to the protocol mentioned above, then concentrated to around 400 μ M long homogeneous filaments formed spontaneously, without the requirement of seeding.

Cryo-electron microscopy

For cryo-EM grid preparation, Quantifoil R1.2/1.3 holey carbon grid (Quantifoil) was glow-

discharged for 60s and plunge-frozen using a Vitrobot (FEI) and was treated with a 5- μ l sample at a concentration of about 2 mg/ml. The high-resolution images were collected under Titan Krios, 300 kV, using Gatan K2 camera (Gatan), and super-resolution mode was used for data collection, and got binned 2 $\times$ when doing motion correction. The sub-frame time was 250 ms so that the total exposure time was 10 s with 40 frames per stack. And the pixel size of the image is 1.10 $\text{\AA}$. The dose rate was 40e/pixel/s.

Data processing

For cryo-EM data, super-resolution image stacks were gain normalized and binned by 2 $\times$ to a pixel size of 1.10 $\text{\AA}$ before drift and local movement correction using MotionCorr2.

Relion 3.0 was used for CTF determination, particle picking, 2D classification, 3D classification, and refinement procedures. Totally, 1424 micrographs were collected and analyzed for CARD8. We manually picked out around 10,000 to 20,000 filaments and windowed out segments of 240 $\times$ 240 pixels, totally yielding 1,201,108 CARD8 particles for 2D classification. After several iterations and bad particles and fewer desirable classes were removed, we used a total of 301,848 particles for 3D classification. Ab initio low-resolution helical structure was generated using a Gaussian cylinder as an initial model. 3D classification and auto-refinement of 98,280 particles were done in Relion3.0 and CisTEM. All 3D refinements were carried out following the gold-standard procedure where the data set was divided into two half-sets. After refinement was converged, a mask of the central 40% segment was calculated and applied to the final data set, and the corrected Fourier shell correlation (FSC) was used to calculate the overall resolution about 3.7 $\text{\AA}$ by using gold standard FSC = 0.143 criterion comparable to measurements used in X-ray crystallography (Scheres and Chen 2012). In the final model, parameters of CARD8-CARD were

calculated to get 5.409Å of the rise per asymmetric unit and an azimuthal rotation per subunit of −99.16°.

Density and model build

The first step was monomer docking. The CARD8 monomer structure was firstly adopted from the RCSB protein data bank (PDB) (PDB ID: 4IKM). And CARD8-CARD monomer residue numbers were renumbered, and several terminal residues were deleted to match the length of the density perfectly. Every individual helix was manually shifted by rigid body movement in Coot according to the locations of three large aromatic side chains. The resolution at 3.7Å allowed us to locate these residues unambiguously.

The second step was to build the tetramer following the rules of Type I, type II, and type III interfaces (a centrally located monomer and three other monomers forming type I, type II, type III interfaces with that specific central one) by visualizing in Chimera. Then CARD8-CARD surface interactions were refined in this way by getting this tetramer. Phenix real-space refinement was done without rigid body refinement, and different parameters were tested, torsion-rotational angle Non-crystallography symmetry (NCS) constraints (Phenix default auto setting) were used at this step to maintain the symmetry of tetramer. In this way, all surfaces of the central monomer could be refined. Then the 12mer model was built. The central monomer from Step 2 was copy-and-pasted to occupy the entire 60 Å segment density. In the end, 12 identical monomers were added to the density. Phenix real-space refinement with default NCS torsion-rotational constraint was performed. After several rounds of rigid body refinement, a model with reasonable statistics was obtained. For issues like relative high clash score and Ramachandran plot outliers, we manually identified those issues and rebuilt the model the repeat the Phenix until satisfactory statistics and

density interpretation was achieved.

Chapter 3. Homological filament structure and heterogenous pair interaction prediction of CARD domains

Chapter 3.1 Overview of the CARD domain in innate immunity

The immune response is the body's defence against outside infectious organisms and other outside invaders. And innate immunity, as one of the two main immunity strategies, refers to the nonspecific immune defence mechanisms that come into play immediately or within hours of an antigen's appearance in the body. It usually depends on a group of proteins and cells that can recognize conserved features of PAMPs or DMAPs and become quickly activated.

How these innate immunity proteins and cells recognize invaders is always one of the hottest points in the immunology field. PRRs, including membrane bound TLRs, cytoplasmic RLRs and NLRs, can detect the conserved molecular patterns from pathogens and convey signals to their specific adapters to induce the innate immune response and other inflammatory reactions. And during every single stage, macromolecular super-complexes can always be found being actively enrolled in all kinds of interactions in contrast with single monomers.

Recently, immunologists have been focused on the death-fold superfamily that can form cascade-like macromolecular complexes called inflammasomes to trigger a series of immune responses in the body. The CARD domain family performs specific functions in the formation of well-regulated oligomeric complexes, by direct interactions between individual CARD monomers. With increasing numbers of studies nowadays on the CARD domain family, more than half of the CARD monomer structures from human have been solved either by X-ray crystallography or NMR

spectroscopy (Table 3.1.1). However, this area still remains a significant challenge to be understood. As mentioned previously, single monomers do not play the lead in all kinds of activities in innate immunity. How CARD domains interact to form macromolecular complexes and lead to inflammasome formation to conduct immune responses in innate immunity is not completely clear.

Table3.1.1 Summary of human solved and unsolved CARD structures

Monomer structure	
Solved	Unsolved
MAVS CARD (PDB ID: 2MS8)	MDA5 1 st CARD
RIG-I 1 st CARD (PDB ID: 4P4H)	MDA5 2 nd CARD
RIG-I 2 nd CARD (PDB ID: 4P4H)	NOD2 1 st CARD
NOD1 CARD (PDB ID: 2B1W)	NOD2 2 nd CARD
Apaf-1 CARD (PDB ID: 2P1H)	NLRC3 CARD
BCL10 CARD (PDB ID: 2MB9)	Caspase2 CARD
Caspase1 CARD (PDB ID: 2H51)	Caspase4 CARD
Caspase9 CARD (PDB ID: 1NW9)	Caspase5 CARD
BIRC2 CARD (PDB ID: 2L9M)	BIRC3 CARD
CARD11 CARD (PDB ID: 2AR9)	CARD16 CARD
NLRP1 CARD (PDB ID: 4IFP)	CARD17 CARD
CARD9 CARD (PDB ID: 6N2M)	Caspase12 CARD
NLRC4 CARD (PDB ID: 4KXF)	DLG5 CARD
RAIDD CARD (PDB ID: 3CRD)	CARD10 CARD
CARD19 CARD (PDB ID: 4DWN)	CARD6 CARD
CARD18 CARD (PDB ID: 1DGN)	CARD14 CARD
CARD8 CARD (PDB ID: 4IKM)	NLRC5 CARD
NOL3 CARD (PDB ID: 4UZ0)	CRADD CARD
ASC CARD (PDB ID: 2KN6)	RbBp-36 CARD
RIP2 CARD (PDB ID: 6GWM)	

	Solved	Unsolved
Count	20	19

CARD domain forming filamentous oligomeric structures follow helical symmetry and are mainly determined by three parameters, which are the rise, rotation angle, and the diameter. When helical symmetry applies to CARD domain complexes, first CARD performs a helical turn with a translation along the axis while the second CARD follows the way the previous one does. The turning is measured by a rotation angle and translation is measured by the rise (Figure 3.1.1C, D)

Based on the existing solved cascade-like structures of CARD domains, we hypothesized that this filamentous module could apply to all CARD domains, and in this way the CARD domain filament-like oligomer complex may be the active module in innate immunity to conduct signaling and lead to immune responses.

To figure out how this filamentous model fits into the real scenario that CARD domains form macromolecular complexes, we have studied what rules and what parameters the CARD home-oligomers follow in studying the existing solved filamentous structure available in the PDB database. Table 3.1.2 lists all names, IDs and helical parameters of filament structures that have been solved. The data of rise and rotation angle was directly from the RCSB Protein data bank, and the diameter was measured in Pymol by calculating the distance between C α of two certain residues.

Table 3.1.2 Summary of solved human CARDs filament structures parameters: rise, rotation angle, and diameter of filaments, PDB ID, and the year of the structure being solved and deposited.

Name	Rise(Å)	Rotation(°)	Diameter(Å)	PDB ID	Year
RIP2	4.94	101.4	80	5yrn	2018
MAVS	5.1	101.1	85	3j6j	2014
CARD9	5.11	101.6	90	6n2p	2019
NLRC4	4.93	100.48	80	6mks	2018
ASC-CARD	5	100.58	80	6n1h	2018
BCL10	5	100.8	90	6bze	2018
Caspase1	5.06	100.23	80	5fna	2016
RIG-I tandem CARD	5	101	85	4p4h	2014

At first, we studied all filament structures of CARD that have been solved. In comparison with more than half of the monomer's structures having been solved for CARD domains, only seven filament structures of CARDs have been solved until now. For the helical parameters, all seven filament structures have a rise of helical turn along the central axis that lies between the range from 4.7Å to 5.3Å, and a rotation angle that lies between the range from 99.5° to 102.5° (Figure 3.1.1A,B, Table 3.1.2). These parameters were obtained from Protein data bank and corresponding papers. The diameter of the filaments is considered as a relatively unimportant parameter because this can be easily obtained via negative stain EM or other methods. Among all of the 7 solved helical filament structures, RIP2-CARD, ASC-CARD, CARD9-CARD, and MAVS-CARD were selected and shown here as examples in Figure 3.1.1D. We can conclude from RIP2-CARD, ASC-CARD, CARD9-CARD, and MAVS-CARD filament structures that although their monomers share a conserved three-dimensional structure (death fold motif), with different helical parameters, we can get more diverse helical structures. And because of these diverse cascade-like helical filament

structures, the innate immunity forms a variety of pathways that do not interfere with each other and conduct various immune response.

Figure 3.1.1 Overview of the helical symmetry and current CARD filaments architectures. (A) Distribution of rise and rotation angle for those CARDs with solved filament structures. (B) Distribution of diameter for CARDs with solved filament structures. (C) Schematic illustration of helical symmetry applied to CARDs. (D) Monomer structures and filament structures of RIP2-CARD(Cyan), ASC-CARD(Green), CARD9-CARD(Pink), MAVS-CARD(Yellow) showing conserved monomer structures and diverse filament structures.

Chapter 3.2 Computational CARD homogenous filament structure prediction

For all symmetric helical modules, we have made two major hypotheses. Firstly, if we can know the helical parameters, rise, rotation angle, diameter or radius of the filament, and the initial position of the individual CARD, we can theoretically design a helical filament structure. Secondly, if we can find a way to evaluate the physical compatibility of the filaments, we may find a way to predict the most stable helical structures. The theories of three types of interactions have been well established. Both electrostatic force and Van der Waals forces are engaged in all CARD-CARD interactions required for CARD filaments. Here is an example of MAVS-CARD to illustrate the residues involved in Type I, Type II, and Type III interaction in Figure 3.2.1A, B.

In general, we used our in-house scripts to build a series of different helical filaments and tried to put it into a parameter filter that can help us to focus on all those possible parameters and remove those that are not. Our built models keep two variables: rise started from 4Å to 5.9Å and rotation angle from 90° to 109°. All other parameters such as diameter or radius of the filament, monomer orientation, and center point of the filaments are all considered as known quantities and keep constant during the simulation.

Our first selection filter is a Monte-Carlo Rosetta score system (local optimization searching) adopted from David Baker's method with our tuned combination of scores. Every time it scores, we make sure it goes through the second selection filter combining with a "relax" module that adopts all-atom force field refinements to get an improved model under the scoring system. Monte-Carlo local simulation was performed for our artificial models. The first and second selection layer

are specialized for selecting the rise and rotation angle of certain helical symmetry. The third selection filter focuses on adjusting the internal orientation between the monomer structure and the filaments elongation axis. Extra scripts adjusting the monomer orientation by flipping the monomer structures and doing a so-called internal rotation were needed. This procedure may need to be repeated several times to get the best selection results. After all selection filters, we run a short period(1ns) of MD simulation to finalize the sidechain rotamers to ensure fewer clashes in our selected model without changing the helical parameters. Short periods of the MD were run under standard solution conditions, and all parameters related to MD sets were listed in the materials and methods as well as in the Appendix.

The overview of the flow chart of our computational design is shown in Figure3.2.1C.

Figure 3.2.1 Schematic of interacting surface and architecture of CARD-CARD filaments and computational design protocol. (A) Schematics of three types of interactions that contribute in the filaments' formation. (B) MAVS-CARD filament structures, showing typical three types of interaction residues that contribute to filament formation. (C) Flowchart illustrating computational design protocol.

CARDs tend to oligomerize into a particular helical symmetrical filament. A helical symmetrized filament can be determined by three parameters, rise, rotation angle, and diameter of the filaments. As previously mentioned, the diameters of filaments can be relatively easily determined by

negative stain electron microscopy so that we did not design another extra variable and presume it is a known quantity; the actual challenge of modelling a filamentous structure is the rise and rotation angle

Here is an example of MAVS-CARD, showing how we screen helical parameters in our first selection layer. The artificial helical filaments were designed with the rise starting from 4Å to 5.9Å, while the rotation angle starting from 90° to 109°. Here in Figure 3.2.2A shows that scores evaluating our artificial models return a continuous low score value region as a function of the helical rise and the rotation angle. With the modified contribution of the scoring system, as mentioned previously, it is feasible to get a focal point among all possible combinations of the helical rise and the rotation angle of helical symmetry for MAVS-CARD. The MAVS-CARD results here show CARD domains have a relatively low energy region, indicating MAVS-CARD domains have a preference in selecting rise and rotation angle when forming oligomeric filaments. The solved cryo-EM MAVS-CARD filament structure (PDB ID: 3J6J) possesses a helical rise of 5.13Å and a rotation angle of 101.1°, which lies within our simulated consistent low score value region and perfectly matches our prediction result. Precisely speaking, a helical rise 5.13Å and a rotation angle of 101.1° lies right in the central area of our selected low energy region if we draw a circle area representing this consistent low score value region.

As previously mentioned, the diameter or radius of the filaments is considered as an easily accessible parameter that we already get before the sampling of the rise and rotations because diameter or radius can be obtained through negative stain EM or other methods in our study. Although we always combine our simulation with biochemistry validation and experiments related,

this does not mean that our approach is insensitive to the diameter or radius of the filaments. Here shows the example of MAVS-CARD filaments when we changed the diameter or radius of the helical filaments of the input and keep other settings the same, the return of value score was rapidly changed, and most likely, the focal point was lost. Figure 3.2.2C shows that there was no longer a consistent, meaningful rise and rotation angle region for MAVS-CARD for the same setting as before when the radius increased by 5 Å for the original MAVS-CARD filaments (Figure 3.2.2B).

We also take into consideration of initial position of the CARD monomer. So how its α -helices oriented was also considered a critical parameter for our initial input. To solve this problem, we successfully adjusted the monomer orientation by making it make an internal rotation selection. The internal rotation selection for the monomer orientation was conducted in x-y, x-z, y-z plane separately. So, each time we could also get a two-dimensional score region, similar scoring results, just like our first- and second-layer selection.

We also used the scoring system to screen the orientation preference one by one to form a large scale and then narrow down to a small focal point. As the focal point region was narrow enough, we can finalize a monomer model or several monomer models that can be used to make rise/rotation selection again and repeat another round of this selection (Figure 3.2.2D).

To finalize the predicted filament structure, these selection layers may need to be applied several times until a reasonable filament structure was selected.

Figure 3.2.2 Helical symmetry parameter score screen return map (grid search) of MAVS-CARD. (A)Parameter screen of helical rise from $4\text{\AA}$ to $5.9\text{\AA}$, rotation angle from 90° to 109° MAVS-CARD shows a continuous low score region and get a focal point applied optimized score contribution. The focal point was labelled in dash circle area. Unit: REU **(B)**Parameter screen of rise/rotation angle with applied different diameter to MAVS-CARD filaments shows a sensitive return of score region to the diameter of filaments. **(C)** Illustration of the Structure model building with enlarged extra $5\text{\AA}$ for radius of filaments of MAVS-CARD. **(D)** Illustration of the definition of internal

rotation in our method and score screen return map of MAVS-CARD in plane of x-y showing how we select models with different helical symmetry parameter applied.

We aim to use the monomer structure to predict the homo-filament structure by using numeric score value as a tool to evaluate helical symmetry parameters. It could serve the function as the Morse Code for individual CARD domain. A thorough simulation of MAVS-CARD has been performed including parameters refinement and internal rotation angle adjustment, to make sure to get the most accurate model structure to test the whole procedure of our simulation and evaluate our simulation procedure. And based on that, we are able to predict the filament structure of CARDS.

Molecular dynamics

A short time MD (1ns) was applied to our simulated structure to perform a local MD conformation to minimize steric clashes without the distinct parameter (rise and rotation angle) change of our filaments structure. The system potential energy (KJ/mol) and *RMSD* (nm) variation process during the MD have been recorded.

Figure 3.2.3 MAVS-CARD Molecular dynamics simulation energy, RMSD, and superimposition. (A) The energy change during the molecular dynamics of MAVS-CARD simulated filaments(1ns). (B) The RMSD change of backbone of MAVS-CARD filaments and sidechain of MAVS-CARD filaments(1ns). (C) Comparison of simulated MAVS-CARD before and after molecular dynamics and comparison of final predicted filament model and filament model in PDB database.

Here in Figure 3.2.3A,B shows the potential energy and *RMSD* (nm) variation process of the MAVS-CARD filament structure. The potential energy and *RMSD* were calculated by “gm_x_energy” and “gm_x_rms”. The potential energy has been stabilized quite rapidly after the start of MD simulation, while the *RMSD* (nm) of both backbone and sidechain during the MD simulation tended to be relatively stable after a short period. The *RMSD* (nm) for the backbone was around 1.9Å and around 2.5Å for sidechain, showing that MAVS-CARD filament structure remained intact and without distinct conformational change. We compared our simulated MAVS-CARD filament structures (both before MD and after MD) and the deposited cryo-electron

microscopy MAVS-CARD filament structure(PDB ID:3J6J). The *RMSD* (nm) of all atoms between one of the simulated MAVS-CARD filament before MD and the deposited MAVS-CARD(PDB ID: 3J6J) was measured around 1.90Å. The *RMSD*(nm) between the simulated MAVS-CARD filament after MD and the deposited MAVS-CARD filament structure(PDB ID:3J6J)was measured 1.14Å. The superimposition of simulated MAVS-CARD (both before and after MD) and the final predicted (after MD)MAVS-CARD filaments models are shown in Figure 3.4C. Based on these data, we can conclude the simulated MAVS-CARD filament structures remain relatively stable after a short period of MD conformation and was quite close to the deposited PDB structure, which was solved by Cryo-Electron microscopy and 3D reconstruction.

Chapter3.3 CARD filament helical symmetry first-layer predictions

CARDs with solved filament structures in PDB database

Apart from MAVS-CARD, other 6 CARDs (RIP2-CARD, CARD9-CARD, ASC-CARD, BCL10-CARD, CASP1-CARD, NLRC4-CARD) with solved filament structures in PDB database (Protein data bank) were calculated scores of helical parameters in our simulation. The monomer structures from PDB were adopted as our initial input for individual CARDs. The experimentally solved structure helical parameters are directly from Protein data bank and are marked with a red circle (Figure3.3.1). Our prediction result shows all these 6 CARDs are likely to form a stable filamentous oligomeric macromolecule and all their experimental helical parameters are within the continuous low energy region.

Figure 3.3.1 All score/relax maps for a screen of applying helical symmetry-rise from 4Å to 5.9Å, rotation angle from 90° to 109° for all other six CARD filaments with solved cryo-EM filament structures in the PDB database

(unit: REU). The experimentally solved structure parameters are marked with red circle, and the data are from Protein data bank as well as corresponding papers.

CARDs without solved filament structures in PDB database

In order to get an overview of the CARD domain assembly and test our hypothesis, we applied the first layer selection system to all CARD domains. For those CARDs with available monomer structures in the PDB database, we input these directly into our simulations. For those CARDs without available monomer structures on the PDB server, we chose to use the I-TASSER server to predict the monomer structures. The predicted monomer structure with the highest confidence score was always adopted as our initial input for individual CARDs. Figure 3.3.2 shows that the helical parameter score returns for all CARDs without solved filament structures in the PDB database.

Some of the CARDs may have monomer structures solved by NMR spectroscopy or X-ray crystallography while the other CARDs not. For those even don't have available monomer structures in the database, I-TASSER predicted models with the highest confidence were adopted as one of the input files and are labelled in Figure 3.3.2. All solved CARDs and unsolved CARD are summarized in Table 3.1.1. And the original monomer orientation was superimposed to MAVS-CARD. These results may not as precise as the previous ones because of the lack of experimental biochemistry data available. According to our simulation, CARD domain such as CASP2(Caspase2), CASP5(Caspase5), CASP9(Caspase9), CRADD, BLRC2, BLRC3, CARD10, CARD16, CARD17, CARD18, APAF1 are likely to form a stable filamentous oligomeric macromolecule because they form a continuous low energy region. CASP4(Caspase4), CASP12(Caspase12), DLG5, CARD6, CARD11, CARD14, NLRC3, NOD1, RBBP5, NOL3 are

less likely to form helical filament structure because they do not form a continuous region (Figure 3.3.2). From the structural perspective, for those CARDs cannot form helical filament structure, the inability to get a continuous low energy region indicates that the CARD filamentous structures in such helical parameters are not stable or cannot exist in a relatively low energy state. Either it is because that the CARDs filament structures cannot be formed by CARD monomers, or the helical pattern of filaments is too dynamic so that the filament structures contain multiple intermediate conformations. This might also indicate the CARDs transduce the immune signal in a different form of high-order macromolecular complexes instead of forming filaments. This happens to those CARDs whose residues contributing to the homotypic Type I, Type II, and Type III interaction on each interface may not be conserved enough or structural compatible to form CARDs filaments. And also, if the interface a of 'upstream' CARD is not structural compatible to 'downstream' CARD interface b, this may indicate the activation mechanism for these two CARDs is not in the form of oligomerized filaments.

From the function perspective, this may indicate that the CARDs containing proteins conduct their function prompted by other domains. For example, ASC contains both a PYRIN domain and a CARD domain. The activation of ASC inflammasome to form the ASC speck involves oligomerization of ASC-PYRIN into filaments and cross-linking of these filaments by ASC-CARD. Homotypic CARD-CARD interaction are between CARD-containing receptors recruiting protein ASC to bridge ASC molecule and activate the nucleation of ASC-PYRIN filaments rather than forming ASC-CARD filaments (Dick, Sborgi et al. 2016).

There may exist inaccuracy of the prediction. Those CARDS cannot get a low energy region may be due to the lack of information on the correct monomer structures of these candidates. And also, the mismatch between the hypothesized helical pattern and the actual helical oligomeric pattern may cause this situation. These candidates may form alternative helical modules. For example, Pyrin domains such as ASC-PYRIN tend to adopt a different format of filaments, while Caspase8 DED domain takes a C3 symmetry instead of C1. There are also possibilities that some of these candidates do form intrinsic dynamic filament structures with dynamic helical parameters.

*Figure 3.3.2 All score/relax map return for a screen of applying helical symmetry-rise from 4Å to 5.9Å, rotation angle from 90° to 109° for all other CARD filaments without solved cryo-EM filament structures in PDB database(unit: REU). The CARDS without predicted monomer structures are labelled with *.*

Prediction of unknown NLRP1-CARD and CARD8-CARD filament structures

In order to validate our computational prediction approach, we use it to predict another two candidates with no previous information about filament structure. We use the available monomer structure in the PDB database to predict the structures of the filaments. At the same time, we also conducted biochemistry experiments and structural studies by using cryo-EM for these two CARD candidates in parallel. The NLRP1-CARD was done by my lab mate Qin, Gong and the CARD8-CARD experiment was done by me and the specific biochemistry work contributed to **Chapter 4**.

During our processing of the data, we applied the simulation first to get a simulated parameter selection before proceeding to the 3D reconstitution of filament structures. Parameters such as the diameter or radius of the filaments can be easily obtained via negative stain EM. With the input and some primary data about these two CARD candidates, our computational approach provided us with information on the preferable rise and rotation angle parameters, which was helpful during the experiment.

We performed the first layer selection in our method. And the first layer selection return for helical parameters about the rise and rotation angle of NLRP1-CARD and CARD8-CARD filaments are shown here in Figure 3.3.3A and Figure 3.3.4A, separately. In our prediction, the monomer structure of NLRP1-CARD (PDB ID:4IFP) and CARD8-CARD (PDB ID:4IKM) from the PDB database were used as input.

NLRP1-CARD prediction

Figure 3.3.3 NLRP1-CARD prediction score and validation structures. (A) Prediction of helical symmetry score return of filaments of NLRP1-CARD with applying helical parameter (rise 4Å ~5.9Å, rotation angle 90° ~109°). (B) 3D densities of 4 groups of NLRP1-CARD filaments were obtained from cryo-EM data (contributed by Qin,Gong). The densities showed NLRP1-CARD filaments possessed a dynamic helical symmetry, and totally four conformations were found after helical symmetry was applied in Relion.

For NLRP1-CARD, we may conclude that our approach may not fit very well for simulating the NLRP1-CARD or NLRP1-CARD may not be a right candidate for our approach. Figure 3.3.3A showed that there was no significant low-value region for helical parameters of the NLRP1-CARD filament, and the result tended to have a very dynamic filament pattern. This may be due to either the displacement of the initial monomer model or the quite dynamic helical pattern formed by the NLRP1 filament structure. However, when compared with the NLRP1-CARD filament structure solved by cryo-EM, it was noticeable that this dynamic pattern of NLRP1-CARD filament matched our simulation results to some extent.

By collecting the cryo-EM data and doing the 3D reconstruction using RELION 3.0, several densities of NLRP1-CARD filaments were obtained, and the result is shown in Figure 3.3.3B. The NLRP1-CARD density and all related work was contributed by Qin,Gong. In this experiment, the NLRP1-CARD filament demonstrated markedly different flexibility. NLRP1-CARD filaments are highly flexible and adopt multiple helical organizations in vitro, as evident by heterogeneous 2D and 3D classes observed during structural analysis. Even with the same helical symmetry parameters, four conformations of NLRP1-CARD filaments were observed, and the rise and rotation angle are quite hard to define. The significant one conformation of oligomerized NLRP1-CARD and possesses a helical rise with 5.36 Å and a rotation angle with 100.82°.

CARD8-CARD prediction

For the CARD8-CARD helical symmetry simulation, it did show some interesting results. The solved CARD8 full-length crystal structure(4IKM) (CARD domain residues: 413-516) was chosen as the monomer model and was applied to our approach. In our prediction, CARD8-CARD can get a relatively low score in a small rise-rotation angle region, and the results are shown in Figure 3.3.4A.

Compared to the results from MAVS-CARD, the energy minimum of CARD8-CARD covered a smaller area. The rise mainly ranged from around 5.2Å to 5.7Å, and the rotation angle ranged from 98° to 101° in the central area. The predicted model lying in the center of this area, as well as the one with the lowest score, had a 5.4Å rise and 99° as the rotation angle, and was left-handed. Several lowest score models were picked out and shown in Figure 3.3.4C. I successfully prepared the CARD8 CARD sample and obtained the density of CARD8-CARD and solved the CARD8-

CARD filament structure in the subsequent experiment by using traditional cryo-EM methods. The solved cryo-EM CARD8-CARD density and cropped-out 12mer CARD8-CARD filament structure is shown in Figure 3.3.4B, C, which adopts near-identical helical parameters of 5.409Å and 99.16°. The CARD8-CARD filament helical symmetry was consistent with our predicted helical parameter results and near the center point of the low-value energy region, representing one of the most favorable parameters selected in our approach. How we solved the CARD8-CARD filament structure, and more details of the CARD8-CARD filament structure is discussed in Chapter 4.

Figure 3.3.4 CARD8-CARD prediction score and validation structures. (A) Prediction of helical symmetry score of filaments of CARD8-CARD with applying helical parameter (rise 4 Å ~5.9 Å, rotation angle 90° ~109°). (B) The side-view, top-view, and the bottom view of the cropped-out density of CARD8-CARD filament (12mer). (C) The cropped-out CARD8-CARD filament (12mer) structure with fitting the phoenix refined CARD8-CARD monomer structure and the most confident predicted CARD8-CARD filament structures with the input-CARD8-CARD monomer structure from PDB database (PDB ID: 4IKM).

Chapter3.4 Heterogenous (pairwise interaction) study elucidates the “access code” between CARD-CARD interactions

Validation of NOD1-/NOD2-CARD interaction with RIP2-CARD

To extend our method to predict heterogenous CARD-CARD interaction models in innate immunity and possibly to cross-validate the experimental biochemistry data, we also applied our computational approach to predict the heterogenous CARD-CARD pairwise interaction.

Similar to the homogenous filament models building, we adopted two different CARD domain monomer structures to fit into the filament model in a certain helical symmetry. According to our hypotheses, all CARD-CARD cross-talks are undergone in the form of filamentous macromolecules.

The aim of studying pairwise CARD-CARD interaction is to elucidate the crosstalk. And potentially we can use it to explore new CARD-CARD interaction pairs not only for predicting the homotypic CARD-CARD interaction for CARD filaments but also to heterogeneous filamentous oligomers. For heterogenous CARD filaments, the ‘upstream’ and ‘downstream’ CARDS adopt a so-called ‘bottom-up’ orientation in which the Type IIb (top) surface of the ‘upstream CARD’ interacts unidirectionally with the Type IIa (bottom) surface of the ‘downstream’ CARD. In other words, the Type IIb (top) interface of the upstream CARD would interact with the Type IIa (bottom) surface of the downstream CARD. Similarly, ‘top-down’ orientation is just the other way, on the opposite to the ‘bottom-up’ orientation. Our first layer selection system combined Rosetta scoring/relax function for the parameter scores, was applied to study the

pairwise interaction of CARDS, including the orientation of the filaments. Figure 3.4.1 shows that we modeled the hypothetical hetero-oligomeric NOD1-CARD-RIP2-CARD filaments, NOD2-2CARDS-RIP2-CARD filaments. As the structure of human NOD2-2CARDS has not been solved yet, we used the I-TASSER server to predict the NOD2-2CARDS structure separately.

As reported, the N-terminal NOD2-CARDa interacts with RIP2-CARD(Gong, Long et al. 2018).

We built up two pairwise structures to run our scoring selection to see whether the energy score is in agreement with the biochemistry data. Every model went through 10 to 15 repeats in order to get a statistical meaningful result in Figure 3.4.1. The score shows that NOD1-CARD and NOD2-2CARDS are both more physically compatible with RIP2-CARD when it builds up a top-down fusion construct, which is in accordance with the biochemistry data and reported paper that RIP2-CARD forms filaments in a top-down pattern (Figure 3.4.1B). This result shows a preferable direction of the interaction pair for building up a top-down fusion construct as well as the bottom-up fusion construct with only a 12mer filamentous oligomer in a certain helical symmetry. This is composed of one-layer of NOD1-CARD with two layers of RIP2-CARD filaments for the NOD1 hypothetical hetero-model, and one-layer of NOD2-2CARDS each and one-layer of RIP2-CARD. If we had a more precise initial model and included more layers (4 CARDS repeat) in the simulation to replicate the real-world scenario, it is highly convincible that the energy score would perform better, and a more significant statistical result would come out and show the preferable direction.

Prediction of MDA5-2CARDS interaction with MAVS-CARD

MDA5-2CARDS and RIG-I-2CARDS can both detect viral double-stranded RNA and crosstalk with MAVS to conduct inflammatory responses to the virus. We selected the solved RIG-I-

2CARDs-MAVS-CARD filament structures with typical type I, type II, type III interfaces to run a test for our developed modified scoring system to simulate the crosstalk between RIG-I-2CARDs and MAVS-CARD.

RIG-2CARDS interacting with MAVS-CARD has been studied, and the hetero-filaments structure has been solved. However, there has been little study on how MDA5-2CARDs crosstalks with MAVS-CARD. Based on the solved heterogenous crystal complex structure of RIG-2CARDs with MAVS-CAR, we hypothesized that there must be a significant energy preference for the formation of filaments in a certain direction (Figure 3.4.1C,D). We adopted the same helical symmetry to predict MDA5-2CARDs-MAVS-CARD filament structures by using our scoring method. Our score indicates that MDA5-2CARDs are more physically compatible with MAVS-CARD when it builds up in a bottom-up fashion, which is in accordance with the biochemistry data and reported paper that MAVS-CARD forms filaments in a bottom-up pattern, in which the Type IIb (top) surface of the 'MDA5-CARDb' interacts unidirectionally with the Type IIa (bottom) surface of the 'MAVS-CARD' to form the hetero-CARDs-filaments(Peisley, Wu et al. 2013, Wu, Peisley et al. 2014).

Figure 3.4.1 Schematic illustration and score of NOD1: RIP2, NOD2: RIP2, RIG-I: MAVS, MDA5: MAVS pairwise interactions. (A) NOD1: RIP2 and NOD2: RIP2 pairwise helical heterogeneous model of complex. (B) Score showing a preference of interaction pair in the correct complex which is in accordance with the experimental biochemistry data (Gong, Long et al. 2018). (C) RIG-I: MAVS and MDA5: MAVS pairwise helical heterogeneous model of complex. (D) Score showing significant preference of interaction pair with the correct interaction surface

in accordance of the previous study of RIG-I: MAVS heterogenous CARD complex and prediction of MDA5: MAVS showing the same helical heterogenous CARD complex as RIG-I: MAVS complex(Wu, Peisley et al. 2014).

Chapter 3.5 Discussion

Inflammatory response and cell death in innate immunity system are mainly regulated by death domain containing proteins. CARD-containing proteins self-associate and bind with other CARD-containing proteins to form oligomerized CARD filaments to be involved in processes relating to inflammation and apoptosis. Recent findings might be general principle in promoting the assembly of filamentous complexes upon stimulation of CARD domain-driven innate complexes that driven by CARD: CARD interactions to form SMOCs and to mediate innate immune signal transduction. These findings triggered both the structural biologists and immunologists' interest that how CARDS formed filaments and transduce immune signals to downstream adapters.

Solving CARD filament structures by cryo-EM may encounter technical difficulties. Firstly, the difficulties to prepare qualified filament recombinant protein samples. CARDS recombinant proteins tend to exist in monomeric form rather than form the filaments. It often requires strict conditions or other tricky experimental methods such as refolding. Furthermore, unlike monomer structures that can be solved using either X-ray crystallography or NMR spectroscopy, it is hard to get crystallized oligomeric filament samples. Secondly, to prepare cryo-EM qualified grids and to take cryo-EM images can be time, workforce, and even money consuming factors. And also, the equipment of high-resolution cryo-EM like Titan Krois can be a substantial financial cost. Thirdly, obtaining the density map from cryo-EM data can be tricky. Experienced researchers need to use a series of software such as Relion to get a high-resolution density. All these difficulties demonstrate why only a few filaments have been solved.

It was hypothesized that the structures of the helical filaments are critical to illustrate the CARD-CARD interaction in innate immunity. With the development of computational prediction algorithms nowadays, the monomer structure prediction is already possible using software or online servers developed by other groups around the world. Our study aims to establish a new computational approach to elucidate the CARD: CARD interactions for CARDs oligomerized complexes by integrating existing mainstream simulation tools. What we wanted to achieve is to predict the CARD filament structure in helical symmetry, which includes both homogeneous CARD filament structures and heterogeneous pairwise CARD-CARD filaments solely based on previous studies and known biophysical knowledge about the compatibility of structures. This includes using the tools such as Rosetta, I-TASSER, and MD software to generate a practical protocol for predicting the CARDs helical assembling. We hope that this computational approach can help structural biologists get a sign of CARDs filaments' helical pattern and pave the way for model-driven quantitative structural analysis with diverse applications in CARD and other Death Domain family studies.”

However, there are also drawbacks and limitations for our computational methods. First of all, there are only 7 filament structures available as our reference. The variable ranges of rise and rotation ($4-5.9\text{\AA}$, $90-109^\circ$) were chosen to include all possible parameter's combinations based on our hypothesis about the filamentous inflammasome structures. Though known experimental structures of human CARDs (totally 7 in Protein data bank) only have a narrow range of parameters, other CARDs may form filaments outside this range. These uncertainties may acquire additional computational resources to perform the entire simulation. And also, our method assumed all CARDs adopt a C1 symmetry to promote the helical assemblies based on the 7 CARDs filament

structures. However, some other members of DD superfamilies may adopt a C3 symmetry, such as previously mentioned Caspase8 DED, these CARDs candidates may form alternative helical modules. The initial input model may need to be three monomers with a certain orientation and the parameters range needs further refinement and cover a larger range of the rise. More importantly, the mismatch between the hypothesized helical pattern and the actual helical oligomeric pattern may cause the inaccuracy to our prediction result.

Secondly, the rise and rotation have been kept during the prediction because they are the key parameters of determining the CARDs filament structures. Our method is for better guiding the structural study of CARDs and mainly serves as a practical protocol to predict the CARDs filaments by using existing tools rather than develop a completely new method to predict everything out of zero. The helical symmetry parameters include rise, rotation angle, diameter of the filament, the initial position or orientation of the CARD monomer. To design a CARD filament model, the diameter of filament is a relatively unimportant parameter-since 1) If we can have the sample, the diameter of filament can be easily obtained by Negative Stain EM but no cryo-EM. 2) The diameter of helical filament ranges from 85Å to 90Å, which is mostly determined by C1 symmetry and the size of CARD monomer structure. In our study, we fix this diameter to 90Å, to order to study the rise and rotation, but at the same time, we also proves that if the diameter varied from the reality too much, the score map will change greatly and most likely lose a focal point and become less reliable in Figure3.2.2B,C by giving an example enlarging a 5Å to diameter of filament. The initial orientation of individual CARD has been fixed when we screen the rise and rotation, yet it was discussed in Figure 3.2.2D to be introduce back to the simulation by doing an internal rotate to adjust the best orientation for individual CARD. For convenience use, the initial

orientation for CARDS with solved structure was superimposed directly from the PDB database due to the lack of information; For those without solved filament structures, the initial CARDS orientation was superimposed into MAVS-CARD. This may also cause inaccuracy to the prediction results. The predicted CARDS monomer structures may not be accurate enough to perform the simulation to predict their possible helical parameters. It may be due to the lack of information on the correct monomer structures of these candidates or the false initial monomer orientation.

These limitations and drawbacks also provide us a direction that how to improve the method in the future. Our method may expand our method to all death domains and other unique anti-parallel α -helix domains and include more helical symmetries such as C3 symmetry to better predict the other domain subfamilies in DD superfamily, such as PYRIN domain. And also, to find a better way to determine the initial orientation rather than fit into those typical CARDS that are already known. Moreover, compare the redefined helical model in different MD forcefield and more analysis, including RMSDs for individual chains and buried area between chains proving the stability of monomers, to get a better predicted helical filament model. A better assessment of MD simulation convergence is needed if our method wants to include extensive MD simulation. Currently only MAVS-CARD has been assessed of the simulation and protein stability as a successful candidate for getting a predicted model that with little RMSD with the PDB database deposited structure. More models need to be further refined to get more successful candidates to prove the feasibility of our method.

Chapter4: Structural Characterization of CARD8-CARD in Innate Immunity

Chapter4.1 CARD8-CARD structural analysis

Overview of the CARD8 in innate immunity

NOD-like receptor (NLR) proteins are cytosolic sensors and activators of the innate immune response. Many NLRs nucleate large macromolecules that can trigger inflammatory cell death and cytokine secretion as a response to the PAMPs or DAMPs.

In innate immunity, the inflammasome complex relies on homotypic and heterotypic interaction between the death domain. It can be grouped into two categories based on the identity of their DDs: CARD and PYRIN domain. CARD8 is implicated in Mendelian auto-inflammatory diseases and plays an important role in innate immunity signaling.

In cultured cells, human CARD8 is involved in pathways leading to activation of caspase1 directly without the need for adaptor ASC(Ball, Taabazuing et al. 2019). It is reported that CARD8-CARD employs the C-terminal CARD to induce the oligomerization of downstream CASP1-CARD. To elucidate the structural basis of these specific CARD-CARD interactions that drive CARD8 inflammasome activation (CARD8-CARD with CASP1-CARD), we sought ways to directly visualize how CARD8-CARD nucleate the oligomerization of their respective transducers and solved the structure of recombinant CARD8-CARD using cryo-EM. CARD8-CARD forms

filaments in a concentration-dependent manner for both CARD8-CARD and CARD8-CARD-SNAP constructs.

In our first attempt, CARD8-CARD-SNAP was chosen to be the ideal construct to solve the structure because, with the existence of SNAP-tag, the FLAG-CARD8-3C-CARD construct is more stable and relatively easier to extend to form a long filament. We did get straight long filament and observed CARD8-CARD-SNAP in cryo-EM. We collected the entire set of cryo-EM data of CARD8-CARD-SNAP and tried to solve the filament structure. However, we found that it is quite challenging to extract 3D information out of the fuzzy 2D classification due to the flexibility of SNAP-tag.

Architecture design of CARD8-CARD construct

We designed some other constructs of CARD8-CARD in order to solve this flexible -SNAP issue. And we successfully got a construct with a FLAG tag in the N-terminal before the CARD8-CARD, followed by an HRV-3C-Protease cleavage site and a SNAP-tag in the C-terminus. 3C site is introduced here to facilitate the downstream removal of the SNAP tag for structural studies. SNAP tag would be on the outer surface of the filament and they might interact with the filament in a non-specific and non-repetitive manner, interfering with the symmetry and helical repetition of the filament. This would then impact the efficiency of downstream helical reconstruction during data processing. Therefore, SNAP tag is removed via 3C cleavage before imaging.

We successfully digested the CARD8-3C-CARD sample and kept the filament straight and intact. The FLAG tag was small enough and could not be observed under cryo-EM. The FLAG-CARD8-

3C-CARD (also referred to as CARD8-CARD alone/without SNAP-tag) filaments are typically ~8 nm-wide, and CARD8-CARD-SNAP filaments are ~12 nm-wide. The architecture of two CARD8 constructs and identical cryo-EM images taken for two CARD8 constructs are shown below in Figure 4.1.1A. Two identical cryo-EM images for both CARD8-CARD-SNAP and CARD8-CARD are shown in Figure 4.1.1B, C. More details related to the plasmid construction, protein purification, oligomerization assay for obtaining the filament, and cryo-electron microscopy are in the *Materials and methods (chapter 2.2)*.

Figure 4.1.1 The architecture of CARD8 and comparison of cryo-EM image of CARD8-CARD-SNAP and CARD8-CARD. (A) Domain organization of CARD8. (B) Cryo-EM raw image of recombinant CARD8-CARD with SNAP tag. (C) Cryo-EM raw image of recombinant CARD8-CARD with 3c-cleaved SNAP tag.

Data processing for CARD8-CARD

Using the helical reconstruction module in RELION 3.0, we picked and sectioned the filaments to get CARD8 particles, followed by iterative rounds of 2D classification, 3D classification, and 3D refinement. Some of the results obtained during 2D classification during the initial rounds of CARD8-CARD are shown in Figure 4.1.2. More details about the data processing are in the *Materials and methods (chapter 2.2)*.

Figure 4.1.2 Cryo-EM structure determination of CARD8-CARD filaments. (A) Representative images of CARD8-CARD filaments by cryo-EM. Green colour indicates the coordinates for manually picked filaments. (B) Power spectrum map of CARD8-CARD cryo-EM micrographs (obtained by periodogram) showing the resolution to which Thon rings are visible. (C) 2D classification of CARD8-CARD filaments. (D) Helical layer line profile from Relion analysis of CARD8-CARD filaments. The last signal peak (from left to right) indicates the position of the last (furthest) layer line, which could be used to calculate helical rise parameter.

Cryo-EM analyses of homogeneously non-bundled CARD8-CARD filaments followed by helical reconstruction yielded densities of 3.7 Å resolution for CARD8-CARD filaments (Figure 4.1.3A). The final polished densities demonstrated a clear α -helical pitch and supported the unambiguous assignment of all bulky side chains (Figure 4.1.3B).

Figure 4.1.3 *CARD8-density and sidechain fitting. (A) FSC plot of final refined CARD8-CARD filament density (EMD-9948) showing that the resolution for the overall density of the entire segment is 3.7 Å, at FSC=0.143 gold standard (B) Post-refinement 3D density of CARD8-CARD filaments.*

CARD8-CARD filament structure analysis

CARD8-CARD filaments were rigid and adopted a single conformation. The overall architecture of CARD8-CARD filaments adopts near-identical helical parameters of 5.409 Å and -99.16° as for CARD8. The cryo-EM raw image and the structure of CARD8-CARD filament, including the top- and cross-section view of the CARD8-CARD 2D average are shown in Figure 4.1.4.

Figure 4.1.4 Cryo-EM structures of CARD8-CARD filaments.(A) Cryo-EM raw image of recombinant CARD8-CARD filaments. (B)2D average of CARD8-CARD filaments. (C) Cross-section and top view of CARD8-CARD filaments. (D) 3D class of helical reconstructed CARD8-CARD filament density. 5 adjacent CARD8-CARD monomers are colored.

Based on our cryo-EM density of CARD8-CARD, all three interfaces (Type I-III) that are characteristic of oligomerized CARD8-CARD filaments could be clearly (Figure 4.1.4B). As compared to other CARD filaments or active signaling complexes, CARD8-CARD filaments demonstrate extraordinarily strong densities at the Type II interface. In the CARD8-CARD filament, two analogous pairs of interacting residues, Tyr527 with Leu512 and Asp525 with Lys508, form junctional interactions across the Type II interface. In CARD8-CARD filament, the ‘upstream’ and ‘downstream’ CARD adopt a so-called ‘bottom-up’ orientation in which the Type IIb (top) surface of the ‘upstream CARD’ interacts unidirectionally with the Type IIa (bottom) surface of the ‘downstream CARD. We conjectured CARD8 to engage their respective downstream CARDS in the same manner. In other words, and the Type IIb (top) interface of CARD8-CARD would interact with the Type IIa (bottom) surface of the downstream CARD. In CARD8-CARD filaments, the Type IIb (top) surface is predominantly negatively charged (Tyr527 and Asp525) whereas the Type IIa (bottom) surface is positive (Lys508 and Leu512) (Figure 4.1.5 E).

Figure 4.1.5 CARD8-CARD filament structure analysis. (A) general 2D representation of CARD8 filament (B) Type I-III interfaces in CARD8-CARD filaments. (C) Type II interface density and residues in CARD8-CARD Type II interface (PDB ID: 6K9F, EMD-9948). Atomic model cartoon and the directly observed density map of CARD8-

CARD dimer in the filament, centered at the Type II junction interface. The four residues, Asp525-Tyr527 on the Type IIb interface and the Lys508-Leu512 on the Type IIa interface were highlighted and shown in the 'stick' display mode (D) Assignment of individual residues and α -helices in CARD8-CARD structure.(E) Electrostatic potential map of the top view and bottom view (Type IIb interface, highlighted in dashed area) of CARD8-CARD filament, negative, red, -10 C/m²; positive, blue, 10 C/m².

Based on these observations, we propose that the Type IIa and IIb surfaces, which are discernible via primary sequence alignment, uncover a structural mechanism by which human CARD8 discriminate between downstream partner CARDS (Figure 4.1.6A). A characteristic tyrosine residue is conserved on the Type IIb (top) surface of CARD8-CARD (Tyr527), ASC-CARD (Tyr187) and CASP1-CARD (Tyr82), playing a major role in stabilizing the Type II junctional interaction. This feature is particularly noteworthy, as Type II interface interactions have been proposed to be involved in heterotypic CARD-CARD interactions between ASC and CASP1(Lu, Li et al. 2016, Li, Fu et al. 2018). CARD8 activates pro-caspase-1 directly and independently of ASC to induce pyroptotic cell death and robust IL-1 secretion in human cell(Ball, Taabazuing et al. 2019). This discovery then enabled us to model of all pairwise CARD-CARD interactions amongst CARD8, ASC (PDB ID:6K99), CASP1(PDB ID:5FNA) across the Type II interface. A striking level of specificity became apparent, as some CARD-CARD pairs were structurally favorable while others were incompatible. CARD8-CARD is highly compatible with CASP1-CARD, as Tyr527 and Asp525 on the CARD8-CARD Type IIb (top) surface form a bridge with CASP1-CARD Type IIa (bottom) surface residues Gln67 and Lys64 (Figure 4.1.6D). And these same two bottom CASP1-CARD residues are responsible for interacting with two other upstream CARDS: ASC-CARD (Type IIb residues: Tyr187 and Gln185) and CASP1-CARD itself (Type IIb residue: Asp80) (Fig. 4.1.6D, Fig. 4.1.6C). CARD8-CARD Type IIb (top) residues lack specific interactions with ASC-CARD Type IIa (bottom) surface (Fig. 4.1.6D). Thus CARD8-CARD is unable to engage ASC-CARD. For a potential CARD-CARD interaction to occur, the Type IIb (top) surface of the upstream CARD must be structurally compatible with the Type IIa (bottom)

surface of the downstream CARD. More generally, these findings imply that the Type IIa and IIb surfaces might work as key modular determinants in other inflammasome CARD assemblies.

Figure 4.1.6 Sequence alignment and Type IIa-b interface for inflammasome CARD assemblies. (A) Sequence alignment of human NLRP1-, ASC-, NLRC4-, CARD8- and CASP1-CARD (B) Type II interface in ASC-CARD filaments (PDB ID: 6K99, EMD-9947). (C) CASP1-CARD filaments (PDB ID: 5FNA, EMD-3241). (D) Differing structural compatibilities of potential pairwise CARD-CARD interactions among human CARD8, ASC, and CASP1.

Chapter 4.2 Discussion

We demonstrate that CARD8 employs oligomerized CARDS as a signaling platform to initiate inflammasome signaling, which is in accordance with our hypothesis that all CARDS tend to oligomerize into a filamentous macromolecule to conduct signaling in innate immunity. And our solved CARD8-CARD filament adopts the helical symmetry of 5.409 Å and 99.16° which matches our simulated CARD8-CARD helical symmetry, which can serve as a cross-validation of our computational approach.

By resolving the cryo-EM structures of CARD8-CARD filaments, we discovered that Type II interface functions driving inflammasome CARD8 assemblies and preferentially engaging CASP1-CARD. CARD8-CARD is structurally compatible with CASP1-CARD and the conserved type II interface interactions were found to be key modular determinants in heterotypic CARD-CARD interactions. This also proves that CARD8 is involved in pathways leading to activation of Caspase1 directly without the need for adaptor ASC. These new findings also suggest that any potential CARD8 involved interaction to occur, the Type IIb (top) surface of the upstream CARD must be structurally compatible with the Type IIa (bottom) surface of the downstream CARD. More generally, the previous solved 7 CARDS filaments, Type I interactions between certain residues were the strongest and the key modular determinants were Type Ia and IIb surfaces, for example, MAVS forms a strong density bridge between Arg43 and Typ56 between two Type I surfaces. These findings of CARD8 inflammasomes also imply that apart from Type I interaction, the Type IIa and IIb surfaces work as key modular determinants in other inflammasome CARD assemblies.

Extending this analysis, Type II surfaces of different inflammasome CARD complexes show a high diversity, leading to distinct Type II CARD-CARD interactions for each CARD. The Type II CARD-CARD interactions can function to propagate CARDs self-assembly, and to form pairwise CARD-CARD specificities with multiple downstream inflammasome sensors, such as adaptors or effectors. And these distinct downstream inflammasome complexes form distinct SMOCs and lead to distinct biological outcomes.

Furthermore, our solved CARD8-CARD filaments structure helps to understand how human inflammasome components assemble to initiate, transduce and propagate pyroptosis signaling, and potentially lead to new drug design for therapeutic purposes.

Chapter 5. Conclusions and future directions

Chapter 5.1 Conclusions

Inflammatory response and cell death in innate immunity system are mainly regulated by death domain containing proteins. CARD-containing proteins self-associate and bind with other CARD-containing proteins to form oligomerized complexes to be involved in processes relating to inflammation and apoptosis. CARD domain-driven innate complexes that driven by CARD: CARD interactions are highly versatile and no general assembly principles for other CARD assemblies have been established yet.

Our computational study aims to establish a new computational approach to elucidate the CARD: CARD interactions for CARDs oligomerized complexes. The method proceeds with an initial CARD monomer model obtained, for example, by fragment-based prediction (I-TASSER, Rosetta) or Protein Data Bank as the initial input. With applying to helical symmetry, a pool of oligomerized CARD model candidates with different helical parameters (rise, rotation) can be automatically screened and the grid search method can reject poorly placed oligomerized models and single out candidates with promising parameters for further processing. For MAVS-CARD, we prove that our method can make a prediction of the oligomerized CARD complexes in helical symmetry with little tolerate *RMSD* and deviation between deposited Cryo-EM structures from Protein Data Bank with Rosetta Monte Carlo refinement and molecular dynamics. And we also demonstrate its functionality by predicting oligomerized CARDs helical parameters and the correct direction of heterogenous CARD: CARD complexes for NOD2-2CARD and RIP2-CARD, RIG-I-

2CARD(MDA5-2CARD) and MAVS-CARD. For all human CARDS (with or without solved CARD filament structures), our method provides a conceptual framework for predicting their possibilities to form helical oligomerized complexes, and if so, their helical parameters (rise, rotation).

Our results thus suggest that CARDS have distinct CARD: CARD interaction mechanisms for CARD assembly can be predicted with computational modelling and known biophysical knowledge, such as the residue-residue attractive/repulsive interaction in a particular force field. The current existing computational approaches, such as ROSIE (Rosetta online server), GalaxyWEB, and I-TASSER server, can predict the monomer structures with high confidence by template-based searching and local optimization methods. However, although some of these tools claim they can predict oligomer structures, their performance turns out to be less reliable. Our computational approach can help structural biologists' study oligomerized CARDS complexes' helical pattern and pave the way for model-driven quantitative structural analysis with diverse applications in CARD and other Death Domain family studies. And Rosetta de novo modelling can reduce the demand for high-resolution data for macromolecule of CARD complexes.

For the structural study of CARD8, we demonstrate that CARD8 employs oligomerized CARDS (helical symmetry, rise 5.409 Å and rotation angle 99.16°, left-handed) as a signaling platform to initiate inflammasome signaling. Extraordinarily strong densities at the Type II interface of CARD8-CARD filaments and highly compatible interfaces with CASP1-CARD suggest the type IIa and IIb surfaces work as key modular determinants in both homotypic and heterotypic CARD: CARD interactions for CARD8 inflammasome CARD assemblies. This discovery provides a

unified conceptual framework to understand how inflammasome components assemble to initiate, transduce, and propagate pyroptosis signaling. As inflammasome dysregulation is increasingly implicated in human diseases, these findings might lead to new avenues to modulate the inflammasome for therapeutic purposes.

Chapter 5.2 Future directions

For the computational prediction approach, we could expand our hypothesis that all members from the Death domain superfamily are conducting inflammatory response and transducing signaling in filamentous macromolecules. It could be a universal rule that all families, including PYRIN domains, death domains, and death effector domains, all adopt similar helical filaments structures regardless that they self-interact (convergent) or interact with other members (divergent). With continuous efforts that scientists put on the death domain, there have been already several filament structures of PYRIN domains that could conduct a similar inflammatory response like CARD domains do. Therefore, for the computational approach, we could expand our method to all death domains and other unique anti-parallel α -helix domains, such as the TIR domain and the BIR domain. Potentially we could apply our approach to predict homogeneous potential filament structures of the other death domain members and the heterogeneous filamentous structures when crosstalks happen in the DD-DD interaction and signal transducing in innate immunity based on known biophysical knowledge about the compatibility of structures. With more and more cases

and filament structures have been solved, the parameters of our computational method can be trained and updated so that the prediction can be more precise.

All other anti-parallel α -helical domains can be potential candidates for our computational approach. We can further hypothesize that all anti-parallel α -helical domains may oligomerize and function similarly in vivo.

Furthermore, evolutionary pattern recognition methods, such as using deep sequencing to identify the conserved sequence pattern, have been a rising technique for structure prediction. By deep sequencing, the protein candidates and digging out the conserved residues pairs in an evolutionary view can generate useful information to rebuild the structures without actually purifying the protein so that it can be the future of the structure prediction. Death Domain members may be good candidates to explore the new techniques.

For the CARD8 study, we could expand our study with more biochemistry experiments to study the function of CARD8 in innate immunity. We could visualize the crosstalk between CARD8 and the ‘downstream’ and ‘upstream’ CARD forming filamentous macromolecules by expressing them with certain tags in 239T cell lines. Studying the full length and other domains of CARD8 can also be part of the expansion plan. Mice models can be used to study the CARD8 function in vivo, as well.

References

- Akira, S. and K. Takeda (2004). "Toll-like receptor signalling." Nat Rev Immunol **4**(7): 499-511.
- Albrecht, M., D. Choubey and T. Lengauer (2005). "The HIN domain of IFI-200 proteins consists of two OB folds." Biochem Biophys Res Commun **327**(3): 679-687.
- Alford, R. F., A. Leaver-Fay, J. R. Jeliazkov, M. J. O'Meara, F. P. DiMaio, H. Park, M. V. Shapovalov, P. D. Renfrew, V. K. Mulligan, K. Kappel, J. W. Labonte, M. S. Pacella, R. Bonneau, P. Bradley, R. L. Dunbrack, Jr., R. Das, D. Baker, B. Kuhlman, T. Kortemme and J. J. Gray (2017). "The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design." J Chem Theory Comput **13**(6): 3031-3048.
- Ball, D. P., C. Y. Taabazuing, A. R. Griswold, E. L. Orth, S. D. Rao, I. B. Kotliar, D. C. Johnson and D. A. Bachovchin (2019).
- Benko, S., J. G. Magalhaes, D. J. Philpott and S. E. Girardin (2010). "NLRC5 limits the activation of inflammatory pathways." J Immunol **185**(3): 1681-1691.
- Beroukhim, R. and N. Unwin (1997). "Distortion correction of tubular crystals: improvements in the acetylcholine receptor structure." Ultramicroscopy **70**(1-2): 57-81.
- Beutler, L. E., L. G. Castonguay and W. C. Follette (2006). "Therapeutic factors in dysphoric disorders." J Clin Psychol **62**(6): 639-647.
- Bianchi, M. E. (2007). "DAMPs, PAMPs and alarmins: all we need to know about danger." J Leukoc Biol **81**(1): 1-5.
- Bowman, G. R. and V. S. Pande (2009). "Simulated tempering yields insight into the low-resolution Rosetta scoring functions." Proteins **74**(3): 777-788.
- Brooks, B. R., C. L. Brooks, 3rd, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus (2009). "CHARMM: the biomolecular simulation program." J Comput Chem **30**(10): 1545-1614.
- Brubaker, S. W., K. S. Bonham, I. Zanoni and J. C. Kagan (2015). "Innate immune pattern recognition: a cell biological perspective." Annu Rev Immunol **33**: 257-290.
- Brunette, R. L., J. M. Young, D. G. Whitley, I. E. Brodsky, H. S. Malik and D. B. Stetson (2012). "Extensive evolutionary and functional diversity among mammalian AIM2-like receptors." J Exp Med **209**(11): 1969-1983.

Burckstummer, T., C. Baumann, S. Bluml, E. Dixit, G. Durnberger, H. Jahn, M. Planyavsky, M. Bilban, J. Colinge, K. L. Bennett and G. Superti-Furga (2009). "An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome." Nat Immunol **10**(3): 266-272.

Caruso, R., N. Warner, N. Inohara and G. Nunez (2014). "NOD1 and NOD2: signaling, host defense, and inflammatory disease." Immunity **41**(6): 898-908.

Celhar, T., R. Magalhaes and A. M. Fairhurst (2012). "TLR7 and TLR9 in SLE: when sensing self goes wrong." Immunol Res **53**(1-3): 58-77.

Chaudhury, S., S. Lyskov and J. J. Gray (2010). "PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta." Bioinformatics **26**(5): 689-691.

Conway, P., M. D. Tyka, F. DiMaio, D. E. Konerding and D. Baker (2014). "Relaxation of backbone bond geometry improves protein energy landscape modeling." Protein Sci **23**(1): 47-55.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman (1995). "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules." J Am Chem Soc **117**(19): 5179-5197.

Cui, S., K. Eisenacher, A. Kirchhofer, K. Brzozka, A. Lammens, K. Lammens, T. Fujita, K. K. Conzelmann, A. Krug and K. P. Hopfner (2008). "The C-terminal regulatory domain is the RNA 5'-triphosphate sensor of RIG-I." Mol Cell **29**(2): 169-179.

De Rosier, D. and A. Klug (1968). "Reconstruction of three dimensional structures from electron micrographs." Nature **217**(5124): 130.

Decherchi, S., M. Masetti, I. Vyalov and W. Rocchia (2015). "Implicit solvent methods for free energy estimation." Eur J Med Chem **91**: 27-42.

Dick, M. S., L. Sborgi, S. Ruhl, S. Hiller and P. Broz (2016). "ASC filament formation serves as a signal amplification mechanism for inflammasomes." Nat Commun **7**: 11929.

Dubochet, J., M. Adrian, J. J. Chang, J. C. Homo, J. Lepault, A. W. McDowell and P. Schultz (1988). "Cryo-electron microscopy of vitrified specimens." Q Rev Biophys **21**(2): 129-228.

Dutta, S., M. B. H and F. B. W (2007). "Using the tools and resources of the RCSB protein data bank." Curr Protoc Bioinformatics **Chapter 1**: Unit1 9.

Egelman, E. H. (2000). "A robust algorithm for the reconstruction of helical filaments using single-particle methods." Ultramicroscopy **85**(4): 225-234.

Egelman, E. H. (2015). "Three-dimensional reconstruction of helical polymers." Arch Biochem Biophys **581**: 54-58.

Eisenbarth, S. C., A. Williams, O. R. Colegio, H. Meng, T. Strowig, A. Rongvaux, J. Henao-Mejia, C. A. Thaiss, S. Joly, D. G. Gonzalez, L. Xu, L. A. Zenewicz, A. M. Haberman, E. Elinav, S. H. Kleinstein, F. S. Sutterwala and R. A. Flavell (2012). "NLRP10 is a NOD-like receptor essential to initiate adaptive immunity by dendritic cells." Nature **484**(7395): 510-513.

Faraldo-Gomez, J. D., L. R. Forrest, M. Baaden, P. J. Bond, C. Domene, G. Patargias, J. Cuthbertson and M. S. Sansom (2004). "Conformational sampling and dynamics of membrane proteins from 10-nanosecond computer simulations." Proteins **57**(4): 783-791.

Fermi E., P. J., Ulam S. (1955). Los Alamos report LA-1940

Fernandez-Leiro, R. and S. H. Scheres (2016). "Unravelling biological macromolecules with cryo-electron microscopy." Nature **537**(7620): 339.

Fernandez-Moran, H. (1960). "Low-temperature preparation techniques for electron microscopy of biological specimens based on rapid freezing with liquid helium II." Ann N Y Acad Sci **85**: 689-713.

Ferrao, R. and H. Wu (2012). "Helical assembly in the death domain (DD) superfamily." Curr Opin Struct Biol **22**(2): 241-247.

Fleishman, S. J., A. Leaver-Fay, J. E. Corn, E. M. Strauch, S. D. Khare, N. Koga, J. Ashworth, P. Murphy, F. Richter, G. Lemmon, J. Meiler and D. Baker (2011). "RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite." PLoS One **6**(6): e20161.

Fontalba, A., V. Martinez-Taboada, O. Gutierrez, C. Pipaon, N. Benito, A. Balsa, R. Blanco and J. L. Fernandez-Luna (2007). "Deficiency of the NF-kappaB inhibitor caspase activating and recruitment domain 8 in patients with rheumatoid arthritis is associated with disease severity." J Immunol **179**(7): 4867-4873.

Franchi, L., T. Eigenbrod, R. Munoz-Planillo, U. Ozkurede, Y. G. Kim, C. Arindam, M. Gale, Jr., R. H. Silverman, M. Colonna, S. Akira and G. Nunez (2014). "Cytosolic double-stranded RNA activates the NLRP3 inflammasome via MAVS-induced membrane permeabilization and K⁺ efflux." J Immunol **193**(8): 4214-4222.

Frank, J. (1975). "Averaging of low exposure electron micrographs of non-periodic objects." Ultramicroscopy **1**(2): 159-162.

Freddolino, P. L., A. S. Arkhipov, S. B. Larson, A. McPherson and K. Schulten (2006). "Molecular dynamics simulations of the complete satellite tobacco mosaic virus." Structure **14**(3): 437-449.

Gong, Q., Z. Long, F. L. Zhong, D. E. T. Teo, Y. Jin, Z. Yin, Z. Z. Boo, Y. Zhang, J. Zhang, R. Yang, S. Bhushan, B. Reversade, Z. Li and B. Wu (2018). "Structural basis of RIP2 activation and signaling." Nat Commun **9**(1): 4993.

Gordon, S. (2002). "Pattern recognition receptors: Doubling up for the innate immune response." Cell **111**(7): 927-930.

Gray, E. E., D. Winship, J. M. Snyder, S. J. Child, A. P. Geballe and D. B. Stetson (2016). "The AIM2-like Receptors Are Dispensable for the Interferon Response to Intracellular DNA." Immunity **45**(2): 255-266.

Haile, J. M. (1978). "Molecular-Dynamics Simulations of Simple Fluids with 3-Body Interactions Included." J Am Chem Soc **175**(Mar): 11-11.

He, S. and S. H. W. Scheres (2017). "Helical reconstruction in RELION." J Struct Biol **198**(3): 163-176.

Hess, B. (2002). "Convergence of sampling in protein simulations." Phys Rev E Stat Nonlin Soft Matter Phys **65**(3 Pt 1): 031910.

Hofmann, K. and P. Bucher (1997). "The CARD domain: A new apoptotic signalling motif." Trends Biochem Sci **22**(5): 155-156.

Hornung, V., J. Ellegast, S. Kim, K. Brzozka, A. Jung, H. Kato, H. Poeck, S. Akira, K. K. Conzelmann, M. Schlee, S. Endres and G. Hartmann (2006). "5'-Triphosphate RNA is the ligand for RIG-I." Science **314**(5801): 994-997.

Hou, F., L. Sun, H. Zheng, B. Skaug, Q. X. Jiang and Z. J. Chen (2011). "MAVS forms functional prion-like aggregates to activate and propagate antiviral innate immune response." Cell **146**(3): 448-461.

Inohara, N. and G. Nunez (2001). "The NOD: a signaling module that regulates apoptosis and host defense against pathogens." Oncogene **20**(44): 6473-6481.

Jin, T., A. Perry, J. Jiang, P. Smith, J. A. Curry, L. Unterholzner, Z. Jiang, G. Horvath, V. A. Rathinam, R. W. Johnstone, V. Hornung, E. Latz, A. G. Bowie, K. A. Fitzgerald and T. S. Xiao (2012). "Structures of the HIN domain:DNA complexes reveal ligand binding and activation mechanisms of the AIM2 inflammasome and IFI16 receptor." Immunity **36**(4): 561-571.

Jones, J. D. and J. L. Dangl (2006). "The plant immune system." Nature **444**(7117): 323-329.

Jorgensen, W. L., D. S. Maxwell and J. Tirado-Rives (1996). "Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids." J Am Chem Soc **118**(45): 11225-11236.

Jounai, N., K. Kobiyama, F. Takeshita and K. J. Ishii (2012). "Recognition of damage-associated molecular patterns related to nucleic acids during inflammation and vaccination." Front Cell Infect Microbiol **2**: 168.

Kagan, J. C., T. Su, T. Horng, A. Chow, S. Akira and R. Medzhitov (2008). "TRAM couples endocytosis of Toll-like receptor 4 to the induction of interferon-beta." Nat Immunol **9**(4): 361-368.

Karplus, M. and J. A. McCammon (2002). "Molecular dynamics simulations of biomolecules." Nat Struct Biol **9**(9): 646-652.

Kawai, T. and S. Akira (2010). "The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors." Nat Immunol **11**(5): 373-384.

Kawasaki, T. and T. Kawai (2014). "Toll-like receptor signaling pathways." Front Immunol **5**: 461.

Keestra-Gounder, A. M., M. X. Byndloss, N. Seyffert, B. M. Young, A. Chavez-Arroyo, A. Y. Tsai, S. A. Cevallos, M. G. Winter, O. H. Pham, C. R. Tiffany, M. F. de Jong, T. Kerrinnes, R. Ravindran, P. A. Luciw, S. J. McSorley, A. J. Baumler and R. M. Tsolis (2016). "NOD1 and NOD2 signalling links ER stress with inflammation." Nature **532**(7599): 394-397.

Kell, A. M. and M. Gale, Jr. (2015). "RIG-I in RNA virus recognition." Virology **479-480**: 110-121.

Kersse, K., J. Verspurten, T. V. Berghe and P. Vandenabeele (2011). "The death-fold superfamily of homotypic interaction motifs." Trends Biochem Sci **36**(10): 541-552.

Khare, S., A. Dorfleutner, N. B. Bryan, C. Yun, A. D. Radian, L. de Almeida, Y. Rojanasakul and C. Stehlik (2012). "An NLRP7-containing inflammasome mediates recognition of microbial lipopeptides in human macrophages." Immunity **36**(3): 464-476.

Khatib, F., S. Cooper, M. D. Tyka, K. Xu, I. Makedon, Z. Popovic, D. Baker and F. Players (2011). "Algorithm discovery by protein folding game players." Proc Natl Acad Sci U S A **108**(47): 18949-18953.

Kim, Y. G., J. H. Park, M. H. Shaw, L. Franchi, N. Inohara and G. Nunez (2008). "The cytosolic sensors Nod1 and Nod2 are critical for bacterial recognition and host defense after exposure to Toll-like receptor ligands." Immunity **28**(2): 246-257.

Kim, Y. K., J. S. Shin and M. H. Nahm (2016). "NOD-Like Receptors in Infection, Immunity, and Diseases." Yonsei Med J **57**(1): 5-14.

Kingsmore, S. F., J. Snoddy, D. Choubey, P. Lengyel and M. F. Seldin (1989). "Physical mapping of a family of interferon-activated genes, serum amyloid P-component, and alpha-spectrin on mouse chromosome 1." Immunogenetics **30**(3): 169-174.

Ko, J., H. Park, L. Heo and C. Seok (2012). "GalaxyWEB server for protein structure prediction and refinement." Nucleic Acids Res **40**: W294-297.

Kohl, A. and M. G. Grutter (2004). "Fire and death: the pyrin domain joins the death-domain superfamily." C R Biol **327**(12): 1077-1086.

Kortemme, T., A. V. Morozov and D. Baker (2003). "An Orientation-dependent Hydrogen Bonding Potential Improves Prediction of Specificity and Structure for Proteins and Protein-Protein Complexes." J Mol Biol **326**(4): 1239-1259.

Kwon, D., J. H. Yoon, S. Y. Shin, T. H. Jang, H. G. Kim, I. So, J. H. Jeon and H. H. Park (2012). "A comprehensive manually curated protein-protein interaction database for the Death Domain superfamily." Nucleic Acids Res **40**: D331-336.

Laroui, H., Y. Yan, Y. Narui, S. A. Ingersoll, S. Ayyadurai, M. A. Charania, F. Zhou, B. Wang, K. Salaita, S. V. Sitaraman and D. Merlin (2011). "L-Ala-gamma-D-Glu-meso-diaminopimelic acid (DAP) interacts directly with leucine-rich region domain of nucleotide-binding oligomerization domain 1, increasing phosphorylation activity of receptor-interacting serine/threonine-protein kinase 2 and its interaction with nucleotide-binding oligomerization domain 1." J Biol Chem **286**(35): 31003-31013.

Latty, S. L., J. Sakai, L. Hopkins, B. Verstak, T. Paramo, N. A. Berglund, E. Cammarota, P. Cicuta, N. J. Gay, P. J. Bond, D. Klenerman and C. E. Bryant (2018). "Activation of Toll-like receptors nucleates assembly of the MyDDosome signaling hub." Elife **7**.

Leaver-Fay, A., M. J. O'Meara, M. Tyka, R. Jacak, Y. Song, E. H. Kellogg, J. Thompson, I. W. Davis, R. A. Pache, S. Lyskov, J. J. Gray, T. Kortemme, J. S. Richardson, J. J. Havranek, J. Snoeyink, D. Baker and B. Kuhlman (2013). "Scientific benchmarks for guiding macromolecular energy function improvement." Methods Enzymol **523**: 109-143.

Lee, M. S. and M. Feig (2003). "New trends in generalized born implicit solvent methods." J Am Chem Soc **225**: U791-U791.

Lee, M. S. and M. Olson (2004). "New hybrid explicit/implicit solvent method for biomolecular simulations." J Am Chem Soc **228**: U507-U507.

Lemaitre, B., E. Nicolas, L. Michaut, J. M. Reichhart and J. A. Hoffmann (1996). "The dorsoventral regulatory gene cassette spatzle/Toll/cactus controls the potent antifungal response in Drosophila adults." Cell **86**(6): 973-983.

- Levinsohn, J. L., Z. L. Newman, K. A. Hellmich, R. Fattah, M. A. Getz, S. Liu, I. Sastalla, S. H. Leppla and M. Moayeri (2012). "Anthrax lethal factor cleavage of Nlrp1 is required for activation of the inflammasome." PLoS Pathog **8**(3): e1002638.
- Li, Y., T. M. Fu, A. Lu, K. Witt, J. Ruan, C. Shen and H. Wu (2018). "Cryo-EM structures of ASC and NLRC4 CARD filaments reveal a unified mechanism of nucleation and activation of caspase-1." Proc Natl Acad Sci U S A **115**(43): 10845-10852.
- Liepinsh, E., R. Barbals, E. Dahl, A. Sharipo, E. Staub and G. Otting (2003). "The death-domain fold of the ASC PYRIN domain, presenting a basis for PYRIN/PYRIN recognition." J Mol Biol **332**(5): 1155-1163.
- Lin, S. C., Y. C. Lo and H. Wu (2010). "Helical assembly in the MyD88-IRAK4-IRAK2 complex in TLR/IL-1R signalling." Nature **465**(7300): 885-890.
- Lindahl, E. and M. S. P. Sansom (2008). "Membrane proteins: molecular dynamics simulations." Curr Opin Struct Biol **18**(4): 425-431.
- Loo, Y.-M. and M. Gale (2011). "Immune Signaling by RIG-I-like Receptors." Immunity **34**(5): 680-692.
- Lu, A., Y. Li, F. I. Schmidt, Q. Yin, S. Chen, T. M. Fu, A. B. Tong, H. L. Ploegh, Y. Mao and H. Wu (2016). "Molecular basis of caspase-1 polymerization and its inhibition by a new capping mechanism." Nat Struct Mol Biol **23**(5): 416-425.
- Luo, D., A. Kohlway and A. M. Pyle (2014). "Duplex RNA activated ATPases (DRAs)." RNA Biology **10**(1): 111-120.
- Maekawa, S., U. Ohto, T. Shibata, K. Miyake and T. Shimizu (2016). "Crystal structure of NOD2 and its implications in human disease." Nat Commun **7**.
- McCammon, J. A., B. R. Gelin and M. Karplus (1977). "Dynamics of folded proteins." Nature **267**(5612): 585-590.
- McEntyre, J. R. and T. J. Gibson (2004). "Patterns and clusters within the PSM column in TiBS, 1992-2004." Trends Biochem Sci **29**(12): 627-633.
- Medzhitov, R. (2001). "Toll-like receptors and innate immunity." Nat Rev Immunol **1**(2): 135-145.
- Medzhitov, R. and C. Janeway, Jr. (2000). "Innate immune recognition: mechanisms and pathways." Immunol Rev **173**(1): 89-97.
- Medzhitov, R., I. R. Mackay, F. S. Rosen and C. Janeway (2000). "Innate Immunity." N Engl J Med **343**(5): 338-344.

Medzhitov, R., P. Preston-Hurlburt and C. A. Janeway, Jr. (1997). "A human homologue of the Drosophila Toll protein signals activation of adaptive immunity." Nature **388**(6640): 394-397.

Meylan, E., J. Tschopp and M. Karin (2006). "Intracellular pattern recognition receptors in the host response." Nature **442**(7098): 39-44.

Milne, J. L., M. J. Borgnia, A. Bartesaghi, E. E. Tran, L. A. Earl, D. M. Schauder, J. Lengyel, J. Pierson, A. Patwardhan and S. Subramaniam (2013). "Cryo-electron microscopy—a primer for the non-microscopist." The FEBS J **280**(1): 28-45.

Molnar, S. P. (1999). "The art of molecular dynamics simulation (vol 76, pg 171, 1999)." J Chem Educ **76**(4): 471-471.

Moretti, R., S. Lyskov, R. Das, J. Meiler and J. J. Gray (2018). "Web-accessible molecular modeling with Rosetta: The Rosetta Online Server that Includes Everyone (ROSIE)." Protein Sci **27**(1): 259-268.

Motta, V., F. Soares, T. Sun and D. J. Philpott (2015). "NOD-like receptors: versatile cytosolic sentinels." Physiol Rev **95**(1): 149-178.

Murali, A., X. Li, C. T. Ranjith-Kumar, K. Bhardwaj, A. Holzenburg, P. Li and C. C. Kao (2008). "Structure and Function of LGP2, a DEX(D/H) Helicase That Regulates the Innate Immunity Response." J Biol Chem **283**(23): 15825-15833.

Nakamura, N., J. R. Lill, Q. Phung, Z. Jiang, C. Bakalarski, A. de Maziere, J. Klumperman, M. Schlatter, L. Delamarre and I. Mellman (2014). "Endosomes are specialized platforms for bacterial sensing and NOD2 signalling." Nature **509**(7499): 240-244.

Nivon, L. G., R. Moretti and D. Baker (2013). "A Pareto-optimal refinement method for protein design scaffolds." PLoS One **8**(4): e59004.

O'Meara, M. J., A. Leaver-Fay, M. D. Tyka, A. Stein, K. Houlihan, F. DiMaio, P. Bradley, T. Kortemme, D. Baker, J. Snoeyink and B. Kuhlman (2015). "Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta." J Chem Theory Comput **11**(2): 609-622.

Ogura, Y., D. K. Bonen, N. Inohara, D. L. Nicolae, F. F. Chen, R. Ramos, H. Britton, T. Moran, R. Karaliuskas, R. H. Duerr, J. P. Achkar, S. R. Brant, T. M. Bayless, B. S. Kirschner, S. B. Hanauer, G. Nunez and J. H. Cho (2001). "A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease." Nature **411**(6837): 603-606.

Ouda, R., K. Onomoto, K. Takahasi, M. R. Edwards, H. Kato, M. Yoneyama and T. Fujita (2011). "Retinoic acid-inducible gene I-inducible miR-23b inhibits infections by minor group rhinoviruses through down-regulation of the very low density lipoprotein receptor." J Biol Chem **286**(29): 26210-26219.

Palm, N. W. and R. Medzhitov (2009). "Pattern recognition receptors and control of adaptive immunity." Immunol Rev **227**: 221-233.

Park, H., J. Ko, K. Joo, J. Lee, C. Seok and J. Lee (2011). "Refinement of protein termini in template-based modeling using conformational space annealing." Proteins **79**(9): 2725-2734.

Park, H. and C. Seok (2012). "Refinement of unreliable local regions in template-based protein models." Proteins **80**(8): 1974-1986.

Park, H. H., Y. C. Lo, S. C. Lin, L. Wang, J. K. Yang and H. Wu (2007). "The death domain superfamily in intracellular signaling of apoptosis and inflammation." Annu Rev Immunol **25**: 561-586.

Park, H. H., E. Logette, S. Raunser, S. Cuenin, T. Walz, J. Tschopp and H. Wu (2007). "Death domain assembly mechanism revealed by crystal structure of the oligomeric PIDDosome core complex." Cell **128**(3): 533-546.

Peisley, A., B. Wu, H. Yao, T. Walz and S. Hur (2013). "RIG-I forms signaling-competent filaments in an ATP-dependent, ubiquitin-independent manner." Mol Cell **51**(5): 573-583.
Pippig, D. A., J. C. Hellmuth, S. Cui, A. Kirchhofer, K. Lammens, A. Lammens, A. Schmidt, S. Rothenfusser and K. P. Hopfner (2009). "The regulatory domain of the RIG-I family ATPase LGP2 senses double-stranded RNA." Nucleic Acids Res **37**(6): 2014-2025.

Poltorak, A., X. He, I. Smirnova, M. Y. Liu, C. Van Huffel, X. Du, D. Birdwell, E. Alejos, M. Silva, C. Galanos, M. Freudenberg, P. Ricciardi-Castagnoli, B. Layton and B. Beutler (1998). "Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene." Science **282**(5396): 2085-2088.

Proell, M., S. J. Riedl, J. H. Fritz, A. M. Rojas and R. Schwarzenbacher (2008). "The Nod-like receptor (NLR) family: a tale of similarities and differences." PLoS One **3**(4): e2119.
Qiao, Q. and H. Wu (2015). "Supramolecular organizing centers (SMOCs) as signaling machines in innate immune activation." Sci China Life Sci **58**(11): 1067-1072.

Qin, H., S. M. Srinivasula, G. Wu, T. Fernandes-Alnemri, E. S. Alnemri and Y. Shi (1999). "Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1." Nature **399**(6736): 549-557.

Roccatano, D., A. Barthel and M. Zacharias (2007). "Structural flexibility of the nucleosome core particle at atomic resolution studied by molecular dynamics simulation." Biopolymers **85**(5-6): 407-421.

Rothenfusser, S., N. Goutagny, G. DiPerna, M. Gong, B. G. Monks, A. Schoenemeyer, M. Yamamoto, S. Akira and K. A. Fitzgerald (2005). "The RNA helicase Lgp2 inhibits TLR-independent sensing of viral replication by retinoic acid-inducible gene-I." J Immunol **175**(8): 5260-5268.

Saito, T., R. Hirai, Y. M. Loo, D. Owen, C. L. Johnson, S. C. Sinha, S. Akira, T. Fujita and M. Gale (2006). "Regulation of innate antiviral defenses through a shared repressor domain in RIG-I and LGP2." Proc Natl Acad Sci U S A **104**(2): 582-587.

Scheres, S. H. and S. Chen (2012). "Prevention of overfitting in cryo-EM structure determination." Nat Methods **9**(9): 853-854.

Schneider, C., A. Teufel, T. Yevsa, F. Staib, A. Hohmeyer, G. Walenda, H. W. Zimmermann, M. Vucur, S. Huss, N. Gassler, H. E. Wasmuth, S. A. Lira, L. Zender, T. Luedde, C. Trautwein and F. Tacke (2012). "Adaptive immunity suppresses formation and progression of diethylnitrosamine-induced liver cancer." Gut **61**(12): 1733-1743.

Seeger, R. and E. G. Krebs (1995). "The MAPK signaling cascade." The FASEB J **9**(9): 726-735.

Simons, K. T., C. Kooperberg, E. Huang and D. Baker (1997). "Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions." J Mol Biol **268**(1): 209-225.

Sun, R., Z. L. Li and T. C. Bishop (2017). "Molecular Dynamics Simulation Workflow for Nucleosome Studies." Biophys J **112**(3): 289a-289a.

Takamatsu, S., K. Onoguchi, K. Onomoto, R. Narita, K. Takahashi, F. Ishidate, T. K. Fujiwara, M. Yoneyama, H. Kato and T. Fujita (2013). "Functional characterization of domains of IPS-1 using an inducible oligomerization system." PLoS One **8**(1): e53578.

Takeuchi, O. and S. Akira (2010). "Pattern Recognition Receptors and Inflammation." Cell **140**(6): 805-820.

Ting, J. P., R. C. Lovering, E. S. Alnemri, J. Bertin, J. M. Boss, B. K. Davis, R. A. Flavell, S. E. Girardin, A. Godzik, J. A. Harton, H. M. Hoffman, J. P. Hugot, N. Inohara, A. Mackenzie, L. J. Maltais, G. Nunez, Y. Ogura, L. A. Otten, D. Philpott, J. C. Reed, W. Reith, S. Schreiber, V. Steimle and P. A. Ward (2008). "The NLR gene family: a standard nomenclature." Immunity **28**(3): 285-287.

Travassos, L. H., L. A. Carneiro, M. Ramjeet, S. Hussey, Y. G. Kim, J. G. Magalhaes, L. Yuan, F. Soares, E. Chea, L. Le Bourhis, I. G. Boneca, A. Allaoui, N. L. Jones, G. Nunez, S. E. Girardin and D. J. Philpott (2010). "Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry." Nat Immunol **11**(1): 55-62.

Trivedi, S. and E. L. Greidinger (2009). "Endosomal Toll-like receptors in autoimmunity: mechanisms for clinical diversity." Therapy **6**(3): 433-442.

Tsuda, M., H. Tozaki-Saitoh and K. Inoue (2012). "Purinergic system, microglia and neuropathic pain." Curr Opin Pharmacol **12**(1): 74-79.

Tyka, M. D., D. A. Keedy, I. Andre, F. Dimaio, Y. Song, D. C. Richardson, J. S. Richardson and D. Baker (2011). "Alternate states of proteins revealed by detailed energy landscape mapping." J Mol Biol **405**(2): 607-618.

Valmiki, M. G. and J. W. Ramos (2009). "Death effector domain-containing proteins." Cell Mol Life Sci **66**(5): 814-830.

Venkataraman, T., M. Valdes, R. Elsby, S. Kakuta, G. Caceres, S. Saijo, Y. Iwakura and G. N. Barber (2007). "Loss of DExD/H box RNA helicase LGP2 manifests disparate antiviral responses." J Immunol **178**(10): 6444-6455.

Ver Heul, A. M., L. Gakhar, R. C. Piper and R. Subramanian (2014). "Crystal structure of a complex of NOD1 CARD and ubiquitin." PLoS One **9**(8): e104017.

Wang, P., S. Zhu, L. Yang, S. Cui, W. Pan, R. Jackson, Y. Zheng, A. Rongvaux, Q. Sun, G. Yang, S. Gao, R. Lin, F. You, R. Flavell and E. Fikrig (2015). "Nlrp6 regulates intestinal antiviral innate immunity." Science **350**(6262): 826-830.

Warshel, A. and M. Levitt (1976). "Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme." J Mol Biol **103**(2): 227-249.

Weber, C. H. and C. Vincenz (2001). "The death domain superfamily: a tale of two interfaces?" Trends Biochem Sci **26**(8): 475-481.

Wu, B. and S. Hur (2015). "How RIG-I like receptors activate MAVS." Curr Opin Virol **12**: 91-98.

Wu, B., A. Peisley, D. Tetrault, Z. Li, E. H. Egelman, K. E. Magor, T. Walz, P. A. Penczek and S. Hur (2014). "Molecular imprinting as a signal-activation mechanism of the viral RNA sensor RIG-I." Mol Cell **55**(4): 511-523.

Wu, C., Z. Su, M. Lin, J. Ou, W. Zhao, J. Cui and R. F. Wang (2017). "NLRP11 attenuates Toll-like receptor signalling by targeting TRAF6 for degradation via the ubiquitin ligase RNF19A." Nat Commun **8**(1): 1977.

Wu, S. and Y. Zhang (2007). "LOMETS: a local meta-threading-server for protein structure prediction." Nucleic Acids Res **35**(10): 3375-3382.

Xiao, T., P. Towb, S. A. Wasserman and S. R. Sprang (1999). "Three-dimensional structure of a complex between the death domains of Pelle and Tube." Cell **99**(5): 545-555.

Yamamoto, M., S. Sato, H. Hemmi, K. Hoshino, T. Kaisho, H. Sanjo, O. Takeuchi, M. Sugiyama, M. Okabe, K. Takeda and S. Akira (2003). "Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway." Science **301**(5633): 640-643.

Yang, J., R. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang (2015). "The I-TASSER Suite: protein structure and function prediction." Nat Methods **12**(1): 7-8.

Yang, J. and Y. Zhang (2015). "I-TASSER server: new development for protein structure and function predictions." Nucleic Acids Res **43**(W1): W174-181.

Yang, J. and Y. Zhang (2015). "Protein Structure and Function Prediction Using I-TASSER." Curr Protoc Bioinformatics **52**: 5 8 1-15.

Yoneyama, M. and T. Fujita (2009). "RNA recognition and signal transduction by RIG-I-like receptors." Immunol Rev **227**(1): 54-65.

Yoneyama, M., M. Kikuchi, K. Matsumoto, T. Imaizumi, M. Miyagishi, K. Taira, E. Foy, Y. M. Loo, M. Gale, Jr., S. Akira, S. Yonehara, A. Kato and T. Fujita (2005). "Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity." J Immunol **175**(5): 2851-2858.

Yoneyama, M., M. Kikuchi, T. Natsukawa, N. Shinobu, T. Imaizumi, M. Miyagishi, K. Taira, S. Akira and T. Fujita (2004). "The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses." Nat Immunol **5**(7): 730-737.

Zhang, L., S. Chen, J. Ruan, J. Wu, A. B. Tong, Q. Yin, Y. Li, L. David, A. Lu, W. L. Wang, C. Marks, Q. Ouyang, X. Zhang, Y. Mao and H. Wu (2015). "Cryo-EM structure of the activated NAIP2-NLRC4 inflammasome reveals nucleated polymerization." Science **350**(6259): 404-409.

Zhang, X., C. Wang, L. B. Schook, R. J. Hawken and M. S. Rutherford (2000). "An RNA helicase, RHIV -1, induced by porcine reproductive and respiratory syndrome virus (PRRSV) is mapped on porcine chromosome 10q13." Microb Pathog **28**(5): 267-278.

Zhang, Y. (2008). "I-TASSER server for protein 3D structure prediction." BMC Bioinformatics **9**: 40.

Zhang, Y., A. Kolinski and J. Skolnick (2003). "TOUCHSTONE II: A New Approach to Ab Initio Protein Structure Prediction." Biophys J **85**(2): 1145-1164.

Zhang, Y. and J. Skolnick (2004). "Scoring function for automated assessment of protein structure template quality." Proteins **57**(4): 702-710.

Zhang, Y. and J. Skolnick (2005). "TM-align: a protein structure alignment algorithm based on the TM-score." Nucleic Acids Res **33**(7): 2302-2309.

Zhang, Y., W. Zhang, S. Ogata, D. Clements, J. H. Strauss, T. S. Baker, R. J. Kuhn and M. G. Rossmann (2004). "Conformational changes of the flavivirus E glycoprotein." Structure **12**(9): 1607-1618.

Zhao, Y., J. Yang, J. Shi, Y. N. Gong, Q. Lu, H. Xu, L. Liu and F. Shao (2011). "The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus." Nature **477**(7366): 596-600.

Zhou, R., A. S. Yazdi, P. Menu and J. Tschopp (2011). "A role for mitochondria in NLRP3 inflammasome activation." Nature **469**(7329): 221-225.

Appendix

Supplement Table S1 1 Data collection, map and model refinement, validation of CARD8 filament structures

<i>Data collection</i>	
Microscope	Titan Krios
Voltage (kV)	300
Detector	K2
Pixel size (Å)	1.1
Defocus range(μm)	-1, -2
Electron Dose (e ⁻ /Å ²)	40
<i>Helical Reconstruction</i>	
Software	RELION 3.0
Segment length (Å)	264
Particles	1201108
Helical rise (Å)	5.409
Helical rotation (°)	-99.16
Resolution (Å)	3.7

<i>Coordinate Refinement (mid-segment 12mer)</i>	
Software	Phenix
Rwork	0.2530
Rfree	0.3017
Model	
Number of residues	1044
B-factor overall	276
<i>R.M.S. deviation</i>	
Bond length (Å)	0.002
Bond angle (°)	0.423
<i>Validation</i>	
Molprobability clashscore	5.41
Rotamer outliers (%)	0
C _β deviation (%)	0
<i>Ramachandran plot</i>	
Favored (%)	100
Allowed (%)	0
Outliers (%)	0

Scripts in Gromacs 4.5.5as below performed the standard procedures and MD simulation:

1. prepare a PDB file of sample molecule that is appropriate to run molecular dynamics (MD) simulation.
2. Run prb2gmx to generate the topology files for the molecule and position restraint files and post-processed structure files. The pdb2gmx is executed by issuing the following

command(The molecule is named 'sample' as a reference):

```
pdb2gmx -f sample.pdb -o sample_processed.gro -water spce -ignh
```

And choose force field 6: AMBER99SB-ILDN force field (Lindorff-Larsen et al., Proteins 78, 1950-58, 2010).

3. Define the box of the unit cell by issuing the command at least 10Å from the box edge

```
editconf -f sample_processed.gro -o sample_newbox.gro -c -d 1.0 -bt cubic
```

4. Fill the box with solvent(water) and solvation is accomplished using genbox command:

```
genbox -cp sample_newbox.gro -cs spc216.gro -o sample-solv.gro -p topol.top
```

5. Add the ions into the solvent and pass the file to genion by running the commands as following (and choose SOL when prompted)

```
gromp -f ions.mdp -c sample_solv.gro -p topol.top -o ions.tpr
```

```
genion -s ions.tpr -o sample-solv_ions.gro -p topol.top -pname NA -nname CL -neutral
```

ions.mdp is a pre-defined parameter file and the details are listed in the **Appendix**.

6. Energy minimization of the sample molecule by running the command:

```
grompp -f minim.mdp -c sample-solve_ions.gro -p topol.top -o em.tpr
```

```
nohup mdrun -v -deffnm em -nt * & (* depends on the computer resources available on the server)
```

minim.mdp is a pre-defined parameter file and the details are listed in the **Appendix**.

7. Equilibration under an NVT ensemble (constant number of particles, volume, and temperature) by running the command as following:

```
grompp -f nvt.mdp -c em.gro -p topol.top -o nvt.tpr
```

```
nohup mdrun -deffnm nvt -nt * & (* depends on the computer resources available on the server)
```

nvt.mdp is a pre-defined parameter file and the details are listed in the *Appendix*.

8. Equilibration under an NPT ensemble (constant number of particles, volume, and temperature) by running the command as following:

```
grompp -f npt.mdp -c nvt.gro -t nvt.cpt -p topol.top -o npt.tpr
```

```
nohup mdrun -deffnm npt -nt * & (* depends on the computer resources available on the server)
```

npt.mdp is a pre-defined parameter file and the details are listed in the *Appendix*.

9. Run molecular dynamics (MD) simulation by the command as following:

```
grompp -f md.mdp -c npt.gro -t npt.cpt -p topol.top -o md.tpr
```

```
mdrun -deffnm md -nt * & (* depends on the computer resources available on the server)
```

md.mdp is a pre-defined parameter file and the details are listed in the *Appendix*.

Molecular dynamics parameters(mdp) option file scripts:

-ions.mdp

; ions.mdp - used as input into grompp to generate ions.tpr

; Parameters describing what to do, when to stop and what to save

Integrator = steep ; Algorithm (steep = steepest descent minimization)

emtol = 1000.0 ; Stop minimization when the maximum force < 1000.0 kJ/mol/nm

emstep = 0.01 ; Energy step size

nsteps = 50000 ; Maximum number of (minimization) steps to perform

Nstlist = 1 ; Frequency to update the neighbor list and long range forces

ns_type = grid ; Method to determine neighbor list (simple, grid)

rlist = 1.0 ; Cut-off for making neighbor list (short range forces)

coulombtype = PME ; Treatment of long range electrostatic interactions

rcoulomb = 1.0 ; Short-range electrostatic cut-off

rvdw = 1.0 ; Short-range Van der Waals cut-off

pbc = xyz ; Periodic Boundary Conditions (yes/no)

-minim.mdp

; minim.mdp - used as input into grompp to generate em.tpr

; Parameters describing what to do, when to stop and what to save

Integrator = steep ; Algorithm (steep = steepest descent minimization)

Emtol = 1000.0 ; Stop minimization when the maximum force < 1000.0 kJ/mol/nm

Emstep = 0.01 ; Energy step size

nsteps = 50000 ; Maximum number of (minimization) steps to perform

nstlist = 1 ; Frequency to update the neighbor list and long range forces

ns_type = grid ; Method to determine neighbor list (simple, grid)

rlist = 1.0 ; Cut-off for making neighbor list (short range forces)

coulombtype = PME ; Treatment of long range electrostatic interactions

rcoulomb = 1.0 ; Short-range electrostatic cut-off
 rvdw = 1.0 ; Short-range Van der Waals cut-off
 pbc = xyz ; Periodic Boundary Conditions (yes/no)

-nvt.mdp

title = OPLS Lysozyme NVT equilibration
 define = -DPOSRES; position restrain the protein
 ; Run parameters
 Integrator = md ; leap-frog integrator
 Nsteps = 50000 ; 2 * 50000 = 100 ps
 dt = 0.002 ; 2 fs
 ; Output control
 Nstxout = 100 ; save coordinates every 0.2 ps
 Nstvout = 100 ; save velocities every 0.2 ps
 Nstenergy = 100 ; save energies every 0.2 ps
 Nstlog = 100 ; update log file every 0.2 ps
 ; Bond parameters
 Continuation = no ; first dynamics run
 constraint_algorithm = lincs ; holonomic constraints
 constraints= all-bonds ; all bonds (even heavy atom-H bonds) constrained
 lincs_iter = 1 ; accuracy of LINCS
 lincs_order = 4 ; also related to accuracy
 ; Neighborsearching
 ns_type = grid ; search neighboring grid cells

```

nstlist = 5 ; 10 fs

rlist = 1.0 ; short-range neighborlist cutoff (in nm)

rcoulomb = 1.0 ; short-range electrostatic cutoff (in nm)

rvdw = 1.0 ; short-range van der Waals cutoff (in nm)

; Electrostatics

Coulombtype = PME ; Particle Mesh Ewald for long-range electrostatics

pme_order = 4 ; cubic interpolation

fourierspacing = 0.16 ; grid spacing for FFT

; Temperature coupling is on

tcoupl= V-rescale ; modified Berendsen thermostat

tc-grps= Protein Non-Protein ; two coupling groups - more accurate

tau_t = 0.1 0.1 ; time constant, in ps

ref_t = 300 300 ; reference temperature, one for each group, in K

; Pressure coupling is off

Pcoupl = no ; no pressure coupling in NVT

; Periodic boundary conditions

Pbc = xyz ; 3-D PBC

; Dispersion correction

DispCorr= EnerPres ; account for cut-off vdW

scheme

; Velocity generation

gen_vel = yes ; assign velocities from

Maxwell distribution

```

```

gen_temp= 300                                ; temperature for Maxwell distribution

gen_seed = -1                                ; generate a random seed


-npt.mdp

Title = OPLS Lysozyme NPT equilibration

define= -DPOSRES                             ; position restrain the protein

; Run parameters

Integrator = md                             ; leap-frog integrator

Nsteps = 50000                               ; 2 * 50000 = 100 ps

Dt = 0.002                                   ; 2 fs

; Output control

Nstxout = 100                               ; save coordinates every 0.2 ps

Nstvout = 100                               ; save velocities every 0.2 ps

Nstenergy = 100                             ; save energies every 0.2 ps

Nstlog = 100                                ; update log file every 0.2 ps

; Bond parameters

Continuation = yes                          ; Restarting after NVT

constraint_algorithm = lincs                 ; holonomic constraints

constraints = all-bonds                      ; all bonds (even heavy atom-H bonds) constrained

lincs_iter = 1                              ; accuracy of LINCS

lincs_order = 4                             ; also related to accuracy

; Neighborsearching

ns_type = grid                              ; search neighboring grid cells

nstlist = 5                                 ; 10 fs

```

```

rlist = 1.0 ; short-range neighborlist cutoff (in nm)
rcoulomb = 1.0 ; short-range electrostatic cutoff (in nm)
rvdw = 1.0 ; short-range van der Waals cutoff (in nm)
; Electrostatics
Coulombtype = PME ; Particle Mesh Ewald for long-range electrostatics
pme_order = 4 ; cubic interpolation
fourierspacing = 0.16 ; grid spacing for FFT
; Temperature coupling is on
Tcoupl = V-rescale ; modified Berendsen thermostat
tc-grps= Protein Non-Protein ; two coupling groups - more accurate
tau_t= 0.1 0.1 ; time constant, in ps
ref_t = 300 300 ; reference temperature, one for each group, in K
; Pressure coupling is on
Pcoupl = Parrinello-Rahman ; Pressure coupling on in NPT
Pcoupltype = isotropic ; uniform scaling of box vectors
tau_p = 2.0 ; time constant, in ps
ref_p = 1.0 ; reference pressure, in bar
compressibility = 4.5e-5 ; isothermal compressibility of water, bar^-1
refcoord_scaling = com
; Periodic boundary conditions
Pbc = xyz ; 3-D PBC
; Dispersion correction

```



```

DispCorr = EnerPres                                ; account for cut-off vdW
scheme
; Velocity generation
gen_vel = no                                       ; Velocity generation is off

-md.mdp
Title = OPLS Lysozyme MD
; Run parameters
Integrator = md                                   ; leap-frog integrator
Nsteps = 500000                                   ; 2 * 500000 = 1000 ps, 1 ns
Dt = 0.002                                         ; 2 fs
; Output control
Nstxout = 1000                                     ; save coordinates every 2 ps
Nstfout = 1000                                     ; save velocities every 2 ps
Nstxtcout = 1000                                  ; xtc compressed trajectory
output every 2 ps
Nstenergy = 1000                                   ; save energies every 2 ps
Nstlog = 1000                                      ; update log file every 2 ps
; Bond parameters
Continuation = yes                                ; Restarting after NPT
constraint_algorithm = lincs                       ; holonomic constraints
constraints = all-bonds                            ; all bonds (even heavy atom-H bonds) constrained
lincs_iter = 1                                     ; accuracy of LINCS
lincs_order = 4                                    ; also related to accuracy

```

```

; Neighborsearching

ns_type = grid                                ; search neighboring grid
cells

nstlist = 5                                  ; 10 fs

rlist = 1.0                                  ; short-range neighborlist cutoff (in nm)

rcoulomb = 1.0                                ; short-range electrostatic cutoff (in nm)

rvdw = 1.0                                    ; short-range van der Waals cutoff (in nm)

; Electrostatics

Coulombtype = PME                            ; Particle Mesh Ewald for long-range electrostatics

pme_order = 4                                ; cubic interpolation

fourierspacing = 0.16                        ; grid spacing for FFT

; Temperature coupling is on

Tcoupl = V-rescale                            ; modified Berendsen thermostat

tc-grps = Protein Non-Protein                ; two coupling groups - more accurate

tau_t = 0.1 0.1                              ; time constant, in ps

ref_t = 300 300                              ; reference temperature, one for each group, in K

; Pressure coupling is on

pcoupl= Parrinello-Rahman                    ; Pressure coupling on in NPT

pcoupltype = isotropic                        ; uniform scaling of box vectors

tau_p = 2.0                                  ; time constant, in ps

ref_p = 1.0                                  ; reference pressure, in bar

compressibility = 4.5e-5                      ; isothermal compressibility of water, bar^-1

; Periodic boundary conditions

```

pbc= xyz ; 3-D PBC
 ; Dispersion correction
 DispCorr= EnerPres ; account for cut-off vdW
 scheme
 ; Velocity generation
 gen_vel = no ; Velocity generation is off

CARD8 sequencing result

CARD8 full length sequence (CARD8-CARD sequence highlighted)

CARD8:431aa

MMRQRQSHYCSVLFLSVNYLGGTFFPGDICSEENQIVSSYASKVCFEIEEDYKNRQFLGPEGNVDVELIDKSTNRYSV
 WFPTAGWYLWSATGLGFLVRDEVTVTIAFGSWSQHLALDLQHHEQWLVGGLFDVTAEPEEAVAEIHLPHFISLQGE
 VDVSWFLVAHFKNEGMVLEHPARVEPFYAVLESFSFSLMGILLRIASGTRLSIPITSNTLIYYHPHPEDIKFHLYLV
 PSDALLTKAIDDEEDRFHGVRLQTSPPMEPLNFGSSYIVSNSANLKVMPKELKLSYRSPGEIQHFSKFYAGQMKEPI
 QLEITEKRHGTLVWDTEVKPVDLQLVAASAPPPFSGAAFVKENHRQLQARMGDLKGVLDLQDNEVLTENEKELVEQ
 EKTRQSKNEALLSMVEKKGDLALDVLFRSISERDPYLVSYLRQQNL

CARD8 isoform A: 537aa

MEKKECPEKSSSSEELPRRDSGSSRNIDASKLIRLQGSRKLLVDNSIRELQYTKTGIFQAEACVTNDTVYRELPC
 VSETLCDISHFFQEDDETEAEPLLFRAVPECQLSGGDIPSVSEEQESSEGQDSGDICSEENQIVSSYASKVCFEIEE
 DYKNRQFLGPEGNVDVELIDKSTNRYSVWFPTAGWYLWSATGLGFLVRDEVTVTIAFGSWSQHLALDLQHHEQWLVG
 GPLFDVTAEPEEAVAEIHLPHFISLQAGEVDVSWFLVAHFKNEGMVLEHPARVEPFYAVLESFSFSLMGILLRIASG
 TRLSIPITSNTLIYYHPHPEDIKFHLYLVPSDALLTKAIDDEEDRFHGVRLQTSPPMEPLNFGSSYIVSNSANLKV
 MPKELKLSYRSPGEIQHFSKFYAGQMKEPIQLEITEKRHGTLVWDTEVKPVDLQLVAASAPPPFSGAAFVKENHRQLQ
ARMGDLKGVLDLQDNEVLTENEKELVEQEKTQSKNEALLSMVEKKGDLALDVLFRSISERDPYLVSYLRQQNL

Primer design:

5'card8_dd_XmaI

Ataa CCCGGG atgGCAGCCTTTGTGAAGGAGAACC

3'card8_dd_EcoRI

Ataa GAATTC ctaCAAATTCTGCTGTCTAAGATAGGACACGAGG

Sequencing result:

101687003_3card8dd_T7Promoter_A02

NNNNNGNNGNNNATNNNCCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCAC
 ATCACCACCACCATCACTCCGCGGCTCTTGAAGTCCTCTTTCAGGGACCCGGGATGGCAGCCTTTGTGAA
 GGAGAACCACCGGCAACTCCAAGCCAGGATGGGGGACCTGAAAGGGGTGCTCGATGATCTCCAAGACA
 ATGAGGTTCTTACTGAGAATGAGAAGGAGCTGGTGGAGCAGGAAAAGACACGGCAGAGCAAGAATGA
 GGCCTTGCTGAGCATGGTGGAGAAGAAAGGGGACCTGGCCCTGGACGTGCTCTTCAGAAGCATTAGTG
 AAAGGGACCCTTACCTCGTGCCTATCTTAGACAGCAGAATTTGTAGGAATTCTGTACAGGCCTTGGCGC
 GCCCGACGTCCGTCGACAAGCTTGCGGCCGCAGAGCTCGCTCTGGTGCCACGCGGTAGTAAAGAAACC
 GCTGCTGCTAAATTCGAACGCCAGCACATGGACAGCTCTACTTCTGCTGCTCTCGAGGCTTAATTAACCT
 AGGCTGCTAAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATA
 ACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATTGGC
 GAATGGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCG
 CTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTCTTCCCTTCCTTCTCGCCACGTTCCGCGGCT
 TTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCGATTAGTGCTTTACGGCACCTCGACCC
 CAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCTTTGAC
 GTTGGAGTCAGTTCTTTATAGTGGACTCTTGTTCCAACCTGGAACACACTCAACCTATCTCGGNTATTCTT
 TGATTATAAGGGATTTGCCGATTCGGCCTTGTTAAAAATGACTGATTAAACAAAATTTACGCGAATTTAA
 CANATATAACNCTNCATTAGGNGGACTTTTCGGGAANGGGCCGGAACCCNATTGTTATTTTCAAACAT
 CAAANNGATCCCCNNGAATAATCTTAAAAACCCNGC

5'3' Frame 2 Flag tag- CARD8

XXXXXXXXL **Stop** NNFV **Stop** L **Stop** EGD IH **Met** AHHHHHSA
 ALEVLFGPG **Met** AAFVKENHRQLQAR **Met** GDLKGVLD
 DLQDNEVL TENEKELVEQE KTRQSKNEALLS **Met** VEK
 KGD LALDVLFRSISERDPYLVSYLRQQNL **Stop** EFCTG
 LGAPDVRRQACGRRARSGATR **Stop** **Stop** RNRCC **Stop** IR
 TPAHGQLYFCCSRGLINLGC **Stop** TKPERKLSWLLPPL
 SNN **Stop** HNPLGPLNGS **Stop** GVFC **Stop** KEELYPDW
 R **Met** GRAL **Stop** RRIKRGGCGGYAQRDRYTCQRPSARS
 FRFLPFLSRHVRRLSPSSSKSGAPFRVPI **Stop** CFTAP
 RPQKLD **Stop** GDGSRSGPSP **Stop** **Stop** TVFAFDVGVTFFI
 VDSCSNWNTLNLSRXFFDYKGFADSALVKND **Stop** LTK
 FTRI **Stop** XI **Stop** XXIRXTFREXAGTXLLFSKHQXDPX
 E **Stop** S **Stop** KP X

Mon Sep 10, 2018 12:15 +0900

New DNA from 1 to 267

Alignment to

101687003_3card8dd_T7Promoter_A02.seq-- Matches:267; Mismatches:0; Gaps:944; Unattempted:0

```

      *      *      *      *      *      *      *      *      *      *
0>~~~~~>0
1>NNNNNNNGNNNNNATNNNNCTCTAGAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGGCACATCACCACCACCATCACTCCGCGGCTCTTGAA>100

      *      *      *      *      *      *      *      *      *      *
1>~~~~~atgGCAGCCTTTGTGAAGGAGAACCACCGCAACTCCAAGCCAGGATGGGGGACCTGAAAGGGGTGCTCGATGATCTCC>79
101>GTCCTCTTTCAGGGACCCGGGATGGCAGCCTTTGTGAAGGAGAACCACCGCAACTCCAAGCCAGGATGGGGGACCTGAAAGGGGTGCTCGATGATCTCC>200

      *      *      *      *      *      *      *      *      *      *
80>AAGACAATGAGGTTCTTACTGAGAATGAGAAGGAGCTGGTGGAGCAGGAAAAGACACGGCAGAGCAAGAATGAGGCCTTGCTGAGCATGGTGGAGAAGAA>179
201>AAGACAATGAGGTTCTTACTGAGAATGAGAAGGAGCTGGTGGAGCAGGAAAAGACACGGCAGAGCAAGAATGAGGCCTTGCTGAGCATGGTGGAGAAGAA>300

      *      *      *      *      *      *      *      *      *      *
180>AGGGGACCTGGCCCTGGACGTGCTCTTCAGAAGCATTAGTGAAAGGGACCTTACCTCGTGCTCTATCTTAGACAGCAGAATTTGtag>267
301>AGGGGACCTGGCCCTGGACGTGCTCTTCAGAAGCATTAGTGAAAGGGACCTTACCTCGTGCTCTATCTTAGACAGCAGAATTTGTAGGAATTCGTACA>400

      *      *      *      *      *      *      *      *      *      *
267>~~~~~>267
401>GGCCTTGGCGCGCCGACGCTCCGTCGACAAGCTTGCGGCCGAGAGCTCGCTCTGGTGCCACGCGGTAGTAAAGAAACCGCTGCTGCTAAATTCGAACGC>500

      *      *      *      *      *      *      *      *      *      *
267>~~~~~>267
501>CAGCACATGGACAGCTCTACTTCTGCTGCTCTCGAGGCTTAATTAACTAGGCTGCTAAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCT>600

      *      *      *      *      *      *      *      *      *      *
267>~~~~~>267
601>GAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTGTGCTGAAAGGAGGAAGTATATCCGGATTGGCGAATGGGACGCGCC>700

      *      *      *      *      *      *      *      *      *      *
267>~~~~~>267
701>CTGTAGCGGCGCAITTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTTCCCT>800

```

102862002_PET47SANPISB8_T7Promoter_G10

NNNNNNNANNNNNNNNTNNNNCTAGAATAATTTTGTTTACTTTAAGAAGGAGATATAC
ATATGGCACATCACCACCACCATCACTCCGCGGCTCTTGAAGTCCTCTTTCAGGGAC
CCGGGTACCAGGATCCGATGCCGTTTGCTGAAGACAAGACCTATAAGTATATCTGCC
GCAATTTTCAGCAATTTTGTCAATGTGGATGTTGTAGAGATTCTGCCTTACCTGCCCTG
CCTCACAGCAAGAGACCAGGATCGACTGCGGGCCACCTGCACACTCTCAGGGAACC
GGGACACCCCTCTGGCATCTCTTCAATACCCTTCAGCGGCGGCCCCGGCTGGGTGGAGT
ACTTCATTGCGGCACTGAGGGGCTGTGAGCTAGTTGATCTCGCGGACGAAGTGGCCT
CTGTCTACCAGAGCTACCAGCCTGAATTCCTTGAAGTCCTCTTTCAGGGACCCATGG
ACAAAGATTGCGAAATGAAACGTACCACCCTGGATAGCCCGCTGGGCAAACCTGGAA
CTGAGCGGCTGCGAACAGGGCCTGCATGAAATTAAACTGCTGGGTAAAGGCACCAG
CGCGGCCGATGCGGTTGAAGTTCCGGCCCCGGCCGCCGTGCTGGGTGGTCCGGAAC
CGCTGATGCAGGCGACCGCGTGGCTGAACGCGTATTTTCATCAGCCGGAAGCGATTG
AAGAATTTCCGGTTCCGGCGCTGCATCATCCGGTGTTCAGCAGGAGAGCTTTACCC
GTCAGGTGCTGTGGAAACTGCTGAAAGTGGTTAAATTTGGCGAAGTGATTAGCTATC
AGCAGCTGGCGGCCCTGGCGGGTAATCCGGCGGCCACCGCCGCCGTTAAAACCGCG
CTGAGCGGTAACCCGGTGCCGATTCTGATTCCGTGCCATCGTGTGGTTAGCTCTAGC
GGTGCAGTTTGGCGGTTATGAAGGTGGTCTGGCGGTGAAAGAGTGGCTGCTGGCCCA

TGAAGGTCATCGTCTGGGTAAACCGGGTCTGGGACCTGCAGGTATAGGCGCGCCAG
GATCACACCACCATCATCACCATTGATAACTCGAGGCTTAATTAACCTAGGCTGCTA
AACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGC
ATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTGTCTGAAAGGAGGAAC
TTNATCCGGATTGGCGAATGGGAACGCCCCCTGGAGCNGGCCCTT

5'3' Frame 2 CARD8-SNAP tag

Met AHHHHHHSAALEVLFGPGYQDP **Met** AAFVKENHR
QLQAR **Met** GDLKGVLDLQDNEVLTEKEKELVEQEK
RQSKNEALLSMVEKKGDLALDVLFRSISERDPYLVSY
LRQQNDYKDDDDKLEFLEVLFGPGSGS **Met** DKDCE
Met KRTTLDSPGKLELSGCEQGLHEIKLLGKGTSAA
DAVEVPAPAAVLGGPEPL **Met** QATAWLNAYFHQPEAI
EEFPVPALHHPVFQQESFTRQVLWKLLKVVKFGEVIS
YQQLAALAGNPAATAAVKTALSGNPVPIIPCHRVVS
SSGAVGGYEGGLAVKEWLLAHEGHRLGKPGGLGPAGI
GAPGSPHHHHH

102916004_CARD8pet47snap4_T7Promoter_C03

NNGTGNANGNNNNNNNCNNCTAGANTAATTTTGTTTAACTTTAAGAAGGAGATATA
CATATGGCACATCACCACCACCATCACTCCGCGGCTCTTGAAGTCCTCTTTCAGGGA
CCCGGGATGGCAGCCTTTGTGAAGGAGAACCACCGGCAACTCCAAGCCAGGATGGG
GGACCTGAAAGGGGTGCTCGATGATCTCCAAGACAATGAGGTTCTTACTGAGAATG
AGAAGGAGCTGGTGGAGCAGGAAAAGACACGGCAGAGCAAGAATGAGGCCTTGCT
GAGCATGGTGGAGAAGAAAGGGGACCTGGCCCTGGACGTGCTCTTCAGAAGCATT
GTGAAAGGGACCCTTACCTCGTGTCTATCTTAGACAGCAGAATTTGTAGGAATTCT
GTACAGGCCTTGGCGCGCCCGACGTCCGTCGACAAGCTTGCGGCCGCAGAGCTCGC
TCTGGTGCCACGCGGTAGTAAAGAAACCGCTGCTGCTAAATTCGAACGCCAGCACA
TGGACAGCTCTACTTCTGCTGCTCTCGAGGCTTAATTAACCTAGGCTGCTAAACAAA
GCCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACC
CCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGTCTGAAAGGAGGAACTATATC
CGGATTGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGT
GGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCG
TTTCTTCCCTTCCTTTCTCGCCACGTTCCGCGGCTTTCCCCGTCAAGCTCTAAATCGG
GGGCTCCCTTTAGGGTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTT
GATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTTTCGCCCT
TTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAGTGAACAACAC
TCAACCCTATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTTTCGGCCTA
TTGGTTAAAAATGAGCTGATTTAACAAAATTTAACGCGAATTTTAACAAANNTTAAC
GCTTACAATTTAGGGGGACTTTTCGGGGAATGTGCCGGAACCCCNNTTTGTTATTTTT
CTAATACATCCAAATTGGATCCGTNNTGAATAATNTTAGAAANNATCGNNCTCAAT

GAACTGCATTTTTCTTNCGGTTTTTCATTCCTNTTTTGAAAACCGTTTNNATGAAGGAA
AANNCCCGGGGGTTCCTAGNNGNGGAAAC

5'3' Frame 3 CARD8 alone

VXXXXX **Stop** XNFV **Stop** L **Stop** EGD IH **Met** AHHHHHHHSA
LEVL FQGPG **Met** AAFVKENHRQLQAR **Met** GDLKGVLD
LQDNEVL TENEKELVEQE KTRQSKNEALLS **Met** VEKK
GDLALDVLFRSISERDPYLVSYLRQQNL **Stop** EFCTGL
GAPDVRRQACGRRARSGATR **Stop** **Stop** RNRCC **Stop** IRT
PAHGQLYFCCSRGLINLGC **Stop** TKPERKLSWLLPPLS
NN **Stop** HNPLGPLNGS **Stop** GVFC **Stop** KEELYPDWR **Met** G
RAL **Stop** RRIKRGGCGGYAQRDRYTCQRPSARSFRFL
PFLSRHVRRLSPSSSKSGAPFRVPI **Stop** CFTAPRPQK
T **Stop** LG **Stop** WFT **Stop** WAIALIDGFSPFDVGVHV
L **Stop** **Stop** WTLVPTGTTLNPISVYSFDL **Stop** GILPISAYW
LK **Met** S **Stop** FNKI **Stop** REF **Stop** QXLTLTI **Stop** GDFSGNVP
EPXFVIFLIHPNWIRXE **Stop** X **Stop** KXIXLNEHF SXGFH
SXFENRXXEGKXPGVPXXGN

Supplementary Figure1. A SDS-PAGE result of recombinant 6his-CARD8 SNAP purification by Ni-NTA resin. Lane flow through column outcome after loading cell lysate; Lanes elution1-5 column outcome after elution with elution buffer that contained a gradient volume of imidazole (100mM, 150mM, 200mM, 250mM, 300mM).

Alexa Fluor 647
BG Stain Native-PAGE Gel
Label CARD8

Supplementary Figure2. A native PAGE gel result for CARD8 with SNAP tag with fluorescent dye BG-Alexa Fluor 647. The refolded CARD8 CARD was in a monomeric form.