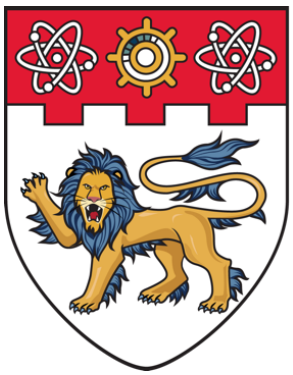

NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

Filamentous oligomer structure reconstruction and a new filament picking method for Cryo-EM micrographs

Xu Chenrui

SCHOOL OF BIOLOGICAL SCIENCES

2021

Filamentous oligomer structure reconstruction and a new filament picking method for Cryo-EM micrographs

Xu Chenrui

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

2021

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.

14 Aug 2021

. . . . . .
Date

. . . . . .
Xu Chenrui

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.

14 Aug 2021

. . . . . .
Date

. . . . . .
Wu Bin

Authorship Attribution Statement

This thesis contains material from two papers published in the following peer-reviewed journals in which I am listed as an author.

Chapter 3.2 is published as Gong, Q., Robinson, K., Xu, C., Huynh, P.T., Chong, K.H.C., Tan, E.Y.J., Zhang, J., Boo, Z.Z., Teo, D.E.T., Lay, K. and Zhang, Y., 2021. Structural basis for distinct inflammasome complex assembly by human NLRP1 and CARD8. Nature communications, 12(1), pp.1-15.DOI: 10.1038/s41467-020-20319-5.

The contributions of the co-authors are as follows:

- Profs Bin Wu, Bruno Reversade and Franklin Zhong provided the initial project direction and edited the manuscript drafts.
- I and Qin Gong, Phuong Thao Huynh, Kelvin Han Chung Chong, Eddie Yong Jun Tan, Jiawen Zhang, Zhao Zhi Boo and Yaming Zhang prepared plasmids or protein samples for structure determination and biochemistry assay, and conducted related experiment.
- I and Qin Gong analyzed sample with EM and collected Cryo-EM data.
- I, Jiawen Zhang and Bin Wu reconstructed Cryo-EM density map.
- I and Bin Wu built protein coordinates and analyzed the structure.
- Kim Robinson, Daniel Eng Thiam Teo, Kenneth Lay, John Soon Yew Lim, Wah Ing Goh and Graham Wright prepared materials for cellular assay and conducted related experiment.

Chapter 3.3 is published as Stsiapanava, A., Xu, C., Brunati, M., Zamora‐Caballero, S., Schaeffer, C., Bokhove, M., Han, L., Hebert, H., Carroni, M., Yasumasu, S. and Rampoldi, L., 2020. Cryo‐EM structure of native human uromodulin, a zona pellucida module polymer. The EMBO journal, 39(24), p.e106807. DOI: 10.15252/embj.2020106807.

The contributions of the co-authors are as follows:

- Prof Bin Wu and Luca Jovine provided the initial project direction and edited the manuscript drafts.
- I and Bin Wu reconstruct electronic density map from Cryo-EM.
- I, Bin Wu and Luca Jovine built the protein coordinates and analyzed the structure.
- Alena Stsiapanava, Sara Zamora-Caballero, Marcel Bokhove, Ling Han, Hans Hebert and Luca Jovine purified sample for Croy-EM data collection and conducted relative validation experiment based structural information.
- Martina Brunati, Céline Schaeffer and Marta Carroni prepare grids for Cryo-EM and collected Cryo-EM data

14 Aug 2021

. . . . . .
Date

. . . . . .
Xu Chenrui

Acknowledgement

My deepest gratitude goes first and foremost to Professor Wu Bin, my supervisor, for his constant encouragement and guidance in both research and life. He is always strict with me in professional standards. Without his consistent and illuminating instruction, this study could not have reached its present form.

Secondly, I would like to express my heartfelt gratitude to my colleagues, Gong Qin, Boo Zhaozhi, Huynh Phuong Thao, Zhang Jiawen, Zhang Yaming, Dr. Kelvin Chong, Chai Yoke Tin helped me in my project. I am also greatly indebted to all the people around me, who have kept company with me in my life and have offered me help in the past one and a half years.

My thanks would also go to my beloved family for their loving considerations and I am thankful for how they gave me great confidence to pursue on throughout these years.

Especially, I want to thank my girlfriend, Ms. Xu Keyun, for her company and care. I feel very fortunate to meet her in my life.

All in all, I thank everyone who played a part in this important stage of my life for their strong support that will help make this study go smoothly. I will now express my sincere thanks to the school of biological science, Nanyang Technological University providing me with a fantastic environment and a great opportunity to conduct my research here.

I will continue to work hard in the future because of those people I loved and loved me.

Table of Contents

List of Figures

List of Tables

List of Abbreviation

Abbreviations	Full name
ALRs	absent-in-melanoma (AIM)-like receptors
ATP	adenosine triphosphate
CCD	charge-coupled device
CLRs	C-type lectin receptors
DD	death domain
DED	death effector domain
iE-DAP	γ-D-glutamyl-meso-diaminopimelic acid
IFNs	type I interferons
LGP2	laboratory of genetics and physiology 2
LPS	bacterial lipopolysaccharide
MAPKs	mitogen-activated protein kinases
MDA5	melanoma differentiation associated factor 5
MDP	muramyl dipeptide
Myd88	myeloid differentiation primary-response protein 88
NF-κB	nuclear factor–kappa B
NMR	nuclear magnetic resonance
NOD	nucleotide oligomerization domain
PGN	peptidoglycan
poly I:C	polyinosinc-polycytidylic acid
PRRs	pattern recognition receptors

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
TLRs	Toll-like receptors
TRAF6	tumor-necrosis factor-receptor-associated factor 6
WHD	winged-helix domain
ADP	Adenosine diphosphate
ASC	Apoptosis-associated, speck-like protein containing a caspase recruitment domain
CARD	Caspase recruitment domain
CARD8	CARD-containing protein 8
Cryo-EM	Cryo-electron microscopy
CT	C-terminal
DAMP	Damage-associated molecular pattern
DPP	Dipeptidyl peptidase
FIIND	Function-to-find domain
GSDMD	Gasdermin D
HD	Helical domain

LRR	Leucine-rich repeat
NACHT	NAIP, CIITA, HET-E, and TP1
NAIP	NLR family of apoptosis inhibitory protein
NBD	Nucleotide-binding domain
NLR	Nucleotide-binding domain and leucine-rich repeat-containing protein
PAMP	Pathogen-associated molecular pattern
PYD	Pyrin domain
UPA	Conserved in UNC5, PIDD and ankyrins
VbP	Val-boro-pro
ZU5	Initially found in Zonula occludens-1 and UNC5
ZP	zona pellucida
UMOD	uromodulin / Tamm–Horsfall protein
α1	α-helix
β1	β-strand
EHP	external hydrophobic patch)
IHP	internal hydrophobic patch
CTP	C-terminal propeptide
CCS	consensus cleavage site (often referred to as CFCS in other ZP module proteins)
EGF	epidermal growth factor-like domains
GPI	glycosylphosphatidylinositol

D8C	cysteine-rich domain
SEC	size-exclusion chromatography
HCE	high choriolytic hatching enzymes
LCE	low choriolytic hatching enzymes
UFE	unfertilized egg coats
UPEC	uropathogenic *E. coli*
DSE	donor-strand exchange
AUC	area under curve
ROC	receiver operating characteristic curve
Cryo-ET	cryo-electron tomography

Summary

Macromolecular protein complexes play important diverse roles in cells. Among various macromolecular complexes, the filamentous oligomer is a special category, which is assembled from identical subunits with helical symmetry. Protein filaments carry out unreplaceable roles in the cellular process. These complexes not only provide mechanical support to cytoskeletons but also contribute towards regulating critical signaling pathways that often determine the live and death fate of the cells. Thus, atomic-resolution structures of these protein filaments help to reveal biological insights of related cellular processes and may serve as druggable targets. As native state protein filaments are reluctant to crystallize, the most common method of structure determination of these filaments is to use Cryo-EM. Thanks to the rapid development of Cryo-EM instruments and reconstruction algorithms, it becomes a viable and efficient way to obtain an atomic-resolution filament structure nowadays. Recognizing the bottleneck step of this process, the filament picking, we developed a rapid and easy-to-use software. Also included in several examples in this thesis, we have demonstrated that we have established a robust workflow in optimizing filamentous protein constructs, preparing filamentous cryo-EM samples, and resolved their atomic structures using cryo-EM helical reconstruction methods.

In one example, Uromodulin (UMOD) is a secretory protein that belongs to the zona pellucida-like domain family. It is the most abundant protein in normal human urine, which forms filament and captures uropathogens, principally serving as decoys to gram-negative bacteria. We solved UMOD native state filament structure with resolution 3.35 Å (EMDB-10553). This high-resolution structure revealed a unique interdomain linker that connects the ZP-N and ZP-C domain of UMOD. This inter-

locked chain architecture confers a tremendous resistance to a hazardous environment, particularly to proteases. Besides, as the first solved ZP domain filament, UMOD filament structure also provided insight on potential ZP4 ZP3 interactions, which is a critical biochemical event involved in the fertilization process.

In another example, Caspase recruitment domains (CARD) also tend to form filaments. They belong to a sub-family of the Death Domain superfamily, whose members play critical roles in cellular apoptotic events and innate immune signaling. CARD complexes are usually stabilized by homotypic interaction among various domains. Atomic-resolution structures of CARD oligomers help to reveal details of signaling regulatory mechanisms, which are highly sought after for drug development and mending-related autoimmunity diseases. NLRC4 belongs to such CARD domain containing receptor family, and it detects bacterial-associated pathogen patterns. Facilitated by NAIP5 and activates downstream ASC-CASP1 inflammasome signaling axis, the CARD domain in NLRC4 relays the activation signal via a cooperative oligomerization process. NLRP1 and CARD8 are two similar innate immune receptors that undergo auto-proteolytic activation, when encountering the right triggering signals from cellular damage or pathogen infection, and release a filament-forming CARD domains. Here, as part of my Ph.D. research, I present structures of NLRC4-CARD, NLRP1-CARD, and CARD8-CARD filaments, at approximately 3.7 Å resolution. These results reveal signaling filament activation mechanisms conducted by specific CARD-CARD interactions.

Chapter 1. Background

1.1 Filamentous Protein Complex

Filamentous protein complexes play a widespread role in the cellular process. For example, the cytoskeleton is mainly composed of the actin filament, microtubule filament, and intermediate filament. These structural filaments mediate cell transport and maintain the shape of eukaryotic cells. Apart from these mechanical support functions, non-cytoskeleton protein filaments also widely exist in cells. These protein filaments participate in many critical cellular processes, like metabolism regulation and signaling transactions, thus related defects may result in severe health problems. Previous studies revealed that the filamentous protein complex involves many diseases, such as autoimmunity disease, neurodegenerative disorder, cancer, and infectious disease (Park and Horton 2019). Apart from independent functions, protein filament can also polymerize and bind with another matrix, like lipid membrane, DNA, or another protein complex (Ghosal and Lowe 2015). Hence, structure study for protein filament provides researchers deeper insight about related disease mechanism and promote drug discovery.

However, before the maturity of Cryo-EM technology, X-ray crystallography dominates protein structure determination, which is not suitable for native protein filament structure determination. Therefore, the high-resolution structure detail of these filamentous protein complexes was rarely reported before. Thanks to the development of Cryo-EM, it becomes possible to get a high-resolution map of a giant molecular complex. In addition, much near-atomic resolution filament structure was reported in

the past decade and revealed significant biochemistry insight for related health problems.

1.2 Helical Reconstruction Method for Cryo-EM

Cryo-EM has become one of the most potent tools for structure biology (Figure 1.2.1). Compared with traditional methods like crystallography and NMR(nuclear magnetic resonance), Cryo-EM provides good preservation and views of native and hydrate state macromolecules, and high-resolution reconstruction for macromolecules complex structure becomes possible (Earl, Falconieri et al. 2017). Nowadays, leading to such near-atomic resolution, Cryo-EM has also become an essential tool for drug discovery.

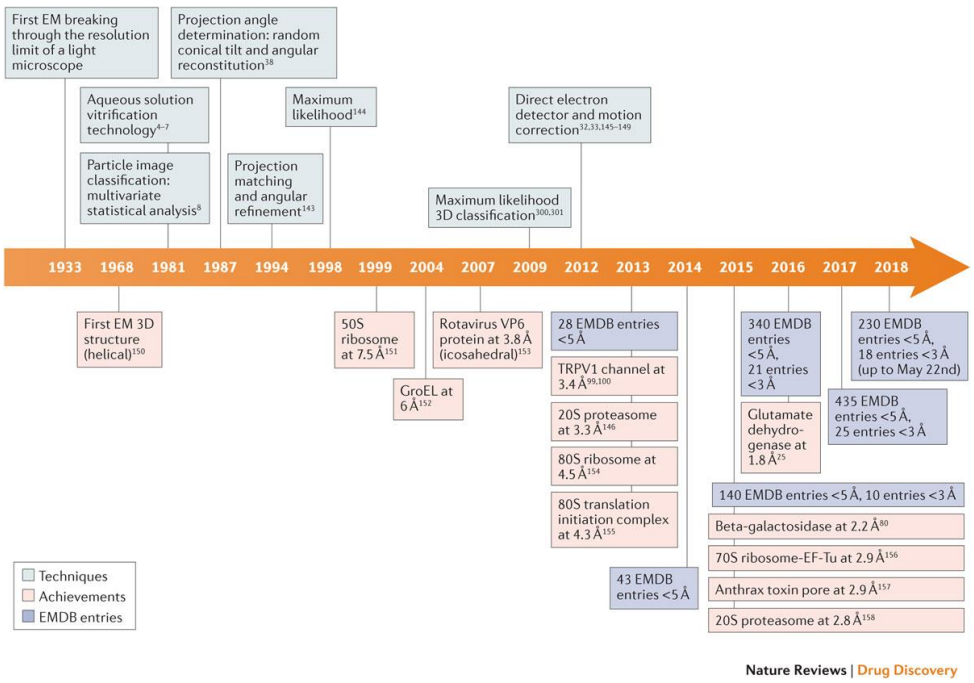

Figure 1.2. 1 Timeline for Cryo-EM development and achievement. Adapted from (Renaud, Chari et al. 2018)

While single particles analysis (SPA) is currently the most common method in Cryo-EM reconstruction, helical reconstruction is also of vital importance. Helical reconstruction is one of the earliest protein structure determination algorithms applied in three-dimension electron microscopy. Although filamentous complex can also be solved with the single particles method, helical reconstruction can achieve higher resolution with fewer data. Single particles analysis requires the orientation of each particle, which is obtained from an alignment task in three-dimension space. However, helical reconstruction only needs to find particle orientation pointed to the central cylinder, as the central axis is fixed.

Furthermore, the helical symmetry of the subunit in filament also promotes the reconstruction performance. Once correct helical twist and helical rise are determined, experimental noise resulted from data collection can be reduced by an averaging operation. Benefiting from this property, the signal-to-noise ratio is significant enhanced during 2D averaging and 3D reconstruction. With this advantage, the structure of the T4 bacteriophage tail was reported in 1970 with the helical reconstruction method, which is the first determined EM 3D structure (De Rosier and Klug 1968).

High-resolution Cryo-EM structure determination is usually comprised of four steps, sample preparation, data collection, Cryo-EM image processing, and coordinates building. During sample and Cryo-Grids optimization, several operations are iteratively performed (Figure 1.2.2). Instrument improvement is the basis of cryogenic electron microscopy. In the past twenty years, hardware, especially electron detectors, develops extremely fast, which results in better micrograph quality and faster data

collection speed (Veesler, Campbell et al. 2013, McMullan, Faruqi et al. 2016). Besides, fast-developing Cryo-EM processing software also contributes significantly to this field. In the first decade of the 21st century, electronic microscopy software is programming-based, which requires professional computational and programming knowledge (Ludtke, Baldwin et al. 1999, Baldwin and Penczek 2007). However, when it comes to the 2010s, more efficient and user-friendly tools are developed—for example, Relion and Cryosparc (Scheres 2012, Punjani, Rubinstein et al. 2017). With high-performance instruments and advanced software, it has become more convenient for structural biology researchers to obtain a high-resolution electron density map. Moreover, the processing pipeline has become more standardized. A standard pipeline of Cryo-EM image processing includes several parts, reference correction, motion correction, CTF (contrast transfer function) estimation, particle picking, 2D classification, 3D classification, 3D refinement, and post-process (Zheng, Palovcak et al. 2017).

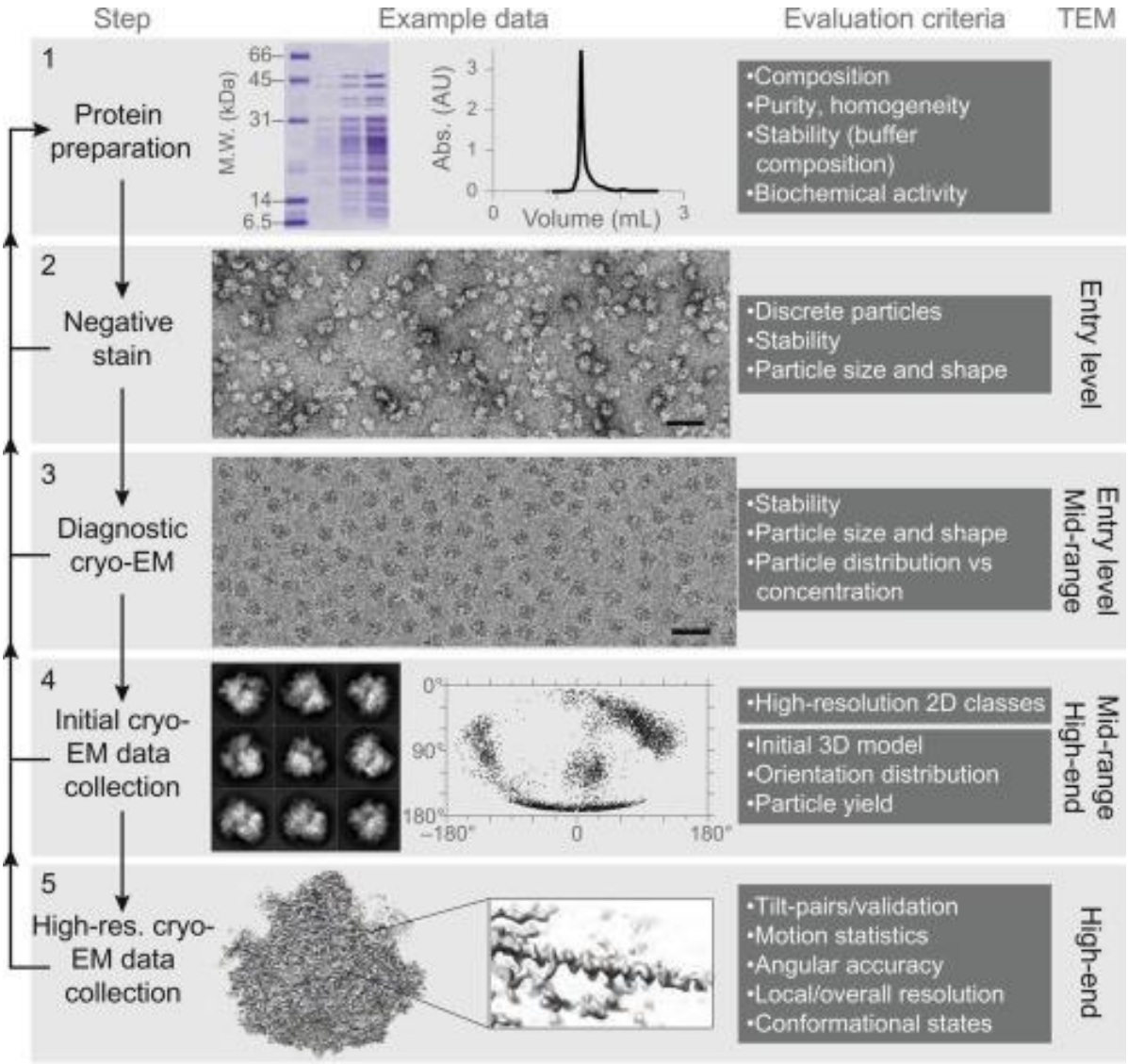

Figure 1.2. 2 Pipeline in Biological Cryo-EM. Adapted from (Passmore and Russo 2016). A standard Cryo-EM structure determination protocol include: Protein sample purification, Negative stain quality checking, Cryo-EM screening quality checking, Cryo-EM data quality checking, final data collection and Data Processing

2D classification, 3D classification, and 3D Refinement are the most compute-bound tasks, which require higher performance computation hardware. Here, 2D classification and 3D Classification can be seen as the Expectation-Maximization task. The goal is to identify 2D averages and 3D density that can represent micrograph

particles and separate homogeneous particles based on the latent statue (2D average and 3D density). Then the homogeneous particles are feed into 3D refinement for high-resolution detail reconstruction. The task that generates the 3D structure of macromolecular from particles micrographs can be described as the Bayesian likelihood optimization process (Punjani, Rubinstein et al. 2017).

$$arg^{max}_{V_{1\dots K}} log\big(p(V_{1\dots K}|X_{1\dots N})\big) =$$

$$arg^{max}_{V_{1\dots K}} \sum_{i=1}^{N} log \sum_{j=1}^{K} \frac{1}{K} \int p\big(X_i, \emptyset_i|V_j\big) d\emptyset_i + \log\big(p(V_{1\dots K})\big)$$

Equation 1. The formula of structure determination from micrograph particles

Here, $X_{1\dots N}$ is the micrograph particles and $V_{1\dots K}$ stands for potential 3D density, which can explain, $X_{1\dots N}$ as much as possible. As all particles are captured in the same plane during data collection, the depth information is fixed. Therefore, the optimization task is iteratively searching $\emptyset_i$ the related orientation to density V_j and reconstruct V_j with $\emptyset_i$ and X_i . This optimization can process in both real space and Fourier space.

1.3 Autopicking Method for Filamentous Sample

Particles picking is an essential step in Cryo-EM image processing. This step determines the quality of particles used in downstream tasks. As the low signal-to-noise ratio nature of Cryo-EM micrographs, many specimen particles are required for high-resolution structure reconstruction. Traditional manually picking is labor- and time-consuming. With the growing number of collected images, automatically picking methods become necessary. In general, there are two types of auto-picking methods, semi-automatically picking and full-automatically picking. Semi automatically picking

relies on the template from either picked coordinate, 2D class average, or 3D map projection, which requires prior information or preprocessing (Bepler, Morin et al. 2019). Template matching based and deep learning based algorithms are usually applied. Full-automatically picking detects particles directly based on Gaussian kernel matching or edge detection method. Several mature picking methods of these two categories were introduced for single particles processing and achieved remarkable performance. However, when dealing with the filamentous specimen, this task becomes more different and challenging. Compared with single particle picking, keeping only the basic coordinate is not enough for the filament picking algorithm. The algorithm should be able to track the helical information between helical segments. Firstly, contours of filament should be detected, and a smooth central axis, which can represent the filament, needs to be extracted from the contour. This process is challenging, especially for some flexible filament.

1.3.1 Existing Filament Picking Methods

Semi automatically filament picking methods base on either template matching or deep learning were introduced previously (Huber, Kuhm et al. 2018, Wagner, Lusnig et al. 2020, Thurber, Yin et al. 2021). These methods all require prior knowledge. As for template-free strategy, there is few choices, highlighting the importance of a fully automatic template-free filament picking method, which can handle realistic Cryo-EM images.

Moreover, particle picking itself is not the only task that we need to care about. The ideal Cryo-EM micrograph is only composed of sample region and clean background

region. The actual images researchers facing are usually more complicated. Apart from a blank background, the undesirable region also includes artifacts and contaminations. High contrast artifacts and contaminations exist in micrographs generated during sample preparation, grid preparation, and data collection. These artifacts and contaminations include the carbon region, the holes edge, ethane ice, crystalline ice, freeze-dried particles, etc. Artifacts and contaminations significantly affect the performance of the picking algorithm and introduce additional false-positive particles in downstream 2D averaging or 3D reconstruction tasks.

As a result, the auto-picking algorithm should do more than sample-background discrimination; it should also exclude the effect of artifacts and contaminations. Algorithms that are designed to detect specific contamination were reported before. The Appion package computes the range of carbon holes through image morphing operation (Lander, Stagg et al. 2009). Alternatively, based on adaptive thresholding and circle fitting, EMPH achieves a similar task (Berndsen, Bowman et al. 2017). Moreover, for high-contrast contamination, the most common trick used in the template-matching method is setting a proper high-pass filter threshold, which is presented in Relion auto-picking job. Nevertheless, many real particles are also filtered at the same time. In addition, it takes time to estimate the suitable threshold numerical value by multiple trials.

1.3.2 Robust Solution to Reduce False Positive Rate

As we see, manually constructed features and prior designed rules can help to reduce the effect of artifacts and contaminations. However, they are far away from a practical

and easy-to-use method. With the development of neuron networks, especially convolution neuron networks (CNN), a great revolution has come into the image processing field. An adequately designed CNN can automatically extract features from labeled data and finish the discrimination task. Once the network training is finished, the inference step is fast and effective. This kind of property indicates CNN may be a modern method for micrograph masking and particles picking. Although different specimen particles may have different patterns, the artifacts and contamination in Cryo-EM micrographs share common patterns. Consequently, a pre-trained artifact and contamination masking CNN is a good solution for undesirable region labeling tasks.

Warp is the first software that ensembles this kind of pre-trained CNN in their picking method, making single particle picking more robust and accurate. Moreover, when it comes to 2020, they separate the masking module and make it an independent library, generating a mask for other usage purposes (Sanchez-Garcia, Segura et al. 2020). Filament auto-picking software should also integrate this method in the current algorithm.

Many software claims that they provide filament autopicking function; however, they cannot correctly mark the right sample region precisely in our experiment, especially for images with bad contrast. A rapid evaluation system for the Autopicking result is absent, which is essential for parameters adjustment. There are also few direct comparison studies for existing autopicking software.

1.4 CARD containing protein and high-resolution reconstruction for NLRC4, NLRP1, and CARD8

1.4.1 Immunity System and Innate Immunity

The primary role of the animal immunity system is to prevent invasion and eliminate pathogens (Janeway 1992, Banchereau and Steinman 1998). Based on the difference of responding mechanisms, the immunity system is composed of innate immunity and adaptive immunity (Janeway 1989, Medzhitov and Janeway 2000) (Figure 1.4.1).

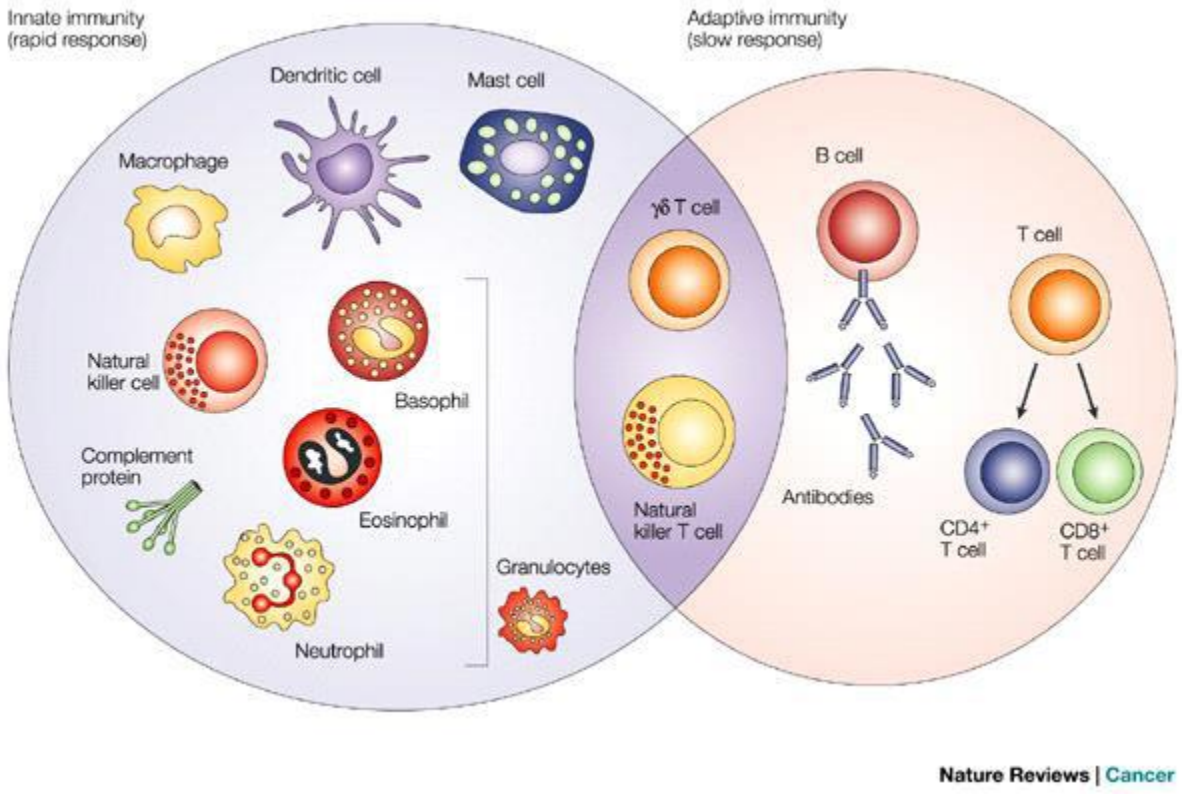

Figure 1.4. 1 Innate and Adaptive Immunity, Adopted from (Dranoff 2004)

B cell or T cell conducts adaptive immunity, which generates specific antibodies to neutralize antigens or induce apoptosis of infected cells (Dudley 1992, Medzhitov, Preston-Hurlburt et al. 1997). Adaptive immunity is memorable, which means pathogens information is stored with memory cells during the process (Demkowicz, Littaua et al. 1996). Innate immunity, also known as non-specific immunity, is quite different from adaptive immunity. However, innate immunity is also vital, as it serves as the first guard for micro-pathogens. Components of innate immunity establish a system to discriminate self-organism and pathogens (Janeway 1992, Medzhitov and Janeway 2000). This discrimination mechanism is achieved by evolutionarily conserved protein family pattern recognition receptors (PRRs). Pathogen-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs) are the important conserved molecular structures that are recognized through PRRs. PAMPs are molecular structures deriving from micro-pathogens like bacteria and viruses. In addition, DAMPs are molecular structures generated from organism damage or dying cells (Janeway 1989, Roh and Sohn 2018). The first identified signaling PRR is from Drosophila melanogaster, whose antimicrobial molecular generation is mediated through Toll (Lemaitre, Nicolas et al. 1996). Pursuing research indicates that homologous genes of Toll also exist in human and act an important role in human innate immunity system. Bacterial lipopolysaccharide (LPS) is a common antigen in human innate immunity and triggers Toll-like receptor activation (Medzhitov, Preston-Hurlburt et al. 1997, Akira and Takeda 2004). After 2000, with rapid growing genome sequencing technology and more immunology studies, more PRRs have been characterized. Till now, well-studied PRRs include NOD-like receptors (NLRs), Toll-like receptors (TLRs), absent-in-melanoma (AIM)-like receptors (ALRs), Retinoic

acid-inducible (RIG)-I-like receptors (RLRs), and C-type-lectin receptors (CLRs) (Meylan, Tschopp et al. 2006) (Table 1.4.1).

Table 1.4.1 Summary of PRRs

PRR type	shared function domain	Ligands	Ligands Sources	localization
NLR	NBD,LRR	PAMPS: iE-DAP, muramyl dipeptides	Bacteria, self	Cytoplasm
		DAMPS: Endogenous danger signals		
TLR	LRR,TIR	PAMPS: Lipoproteins, DNA, RNA, endotoxin	Bacteria, viruses, parasites, self	Plasma membrane
		DAMPS: Endogenous danger signals		
ALR	PYRIN,HIN	DAMPS: Endogenous danger signals	self	Cytoplasm
RLR	DExD/H	PAMPS: dsRNAs	RNA viruses	Cytoplasm
CLR	CLEC	PAMPS: beta-glucans	fungi	Plasma membrane

Table 1.4. 1. NBD (nucleotide-binding domains), LRR (leucine-rich repeat), CARD (caspase activation and recruitment domains), TIR (Toll/interleukin-1 receptor), HIN (Hematopoietic expression, interferon-inducible nature, and nuclear localization), CLEC(C-type lectin)

1.4.2 Discovery of Pattern Recognition Receptors (PRRs)

NLRs, known as Node-like receptors was firstly discovered as immunity components in plants, which protect plants from microbial and parasites (Jones and Dangl 2006). Later, homologs of NLRs were identified in others species, like sea urchin and basal metazoans (Inohara and Nunez 2001, Hibino, Loza-Coll et al. 2006). This result indicates that NLRs in evolutionary conserved from primitive organisms to vertebrates, as its critical role in antigens defense. Moreover, there are 23 NLRs genes in humans, in contrast with 34 homologous NLRs genes in mice (Franchi, Warner et al. 2009).

As an essential part of human PRRs, NLRs act as cytosolic sensors and activators in innate immunity. After detecting danger-associated ligands and micro pathogens, NLR proteins assemble a large macro molecular complex called inflammasome. Inflammasome work as signal transduction center, which triggers cytokine maturation and inflammatory cell death; pyroptosis. Assembling the supermolecular signaling complex is a typical pattern in the vertebrate innate immunity system (Lin, Lo et al. 2010, Hou, Sun et al. 2011). Similarly, the assembling of inflammasome complex also relies on death domain proteins (DDs). Death domains containing proteins are involved in the whole pyroptotic signaling initiation, amplification, propagation through both homotypic or heterotypic interactions (Wen, Miao et al. 2013, Lu, Magupalli et al. 2014). Death domain containing proteins that play a role as the sensor in inflammasome assembly can be classified into two groups, CARD (caspase activation and recruitment domain) and PYD (PYRIN domain) (Hoffman, Mueller et al. 2001, Ogura, Sutterwala et al. 2006, Chen, Shaw et al. 2009). In inflammasome initialization stage, the preliminary signaling filament is composed of these domains.

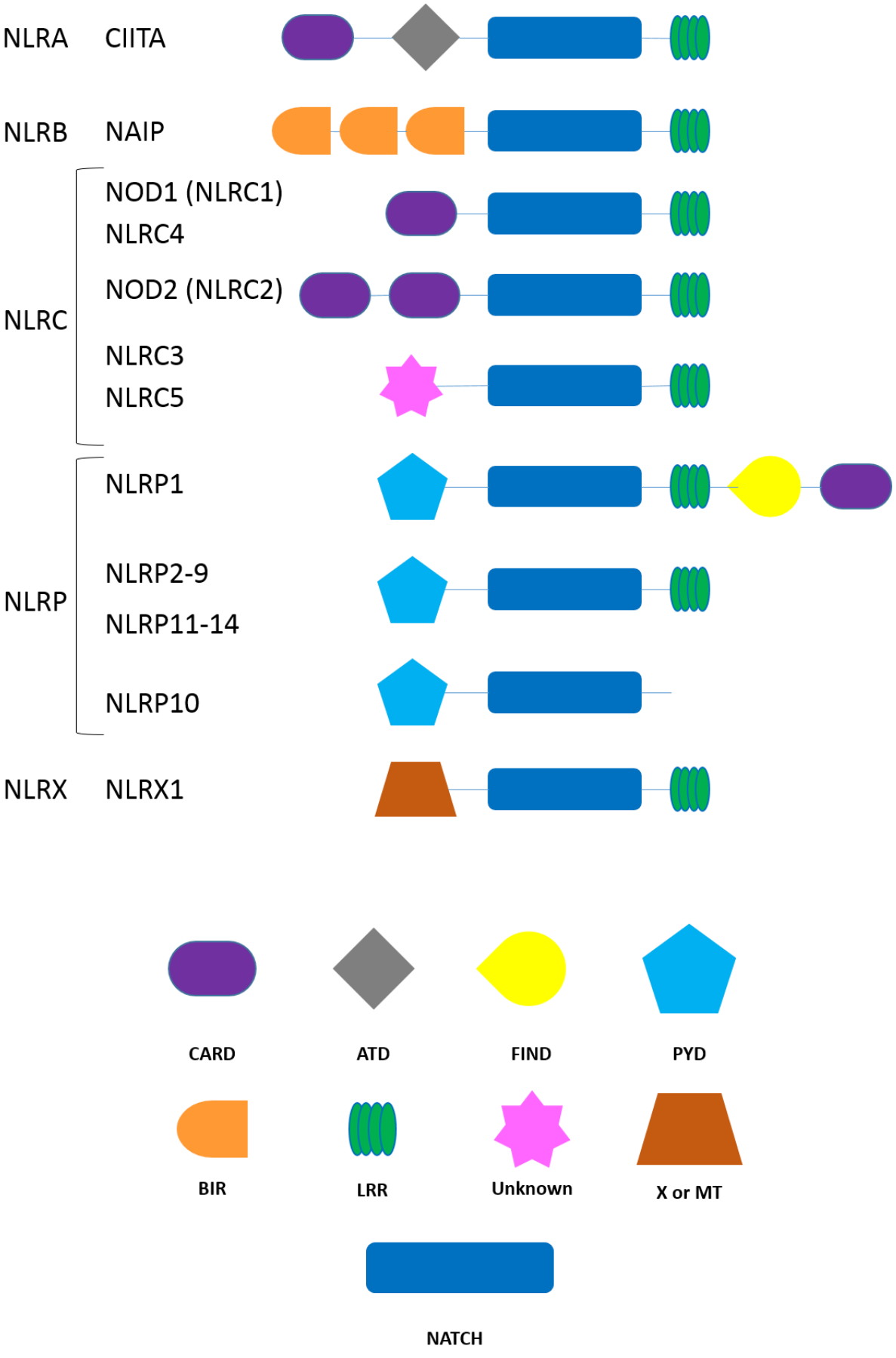
NLRA
CIITA
NLRB
NAIP
NLRC
NOD1 (NLRC1)
NLRC4
NOD2 (NLRC2)
NLRC3
NLRC5
NLRP
NLRP1
NLRP2-9
NLRP11-14
NLRP10
NLRX
NLRX1
CARD
ATD
FIND
PYD
BIR
LRR
Unknown
X or MT
NATCH

Figure 1.4. 2 Members of NLPs and their domain architecture.

Suggesting from the domain architecture, NACHT domain and LRR are the most conserved and prevalent components in intracellular NLRs (Carneiro, Travassos et al. 2007) (Figure 1.4.2). The pattern of NLRs domain architecture also demonstrates that there are three functional domains among NLRs (Figure 1.4.2). It is classified into three parts an N-terminal function domain, a central NACHT domain, and a C-terminal LRR domain. Here, the N-terminal functional domain work as a domain binding effector, which provides protein-protein interaction sites. Apart from CARD and PYD mentioned above, these N-terminal functional domains also include acidic transactivation domain (ATD), baculovirus inhibitor repeat (BIR), and X (function not defined) (Figure 1.4.2). The central NACHT domain plays a role in self-oligomerization and nucleotide binding. As for the C-terminal LRR domain works as a ligands recognizer and detector for conserved antigens (Franchi, Warner et al. 2009, Kim, Shin et al. 2016).

The N-terminal function domain is critical for inflammasome formation and signal propagation. With the diversity of N-terminal functional domain, NLRs can be classified into different subfamilies: NLRA (ATD containing NLR), NLRC (CARD containing NLR), NLRP (PYD containing NLR), NLRB (BIR containing NLR), and NLRX (unknown domain containing NLR) (Ting, Lovering et al. 2008). PYD and CARD are the most common N-terminal functional domains in NLRs, as they exist for more than 50% of NLRs in humans and mice.

Identified NLRC subfamilies include NOD1 (NLRC1), NOD2 (NLRC2), NLRC3, NLRC4, and NLRC5 (Saxena and Yeretssian 2014). The featured N-terminal CARD

domain is associated with caspase1 or kinase recruitment which mediate signal transduction in innate immunity (Carneiro, Travassos et al. 2007). Although NLRC3 and NLRC5 does not contain N-terminal CARD domain, they are homologous to NLRCs and share structural similarity with them. That is why they are still classified into the NLRC subfamily. NOD1 and NOD2 work as sensors for peptidoglycans of invading Gram-negative bacteria and Gram-positive bacteria (Motta, Soares et al. 2015). The abnormal active immune response is harmful to the organism; consequently, the NOD1 NOD2 signaling network keeps a subtle balance between over-activated and effective defense. Previous research has shown that much auto-immunity disease derives from polymorphisms and mutations of NOD1 NOD2 (Carneiro, Travassos et al. 2007, Motta, Soares et al. 2015). NLRC3 is reported as an inhibitor for T-cell and TRAF6 activation (Schneider, Zimmermann et al. 2012). NLRC5 is a negative regulator for the inflammatory pathway (Benko, Magalhaes et al. 2010) (Figure 1.4.3). NLRC4 is a typical member of the NLRC subfamily and is related to T3SS and flagellin detection and we will talk about detail at later chapter.

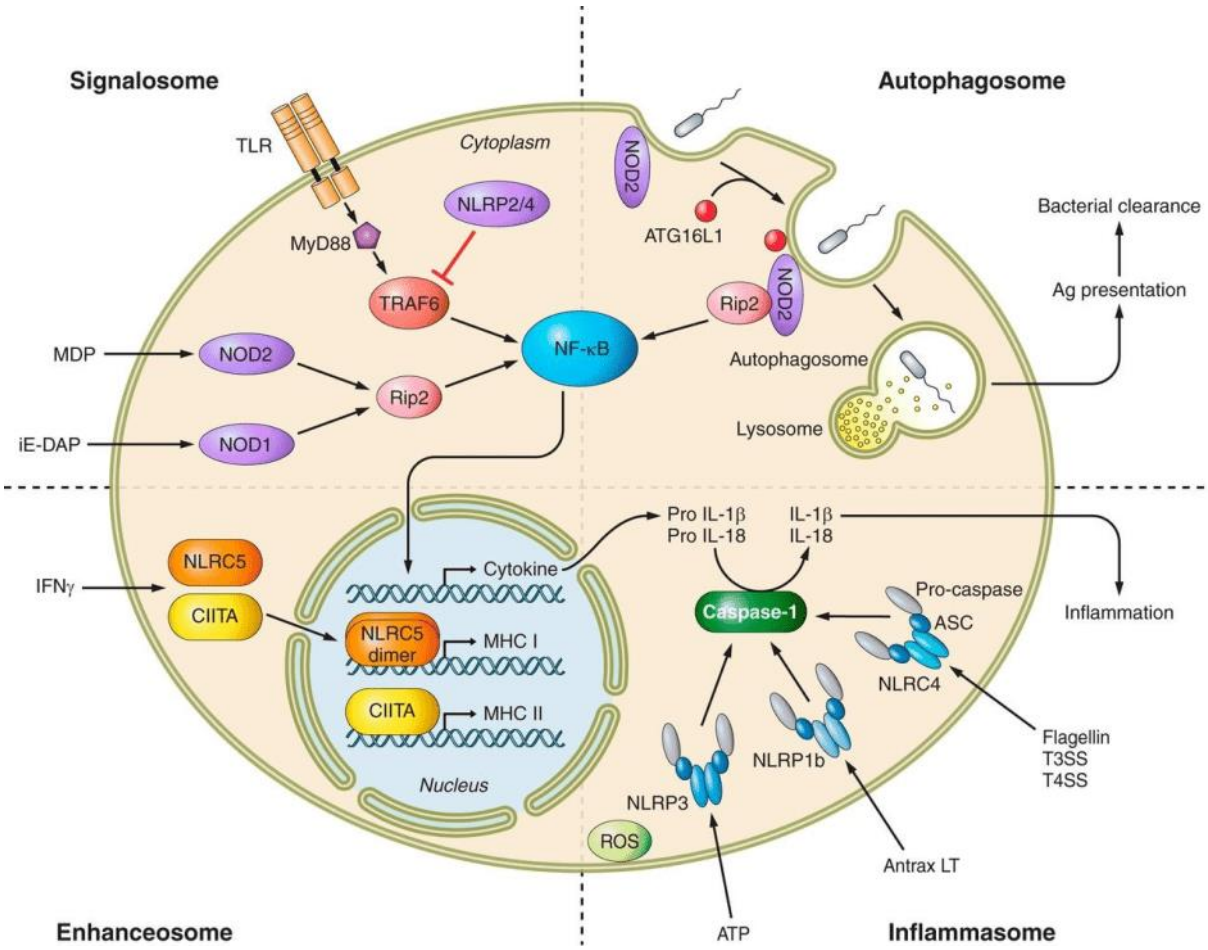

Figure 1.4. 3 The overview of NLR mediated signaling pathways, adapted from (Motta, Soares et al. 2015).

In contrast with the NLRC subfamily, whose featured domain is CARD, the N-terminal functional domain of the NLRP subfamily is PYD. NLRP subfamily has more and diverse homologs among different species (Proell, Riedl et al. 2008). NLRP1, NLRP3, and NLRP6 are expressed in many tissues and participate in inflammasome formation. This inflammasome is associated with caspase1 recruitment and induce IL-18 and IL-1β release. (Zhou, Yazdi et al. 2011, Levinsohn, Newman et al. 2012, Wang, Zhu et al. 2015). NLRP10 is unique in the NLRP subfamily, as it lacks the LRR domain. Recent research reports that NLRP10 is a negative regulator for inflammasome and plays a role in adaptive immunity activation (Eisenbarth, Williams et al. 2012). NLRP2

plays a role in mouse embryo development and negatively mediates the NF-κB pathway (Fontalba, Gutierrez et al. 2007, Peng, Chang et al. 2012). NLRP12 can suppress colon inflammation and tumorigenesis by inhibiting the NF-κB pathway and AKT pathway (Allen, Wilson et al. 2012). NLRP4 was found binding with beclin1 for negative regulation to autophagy activity (Jounai, Kobiyama et al. 2011). As for NLRP5 and NLRP7, they are related to mitochondrial activity and lipopeptides recognition (Fernandes, Tsuda et al. 2012, Khare, Dorfleutner et al. 2012). NLRP11 takes part in ubiquitin ligase related TRAF6 degradation, which mediates TLR-dependent signaling transduction (Wu, Su et al. 2017). More research also indicates that the NLRP subfamily works in limited function diversity, as it suffers selective constraints during evolution. For NLRBs, there is only one subfamily for it: NAIP. NAIP has multiple homologs in mice; however, there is only one NAIP gene in humans. NAIP5 in human work as a sensor for bacterial ligands and activate NLRC4 for downstream signal transduction. The only member of NLRA is CIITA (MHC-II transactivator), which is the regulator to MHC-II.

NLRs are in the auto-inhibited state before being activated by pathogens or damaged ligands, recognized through the LRR domain. Activated LRR then disassociated from NATCH, which results in NLRs transformation auto-inhibited state to active state. Then activated NLRs form inflammasome through self-assemble or recruiting other NLRs to transduce the signal to downstream pathways. Identified downstream effector enzymes include MAPKs, NF-κB, and caspase-1. These active effectors then induce maturation and release of precursor inflammatory cytokines or chemokines. Typical cytokines include interleukin 1β, interleukin 18, and Gasdermin D, which initializes pyroptosis (Kim, Shin et al. 2016).

NOD1 is the first discovered NLRs in the discovery of NLRs, and NOD2 is identified through homologous alignment. NF-κB and MAPKs related inflammation formation dependent on NO1, NOD2 mediation. Upon binding with agonists, NOD1 and NOD2 take on an active conformation and result in downstream RIP2 oligomerization, which then triggers NF-κB and MAPKs activation (Ogura, Inohara et al. 2001, Caruso, Warner et al. 2014) (Figure 1.4.3). Caspase-1 mediate another inflammation related signaling pathway whose downstream cytokines include IL-1β and IL-18 (Srinivasula, Poyet et al. 2002, He, Mekasha et al. 2010). Due to the N-terminal CARD domain, Caspase-1 activation depends on either direct CARD-CARD interaction or ASC mediated indirect interaction, which work as an adaptor. Specifically, NLRC4 can activate caspase-1 directly (Vance 2015). NLRP1 and NLRP3 need the assistance of ASC (He, Mekasha et al. 2010) (Figure 1.4.4).

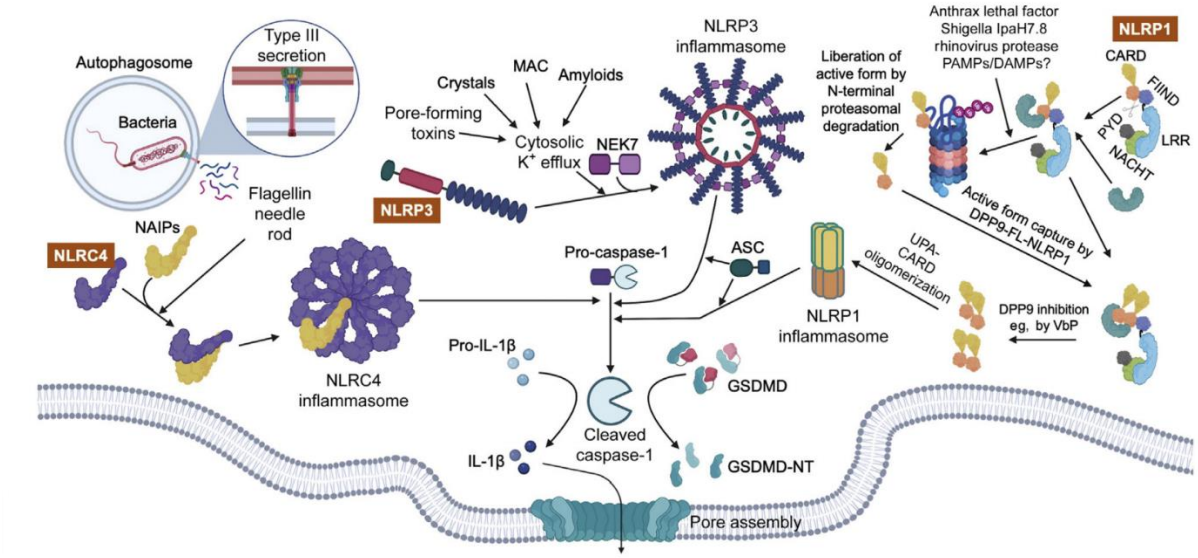

Figure 1.4. 4 The NAIP-NLRC4, NLRP3, and NLRP1 inflammasome pathways., Adopted from (Wang, Sharif et al. 2021)

1.4.3 NLRC4 and NLRC4-based Inflammasome

CARD family (Hofmann and Bucher 1997) is subfamily of death domain superfamily (Park, Lo et al. 2007). Members of the CARD family play essential roles in innate immune response signaling pathways, where CARD containing proteins pass the signal through forming filament-based inflammasomes. NLRC4, as a typical member of the CARD family, follows the same signaling transaction pattern.

In the history of NLRC4 discovery, it was first classified as an apoptotic protein named as ICE-protease Activating Factor 1, as NLRC4 can activate caspase-1 (Poyet, Srinivasula et al. 2001). Then it is discovered that its domain architecture is more similar to NLRs; the name is changed to NLRC4 (Sutterwala, Mijares et al. 2007). Animal experiments demonstrate that NLRC4 is associated with the detection of flagellin node protein of bacterial secretion systems, which may come from *S.Typhimurium (PrgJ), E.coli (EprJ), B.pseudomallei (BsaK), Shigella flexneri (MxiI), and Pseudomonas aeruginosa (PscI)* (Miao, Alpuche-Aranda et al. 2006, Miao, Mao et al. 2010). The direct stimulator of NLRC4, however, is not observed. Later, NAIP (neuronal apoptosis inhibitory protein) family is identified, which serves as sensor for pathogens (Kofoed and Vance 2011). NAIP can bind with conserved components of bacteria like PrgJ and flagellin and then recruit NLRC4 to form inflammasome (Miao and Warren 2010). There are seven NAIP encoding genes in mice; however, only one in humans (Yang, Zhao et al. 2013). Different NAIPs has diverse stimulus. For example, NAIP1 and NAIP2 are sensors for inner rod protein from bacterial T3SS. Furthermore, NAIP5 and NAIP6 are found to bind to bacterial flagellin from *Salmonella typhimurium* FliC (Zhao, Yang et al. 2011). It is currently found that NAIP2 and NAIP5

can activate NLRC4 directly, and two different associated inflammasomes are identified (Gong and Shao 2012). Additionally, the roles of Pkcδ in NLRC4 inflammasome assembly are still unclear. Current evidence shows, in the presence of T3SS pathogens, Pkcδ can induce the phosphorylation of NLRC4 at the site of Ser533 (Qu, Misaghi et al. 2012, Suzuki, Franchi et al. 2014) (Figure 1.4.5).

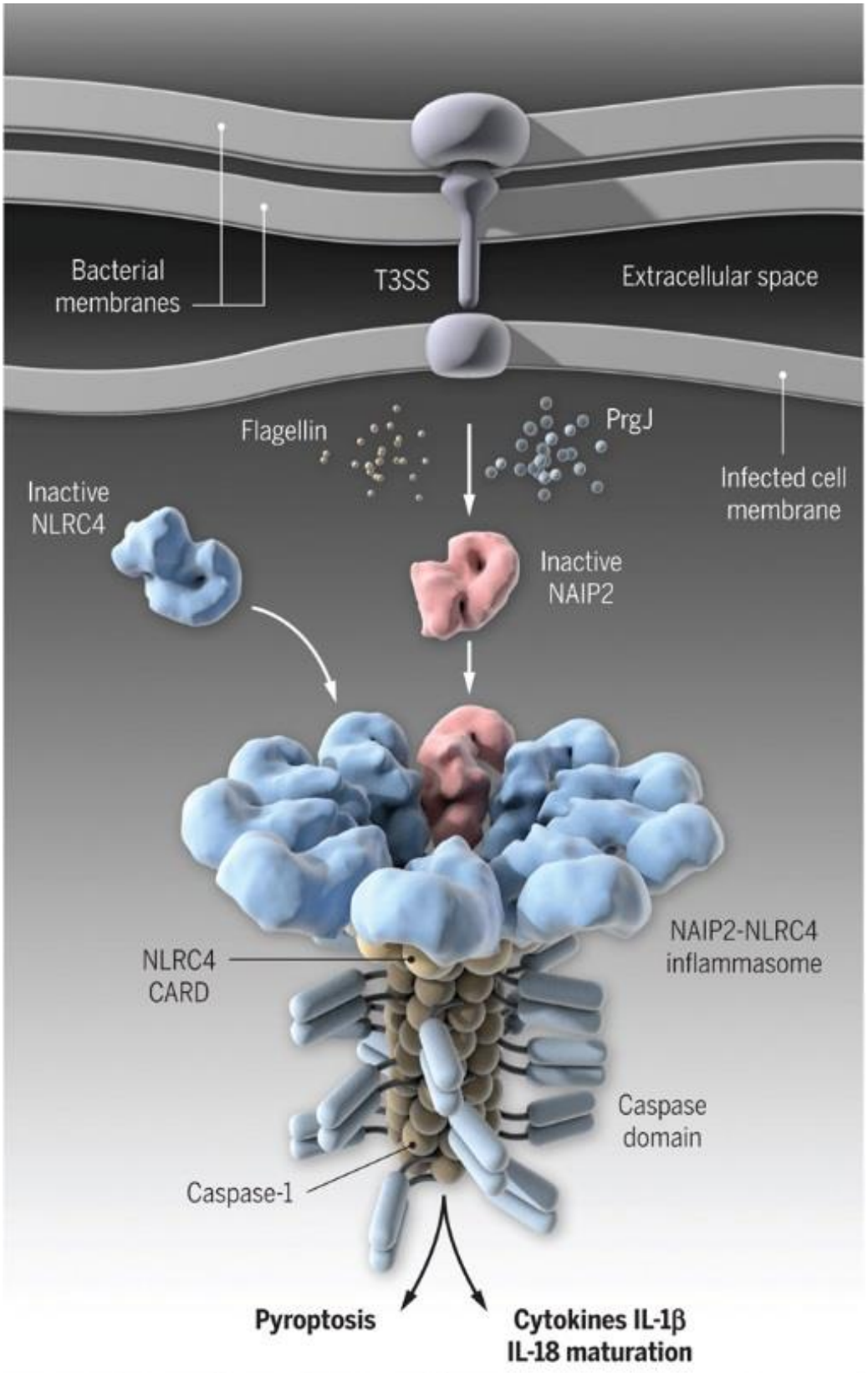
Bacterial membranes
T3SS
Extracellular space
Flagellin
PrgJ
Infected cell membrane
Inactive NLRC4
Inactive NAIP2
NAIP2-NLRC4 inflammasome
NLRC4 CARD
Caspase domain
Caspase-1
Pyroptosis
Cytokines IL-1β IL-18 maturation

Figure 1.4. 5 Activation mode of NLRC4 inflammasome, Adopted from (Liu and Xiao 2015)

Due to the vital role of NLRC4 in innate immunity, its gene mutation is related to many enterocolitis and auto-inflammation diseases (Canna, de Jesus et al. 2014, Kitamura, Sasaki et al. 2014, Romberg, Al Moussawi et al. 2014). For example, mutation p.V341A happened at NLRC4 HD1 domain will lead to cases of autoinflammation and enterocolitis, which is recorded from a patient family (Romberg, Al Moussawi et al. 2014). Another record shows that patient with mutation p.T337S on NLRC4 NBD domain suffers from recurrent fever, splenomegaly and other symptoms (Canna, de Jesus et al. 2014). Furthermore, patients from one family suffer from dermatitis and arthritis because of missense mutations on NLRC4 (Kitamura, Sasaki et al. 2014). These results indicate that gain-of-function NLRC4 mutations result in over-activation of NLRC4 related inflammasome formation, increasing chemokines' production (Rathinam, Vanaja et al. 2012). In order to develop a therapy for these diseases, it is meaningful to reveal the activation mechanism of NLRC4.

Structural study for NLRC4 is also increasing a lot in the past decade. With CARD domain deletion construct NLRC4Δ, PrgJ activated NAIP2/NLRC4 complex and Flic activated NAIP5/NLRC4 complex structure was solved with Cryo-EM (Zhang, Chen et al. 2015). Both NLRC4 inflammasomes exhibit disk-like structures with 10, 11, or 12 subunits (Zhang, Chen et al. 2015, Li, Fu et al. 2018).

As the functional domain of NLRC4, the existing of CARD domain makes full-length NLRC4 exhibit heterogeneous complexes conformation, which is not suitable for Cryo-EM structure determination (Li, Fu et al. 2018). Attempt for NLRC4 full-length structure

determination was reported at 2015. In this report, flagellin-induced NAIP5/NLRC4 complex was observed with Cryo-ET (Diebolder, Halff et al. 2015). This work achieves a resolution of 40 Å due to the resolution limitation Cryo-ET. However, this is still significant progress. The architecture shows that NATCH and LRR domains compose the outer disk of complex, and CARD filament works as the central column. The helicity of the NLRC4 inflammasome appears in this map, with a diameter of 28nm and pitch of 6.5nm. Based on the estimation, a coarse helical parameters rotation angle of 30.9° and twist of 5.57Å is calculated. Compared with CARD deletion NAIP5/NLRC4 complex, significant conformation change happened on NATCH and LRR domain, shown in Figures 1.4.6 below (Diebolder, Halff et al. 2015, Zhang, Chen et al. 2015). Compared with apoptosome, this extraordinary conformation change allows inflammasome to extend longer (Diebolder, Halff et al. 2015).

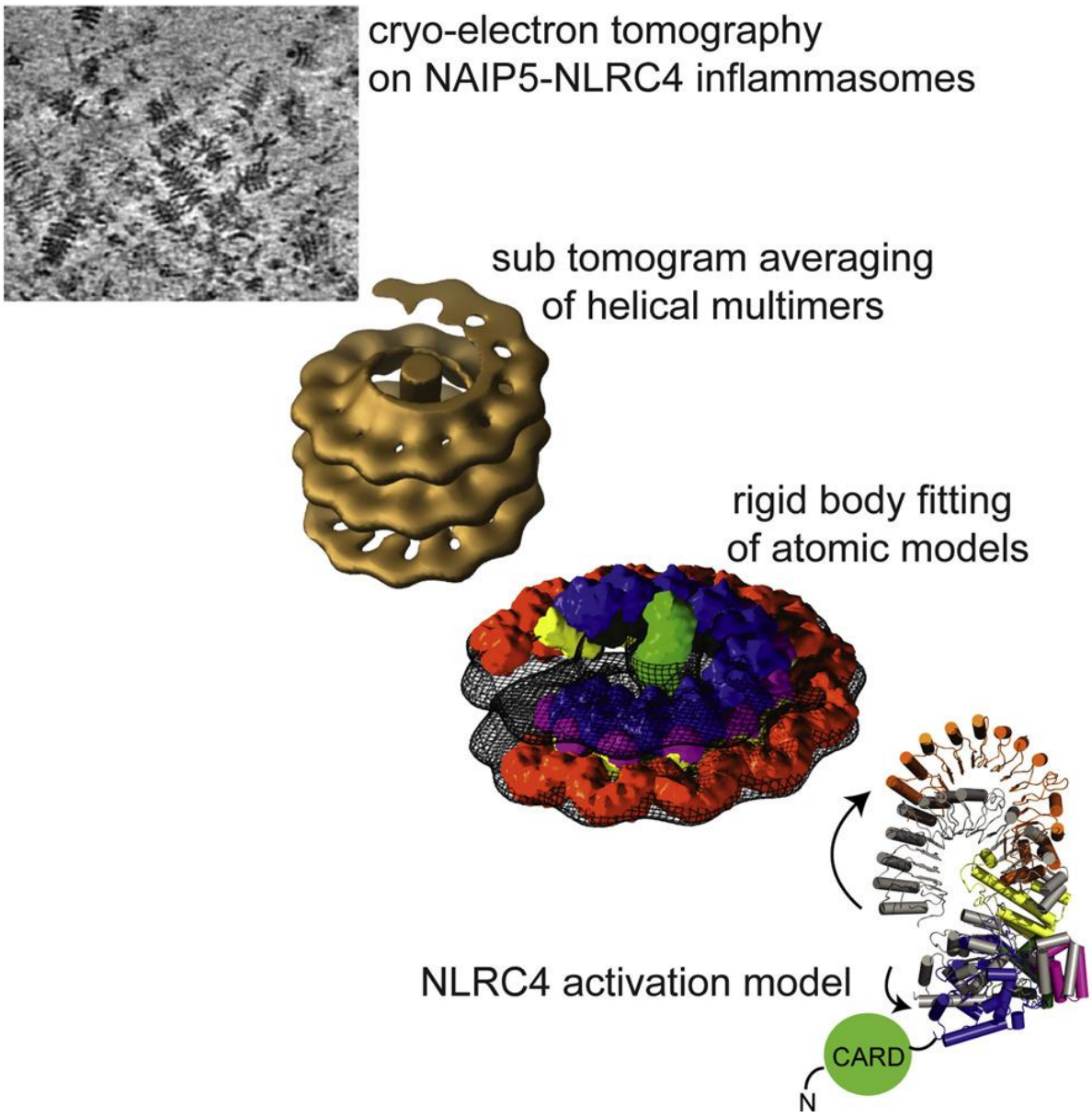

Figure 1.4. 6. Cryo-ET density map of Full-length NAIP5-NLRC4 inflammasome, stimulated by flagellin, adapted from (Diebolder, Halff et al. 2015)

Human NLRC4 contains an N-terminal CARD domain, a NATCH (central nucleotide binding) domain, and a C-terminal LRR(leucine rich repeat domain). The nucleation of NLRC4 is mediated by the N-terminal CARD domain (Figure 1.4.7A). In human macrophages, the canonical NLRC4 signaling pathway is triggered by bacterial flagellin and T3SS (type III secretion system) detection (Zhao, Yang et al. 2011). When

these antigen are processed by macrophage, T3SS needle proteins bind to NAIP (NLR family, apoptosis inhibitory protein) (Yang, Zhao et al. 2013).

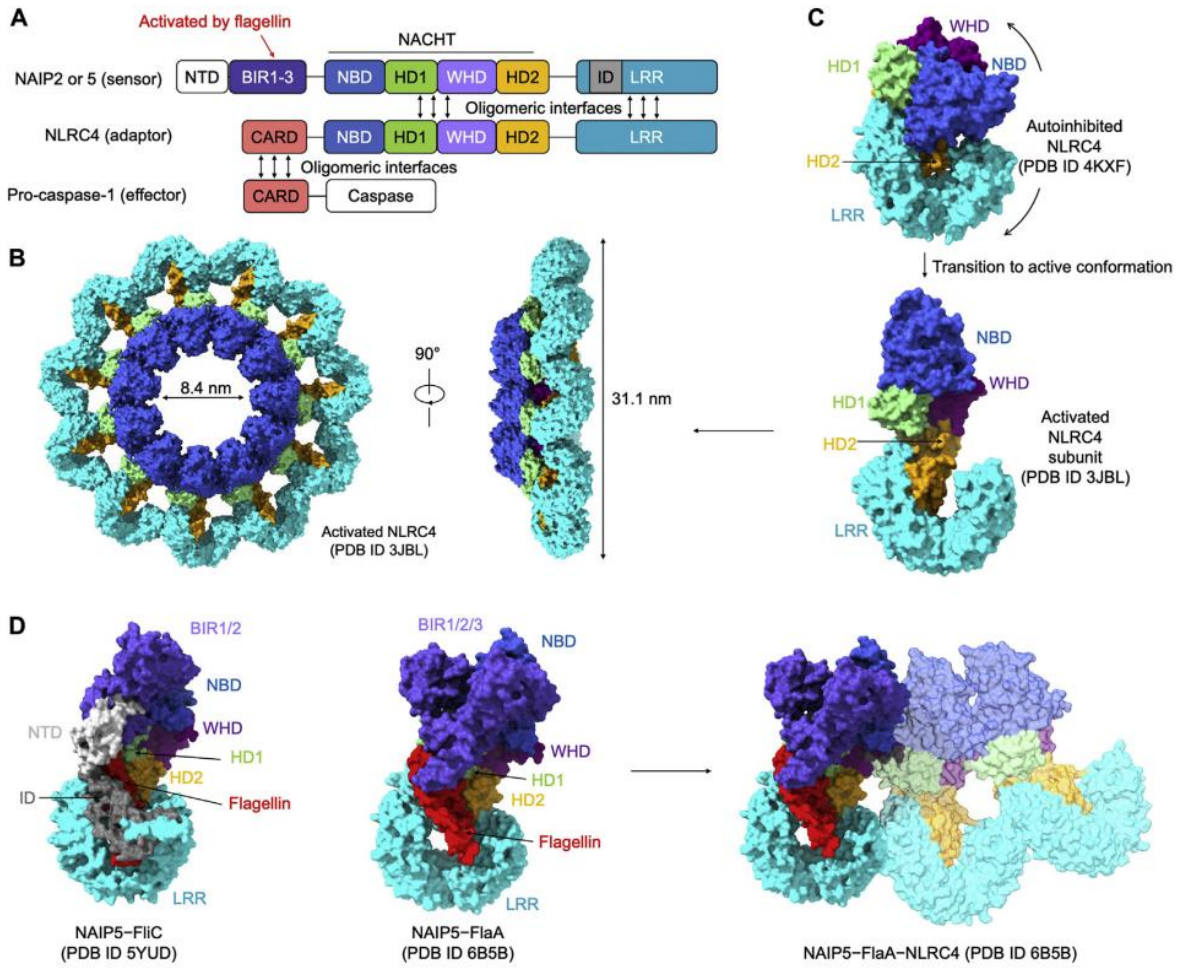

Figure 1.4. 7. Structural insights on NAIP-NLRC4 inflammasomes. (A) Domain architecture for NAIPs and NLRC4. (B) Active NLRC4 inflammasome. (C) Auto-inhibited NLRC4. (D) NAIP5 activation. Adapted from (Wang, Sharif et al. 2021)

Activated NAIP then recruits NLRC4 and induces signaling filament formation of NLRC4 inflammasome (Tenthorey, Haloupek et al. 2017, Duncan and Canna 2018). Monomeric NLRC4, which stays in autoinhibition, forms a bow-like complex (Figure 1.4.7). In further downstream signaling pathways, NLRC4 is able to activate caspase-

1 directly, which triggers cell pyroptosis or the release of cytokines. Although NLRC4 can recruit caspase-1 directly, NLRC4 can also activate ASC oligomerization, as both NLRC4 and ASC contain CARD domains. As typical CARD-CARD interaction, NLRC4 CARD-CARD interaction is also achieved by multiple interface and charge reversal contact between critical residues. Mouse NLRC4 CARD structure was obtained with resolution 3.2A, which is solved by crystallography (Hu, Yan et al. 2013). NLRC4 CARD monomer structurer and signaling filament structure was not solved until 2018.

1.4.4 Unique Architecture of NLRP1 and CARD8

NLRP1 is one of the earliest identified inflammasome protein components. In 2002, NLRP1 and its adaptor protein ASC was discovered with caspase-1 inflammasome (Martinon, Burns et al. 2002). Different from other NLRPs proteins, NLRP1 is also CARD-containing protein, which can be activated by many stimuli, including lethal toxin from Bacillus anthracis, Toxoplasma gondii, muramyl dipep-tide, and host intracellular ATP depletion (Boyden and Dietrich 2006). Apart from the C-terminal CARD domain, like other NLRs, NLRP1 contains an N-terminal PYD, a central NBD domain, a short LRR, and a FIIND. Significantly, different from many other CARD-containing proteins, NLRP1 functional CARD domain is located at the C terminal, which indicates a different activation mechanism for NLRP1 inflammasome. Furthermore, FIIND domain is a unique domain among NLRs, which only exist in NLRP1, CARD8, and a p53 induced protein (Janssens and Tinel 2012) (Figure 1.4.8 A). FIIND can be further defined as a tandem architecture of ZU5 and UPA. ZU5 is

firstly identified in UNC5 and Zonula occludens-1 (D'Osualdo, Weichenberger et al. 2011).

Compared with NLRP1, CARD8 owns a simpler domain architecture. CARD8 only contains a FIIND and a C-terminal CARD (Dieude, Guedj et al. 2011) (Figure 1.4.9 A). In addition, CARD8 FIIND is highly similar to NLRP1 FIIND, which is also composed of ZU5 and UPA (D'Osualdo, Weichenberger et al. 2011). Although CARD8 is able to form inflammasome, CARD8 is not classified into NLRs, whose members usually serve as sensor for inflammasome assembly, lacking the characterized NATCH and LRR domain. CARD8 function and its stimuli is not revealed clearly; however, the previous study shows that CARD8 works as a specific suppressor for NLRP3 inflammasome assembly (Mao, Kitani et al. 2018, Abate, Blomgran et al. 2019). A more recent and notable study shows that CARD8 is associated with the detection of HIV-1 activity (Wang, Gao et al. 2021).

Distinct from other NLRs, NLRP1 and CARD8 own the unique domain FIIND, which contributes to NLRP1 and CARD8 auto-inhibition and is not thoroughly studied. Recent research work shows auto-cleavage of FIIND is necessary for related inflammasome formation (D'Osualdo, Weichenberger et al. 2011, Mitchell, Sandstrom et al. 2019). And this auto-cleavage is mediated by DPP8 and DPP9. DPP8 and DPP9 (dipeptidyl peptidase) activity are highly related to NLRP1 and CARD8 activation. Known DPP8 and DPP9 inhibitor, such as VbP (Val-boroPro), is able to trigger the exceptional activity of NLRP1 and CARD8 inflammasome. For example, CARD8 abnormal assembly in human macrophages and NLRP1 abnormal assembly in airway epithelial cells and skin (Johnson, Taabazuing et al. 2018, Zhong, Robinson et al. 2018,

Johnson, Okondo et al. 2020, Linder, Bauernfried et al. 2020, Robinson, Teo et al. 2020). Many biochemistry studies reveal that DPP8 and DPP9 bind to NLRP1 and CARD8 directly (Zhong, Robinson et al. 2018, Griswold, Ball et al. 2019). And very recent structural studies reveal the structure detail of DPP9/NLRP1 and DPP9/CARD8 complex (Hollingsworth, Sharif et al. 2021, Huang, Zhang et al. 2021, Sharif, Hollingsworth et al. 2021) (Figure 1.4.8 B, Figure 1.4.9 D). Both DPP9/NLRP1 and DPP9/CARD8 complex exhibit a ternary form, including a DPP9 unit and two CARD8/NLRP1 FIIND segments. Both subdomains of FIIND; ZU5 and UPA, contribute to the formation of this complex. Although the mechanism is not revealed clearly, ZU5 here play a role of negative regulator for NLRP1 activation neither with DPP9 or without DPP9 (Huang, Zhang et al. 2021). DPP9/CARD8 complex is topologically similar to DPP9/NLRP1 complex; however, CARD8 UPA-CARD does not firmly interact with the active site of DPP9 (Sharif, Hollingsworth et al. 2021). These findings also validate the hypothesis that DPP9 works as an inhibitor for NLRP1/CARD8, and explain why DPP9 inhibitors is able to trigger abnormal NLRP1/CARD8 activation, shown as docking of VbP into DPP9/NLRP1 complex (Figure 1.4.8 C).

During NLRP1 and CARD8 activation, autoproteolysis happens on the ZU5 and UPA linker, which releases active UPA-CARD fragments. After autoproteolysis, activated NLRP1 and CARD8 segments are both able to transduce pyroptosis signal. A recent report demonstrates that NLRP1 and CARD8 are involved in the CARD domain based inflammasome assembly (Finger, Lich et al. 2012, Zhong, Mamai et al. 2016, Johnson, Taabazuing et al. 2018, Zhong, Robinson et al. 2018) (Figure 1.4.9 B), and disorder of these two proteins are associated with many autoinflammatory diseases,

emphasizing the necessity for NLRP1 and CARD8 study (Levandowski, Mailloux et al. 2013, Zhong, Mamai et al. 2016, Grandemange, Sanchez et al. 2017, Mao, Kitani et al. 2018). Due to the unusual domain architecture, the downstream activation mechanism of NLRP1 inflammasome is quite distinct. Contrast with NAIP/NLRC4 activation signaling pathway. NLRP1 is not recruited directly by specific sensor proteins, whereas NLRP1 filament activation is induced by pathogen enzymatic activity. Usually, downstream NLRs-caspase-1 recruitment is achieved by either direct CARD activation or indirect PYD-ASC activation. It is found that NLRP1 inflammasome induces activation of ASC, which is achieved by the C-terminal CARD rather than N-terminal PYD (Chavarria-Smith and Vance 2015). This signaling transduction relies on CARD filamentous complex; however, the filament assembly mechanism is not precise before. Although NLRP1 and CARD8 are similar in domain architecture, the signaling transduction dynamic seems not identical. The role of ASC, an adaptor protein, in signal transduction is not clear in both CARD-base inflammasome, as CARD domain can recruit caspase-1 directly (Duncan and Canna 2018). A very recent report shows that, in contrast with NLRP1, CARD8 can recruit caspase-1 without ASC acting as an adaptor (Ball, Taabazuing et al. 2020).

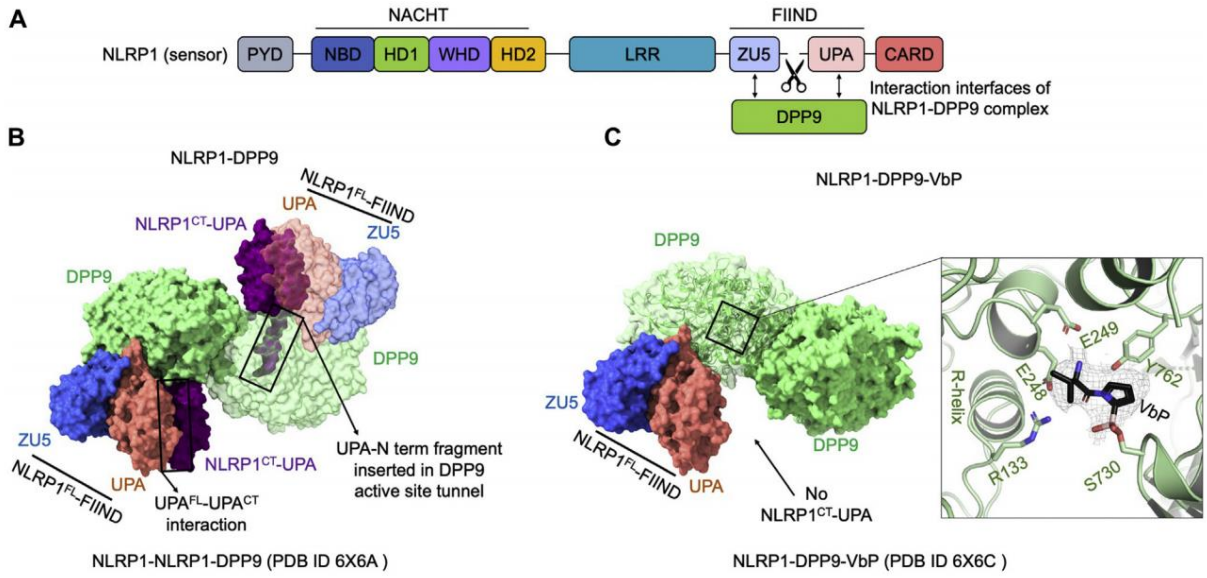

Figure 1.4. 8. Structural insights on the NLRP1 inflammasome. (A) domain architecture of NLRP1. (B)NLRP1-DPP9 complex. (C) NLRP1-FIIND and DPP9 interface. Adapted from (Wang, Sharif et al. 2021)

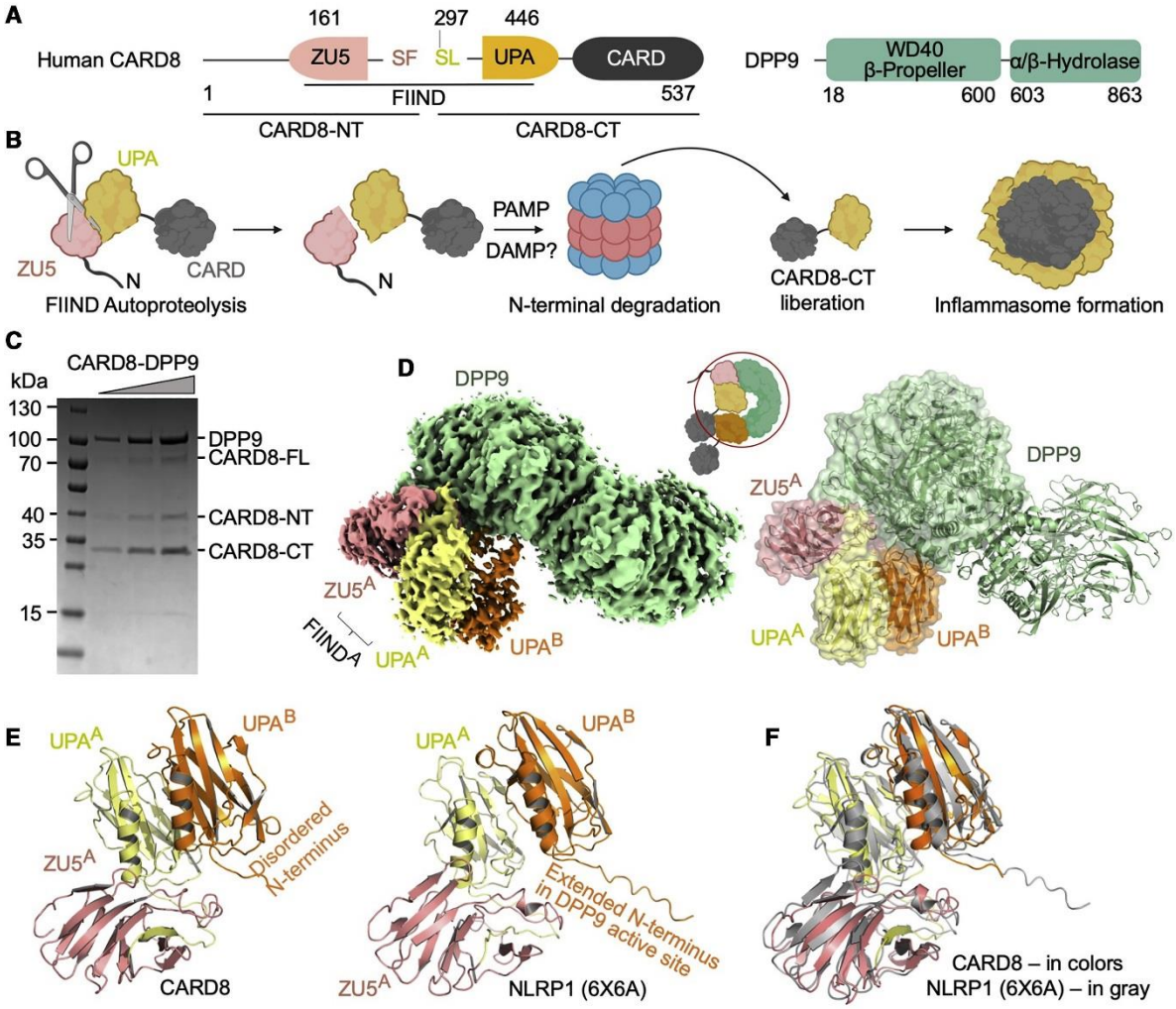

Figure 1.4. 9 Complex of CARD8 dimer and DPP9. (A) domain architecture of human CARD8 and DPP9. (B) Demonstration of CARD8 autoproteolysis and inflammasome assembly. (C) SDS PAGE of CARD8 AND DPP9 expression. (D) Electron density map of CARD8 dimer and DPP9 complex. (E) Protein coordinates of CARD8-DPP9 complex and NLRP1-DPP9 complex, side by side comparison. (F) Superimposed comparison between CARD8-DPP9 complex and NLRP1-DPP9 complex. Adopted from (Sharif, Hollingsworth et al. 2021)

1.5 Accurate CARD-CARD Interaction Prediction

CARD-CARD interaction is of vital importance for inflammasome oligomeric signaling scaffolds. CARD-CARD interaction between identical CARD involves the forming and extending of the signaling complex, which provides insight on the assembly mechanism. Moreover, cross-talk between different CARD domains participate in signaling transduction among different components in innate immune response signaling pathways. Identification of CARD-CARD interaction improves our understanding of innate immune responses (Qin, Srinivasula et al. 1999).

However, there are 28 CARD domain containing proteins in CARD family, which means there will be hundreds of potential interaction pairs. The time and resource consumption of biochemistry assays for every interaction pair will be massive, highlighting the screening for potential interaction CARD-CARD pairs in silico at the first stage. Protein-protein interaction prediction is an essential and classical field in the development of biology. Many computational methods were designed to solve this problem. For proteins with high-resolution structure, structure information can be included for prediction. Besides, machine learning technology, especially deep learning, has become very popular (Sun, Zhou et al. 2017). With the establishment of large protein-protein interaction database. It becomes possible to train machine learning model like neuron network, which requires a large amount of data. However, because of the complexity of CARD-CARD interaction, these methods do not work correctly on them (Zhang, Petrey et al. 2012). Although CARD domain secondary structures are conserved, CARD-CARD interaction relies on subtle residue contact between interfaces, including the featured Type I II III interfaces in Death Domain protein. Single mutation on the interface is sufficient to result in remarkable changes. Furthermore, two residues contacted on the interface may be far away in the amino

acid sequence, which means the pattern of CARD-CARD interaction is spatial-distributed. Recognizing these challenges, well-designed feature engineering, and data augmentation methods should be taken into consideration.

1.6 ZP Domain and UMOD Filament

1.6.1 ZP Module Proteins and ZP Domain

Zona pellucida, which is also called egg coat, is an extracellular matrix that surrounds mammalian oocytes. It is a glycoprotein layer that plays a critical role in fertilization and early embryo development. ZP proteins also belong to the immunoglobulin superfamily, whose members widely exist in human genomes and play a critical role in the immunity system (Barclay 1999). In the identification of ZP domain, ZP2 and ZP3 are discovered as vital components of zona pellucida, and their featured conserved domain is named ZP domain, which is conserved 260 amino acid sequences (Jovine, Qi et al. 2002). Multiple ZP domain proteins participate in the formation of zona pellucida and serve different functions. The featured ZP domain is the most critical functional domain containing 10 or 12 conserved Cys residuals. ZP domain is usually composed of ZP-C and ZP-N subdomain, which are connected through a short linker. Few ZP domain proteins contain only the ZP-N subdomain and are observed to form polymers through self-assembly in vitro (Jovine, Janssen et al. 2006, Monne, Han et al. 2008). As most ZP proteins work as secretive proteins, apart from the featured ZP domain, additional signaling peptide locate at N-terminal. ZP

proteins can form a filament structure, which contains repeat units. These filaments then can form a more complex structure through interconnection (Greve and Wassarman 1985, Green 1997). For ZP proteins in mammalian cells, ZP2 and ZP3 form filament and then crosslink by ZP1. This architecture composes the matrix of egg coat (Jovine, Darie et al. 2005, Han, Monne et al. 2010). VEbeta and VEgamma, which are vitelline envelope proteins in the trout, form similar structures (Darie, Janssen et al. 2008). Isolated UMOD ZP domain is sufficient to form filament-like structure, suggesting that the ZP domain is the main functional domain to promote filament formation (Jovine, Qi et al. 2002, Jovine, Janssen et al. 2006).

Among ZP domain members, functional ZP domain is usually associated with post-translational modification, including proteolytic digestion and glycosylation through linking oligosaccharides to asparagine- (N-) and/or serine/threonine- (O-). These modifications assign more complex functions for ZP proteins; however, they increase the difficulty of structure determination simultaneously. The native functional state of ZP protein is usually generated with highly specialized cell lines, which is hard to culture in vitro. With more and more related research reported, nowadays, several ZP domain family members are identified. ZP domain family members are widely expressed in neural and differentiated epithelial tissue and playing important diverse roles. Much independent research evidence for the function of the ZP domain demonstrates that ZP domain protein participates in cell differentiation, and mutations on the ZP domain gene are associated with human genetic disease. For example, in mammals, the ZP domain protein regulates the function of specific kinds of neural cells. DMBT1 is one of these proteins associated with sound transmission between neural cells (Jovine, Darie et al. 2005, Ligtenberg, Veerman et al. 2007). In the cochlea,

mutation of DMBT1 leads to an abnormal extracellular matrix. This apical gel matrix is vital for normal cochlea function (Petit 2006, Friedman, Dror et al. 2007). Endoglin/C105 is another ZP domain member, which is mainly generated from endothelial cells. It works as a regulator for the TGF-β signaling pathway (Massague and Chen 2000, Wong, Hamel et al. 2000). Previous research also reported that endoglin participates in blood vessel formation. Apart from proangiogenic function, it is reported that endoglin mediates epithelial mesenchymal transition through regulating malignant progression (Bernabeu, Lopez-Novoa et al. 2009).

1.6.2 Structural study of UMOD

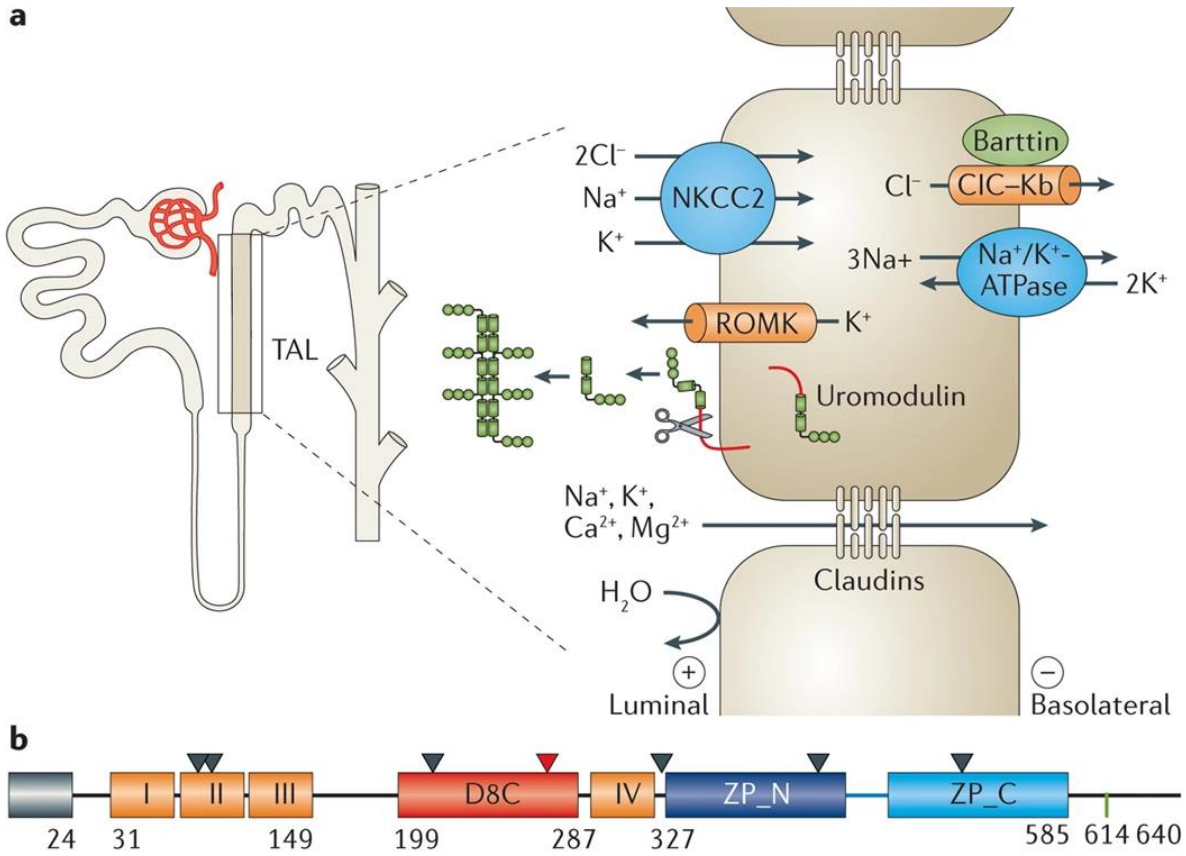

Figure 1.6. 1 UMOD produced in kidney cell and predicted domain of UMOD. Adopted from (Devuyst, Olinger et al. 2017)

UMOD (Uromodulin), sometimes also called THP (Tamm–Horsfall protein), is a secretion protein and is the most abundant protein in the urine. UMOD is also a member of the ZP domain family. ZP-N and ZP-C domain located at the C-terminal, which are functionally distinct but topologically similar IG-like (immunoglobulin-like) domains (Jovine, Qi et al. 2002, Jovine, Qi et al. 2004, Bokhove, Nishimura et al. 2016, Bokhove and Jovine 2018). Apart from the C-terminal ZP domains, UMOD also contains 3 EGF-like domains (epidermal growth factor-like) EGF I II III, an unclear-function named D8C, which contains multiple cysteine, the fourth EGF domain (EGF IV), a CCS/CFCS (consensus cleavage site) and C-terminal propeptide (Serafini-Cessi, Malagolini et al. 2003, Rampoldi, Scolari et al. 2011, Bokhove, Nishimura et al. 2016) (Figure 1.6.1 B). ZP-N and ZP-C domains are the main component that composes UMOD core filament. As mentioned above, among ZP domain family, ZP-N and ZP-C are topologically related; however they are featured through distinct disulfide bond intramolecular network. Furthermore, in the non-polymeric state, the interdomain linker, connecting ZP-N and ZP-C, exhibits either rigid or flexible conformation and lead to a different relative position of ZP-N and ZP-C, which were presented in previous crystallography study (Bokhove and Jovine 2018). Taking ZP3 as an example, the non-polymeric ZP3 exhibits an antiparallel homodimer conformation, where each ZP-N domain is placing against the ZP-C domain of the identical molecule and the ZP-C from the other molecule (Han, Monne et al. 2010). Here the intramolecular linker exists as disorder loop. However, in the precursor of UMOD, the interdomain linker exhibit rigid conformation and pack against ZP-C, which forms a structure of α1(α-helix) and β1(β-strand). Due to this rigid connection, in the precursor UMOD dimer, ZP-N domain is oriented to pack against the ZP-N domain of

the other molecule (Bokhove, Nishimura et al. 2016, Bokhove, Sadat Al Hosseini et al. 2016). Although different in linker conformation, ZP-C βG, the last β-strand, of both UMOD and ZP3, which is principally regarded as EHP (the external hydrophobic patch), belongs to polymerization-blocking CTP (C-terminal propeptide). Filamentous formation relies on the release of this C-terminal propeptide, which is mediated by protease (Jovine, Qi et al. 2004, Schaeffer, Santambrogio et al. 2009). Significantly, filament formation of both UMOD and mammalian egg coat is associated with precursor anchoring to the cell membrane (Jovine, Qi et al. 2002, Brunati, Perucca et al. 2015). Nevertheless, for most ZP module proteins, the dynamic of ZP filament formation and how the precursor molecule dissociation stimulates oligomerization is still ambiguous.

Compared with other ZP proteins, UMOD is naturally abundant in the urine and easier to purify. Consequently, UMOD was studied through electron microscopy previously (Tamm and Horsfall 1950, Serafini-Cessi, Malagolini et al. 2003). In the identification of UMOD, it is first characterized as a principle element of hyaline casts (Rovida 1873), and then recognized as a viral hemagglutination inhibitor (Tamm and Horsfall 1950, Serafini-Cessi, Malagolini et al. 2003). UMOD is primarily expressed by the thick ascending limb of Henle's loop cells. Secreted precursor UMOD is highly glycosylated and GPI-anchored (glycosylphosphatidylinositol), which also owns an intramolecular disulfide bond network. Separation of UMOD from CTP relies on the digestion of CCS, which is mediated by hepsin protease. This dissociation then activate assembly of UMOD filament (Schaeffer, Santambrogio et al. 2009, Brunati, Perucca et al. 2015).

1.6.3 Function of Uromodulin

As mentioned above, in the kidney, the thick ascending limb of the Henle's loop specifically expresses UMOD. During expression, once UMOD comes to the apical membrane, it is released into the tubular lumen. Mature UMOD is able to form filament, and single filament then composes a more complicated 3D matrix, which shows a gel-like property. It is reported that native state UMOD matrix participates in uropathogenic defense through binding to UPEC (uropathogenic E. coli), (Rampoldi, Scolari et al. 2011). Recent Cryo-ET, which combines tomography and electronic microscopy technology, also shows UMOD filament binds to bacterial pilus. However, lacking a high-resolution map, the detail of the UMOD-pilus contact information is not enough. Moreover, the Cryo-ET filament model presented in that study is not accurate because of the resolution limitation (Weiss, Stanisich et al. 2020). In addition, this mature and native state UMOD matrix mediates urine concentration and salt transport. Its gel property achieves water impermeability of the thick ascending limb of Henle’s loop.

Moreover, it is also reported that UMOD gene mutation may lead to chronic kidney disease (CKD) (Devuyst, Olinger et al. 2017). However, although common variants of UMOD are highly related to the risk of CKD, increasing concentration of UMOD monomer circulating in serum and mediating global and renal oxidative stress are associated with lower risk for cardiovascular disease and mortality in seniors (LaFavers, Macedo et al. 2019, Steubl, Buzkova et al. 2020). Furthermore, UMOD filament is also able to reduce nephrolithiasis and play a role in kidney innate immunity (Serafini-Cessi, Malagolini et al. 2003, Devuyst, Olinger et al. 2017, Weiss, Stanisich et al. 2020). The previous study shows that UMOD gene deficiency does not lead to

severe developmental defects. However, UMOD null mice suffer from urinary stones and nephrocalcinosis and are more fragile for microbial infections (Mo, Zhu et al. 2004, Weichhart, Haidinger et al. 2008). These diverse and critical biological functions of UMOD highlight the importance of revealing dynamic and structural detail of UMOD filament, which also helps to elucidate the general assemble principle of ZP domain filament.

Chapter 2. Materials and Methods

2.1 Molecular Cloning and Mutagenesis

2.1.1 Recombinant protein plasmids

For recombinant CARD proteins, NLRP1 constructs (985-1473, 1213–1473, 1374-1466, 1374–1473) with or without C-terminal SNAP tag (NEB), CARD8-CARD (451–537) with or without C-terminal SNAP tag, ASC-CARD (1–102) with N-terminal SNAP tag, CASP1-CARD (1–92) with or without C-terminal SNAP tag, NLRC4-CARD (1–102) with or without C-terminal SNAP tag. were cloned into pET47 vector in between XmaI and XhoI restriction sites. A 3 C protease cleavage site was inserted in between the CARD domain and the SNAP tags in case SNAP needs to be removed for biochemical analysis. Site-directed mutagenesis was performed using the KAPA Hifi PCR kits (KAPA Biosystems).

2.1.2 Site directed plasmid mutagenesis

For all the mutagenesis work in this thesis, KAPA Hifi PCR KITS (KAPA Biosystems) was used to set up PCR reaction with primers from Integrated DNA technologies (IDT) and targeted template plasmids. The 13 µl reaction volume contains ~50 ng template plasmid, 0.5 µl of forward and reveres primers at 10 µM, 3 µl KAPA KITS reagents and topped up with water. The template double DNA strands were separated by heat to 95ºC for 3 to 5 minutes, following with 20-30 amplification cycles. The cycle includes 30 seconds denaturation at 95 ºC, 45 seconds annealing at Tm temperature ± 5 ºC and 5 minutes extension at 72 ºC. The final extension was at 72 ºC for 10 minutes. After PCR, 1 µl Dpn1 enzyme (New England biolab) was added to 5 µl reaction sample

with 10µl water and incubated at 37 ºC for two hours. The reaction was transformed into transformation competent E. coli - DH5α. The transformation reaction was recovered by adding 0.5ml LB and incubated at 37 ºC for 1 hour. The sample was plated on LB agar plate supplemented with respective antibiotic and incubated at 37 ºC overnight. The positive colonies were picked next day and grown in 3 ml of LB with respective antibiotic. Plasmids were extracted the following day and mutagenesis was confirmed by sequencing.

2.2 Protein Expression and Purification

For the expression of the His-tagged recombinant proteins, the constructed plasmids were transformed into BL21(DE3) (NEB) and grown at 37 °C for 16–18 h. Then the starting culture was transferred to a larger volume of lysogeny broth, which was grown until an OD600 of ~0.6–0.8, then the temperature was lowered to 16 °C, and the cells were induced with 0.5 mM IPTG and grown overnight. Bacteria were harvested at 4600 × g (Avanti JXN series, Beckman Coulter) for 10 min and then resuspended with lysis buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole, and 10% glycerol). Cells were then lysed by Emulsiflex-C3 and centrifuged at a speed of 30,000 × g (Avanti JXN series, Beckman Coulter) for 20 min at 4 °C to separate supernatant and pellet. The supernatant was passed through a pre-equilibrated column containing Ni-NTA agarose beads (ThermoFisher Scientific) by gravity. The column was then washed with lysis buffer containing 20 mM imidazole to remove nonspecific binding proteins. His-tagged proteins were eluted with elution buffer that contained 300 mM imidazole. To further purify our recombinant proteins, size-exclusion chromatography (Superdex 200 increase, GE Healthcare) was performed in

buffer A (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 1 mM DTT). At this step, the purified proteins were the oligomeric pre-activated seed fraction. For monomeric CARD-SNAP or SNAP-CARD domain proteins, it was prepared following the protocol as described here. Concentrated and purified soluble filament seed at around 10 mg/ml 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. In the absence of stimuli, refolded SNAP-ASC-CARD protein remained monomeric for 24 h, and CASP1-CARD-SNAP remained monomeric or lower oligomeric for about 12 h. Recombinant SNAP-NLRP1-FIIND^{UPA}-CARD and SNAP-CARD8-FIIND^{UPA}-CARD were expressed in similar ways as the CARD domain constructs. Notably, most of the protein exist in cellular pellets after the Emulsiflex step, thus, direct pellet wash with lysis buffer with an additional 1% Triton-X100, followed by 6 M Guanidinium is necessary to obtain solution proteins. Chemical refolding following the same dialysis protocol, followed by 3 °C cleavage of the SNAP tag, was used to obtain homogenous oligomeric population of FIIND^{UPA}-CARD complexes at various concentrations.

2.3 Oligomerization Assay

SNAP-ASC-CARD and CASP1-CARD-SNAP monomers that were labeled with Alexa Fluor 647 (New England Biolabs) were co-incubated with CARD domain only filament seeds (CARD8-CARD without SNAP, NLRP1-CARD without SNAP, NLRC4-CARD without SNAP) separately. One microliter of 100 μM BG-Alexa dye stocks (dissolved in water containing 5% dimethyl sulfoxide) were added into every 100 μl of pre-labeling

protein stock at a concentration of 20 μM. Fluorescently labeled monomeric SNAP-ASC-CARD (10 μM) and CASP1-CARD-SNAP (10 μM) was mixed with 20 or 5 μM seed oligomer for 15 min at room temperature before analysis by Bis-Tris native PAGE gel (Life) or 15% sodium dodecyl sulfate polyacrylamide gel electrophoresis gel. The fluorescently labeled gel was run at a voltage of 200 V for 60 min. Once the EMSA was completed, they were visualized by Typhoon FLA7000 (GE Healthcare) with 100-μm pixel size at channel 647 using PTM 600. To form ASC-CARD and NLRC4-CARD long filament (200–500 nm), ~20 μM monomeric SNAP-tagged protein was mixed with unlabeled, wild-type SNAP-tagged ASC-CARD and NLRC4-CARD seeds at a molar ratio of 20:1 for overnight. Later on, in-house-prepared MBP-3C-protease was added to the fully extended filament with a final concentration of 0.5 μM. The digestion was performed at 4 °C for 8 h, before further size-exclusion purification to separate the core filament, the cleaved-off SNAP tags, and proteases. Final extended ASC-CARD and NLRC4-CARD thin filament were then purified using size-exclusion column Superose 6 (GE Healthcare), before structural analysis. To form long NLRP1-CARD and CARD8-CARD filaments, N-terminal His-tagged CARD domain alone constructs were expressed and purified according to the protocol mentioned above, then concentrated to ~400 μM, homogeneous long filaments formed spontaneously, without the requirement of seeding.

2.4 Negative Stain Electron Microscopy

The protein samples are diluted to suitable concentration (~0.02mg/ml for single particle, ~0.1mg/ml for oligomeric complexes) and absorbed onto a carbon-coated

formvar film which is attached to a metal specimen grid (Electron Microscopy Sciences). The surface of the carbon grid is hydrophobic and is difficult to bind with charged samples. So, the grids glow discharged before applying samples onto the grid surface, in this way, electrons are hit the carbon to make the carbon surface hydrophilic in the environment of vacuum chamber of a vacuum evaporator. After about 60 seconds, extra samples are blotted off by filter paper and immediately the grid is covered with 4 µl staining solution (2% Uranyl acetate) for another 60 seconds. Extra staining buffer are removed by filter paper. The prepared grids are air-dry for 10 mins and stored in a grid box and labelled properly. These grids were imaged on the Tecnai 12 electron microscope and recorded with an Eagle CCD camera operating at 120kV (Nanyang Technological University Microscopy Facility).

2.5 Data Collection with Cryo-EM

For cryo-EM grid preparation, Quantifoil R1.2/1.3 holey grid (Quantifoil) was glow-discharged for 60 s and plunge-frozen using a Vitrobot (FEI) to be treated with 5-µl sample at a concentration of ~2 mg/ml. Final high-resolution images were collected at Titan Krios, 300 kV, using Gatan K2 camera (Gatan), super-resolution mode for data collection, and binned 2× when doing motion correction. The sub-frame time is 250 ms and total exposure time is 10 s, this will give 40 frames per stack. The pixel size on the final image is 1.10 Å. The dose rate is 4 e/pixel/s.

2.6 Structure Determination and Model building

For cryo-EM data, super-resolution image stacks were “gain-normalized” and binned by 2× to a suitable pixel size prior to drift and local movement correction using MotionCorr2. Date processing software Relion 3.0 was used for CTF determination, particle picking, 2D classification, 3D classification, and refinement procedures.

Unpolished maps were used for model building. Firstly, the monomer model is fetched from the PDB database (Protein Data Bank) or protein structure prediction server, which is used as a start point for complex model building. Firstly, monomers are fit into the map rigidly, followed by manual editing. Phenix real-space refinement with default NCS torsion-rotational constraint was performed to this preliminary model. After several rounds of rigid body refinement, a model with reasonable statistics was obtained. Then in the final refinement. Specific issues, like high clash score and Ramachandran plot outliers after Phenix real-space refinement, were manually identified and rebuilt, and then Phenix real-space refinement was repeated until satisfactory statistics and density interpretation was achieved.

2.7 Random Forest and hierarchical clustering

Protein sequence data is fetched from Uniprot. Python machine learning library scikit-learn was used in this step. The pipeline of the model training includes feature extraction, model training, hyper-parameters grid search, cross-validation. In feature extraction, the amino acid sequence is encoded, and the feature vector is represented by contacting vectors of two proteins in one interaction pair. Output was the possibility

of interaction between two proteins. Rstudio generated the hierarchical clustering heat map with command *heatmap*.

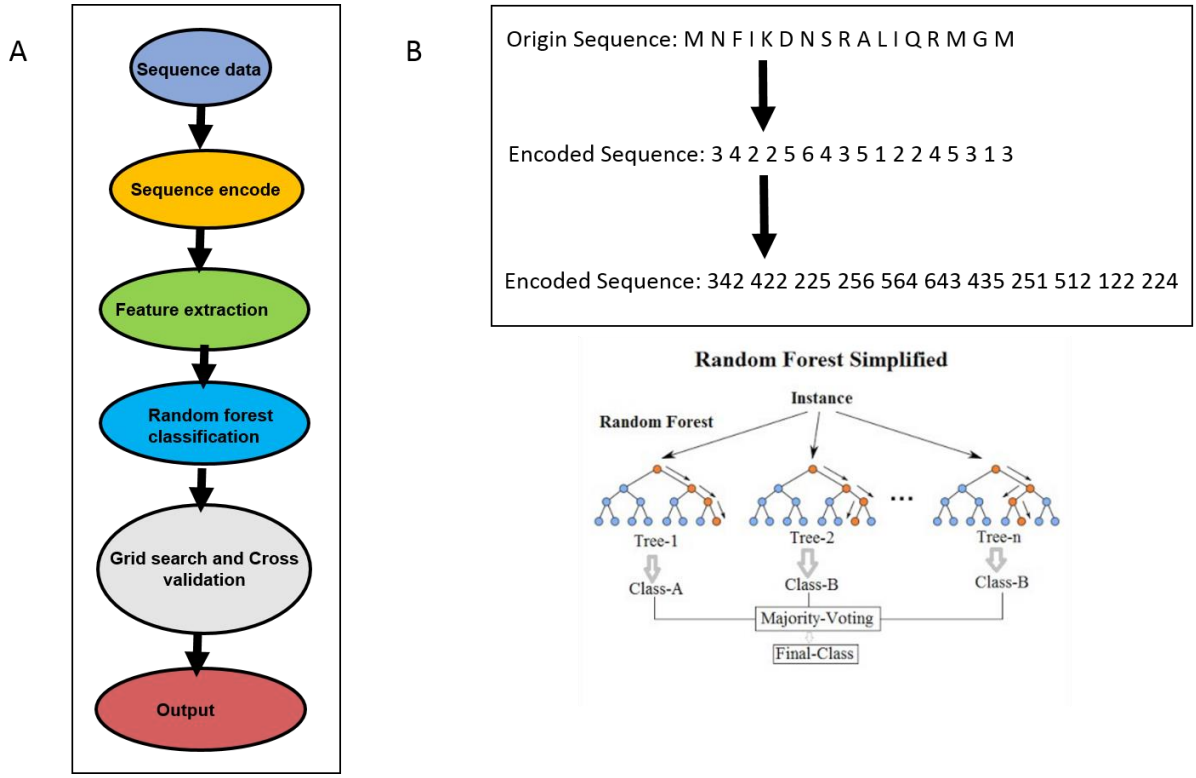

Figure 2.7. 1. Illustration of feature engineering and Random Forest. (A) Pipeline of our method. (B) Feature extraction demonstration

38 CARD-containing proteins used in this experiment shows in **Table 2.7.1**.

Table 2.7. 1 ***Indexing of CARD proteins***

01.Apoptotic_protease-activating_factor1(Apaf1)
02.Receptor-interacting serine/threonine-protein_kinase2(RIP2)
03.Nucleolar_protein3(NOL3)
04.B-cell_lymphoma/leukemia_10
05.Human_Caspase_1

06.Human_Caspase_2
07.Human_Caspase_4
08.Human_Caspase_5
09.Human_Caspase_9
10.CARD18
11.Baculoviral_IAP_repeat-containing_protein_3(CIAP2)
12.Baculoviral_IAP_repeat-contain_protein_2(CIAP1)
13.CARD16
14.CARD17
15.Inactive_caspase_12
16.Disk_large_homolog 5
17.CARD10
18.CARD6
19.CARD14
20.CARD11
21.NLRP1_CARD
22.CARD9
23.NOD2_1st_CARD(CARD15)
24.NOD2_2nd_CARD(CARD15)
25.NLRC4
26.ASC_CARD
27.NOD1(CARD4)
28.CARD8

29.CASP2 and RIPK1 domain containing adaptor with death domain-1st
30.CASP2 and RIPK1 domain containing adaptor with death domain-2nd
31.Death domain-containing protein CARDD
32.PRb-interacting_protein_RbBP-36
33.Death_adaptor_molecule_RAIDD-2-1st_CARD
34.Death_adaptor_molecule_RQIDD-2-2nd_CARD
35.RIG-I-1st_CARD(1-100)
36.RIG-I-2nd_CARD(100-190)
37.MDA5(28-107)
38.MDA5(107-220)

2.8 Filament Auto-Picking with OpenCV

Our fully automatically picking software is composed of two parts, filament identification, and contamination mask ensemble. This method is programmed in python3 environment with advanced library OpenCV, scikit-image, NumPy, fil_find.

Filament picking pipeline includes image preprocess, pseudo-filament kernel matching, image binary thresholding, filament identification, contours fitting, and coordinates selection. Contamination mask is generated from a pre-trained neuron network and

used as an optional input for the image binary thresholding step. The default format of the output coordinate is Relion STAR format and is readily compatible with Relion.

The details of these algorithms will be discussed in Chapter 3.4, and the dataset that used to test this method was collected at NTU and Karolinska Institute.

Chapter 3. Results

3.1 Structure study of NLRC4-CARD filament

3.1.1 Molecular structure of NLRC4 filament signaling complex

NLRC4 comprises a functional N-terminal CARD domain, a NATCH domain located at central, and a C-terminal LRR domain. Here, the N-terminal CARD domain is involved in the signaling filament aggregate and extending. Since the central filament part is smeared in full-length NLRC4 inflammasome electron map, which is reported previously (Tenthorey, Haloupek et al. 2017). An NLRC4 CARD alone construct is designed. As homogeneity and rigidity are essential for high-resolution structure determination through Cryo-EM. The first step is to obtain straight and homologous filament. After iteratively amino acid sequence optimization, the NLRC4 CARD boundary was cautiously truncated to 1-102. Besides, a flag tag and a tag with 13 residues were inserted into the N terminal of NLRC4-CARD-SNAP fusion protein for better stability.

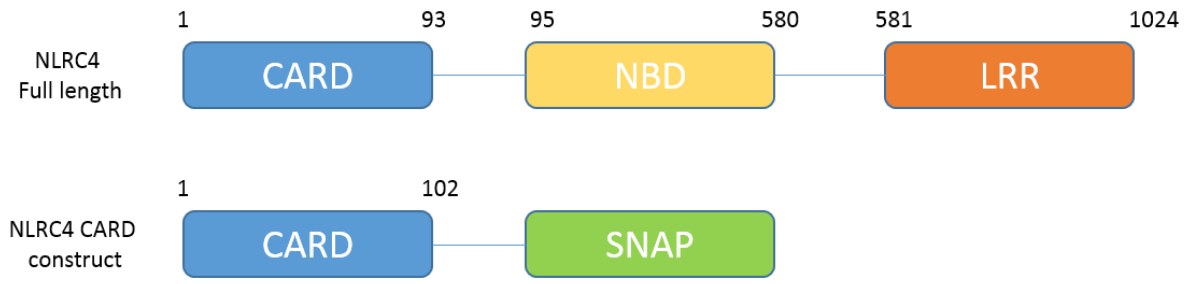

Figure 3.1. 1. NLRC4 architecture and NLRC4 CARD SNAP construct

SNAP-tag contributes to stabilize NLRC4 CARD filament and works as an indicator for oligomerization assay. However, the SNAP-tag prevents NLRC4 from forming longer filament and blocks detail of core filament (Figure 3.1.2A). After digestion and removal of SNAP-tag, the thin and straight filament was produced (Figure 3.1.2B). The same sample was used to prepare the Cryo-EM grid, and data was collected with FEI TITAN KRIOS electron microscope.

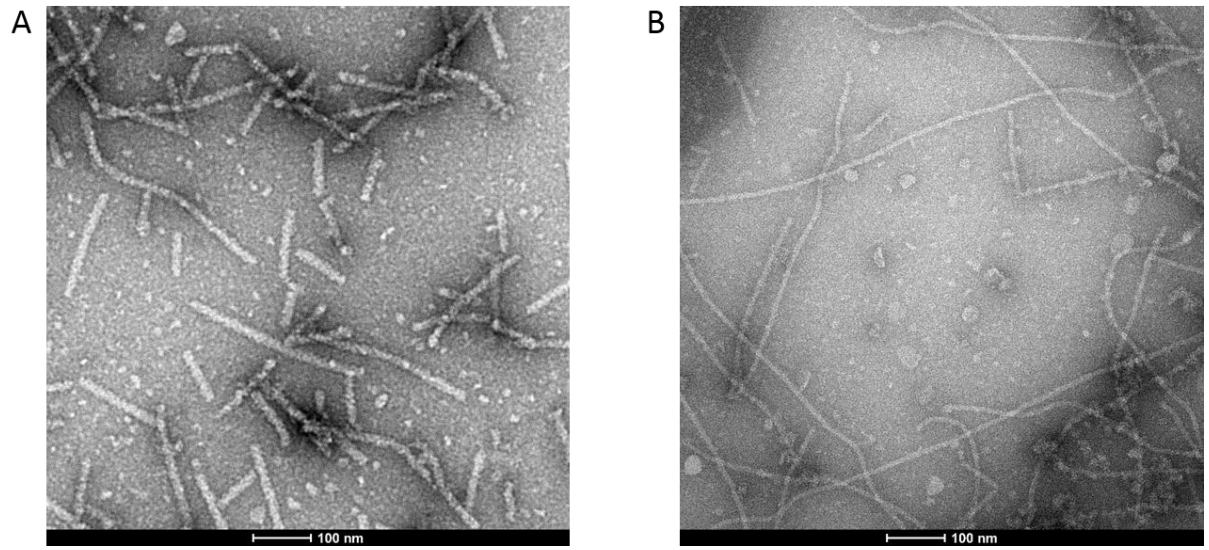

Figure 3.1. 2. (A) NLRC4-SNAP protein under negative staining. (B) Negative stain result of NLRC4 CARD without SNAP tag

Image processing and electron density reconstruction was processed in Relion3. Preprocess steps, including motion correction and CTF estimation, were finished firstly. Good images contrast and the high-frequency space region in the power spectrum show that high-resolution detail is captured during data collection (Figure 3.1.3B). Roughly 310k particles were manually picked from raw images and feed into the 2D classification step. After 2D classification, we get many excellent 2D averages of filament structure (Figure 3.1.3C). Helical layer line profile generated from 2D average

visualizes its diffractions patterns. Peaks in helical layer line profile resulted from the repeated feature in the 2D average, which gives an estimate for alpha-helix rotation angel and step size. In addition, it represents the best resolution level this class can reach (Figure 3.1.3D). Based on the quality of the 2D average class, 100k particles were selected for 3D classification.

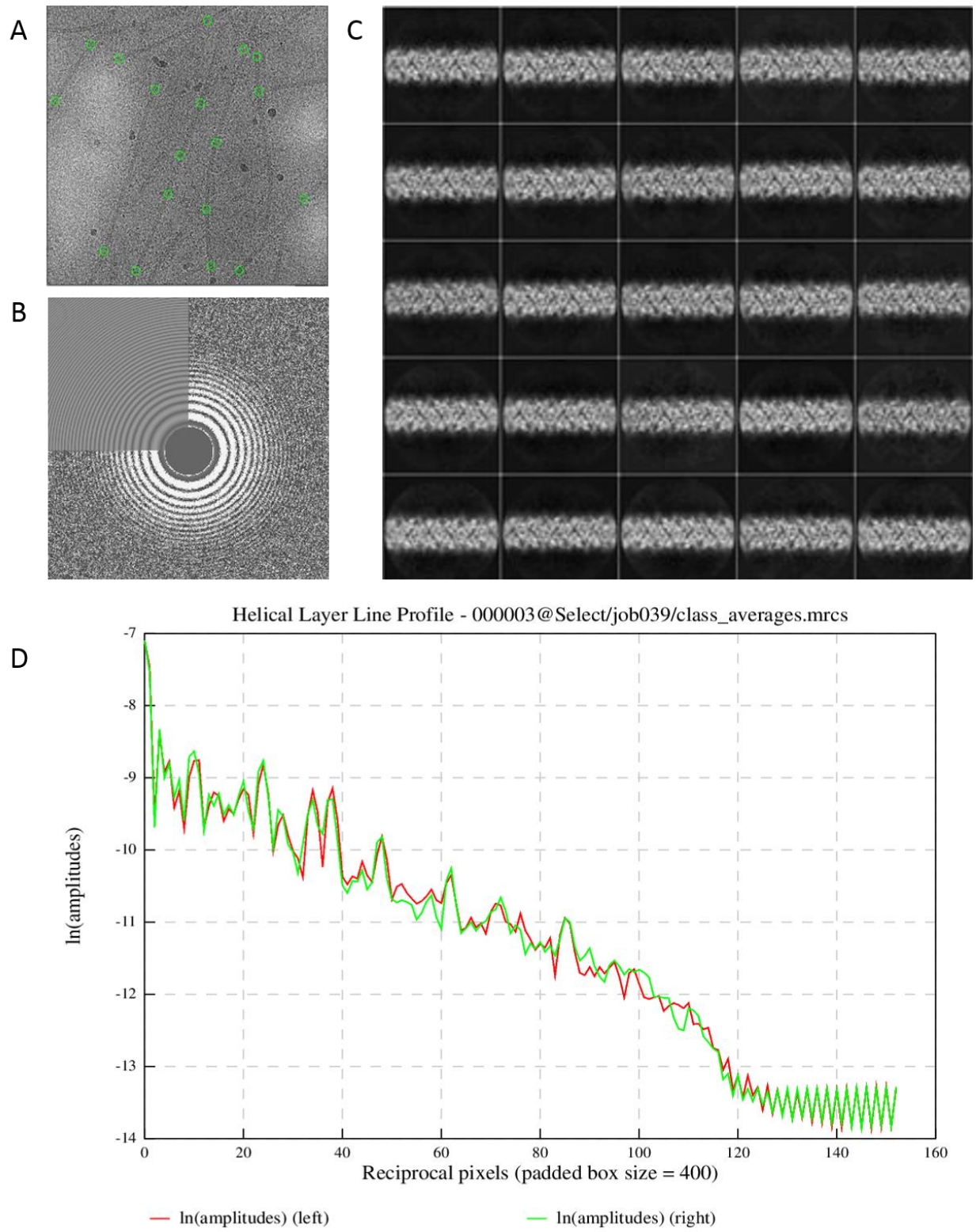

Figure 3.1. 3. (A) NLRC4 CARD filament under cryo-EM. (B) Power spectrum of Cryo-EM images. (C) 2D classification of average of NLRC4 CARD filament. (D) Helical layer Line profile for 2D class average images.

With 3D classification and 3D auto refinement, rotation angle and step size of NLRC4 CARD filament were determined with grid search (Figure 3.1.4B), for the final density was reconstructed with rotation angel –100.55 and step size 5.36. After masking and final polishing, the resolution came to 3.3 Å. FSC gold standard curve shows that cut-off happens at 1/resolution 0.25, corresponding to resolution 3.3 Å (Figure 3.1.4A). Although helical symmetry is applied in filament reconstruction, resolution at the central region of filament is usually higher, as the central part is applied higher weight during reconstruction. The local resolution map illustrates this fact (Figure 3.1.4C). The resolution is not evenly distributed across the whole density; however, the central part is 3.3 Å. As a result, the central part is used for protein coordinates building.

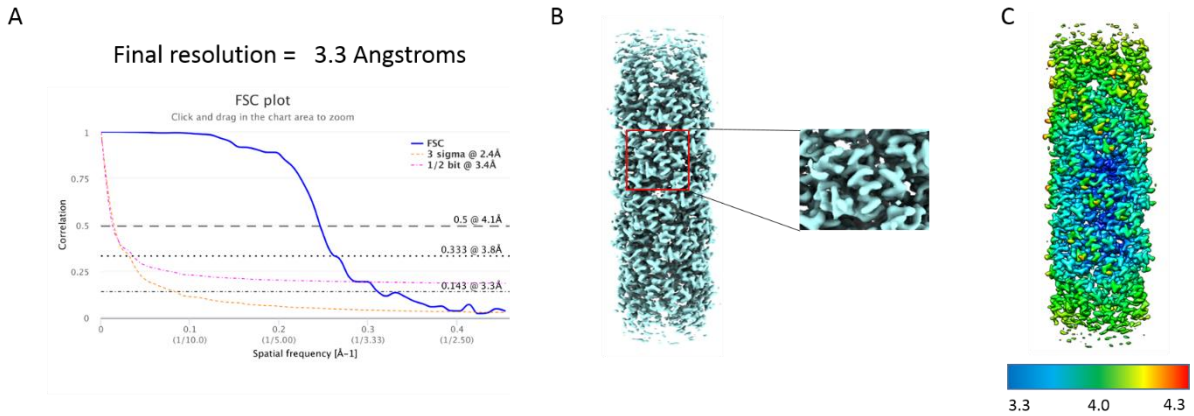

Figure 3.1. 4. (A) FSC gold standard curve. (B) NLRC4 CARD filament electron density model after polishing

As a previous study has revealed, activated NAIPs recruit NLRC4 to form signaling filament. In NLRC4 oligomerization, CARD is the functional domain that recruits other NLRCR4 and self-assembly. In Figure 3.1.5, the NLRC4 CARD core filament, colored blue, was inserted into NAIP2/NLRC4-CARDΔ oligomer (EMD-6458), colored silver.

Our CARD filament perfectly matches the hole which derives from CARD deletion. (Figure 3.1.5) This architecture is consistent with the previous NLRC4 activation hypnosis model.

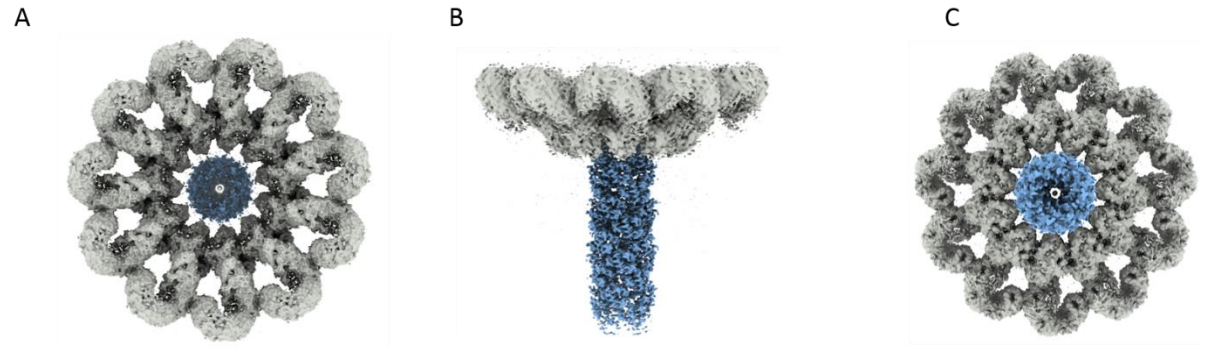

Figure 3.1. 5. NLRC4 CARD filament match NAIP2/NLRC4 oligomer, A top view, B side view, C bottom view

Based on this high-resolution density, we build NLRC4-CARD oligomer coordinates. Typical CARD-CARD interaction is composed of three interaction interfaces. The definition criteria are shown in Figure 3.1.6A. Each NLRC4 CARD monomer interacts with the surrounding six NLRC4 CARD subunits. We show the contact detail of these three interfaces (Figure 3.1.6B).

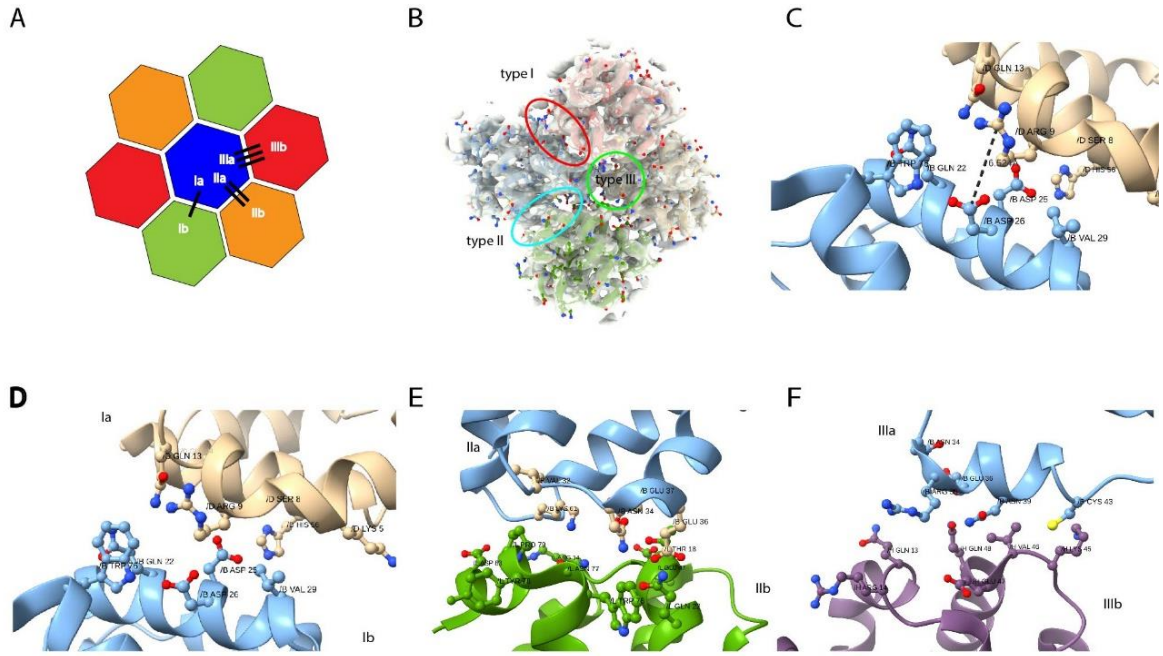

Figure 3.1. 6. CARD-CARD interaction interface, (A) Model of three kinds of CARD-CARD interaction, (B) overview of three interfaces and their position, (C) critical residue contact in the type-I interface, (D) zoomed-in view of the type-I interface, (E) zoomed-in view of the type-II interface, (F) zoomed-in view of the type-III interface

3.1.2 Structure analysis and biochemical assay

Biochemistry assay validates the model built from our high-resolution map; CARD-CARD interaction at interfaces is achieved by complementary charged residues. We design a charge reversal mutation construct based on the contact residues between interfaces of this high-resolution structure. R9D, D26R, and R9D/D26R NLRC4-CARD mutant proteins were constructed. As single mutation R9D leads to function loss, and single mutation D26R leads to partial function loss. Double mutation R9DD26R

recovers the self-assembly ability of NLRC4 (Figure 3.1.7B). This result indicates the importance of the R9 and D26 contact points.

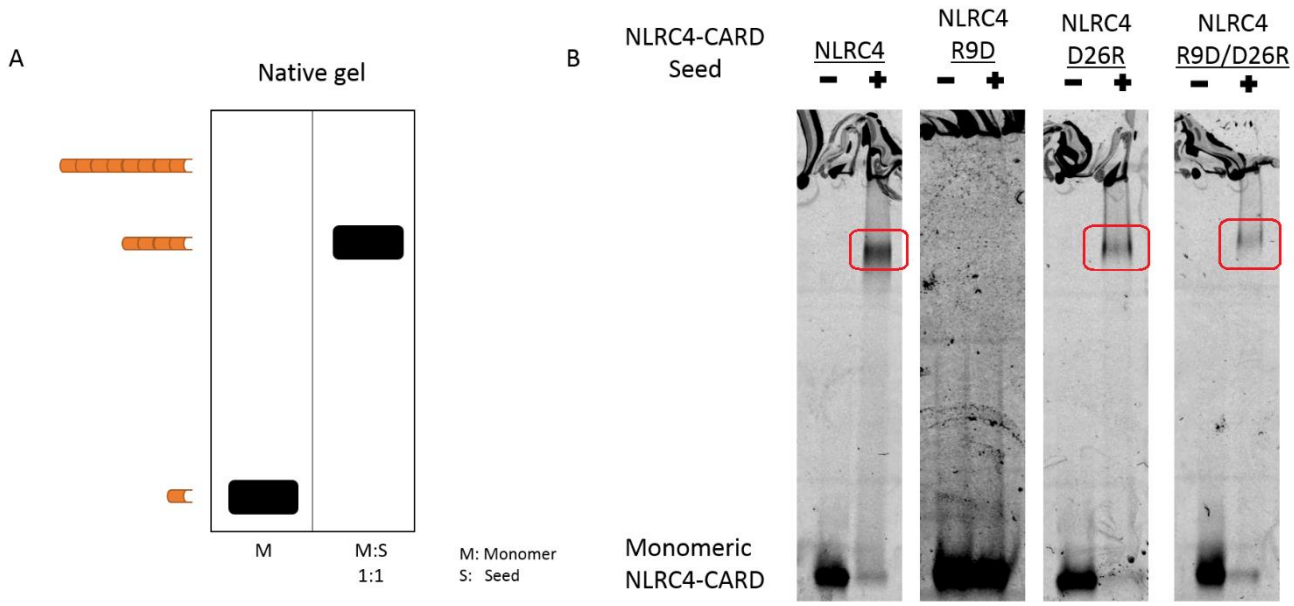

Figure 3.1. 7. (A) Seed-induced oligomerization demonstration. (B) NLRC4 charge reversal mutagenesis. It demonstrates that R9D mutation altered NLRC4-CARD oligomerization capacity, whereas double mutation R9D/D26R recovers it.

Electrostatic interaction is of vital importance for type II interaction. The charge distribution of the NLRC4 CARD monomer is polarized (Figure 3.1.8B), which is the basis for complementary charge interaction. Electrostatic potential distribution of NLRC4 CARD filament show below (Figure 3.1.8). The complementary electrostatic potential distribution between the top and bottom interface explains the strong internal cohesive force inside the NLRC4 CARD filament. NLRC4 CARD filament can trigger other CARD containing protein, which has similar bottom interface patterns through top-down direction activation.

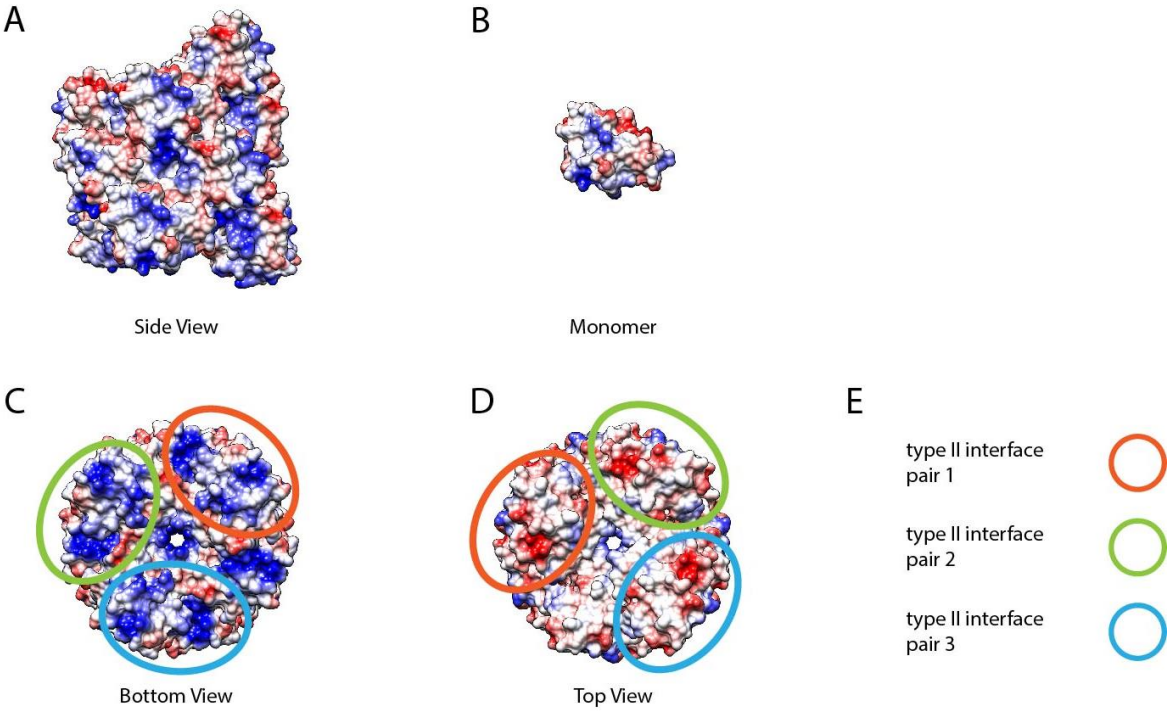

Figure 3.1. 8. (A) Side view of NLRC4 filament electrostatic potential distribution. (B) Side view of NLRC4 CARD monomer electrostatic potential distribution. (C) Bottom view of NLRC4 filament electrostatic potential distribution. (D) Top view of NLRC4 filament electrostatic potential distribution. (E) Indicator for Type II interfaces

3.1.3 Conclusion and Summary

As one of the canonical inflammation signaling pathways, NLRC4 is well studied in the past decades for both functionally and structural aspects. The high-resolution structure of native NLRC4-CARD filament concludes the structural study. Our NLRC4 filament structure demonstrates the featured Type-I II III interface among death domain proteins, thereby revealing the structural basis for NLRC4 filament extending and signaling transduction. Due to the N-terminal CARD domain, NLRC4 can stimulate

downstream effectors in diverse ways by activating either adaptor ASC or caspase-1 directly. Here Type II interface dominates this signal transduction progress. Critical residues located there and the Coulombic distribution on the exposed surface comprise the structural basis for this flexible activation capacity of NLRC4.

Table 3.1. 1. Data collection, image processing, and coordinates refinement statistic for NLRC4

Data Collection	
Microscope	Titan Krios
Voltage (kV)	300
Detector	K2
Pixel size (Å)	1.1
Defocus range (μm)	-1 to -2
Electron Dose (e-/A2)	40
Movie Frames	40
Helical Reconstruction	
Software	RELION 3.0.4
Particles	382715
Helical rise (Å)	5.17
Helical rotation (°)	-100.55
Resolution (Å)	3.3
Coordinate Refinement	
Software	Phenix
Resolution Cut Off	3
Rwork	0.2571
Rfree	0.3057
Model	
Oligomer	center 12mer
Number of residues	1020
B-factor overall	248.57
R.M.S. deviation	
Bond length (Å)	0.002
Bond angle (°)	0.364
Validation	
Molprobity clashscore	0.24
Rotamer Outliers (%)	0
Cβ deviation (%)	0
Ramachandran plot	
Fovored (%)	98.1
Allowed (%)	1.9
Outliers (%)	0

Deposit	
PDB code	6K8J
EMDB code	9946

3.2 NLRP1-CARD filament self-assemble mechanism study and structure characterization

3.2.1 The function of NLRP1 in inflammasome formation

The primary content of this part is the structure study of NLRP1 and CARD8 based inflammasome assembly through FIIND^{UPA} and CARD domain which is auto-proteolysis dependent. Our work shows, both FIIND^{UPA} and CARD participate in NLRP1 inflammasome assemble regulation. NLRP1 is able to form an active signaling filament through self-assemble in vitro, then trigger ASC oligomerization and downstream immune responses. As for FIIND^{UPA}, although it does not participate in ASC recruitment directly, existing of FIIND^{UPA} promotes NLRP1-CARD oligomerization and filament assemble. NLRP1-FIIND^{UPA} alone construct tends to form ring shape structure; however, NLRP1-FIIND^{UPA}-CARD recombinant protein forms filamentous complex, composed of inner CARD filament and outer FIIND^{UPA} ring. Structural analysis shows that, a linker of 9-22 residues between FIIND^{UPA} and CARD results in helicity difference of the inner filament and outer ring. We solved the structure of NLRP1-CARD and CARD8-CARD filament with Cryo-EM, which reach a resolution of 3.7 Å. Surface electrostatic and hydrophobicity analysis identified the critical structure motifs that contribute to heterotypic CARD-CARD interactions, including the CARD domain of NLRP1, ASC, CARD8, caspase-1. These results give insights into

why NLRP1 inflammasome relies on ASC as signal adaptor, whereas CARD8 can activate caspase-1 directly, bypassing ASC. Our works reveal different activation modes between NLRP1-ASC and CARD8-caspase-1 from structural level, and suggest different electrostatic and hydrophobicity distributions of CARD associated with specific inflammasome, which is consistent with current and related reports (Hollingsworth, Sharif et al. 2021, Huang, Zhang et al. 2021).

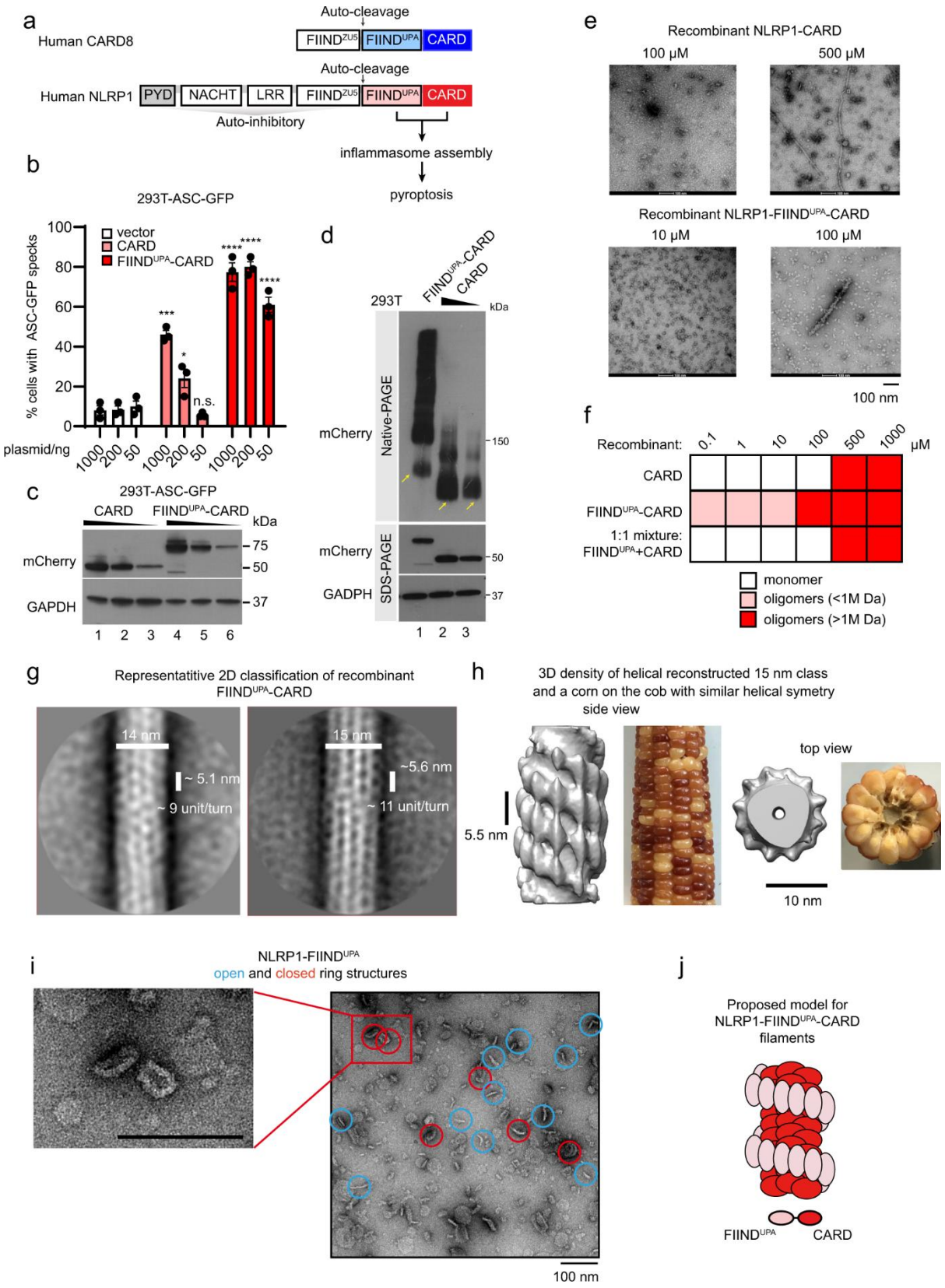

Figure 3.2. 1: Inflammasome complexes of human NLRP1. (a) domain architecture of NLRP1 and CARD8, functional domain is highlighted with red (CARD) and blue (ASC), respectively.

(b) HEK293T-ASC-GFP cells assay result, the y-axis shows the percentage of ASC-CFG specks derived from NLRP1-FIIND^{UPA}-CARD (labeled with mCherry, red) or NLRP1-CARD (labeled with mCherry, light red). The cultured cells are fixed 24 hours for fluorescence microscopy after transfection. Data demonstrate with mean ± standard-deviation format. One-way ANOVA is used to calculate the statistical significance values, with biological replicates n=3. '/**/****' here indicates P-value < 0.05, 0.01, and 0.0001, respectively. (c) Western blot results that correspond for NLRP1 mCherry-tagged construct are shown in Figure 3.2.1B. (d) Corresponding blue-native gel blot for NLRP1-FIIND^{UPA}-CARD (labeled with mCherry) and NLRP1-CARD (labeled with mCherry). Monomeric samples are pointed out with Yellow arrows. (e) Negative staining EM results for construct NLRP1-CARD and NLRP1-FIIND^{UPA}-CARD with different concentrations. (f) Oligomerization assay results from different NLRP1 based constructs at different concentrations. (g) The most representative 2D average images of negative staining EM, from which distinguishable patterns of NLRP1-FIIND^{UPA}-CARD filament appear. (h) 3D density side view and top view of NLRP1-FIIND^{UPA}-CARD filament with estimated resolution 25 Å. Helical segment used for reconstruction is selected from the 150A 2D class. The pattern appears to be similar to corn cobs. (i) Zoomed-in view of negative staining EM images for NLRP1-FIIND^{UPA} (tagged with SUMO). The size of the ring-like complex is around 20nm. (j) Illustration of the hypothesized model of NLRP1 FIIND^{UPA}-CARD filament.*

3.2.2 NLRP1-FIIND^{UPA} domain facilitates NLRP1-CARD oligomerization

In a previous study, it was demonstrated that NLRP1-FIIND^{UPA}-CARD and CARD8-FIIND^{UPA}-CARD, which shows the auto-cleavage property, can trigger pyroptosis in human cells which are cultured in vitro (Finger, Lich et al. 2012, Zhong, Mamai et al. 2016, Sandstrom, Mitchell et al. 2019) (Figure 3.2.1a). In order to identify the exact role of FIIND^{UPA} domain, FIIND^{UPA} is expressed in HEK239T cells, engineered to express GFP-tagged full-length ASC. In agreement with the previous report, NLEP1-CARD overexpressed in HEK239T cells is able to induce ASC-GFP specks. However, FIIND^{UPA} triggers more ASC-GFP specks compared CARD alone assay (Zhong, Mamai et al. 2016, Sandstrom, Mitchell et al. 2019). This difference becomes more significant at low protein expression level (Figure 3.2.1b, 50 ng; Figure 3.2.1c, lane 3 and 6). With a low expression level, although most of NLRP1-CARD (roughly 90%) exhibit monomeric form (Figure 3.2.1d, lane 2 and 3), there are still some functional NLRP1-CARD oligomers, which again shows sufficient activation capacity. FIIND^{UPA}-CARD, compared with CARD alone, exhibits more effective activation capacity. With similar expression level, FIIND^{UPA}-CARD forms more oligomers (Figure 3.2.1d, lane 1 vs. 2 and 3). These assay results demonstrate, although NLRP1-CARD has sufficient self-assemble and activation capacity for ASC, existing of FIIND^{UPA} significantly decreases yield concentration for oligomerization and promotes activation capacity.

Furthermore, FIIND^{UPA}-CARD, NLRP1-CARD, and FIIND^{UPA}, CARD mixture with a 1:1 molar ratio was observed under negative stain-electron microscopy (EM). Consistent with the result of native PAGE in 293 T cells, when expressed with CARD in a single polypeptide (Figure 3.2.1f), NLRP1-FIIND^{UPA} is able to decrease the concentration needed for CARD oligomerization by around 50-fold (Figure 3.2.1e, f). Under limited concentration (roughly below ~200 μM), NLRP1-CARD alone construct cannot form

rigid oligomers neither in isolation form nor incubated with FIIND^{UPA}. However, prolonged and rigid filament forms when concentration goes beyond 500 μM (Figure 3.2.1e, right, Figure 3.2.1f). These negative staining EM results indicate that, although NLRP1-CARD innately owns limited but sufficient capacity for self-oligomerization, FIIND^{UPA} promotes NLRP1-CARD self-recruitment capacity through some catalyst mechanism.

In order to verify this hypothesis, related NLRP1 fragments construct is built and purified, including CARD alone, FIIND^{UPA} alone, FIIND^{UPA}-CARD, and FIINDfl-CARD (full-length FIIND-CARD) (Figure 3.2.12a). Recombinant proteins oligomerization assay generated from both mammalian and bacterial expression shows, FIIND^{UPA}-CARD can form long but heterogenous filament with a diameter around 14 nm (Figure 3.2.12a, b, e). FIIND^{UPA} alone and FIINDfl-CARD are not able to form filament (Figure 3.2.12c, d). Consistent with HEK293 T cell assay, recombinant NLRP1-CARD alone construct is able to form filaments in vitro, which is similar to another CARD-based signaling filament in the innate immunity system. This NLRP1-CARD alone filament owns diameters around 8 nm (Figure 3.2.12b, e), which is remarkably thinner than the FIIND^{UPA}-CARD filament.

In order to reveal the structural mechanism of the phenomenon, high-resolution maps of NLRP1-FIIND^{UPA}-CARD and the homolog fragment of CARD8 were necessary. Construct optimization was performed to identify the role of linkers between FIIND^{UPA} and CARD and optimal domain boundary, facilitating structure determination. We evaluate the function of linkers through truncation analysis. These results show that eight amino acids are located at NLRP1-CARD (1466–1473), 11 amino acids located

at the linker region (1360–1370), and 11 residues located at NLRP1-FIIND^{UPA} (1213-1223) are not necessary for NLRP1-FIIND^{UPA}-CARD oligomerization and NLRP1-ASC activation. Construct with shorter linkers does not demonstrate better homogeneity of filament assembled in vitro. Consequently, NLRP1-FIIND^{UPA}-CARD (1213–1473) was used to further filament preparation and structural studies. 2D averages from negative stain EM images show different patterns, suggesting different helicity and diameters of the filament (Figure 3.2.1g). The 2D class with the most particles is chosen for 3D reconstruction, an estimated helical parameter of helical rise 5.6 nm and 11-12 subunit per turn is calculated during the iterative 3D classification task. This estimation is quite similar to NLRC4 inflammasome, which is also a disk-like complex with 11-12 subunits (Diebolder, Halff et al. 2015). NLRP1-FIIND^{UPA}-CARD filament contains a two-layer structure. By adjusting the contour level of the density map, the inner filament shows a diameter of 8 nm, which is similar to the NLRP1-CARD alone filament (Figure 3.2.1h). When expanding the density map to the proper contour level, the detail of the outer layer shows up with a diameter of 15nm. A corn comb-like pattern appears on the filament's surface, resulting from helicity (Figure 3.2.1h). These results indicate a two-layer architecture for the inflammasome sensor, which contains FIIND. This kind of two-step recruitment is similar to NLRC4 and NLRP6 inflammasome (Li, Fu et al. 2018, Shen, Lu et al. 2019). In the preliminary step, NBD (nucleotide-binding domain) performs self-assemble and then initializes CARD or PYD oligomerization. The size of the central hole, which is surrounded by the ‘corn cob’, matches the inner core filament. In addition, it is consistent with recombinant CARD alone. These results suggest that the filamentous FIIND^{UPA}-CARD oligomer comprises two parts: the rigid inner CARD filament and the surrounding spiral FIIND^{UPA} (Figure 3.2.1h).

In vitro, NLRP1-FIIND^{UPA} with SUMO tag construct form both ring-like and filamentous oligomers simultaneously, which again validates the two-layer architecture. The close form rings own a diameter of approximately 17 nm (Figure 3.2.1i, red circles). Because of the presence of the SUMO tag, the calibrated diameter is similar to the outer diameter of NLRP1-FIIND^{UPA}-CARD filament, which shows the corn cob pattern (Figure 3.2.1h–j). As CARD8 contains homologs FINND, CARD8 FIIND^{UPA}-CARD, and FIIND^{UPA} domains were used to process similar biochemical analysis. These results demonstrate that FIIND^{UPA} of CARD8 may perform similar functions as NLRP1. Although more heterogeneous, CARD8-FIIND^{UPA} recombinant protein still own the capacity to form the filamentous complex. In order to get a more rigid and homogenous filament of CARD8-FIIND^{UPA}-CARD, we optimize the sequence of the construct. With protein engineering, a linker of 14 amino acids between UPA and CARD is modified. This modification does not influence the capacity of CARD8 to trigger caspase1 CARD oligomerization. Both linker-modified and wild-type constructs are observed under negative staining EM. Image taken from these samples is used to perform 2D classification analysis. The 2D average shows a clear two-layer filament pattern. Particles from the best 2D classes are selected and feed into the next round helical symmetry free 3D classification. Although suffering from dramatic heterogeneity, the overall shape of CARD8-FIIND^{UPA}-CARD oligomer is quite similar to NLRP1-FIIND^{UPA}-CARD filament. When applying helical symmetry during 3D reconstruction, the density map becomes clearer. The top view of this CARD8-FIIND^{UPA}-CARD map supports the two-layer filament mode and is very similar to that of untagged CARD8-FIIND^{UPA}-CARD after refolding. Consequently, although CARD8-FIIND^{UPA}-CARD filament is more heterogeneous, it is able to form this featured two-layer architecture, which is very similar to that of NLRP1-FIIND^{UPA}-CARD. These results highlight the function of

FIIND^{UPA}, as it is necessary for NLRP1 and CARD8 CARD filament formation. FIIND^{UPA} owns a higher capacity of self-assemble at low concentration. This property aid preliminary NLRP1 and CARD8 oligomerization and initialize CARD oligomerization, which then transduces pyroptosis signal to downstream effectors.

3.2.3 Structure determination for NLRP1-CARD filament and CARD8-CARD filaments with Cryo-EM: the inner core column of FIIND^{UPA}-CARD filamentous complex

Although NLRP1 and CARD8 are homologous proteins, due to their featured C-terminal FIIND-CARD architecture, recent reports suggest that they may lead to a different signaling outcome. NLRP1 trend to induce ASC oligomerization, while CARD8 is able to activate caspase-1 directly, bypassing ASC (Ball, Taabazuing et al. 2020). However, the mechanism is still unclear. Our previous negative staining EM results demonstrate that NLRP1-CARD is able to perform self-assemble to some extend and transduce downstream signaling (Figure 3.2.1b). Therefore, we propose a hypothesis that this functional difference between NLRP1-CARD and CARD8-CARD is derived from the intrinsic structural difference. To validate this hypothesis, we prepare NLRP1-CARD and CARD8-CARD filament samples and solve the structure with Cryo-EM. The resolution of both samples comes to 3.7 Å measured by Gold standard FSC, without masking (Figure 3.2.2a–i). 2D classification and 3D classification results from RELION 3.0 indicate that NLRP1-CARD oligomers with multiple conformation are generated during self-assemble, as there are many heterogeneous 2D averages and 3D density maps (He and Scheres 2017). From multiple heterogeneous 3D classes, particles from the best class is selected for high-resolution 3D refinement, based on estimated resolution and density quality (Grant,

Rohou et al. 2018) (Figure 3.2.2b–e). NLRP1-CARD construct that removes the last eight amino acids (1466–1473) achieves better rigidity and homogeneity.

Subsequent biochemical assays with HEK293T-ASC-GFP cells also supports that this construct is active state and can trigger ASC oligomerization. Consistent with our above negative staining EM result, this high-resolution NLRP1-CARD filament density is able to fit into the core column of NLRP1-FIIND^{UPA}-CARD filament density—both the shape and diameter match perfectly (Figure 3.2.2e). As for CARD8-CARD, it also forms filamentous complex with a diameter around 8 nm, which is more rigid and homogenous than NLRP1-CARD filament (Figure 3.2.2f–h). After post-process and polishing, both filament density shows a clear secondary structure of α-helix and bulky side chains (Figure 3.2.2d, i).

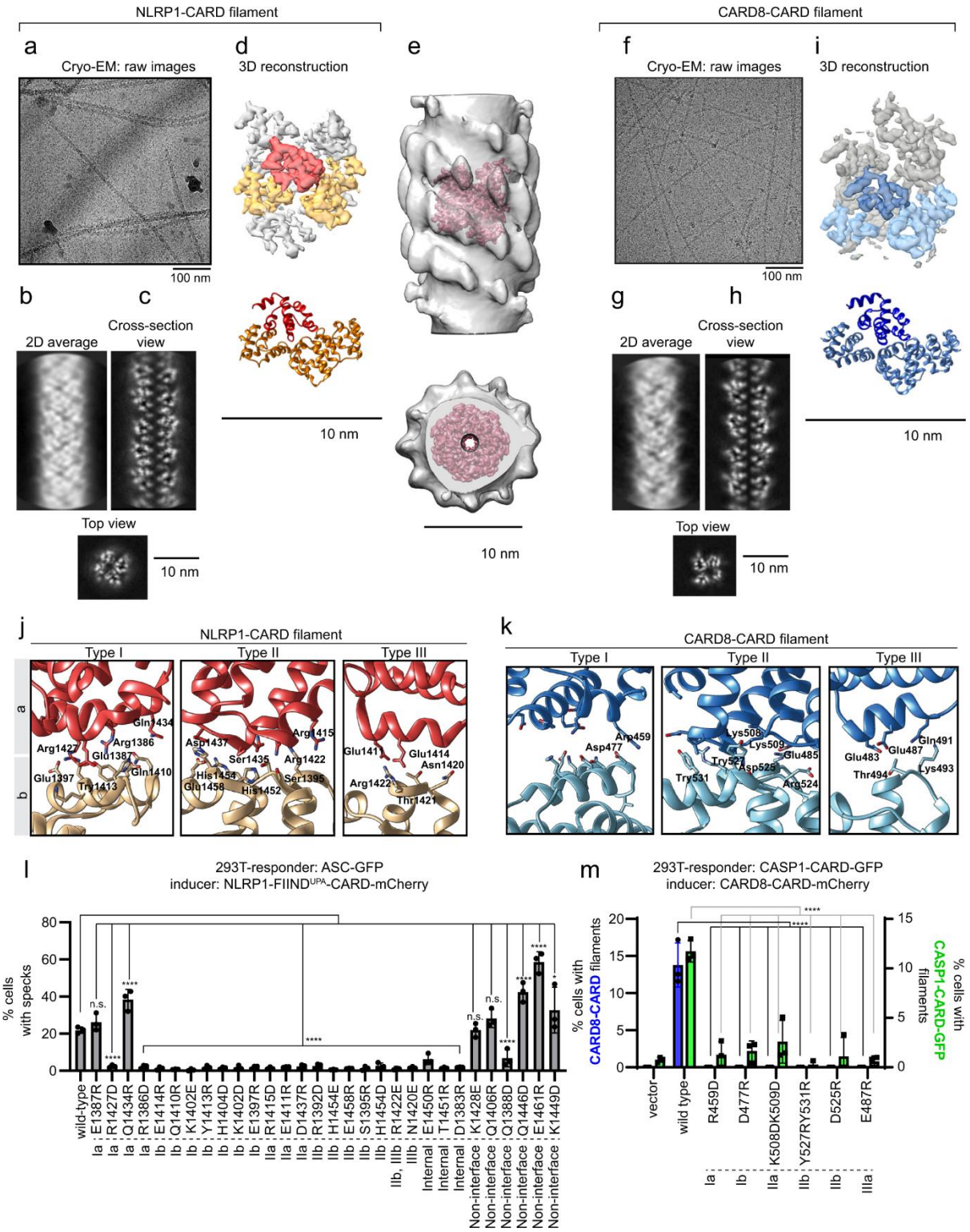

Figure 3.2. 2: High-resolution structure for NLRP1-CARD filament and CARD8-CARD filament. (A) Images of NLRP1-CARD filament under Cryo-EM (B) Representative 2D class average from NLRP1-CARD filamentous complex (C) Top view and side projection view from NLRP1 filamentous complex. (D) NLRP1-CARD filament electronic map from Croy-EM

reconstruction and molecular coordinates built from the map. NLRP1-CARD subunit labeled with different colors based on their roles, priory defined in Type I II III interactions. "a" monomer is labeled with light red and "b" monomer is labeled with orange. (E) Near atomic resolution, the electronic map of NLRP1-CARD perfectly fits into the inner core filament of the previously mentioned NLRP1-FIIND^{UPA}-CARD filament density. (F) Images of CARD8-CARD filament under Cryo-EM. (G) Representative 2D class average from CARD8-CARD filamentous complex. (H) Top view and side projection view from CARD8 filamentous complex. (I) CARD8-CARD filament electronic map from Croy-EM reconstruction and molecular coordinates built from the map. NLRP1-CARD subunit labeled with different colors based on their roles, priory defined in Type I II III interactions. "a" monomer is labeled with blue and "b" monomer is labeled with light blue. (J) Priory defined canonical interaction in NLRP1-CARD filamentous complex, Type I II III interaction. (K) Priory defined canonical interaction in CARD8-CARD filamentous complex, Type I II III interaction. (L) Mutagenesis analysis of important residues located at NLRP1-CARD interfaces and their influence on ASC-GFP specks which is trigger by NLRP1-FIIND^{UPA}-CARD (labeled with mCherry) in hek293t cells. Data demonstrate with mean ± standard-deviation format. One-way ANOVA is used to calculate the statistical significance values, with biological replicates n=3. '/**/****' here indicates P-value < 0.05, 0.01, and 0.0001, respectively. (M) Mutagenesis analysis of important residues located at CARD8-CARD interfaces and their influence on CARD8-CARD filament self-assembly (blue) and Caspase-1-CARD-GFP polymerization (green) in hek293t cells. Data demonstrate with mean ± standard-deviation format. One-way ANOVA is used to calculate the statistical significance values, with biological replicates n=3. '*/**/****' here indicates P-value < 0.05, 0.01, and 0.0001, respectively.*

Helical parameters of these two filaments are determined during 3D refinement; NLRP1 owns twist angle −100.8 degree and helical rise 5.36A, CARD8 owns twist angle −99.16 degree and helical rise 5.409A (Figure 3.2.2d, i). This suggests that NLRP1 and CARD8 are assembling through a similar mechanism. Based on protein coordinates built from our high-resolution density map, the featured Type I–III interfaces in death domain interaction can be easily identified (Figure 3.2.2j–k) (Lin, Lo et al. 2010, Lu, Li et al. 2016, Wang, Qiao et al. 2017, Li, Fu et al. 2018). Type I–III interfaces were first defined to describe the MyDDosome complex and then used as a convenient naming system to describe interactions surfaces between death domains, such as the MAVS filament CARD-CARD interface. To define Type I–III interfaces, firstly, the CARD monomer is fixed in position; alpha-helix pointing downward and sixth alpha-helix pointing upward. Then the monomer on the top is named as ‘a’ monomer, and the other three surrounding monomers are named as ‘b’ monomers. Next, the order I, II or III is determined by the left to right orientation of the ‘b’ monomer. With this order, each ‘b’ monomer owns a unique interface with ‘a’ monomer, which are called Type I, Type II, and Type III, respectively. With the above definition, six distinct interfaces are identified. Here ‘a’ monomer owns Type Ia, Type IIa, and Type IIIa surfaces, and ‘b’ monomers own Type Ib, Type IIb, and Type IIIb surfaces. Because of the structural similarity among death domain proteins, nearly all interfaces involved in death domain filament formation can be classified with this definition (Figure 3.2.2d, i). Type II interfaces of NLRP1-CARD and CARD8-CARD filament may contribute the most to filament formation, as it shows greater observable and significant electrostatic interaction, compared to Type I and Type III interface (Figure 3.2.2j, k). This property differs from other death domain complexes. The effect is more significant on CARD8-

CARD, as the distance between Type Ia and Ib interfaces of CARD8 is quite large (Figure 3.2.2j, k).

In order to validate our structure, single point mutagenesis assay was performed on critical contact residue at type I–III interface, guided with our high-resolution model. Predicted loss-of-function mutations at the Type I–III interface indeed prevents filament assembly. In addition, in ASC-GFP or caspase-1-CARD-GFP assay, specks/ filaments significantly decrease because of these function-loss mutations (Figure 3.2.2l, m). However, mutations on non-interface regions do not affect. The above results again support the hypothesis that signaling filament formed by CARD is essential for NLRP1 and CARD8 related inflammasome activation. However, there are exceptions. Mutations E1387R and Q1434R, which are located at NLRP1 Type Ia interface, still partially reduce the amount of NLRP1-CARD filament formed in vitro, although they did not influence HEK293T ASC-GFP speck formation (Figure 3.2.2l). We suppose that although the NLRP1 Type Ia interface cannot eradicate NLRP1-CARD assembly ability, it is crucial for NLRP1-ASC activation.

3.2.4 FIIND^{UPA} arrangement in the outer layer of FIIND^{UPA}-CARD filamentous complex

In the next step, we scan functional mutants to identify the crucial structural motif contributing to NLRP1-FIIND^{UPA} self-assembly. Based on previously reported and homologous UPA structure (PDB 3G5B) (Wang, Wei et al. 2009), nine SUMO-NLRP1-FIIND^{UPA} mutants were generated; mutations were predicted to disturb critical sidechain residues. A native gel shows that P1278A P1279G (PP1278AG, colored in

red) mutations have the greatest influence on SUMO-NLRP1-FIIND^{UPA} oligomerization (Figure 3.2.3a); as compared to the other four mutants R1240D, RK1260DD, R12785D, and F1320A (colored in orange). The later four mutants subtly influence NLRP1-FIIND^{UPA} oligomerization. Consistent with our previous conclusion that UPA promote NLRP1 oligomerization, demonstrated with an ASC-GFP specks induction assay, all mutants weaken the ability of NLRP1-FIIND^{UPA}-CARD to trigger ASC oligomerization, among which PP1278AG has the most significant influence (Fig, 3b, c).

Based on the above results, we can propose a large hydrophobic surface patch located around amino acids 1278-1279 (red), which promote self-assembly in a “side-by-side” arrangement (Figure 3.2.3d). This results in the two-layer architecture of FIIND^{UPA}-CARD filaments (Figure 3.2.1j). Very recent reported NLRP1-DPP9 complexes, which contain the FIIND^{UPA}-dimer, also show this organization (Hollingsworth, Sharif et al. 2021, Huang, Zhang et al. 2021). CARD8 also demonstrates similar architecture (Figure 3.2.3g).

This model also gives an explanation for the necessity of linker optimization between UPA and CARD to get stable NLRP1 and CARD8 filament. Based on the distance measured from the density, the optimal UPA-CARD linker should have a length of ~25~40 Å (Figure 3.2.3f). Linker shorter than this range would lead to spatial constraint and negatively influence filament extension. However, if the linker is too long, it may result in extra flexibility of the outer layer FIIND^{UPA}, leading to more heterogeneous conformation. Multiple sequence alignment results demonstrate that a linker of 9–20

amino acid should exist for NLRP1 homologs of both human and rodent, which suggest the architecture of FIIND-linker-CARD is conserved.

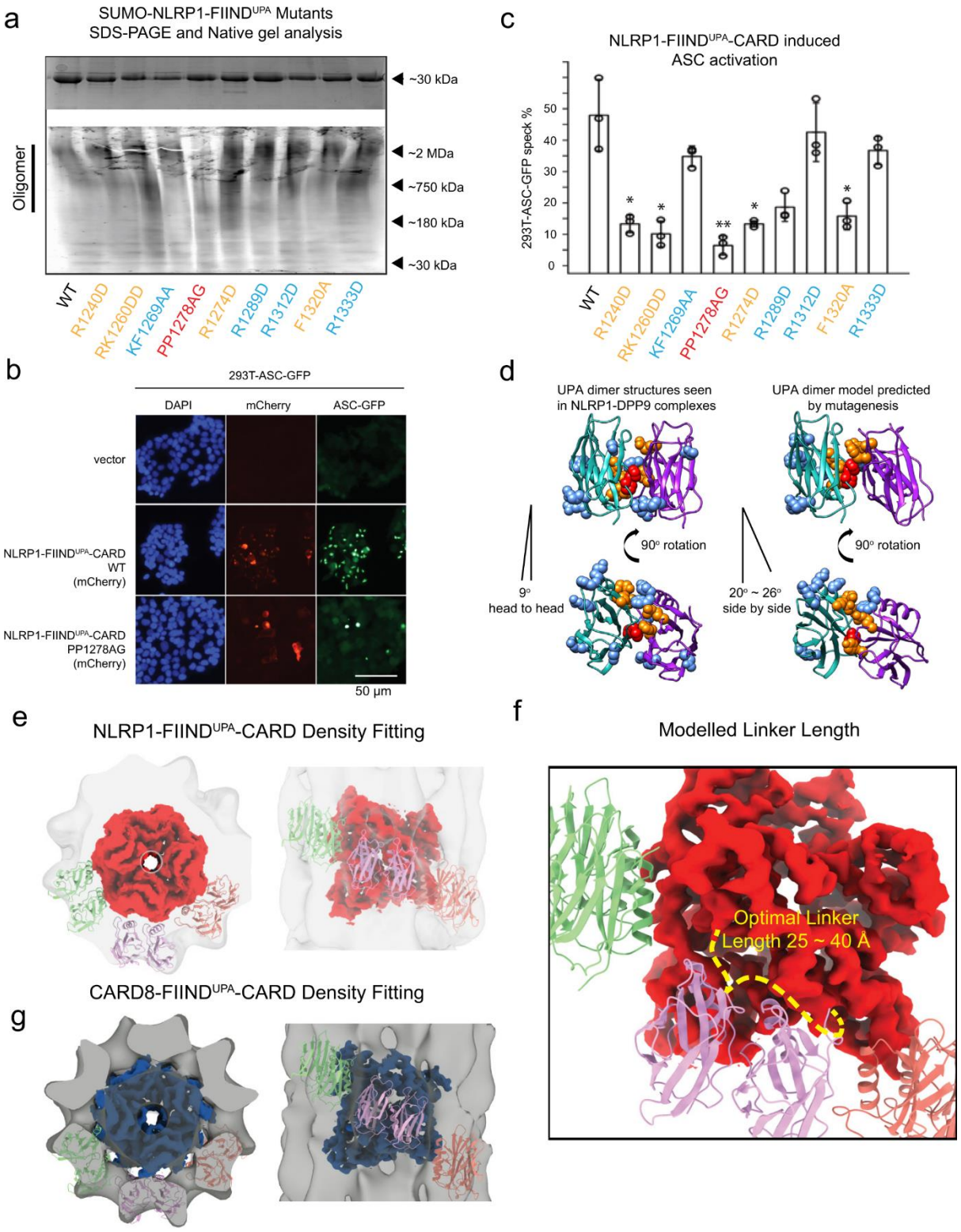

Figure 3.2. 3: Illustration of the two-layer architecture of NLRP1- FIIND^{UPA}-CARD and CARD8 FIIND^{UPA}-CARD filamentous complex. (A) Top part is the SDS page results for NLRP1-FIIND^{UPA}-CARD mutants constructs. And bottom part is the native gel page results for SUMO-NLRP1-FIIND^{UPA} recombinant protein, which is stained through Krypton. (B) 293T-ASC-GFP cells assay for ASC-GFP induction through NLRP1-FIIND^{UPA}-CARD wild-type construct and PP1278AG mutants. (C) HEK293T-ASC-GFP cells assay result, the y-axis shows the percentage of ASC-CFG specks resulting from different NLRP1-FIIND^{UPA}-CARD constructs. The cultured cells are fixed 24 hours for fluorescence microscopy after transfection. Data demonstrate with mean ± standard-deviation format. One-way ANOVA is used to calculate the statistical significance values, with biological replicates n=3. '/**/****' here indicates P-value < 0.05, 0.01, and 0.0001, respectively. (D) Dimer structure of FIIND^{UPA} solved by other groups and modeled polymer interface for NLRP1-FIIIND^{UPA}, which are demonstrated through side view and top vies. Important residues that may play a critical role in homotypic polymerization were labeled as spheres. Here, resides R1240, RK1260, R12785, and F1320 are labeled with orange color, and PP1278-1279 are highlighted with red color. (E) Projection of NLRP1-CARD filamentous complex from top and side, and three predicted FIIND^{UPA} dimer, which fits into the negative stain EM filament density of NLRP1-FIIND^{UPA}-CARD. (F) A zoom-in view of the modeled conformation of the UPA-CARD linker. When the linker owns a length of 25-50 Å, the two-layer architecture will be stable. (G) Projection of CARD8-CARD filamentous complex from top and side, and three predicted FIIND^{UPA} dimer, which fits into the negative stain EM filament density of CARD8-FIIND^{UPA}-CARD.*

3.2.5 Polymers of NLRP1-CARD and CARD8-CARD owns different Coulombic distribution surface, however, share similar helical parameters

Although owning almost the same helical parameters, electrostatic potential distribution on the surface of NLRP1-CARD and CARD8-CARD is significantly different (Figure 3.2.4a–f), which cannot be observed from their CARD alone structure (Jin, Curry et al. 2013, Jin, Huang et al. 2013). NLRP1-CARD filaments have a more balanced charge distribution on the axial surface, similar to other published death domain complexes (Figure 3.2.4a) (Lin, Lo et al. 2010, Lu, Magupalli et al. 2014, Wu, Peisley et al. 2014). Compared with NLRP1, the positively charged region of the CARD8-CARD monomer is hidden inside the filamentous complex, which results in a negative-charge exposed surface (Figure 3.2.4b). As our high-resolution structure shown above, the Type I interface is critical for CARD filament assembly; however, it does not play a major role at the observed surface density of either NLRP1 or CARD8 filament. For both NLRP1 and CARD8 filament, the charge distribution of the exposed surface is mainly due to the strong Type II interface density. At the Type II interface of NLRP1-CARD filament, there are two negatively charged residues, Asp1437, and Ser1435, locate at Type IIa surface, and two positive charge residues, His1454, and His1452, locate at Type IIb surface. This pair provides strong electrostatic interaction (Figure 3.2.4c). As for CARD8, its Type IIb surface (Tyr531, Tyr527, and Asp525) exhibits partial negative charge, and positively charged residues (Lys508, Lys509, and Leu512) locate at Type IIa surface (Figure 3.2.4d).

This intrinsic structural difference results in the subtle difference of helical symmetry; roughly 2 degrees of rotation angle and 15 degrees of monomer position-rotation angle. This is hard to be interpreted from the amino acid of sequence and sequence alignment. One of the most exciting findings from surface charge distribution is that NLRP1-CARD filament top surface charge distribution is similar to that of ASC-CARD

filament top surface (ASC-CARD filament solved in-house, PDB-6K99, EMD-9947, Figure 3.2.4e, g, left). As for CARD8-CARD, it is more similar to caspase-1-CARD (Figure 3.2.4f, h, left). Consequently, ASC-CARD bottom surface is charge complementary to the top surface of NLRP1-CARD (compare Figure 3.2.4e, g, right); whereas the bottom surface of caspase-1-CARD preferentially interacts with CARD8-CARD top surface (compare Figure 3.2.4f, h, right). The affinity difference between NLRP1 and ASC, versus CARD8 and caspase-1, interfaces give a potential explanation for different activation modes of NLRP1 and CARD8, which is observed in our cellular assay above.

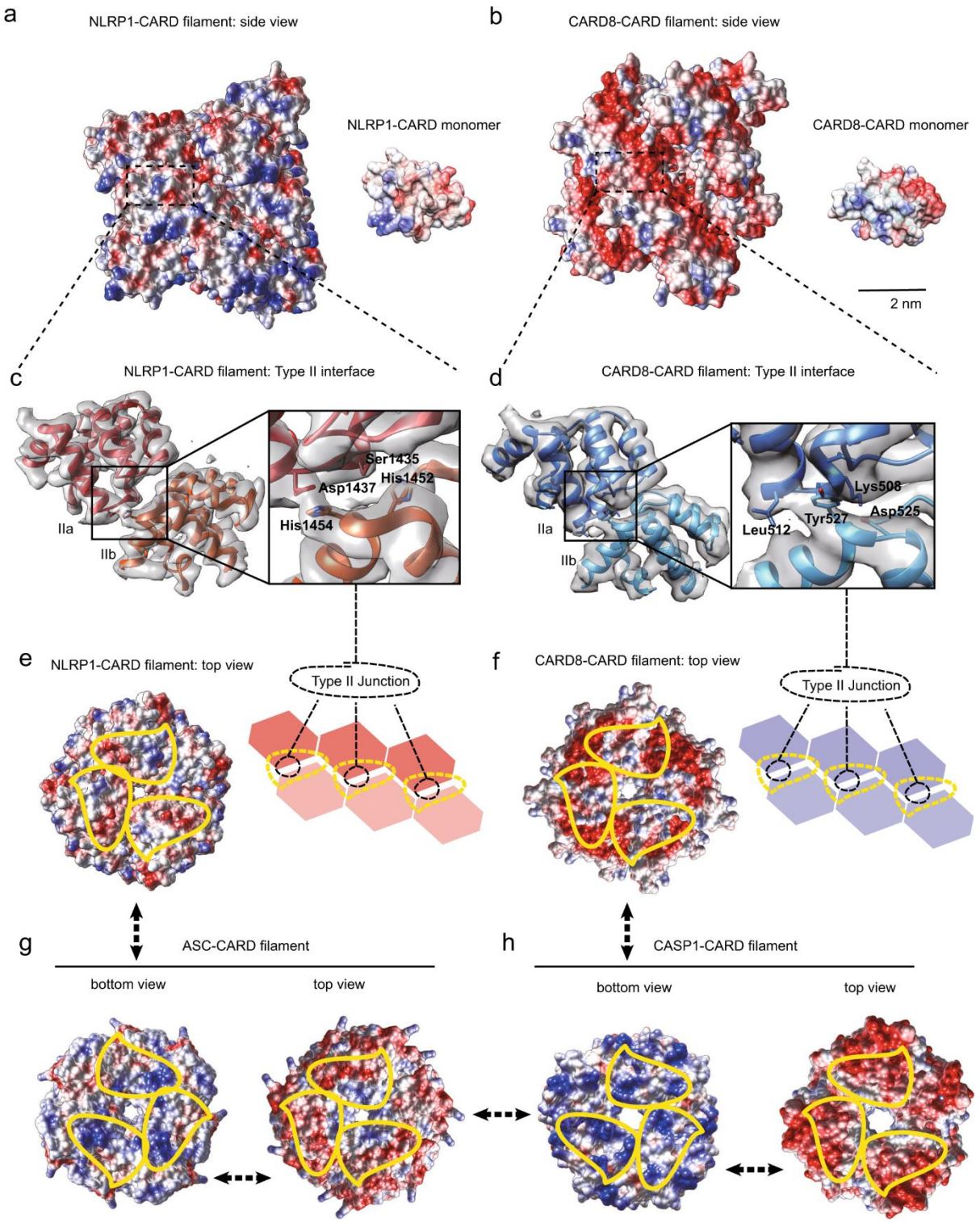

Figure 3.2. 4: the polymeric interface of CARD domains that involve in inflammasome formation. (A) The left part demonstrates side view of the coulombic surface of NLRP1-CARD filament segment, representing the electrostatic potential distribution generated through Chimera, color gradient from red (−10 C/m2) to blue (10 C/m2). The right part shows the

Coulombic surface of NLRP1-CARD monomer, placed in the same orientation with the left filament segment oligomer. (B) The left part demonstrates side view of the coulombic surface of CARD8-CARD filament segment, representing the electrostatic potential distribution generated through Chimera, color gradient from red (−10 C/m2) to blue (10 C/m2). Right part shows the coulombic surface of CARD-CARD monomer, which is placed in the same orientation as the left filament segment oligomer. (C) NLRP1-CARD dimer high-resolution model demonstrated with cartoon and the corresponding electronic density map, which is isolated from the filament model. Type II interfaced is used as entering to place the dimer. Four critical residues located at the Type II interface that are of vital importance is highlighted with "stick" mode in the demonstration. That includes Type IIb residues His1452 and His1454 and Type IIa residues Ser1435 and Asp1437. (D) CARD8-CARD dimer high-resolution model demonstrated with cartoon and the corresponding electronic density map, which is isolated from the filament model. Type II interfaced is used as enter to place the dimer. Four important residues located at Type II interface that are of vital importance is highlighted with "stick" mode in the demonstration. That includes Type IIb residues Asp525 and Tyr527 and Type IIa residues Lys508 and Leu512. (E) Coulombic surface of NLRP1-CARD filament top view, where Type IIb conjunction interface is highlighted with a yellow circle. (F) Coulombic surface of CARD8-CARD filament top view, where Type IIb conjunction interface is highlighted with a yellow circle. (G) Coulombic surface of ASC-CARD filament bottom and top view, where Type IIa conjunction interface (bottom) and Type IIb conjunction interface (top) are highlighted with a yellow circle. (H) Coulombic surface of caspae1-CARD filament bottom and top view, where Type IIa conjunction interface (bottom) and Type IIb conjunction interface (top) are highlighted with a yellow circle.

3.2.6 The subtle difference in the structure between NLRP1-CARD and CARD8-CARD oligomers leads to distinct downstream effector CARD

In order to identify a determinant region on the interface that results in different activation modes via ASC and caspase-1, we move attention to the Type II interface of the respective CARD filament. The difference of surface charge distribution for NLRP1-ASC and CARD8-caspase-1 interaction is mainly derived from this interface (Figure 3.2.2j, k; Figure 3.2.4a–f). Consequently, we propose a hypothesis that the Type II interface not only contributes to NLRP1 and CARD8 self-assembly but also plays a critical role in signal transduction from “activator” CARD (NLRP1 and CARD8) to “receiver” CARD (ASC and caspase-1.) To validate this hypothesis, firstly, we compare the Type II interface based on previously published ASC-CARD and caspase-1-CARD filament structure. A notable residue pair exists on the Type II interface of ASC and caspase-1 filament, which is similar to that of NLRP1 and CARD8 CARD filament (compare Figure 3.2.5a, b to Figure 3.2.4c, d). This was analyzed in multiple previous publications (Wu, Peisley et al. 2013, Peisley, Wu et al. 2014, Cheng, Hong et al. 2016, Lu, Li et al. 2016, Li, Zhou et al. 2017, Gong, Long et al. 2018, Li, Fu et al. 2018, Matyszewski, Zheng et al. 2018, Holliday, Witt et al. 2019). The critical residue contact pair in ASC-CARD filaments comprises Type IIb surface residues; Gln185 and Tyr187, and Type IIa surface residues; Ala168 and Asn170 (Figure 3.2.5a). As for caspase-1 CARD filament, this interface pair is achieved by Type IIb residues; Asp80 and Tyr82, and Type IIa surface residues; Lys64 and Gln67 (Figure 3.2.5b). This critical “bridge-like” contact pair becomes more apparent when the existing caspase-1 filament model is further refined. In conclusion, it seems that the “bridge-like” contact pair is a common feature in CARD-based inflammasome complexes.

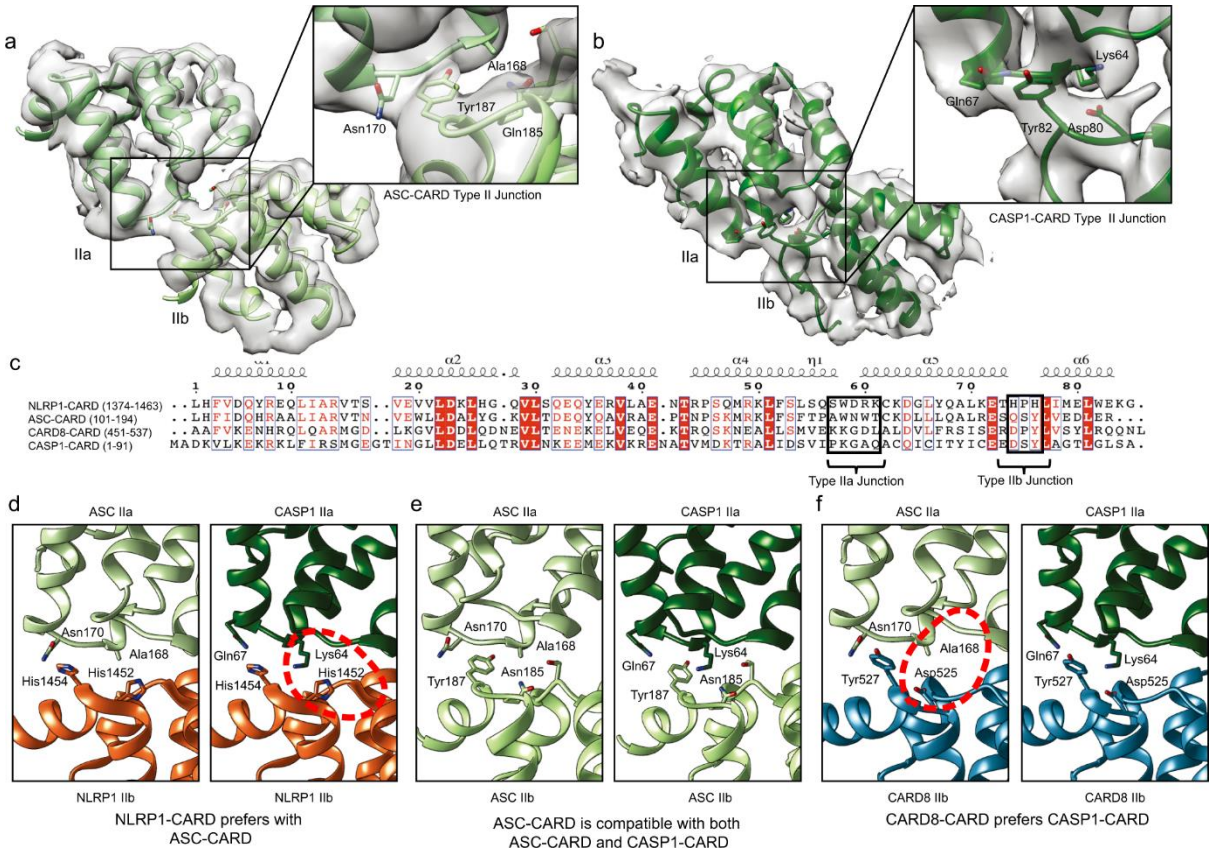

Figure 3.2. 5: Type II interface and important residues located on it are critical for inflammasome signal transduction. (A) Important residues located at Type II interface of ASC-CARD filamentous complex (model source: PDB-6K99, density map source: EMD-9947) (B) Important resides located at Type II interface of caspase-1-CARD filamentous complex (model source: PDB-5FNA, density map source: EMD-3241), refitting result. (C) Multiple sequence alignment analysis of human NLRP1-CARD, ASC-CARD, CARD8-CARD, and caspase-1-CARD. Critical amino acids located at the Type II interface are labeled with black rectangles. (D) Comparing of NLRP1-ASC interface and NLRP1-caspase-1 interface, potential clashes are labeled with red circles. (E) Comparing of ASC-ASC interface and ASC-caspase-1 interface, possible clashes are labeled with red circle. (F) Comparing of CARD8-ASC interface and CARD8-caspase-1 interface, potential clashes are labeled with red circle. Possible clashes are labeled with red circle.

After further model refinement, we demonstrate the critical reside pairs on the Type II interface of NLRP1, CARD8, ASC, and caspase-1 (Figure 3.2.5c–f). Here, NLRP1, CARD8, and ASC models are from our deposited PDBs (6K7V, 6K9F, 6K99), and the model of caspase-1 is fetched from PDB 5FNA, which is then locally refined using CARD8-CARD filament as reference. With this comparison, the specificity is more apparent; some interactions are more favored; however, others demonstrate partial incompatibility. At the NLRP1-ASC interface, there exists an amino acid combination motif HPH (1452-1454) on the NLRP1 type IIb interface and residue combination motif AWN (168–170) on ASC IIa interface. These two motifs achieve ionic interactions on this interface (Figure 3.2.5c, d, left panel). In contrast, the HPH motif at NLRP1 clashes with residues motif KGAQ (64–67) at caspase-1 Type IIa interface (Figure 3.2.5c, d, right panel, circle). That explains why the NLRP1-CARD oligomer is able to recruit ASC whereas it is not able to activate caspase-1 directly. (Figure 3.2.5c, d).

On the other hand, the CARD8-CARD complex owns a residues motif DPY (525–527) on its Type IIb interface. This motif is favorable for residue Lys64 of caspase-1, which locates at the Type IIa interface. (Figure 3.2.5c, f, right panel). In contrast, the CARD8 DPY motif is not favorable enough to establish interaction with ASC Type IIa interface, as ASC Ala168 cannot pair with CARD8 Asp525 (Figure 3.2.5f, left panel). Consequently, from a structural view, CARD8-CARD oligomer is more compatible with caspase-1-CARD oligomer than ASC-CARD. Results above demonstrate that charge distribution and critical motif at Type II interface determine the binding specificity among NLRP1-ASC activation and CARD8-caspase-1 activation.

This structural analysis is able to explain why, in some CARD-based inflammasome signaling pathways, ASC is an essential adaptor, whereas not in some others. This process includes the ASC-CARD oligomerization step when receiving sensor signal and the caspase-1 activation step. The structural basis is that there exists a unique amino acid motif QSY (185–187) at the Type IIb surface of ASC, which determines different activation affinities for NLRP1-CARD and CARD8-CARD, respectively. Furthermore, this ASC-CARD Type IIb QSY motif is favorable for Type IIa motifs of caspase-1. For example, AWN and KGAQ motif (Figure 3.2.5e). More evidence supporting this hypothesis appears with additional sidechain features. At Type IIb surface, compared with significant negative Asp residue on CARD8-CARD, ASC-CARD owns a less-charged residue Asn, which enhances interaction with Ala168 on ASC-CARD Type IIa surface. Additionally, compared to Tyr527 located at CARD8 Type IIb surface, at corresponding position on ASC-CARD type IIb interface, Tyr187 exhibit a more inwards orientation, which is more effective in engaging ASC-CARD Type IIa interface and caspase-1-CARD interface (Figure 3.2.5e).

In order to reveal the activation specificity between different CARD domains, an in vitro oligomerization assay was designed, which uses recombinant ASC-CARD, and caspase-1-CARD labeled with a fluorescent protein. Agreeing with the structural detail, CARD8-CARD is able to trigger caspase-1-CARD oligomerization, which is supported by the native gels result as the bands migrate slower (Figure 3.2.13). A similar pattern is the same for NLRP1-ASC activation. Then 3D-SIM (3-dimensional structured illumination microscopy) was used to image the CARD-CARD interaction interface of NLRP1-ASC and CARD8-caspase-1 within cultured HET239T cells. Conspicuous

filament bundles of mCherry-tagged NLRP1-CARD and CARD8-CARD are formed within the cytosol (Figure 3.2.6a, red). Moreover, as expected, GFP-tagged ASC and caspase-1 filament are formed and intertwined with their activator, respectively (Figure 3.2.6a, green). Given the theoretical resolution of around 110 nm in width, we hypothesis that the thinnest observed bundle complex was roughly six times the width of the mCherry tag labelling individual filaments.

Among these imaging regions, many instances demonstrate an “emanating” pattern: the receiver CARD filament (ASC and caspase-1) is extended from its activator (NLRP1 and CARD8) (Figure 3.2.6a). This seeding phenomenon was not only shown on CARD alone construct. Full-length NLRP1 and CARD8 activated by Talabostat can form filament (Zhong, Mamai et al. 2016, Mitchell, Sandstrom et al. 2019, Sandstrom, Mitchell et al. 2019) is also shown simultaneously with ASC-CARD and caspase-1-CARD filaments, respectively. Furthermore, full-length NLRP1 activated by Talabostat or its active fragment NLRP1-FIIND^{UPA}-CARD cannot stimulate caspase-1-GFP oligomerization without adaptor ASC. As compared, activated full-length CARD8 and its active fragment CARD8-FIIND^{UPA}-CARD could trigger ASC-GFP specks in HEK293 T cells. In conclusion, the above cellular assay gives evidence that NLRP1-CARD and CARD8-CARD oligomers formed in vivo differentially recruit ASC or caspase-1 through heterotypic CARD-CARD interaction.

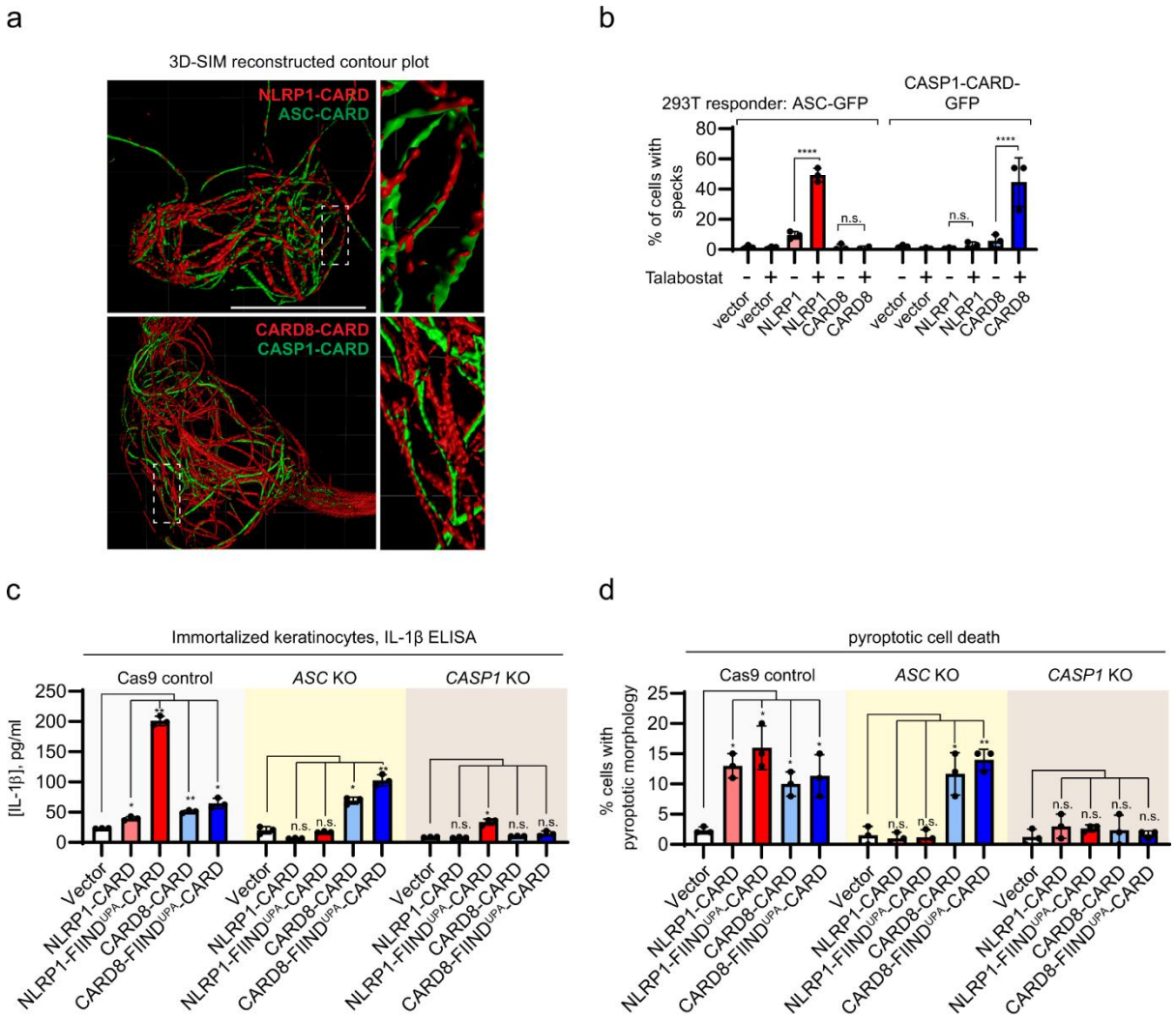

Figure 3.2. 6: Visualization of NLRP1 and CARD8 assembling in mammalian cells. (A) Imaging of NLRP1-CARD (labeled with mCherry) filament (red) and CARD8-CARD (labeled with mCherry) and filament (red) with ASC-CARD-GFP specks (green) and caspase-1-CARD-GFP filament (green) through Structured illumination microscopy. The Zoom-in region is illustrated with a white rectangle. (B) HEK293T assay result, the y-axis shows the percentage of ASC-CFG specks or caspase-1-CARD-GTP filament resulting from Talabostat-treated (three μM) NLRP1 or CARD8. Data demonstrate with mean ± standard-deviation format. One-way ANOVA is used to calculate the statistical significance values, with biological replicates n=3. '/**/****' here indicates P-value < 0.05, 0.01, and 0.0001, respectively. (C) Comparing IL-1β secretion in control, ASC knocked-out and caspase-1 knocked-out keratinocytes. Different NLRP1 and CARD8 constructs are transfected into these 3 types of*

cells. Data demonstrate with mean ± standard-deviation format. One-way ANOVA is used to calculate the statistical significance values, with biological replicates n=3. '/**/****' here indicates P-value < 0.05, 0.01, and 0.0001, respectively. (D) Keratinocytes transfection results, the y-axis shows the percentage of cells with pyroptotic morphology resulting from different NLRP1 and CARD8 constructs. Data demonstrate with mean ± standard-deviation format. One-way ANOVA is used to calculate the statistical significance values, with biological replicates n=3. '*/**/****' here indicates P-value < 0.05, 0.01, and 0.0001, respectively.*

Our work further corroborates that CARD8 is an exceptional human inflammasome sensor that can activate caspase-1 directly whereas cannot activate ASC (Ball, Taabazuing et al. 2020). When endogenous ASC (PYCARD) is deleted with CRISPR/Cas9, immortalized human keratinocytes do not suffer from increasing endogenous IL-1β secretion (Figure 3.2.6c, red bars) and pyroptotic cell death (Figure 3.2.6d, red bars), which should be resulted from overexpressed NLRP1-CARD or NLRP1-FIIND^{UPA}-CARD fragment. As for CARD8, pyroptosis relying on CARD8 does not affect significantly by ASC knockout in keratinocytes (Figure 3.2.6c, d, blue bars). These results again validate the model. NLRP1 needs ASC as adaptor in human inflammasome signal transduction, whereas CARD8 can recruit caspase-1 directly independent of ASC.

In the end, a "seeding" experiment was performed. We use naked NLRP1-CARD and CARD8-CARD oligomers (seeds) to trigger ASC-SNAP and caspase-1-CARD-SNAP oligomerization, respectively. With these constructs, downstream ASC and caspase-

1 filament will be apparently thicker because of SNAP-tag. Consequently, we can easily differentiate “naked” NLRP1-CARD and CARD8-CARD filament from ASC and caspase-1. The pattern that thicker filament (ASC-CARD and caspase-1-CARD) extended from thin seed filament (NLRP1-CARD and CARD8-CARD) was observed multiple times in several independent experiments. This result supports our structure model where different Type II interface patterns lead to NLRP1-ASC and CARD8-caspase-1 activation specificity.

3.2.7 NLRP1 self-assembly study and structure determination

NLRP1 CARD domain undergoes concentration-dependent self-assembly. NLRP1 CARD exists in a monomer state in low concentration; however, NLRP1 CARD filament was observed when concentration was increased. With concentration further increasing, NLRP1 CARD filament bundles appeared. These oligomer bundles can be removed by centrifugation (Figure 3.2.7).

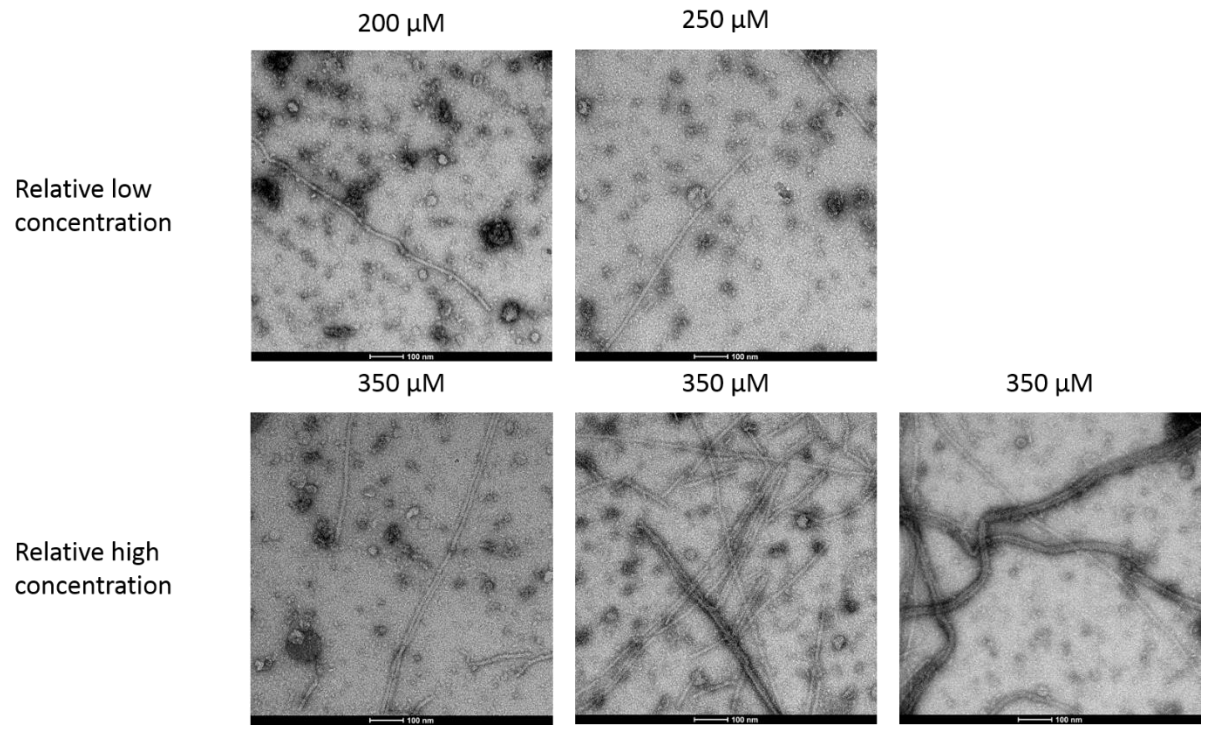

Figure 3.2. 7. Concentration-dependent self-assembly of NLRP1 CARD, as concentration increases, NLRP1 CARD tends to from filament bundles

NLRP1 CARD data collection was processed with Titan Krois. 600,000particles were manually picked for the 2D average, and 310,000 particles were selected from 2D classification (Figure 3.2.8). After 3D classification and 3D refinement, a density of 3.7 Å was generated (Figure 3.2.9). Based on this oligomer density, the NLRP1 CARD filament model was built in Chimera.

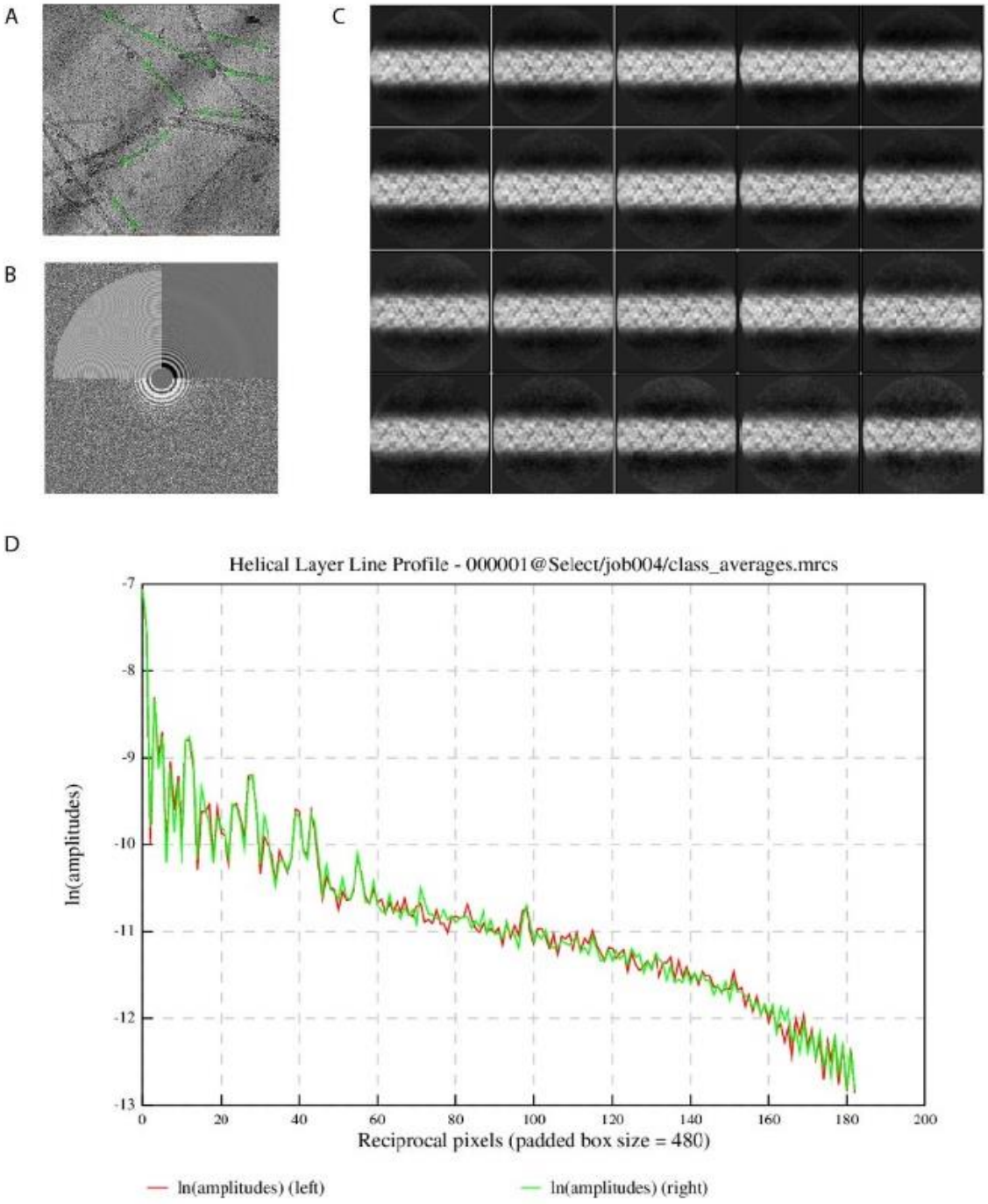

Figure 3.2. 8. (A) Raw image of NLRP1 CARD filament. (B) Power spectrum of the micrograph. (C) Representative 2D average from 2D classification (D) helical profile of one typical 2D average

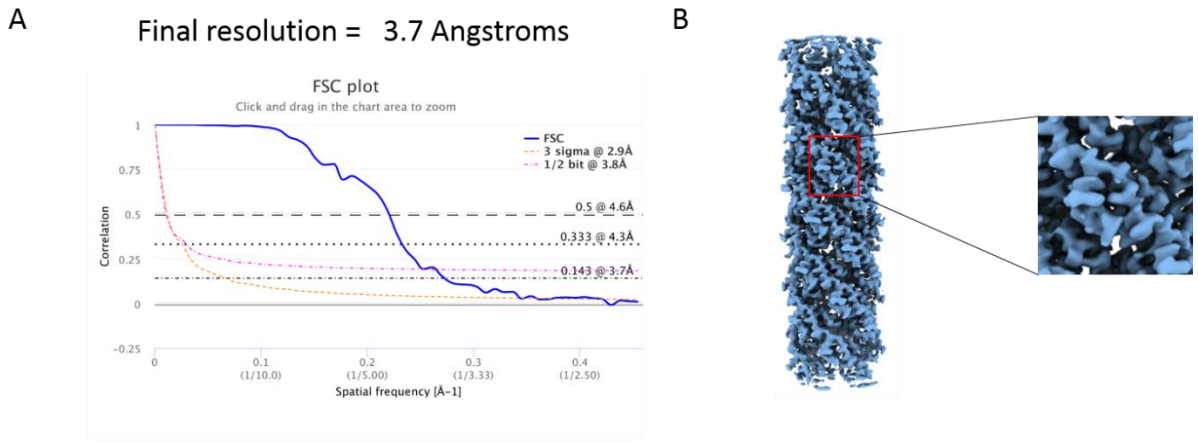

Figure 3.2. 9. (A) FSC gold standard curve for NLRP1 filament map. (B) NLRP1 CARD filament electron density model after polishing

CARD8 filament was processed with similar method, and the non-polished map exhibit a resolution of 4.1 Å (Figure 3.2.10A). Density map in the right panel also demonstrate corresponding high-resolution details (Figure 3.2.10B).

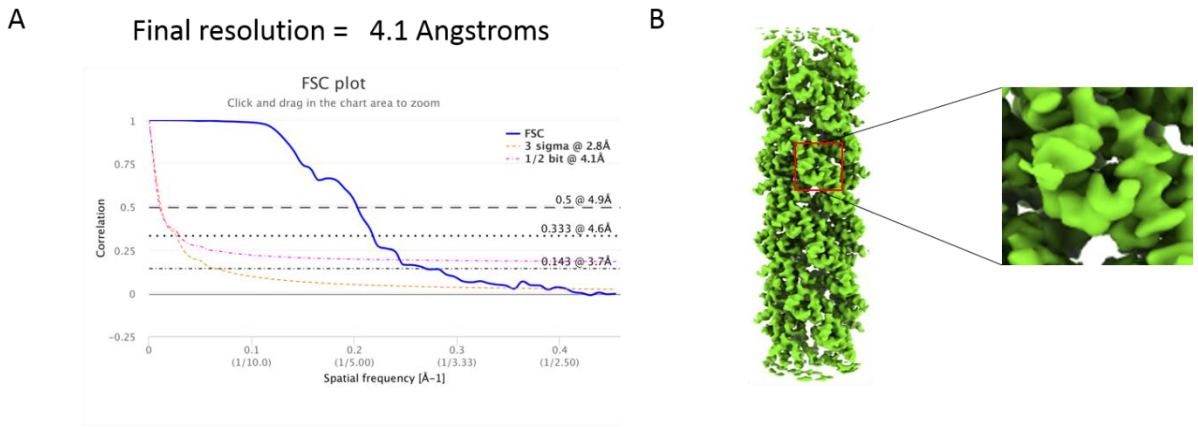

Figure 3.2. 10. (A) FSC gold standard curve for CARD8 filament map. (B) CARD8 CARD filament electron density model after polishing

Table 3.2. 1. Data collection, image processing, and coordinates refinement statistic for NLRP1

Data Collection	
Microscope	Titan Krios
Voltage (kV)	300
Detector	K2
Pixel size (Å)	1.1
Defocus range (μm)	-1 to -2
Electron Dose (e-/A2)	40
Movie Frames	40
Helical Reconstruction	
Software	RELION 3.0.4
Particles	422388
Helical rise (Å)	5.36
Helical rotation (°)	-100.821
Resolution (Å)	3.7
Coordinate Refinement	
Software	Phenix
Resolution Cut Off	3
Rwork	0.2548
Rfree	0.2846
Model	
Oligomer	center 12mer
Number of residues	1008
B-factor overall	360
R.M.S. deviation	
Bond length (Å)	0.005
Bond angle (°)	0.767
Validation	
Molprobity clashscore	2.06
Rotamer Outliers (%)	1.5
Cβ deviation (%)	0
Ramachandran plot	
Fovored (%)	96.5
Allowed (%)	3.5
Outliers (%)	0
Deposit	
PDB code	6K7V
EMDB code	9943

Table 3.2. 2. Data collection, image processing and coordinates refinement statistic for CARD8

Data Collection	
Microscope	Titan Krios

Voltage (kV)	300
Detector	K2
Pixel size (Å)	1.1
Defocus range (μm)	-1 to -2
Electron Dose (e-/A2)	40
Movie Frames	40
Helical Reconstruction	
Software	RELION 3.0.4
Particles	1201108
Helical rise (Å)	5.409
Helical rotation (°)	-99.16
Resolution (Å)	3.7
Coordinate Refinement	
Software	Phenix
Resolution Cut Off	3
Rwork	0.253
Rfree	0.3017
Model	
Oligomer	center 12mer
Number of residues	1044
B-factor overall	276
R.M.S. deviation	
Bond length (Å)	0.002
Bond angle (°)	0.423
Validation	
Molprobity clashscore	5.41
Rotamer Outliers (%)	0
Cβ deviation (%)	0
Ramachandran plot	
Fovored (%)	100
Allowed (%)	0
Outliers (%)	0
Deposit	
PDB code	6K9F
EMDB code	9948

3.2.8 Discussion and Summary

Taken together, our high-resolution structure, biochemical assay, cellular assay suggests a model for inflammasome activation of NLRP1 and CARD8 (Figure 3.2.11).

Here FIIND domain auto-cleave plays an “initializer” role to liberate the FIIND^{UPA} domain, which is shown to be necessary for NLRP1 activation. Some ligand and 3C protease (Robinson, Teo et al. 2020) are able to activate NLRP1 from an auto-inhibition state. Firstly, FIIND^{UPA} performs self-assembly and forms ring-like oligomers; this conformation then promotes adjacent CARD oligomerization. This two-step self-assembly lowers the necessary concentration for CARD oligomerization and promotes the further extension of FIIND^{UPA}-CARD filaments (Figure 3.2.11a). CARD8-FIIND^{UPA}-CARD oligomerization mechanism can also be explained through our model (Figure 3.2.11b).

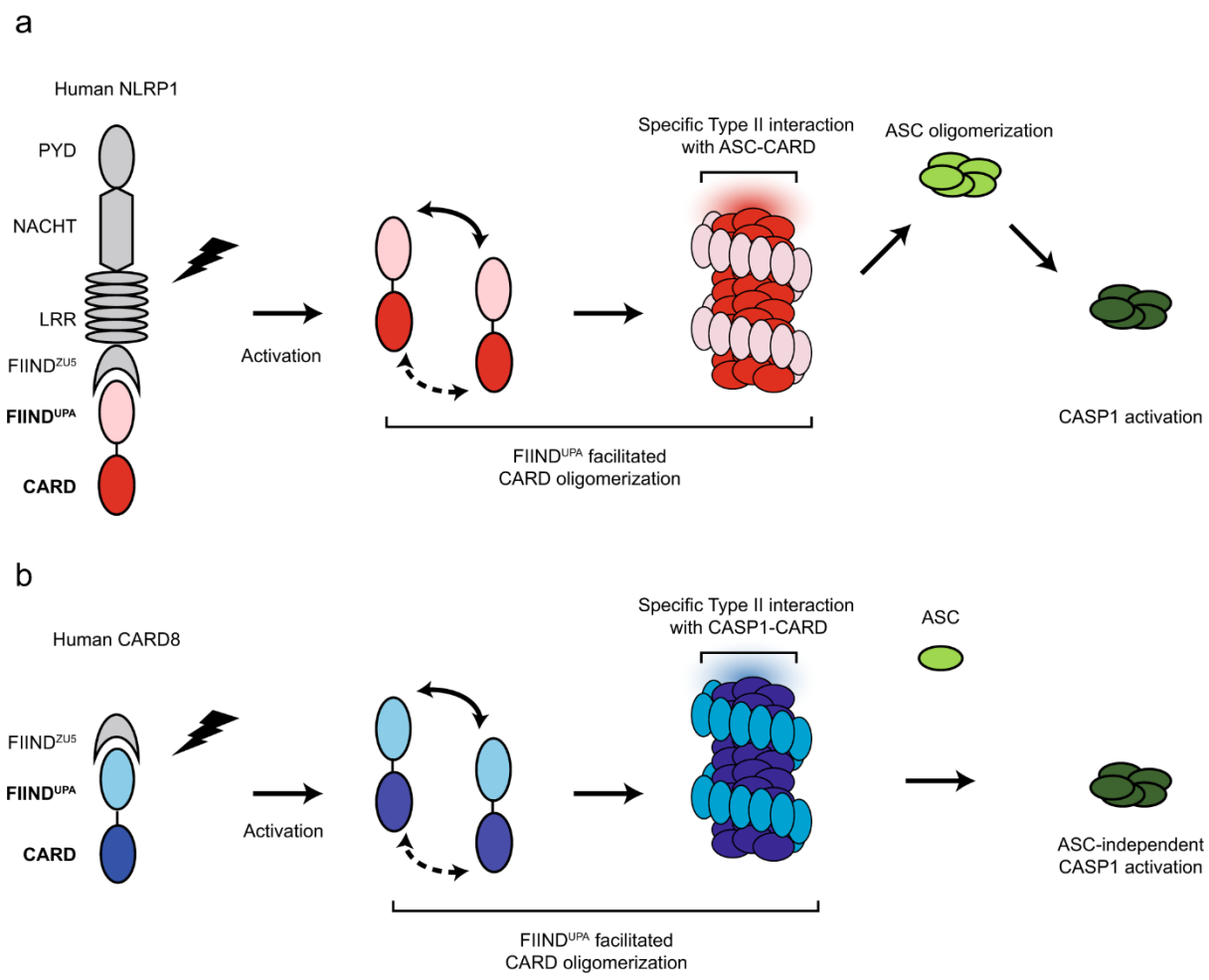

Figure 3.2. 11: Mechanism model of NLRP1 and CARD8 related inflammasome formation, which is conducted through two steps. At a resting stage, NLRP1(A) and CARD8(A) stay in an

auto-inhibited state. Once receiving through DAMP and PAMP signals, the N-terminal proteolytic domain is activated, and FIIND^{UPA} is able to promote self-assembly of CARD, resulting in the formation of the filamentous complex with two-layer architecture. Active NLRP1-CARD filament and CARD8-CARD filament then trigger ASC and caspase-1, respectively, leading to the formation of different pyroptosis complexes.

Then we solve the structure of NLRP1-CARD and CARD8-CARD filament with Cryo-EM. Although these filaments are similar in helical parameters, there is a subtle but essential difference in the rotation angle for the respective monomer. Besides, the charge distribution on the Type II interface is also distinct. With further analysis, we found that the specificity of NLRP1-ASC and CARD8-caspase-1 activation is due to the critical motifs at the Type II interface. In particular, we explain the structural basis of why ASC is essential for NLRP1 inflammasome activation, whereas CARD8 does not need it. Therefore, each CARD is a unique and non-interchangeable role in different inflammasome. To specify the composition of distinct inflammasome complexes, each CARD possesses intrinsic abilities to bind specific downstream effector CARDs. It is understood that different inflammasome complexes is able to trigger different biological outcomes, which is consistent with previous findings. Cell death and IL-1 cytokine secretion regulation require diverse pyroptotic signaling strengths (Broz, von Moltke et al. 2010, Dick, Sborgi et al. 2016). It is an exciting finding that the importance of Type II interface is different for various inflammasome complexes. Although NLRP1 and CARD8 inflammasome rely more on Type II interface, other CARD domain inflammasome complexes like MAVS (EMD-5925) (Wu, Peisley et al. 2014), BCL10 (EMD-7314) (David, Li et al. 2018), RIPK2 (EMD-4399,

6842) (Gong, Long et al. 2018, Pellegrini, Desfosses et al. 2018) rely more heavily on Type I interface. This highlights the importance of the atomic resolution model in the complex macromolecular study. As the macromolecular complexes assembly is critical for signal transduction in innate immune and cell death (Kagan, Magupalli et al. 2014), our work suggests that future work should focus on the general assembly mechanism of other large molecular complexes, which is involved in these crucial processes.

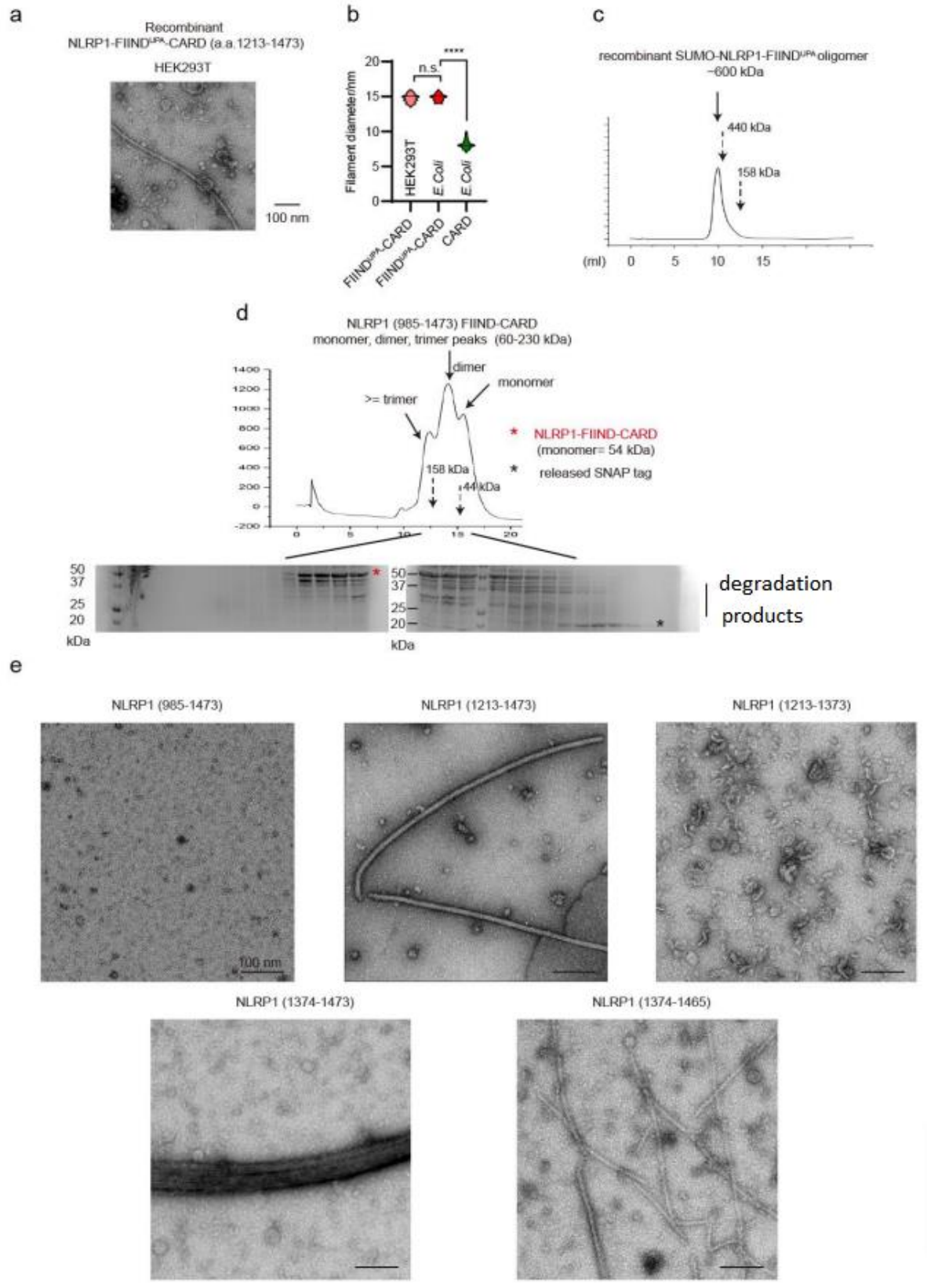

Figure 3.2. 12. Oligomerization assay for NLRP1 recombinant protein. (A) NLRP1-FIIND^{UPA}-CARD filament purified from HEK293T cells and its negative staining imaging. (B) Summary

for the dimension of NLRP1 filament. (C) Superdex S200 increase column size exclusion profile for SUMO-NLRP1-FIIND^{UPA}. (D) sample separated with size exclusion and their corresponding SDS-PAGE results. (E) Negative staining imaging for NLRP1 with domain different boundary.

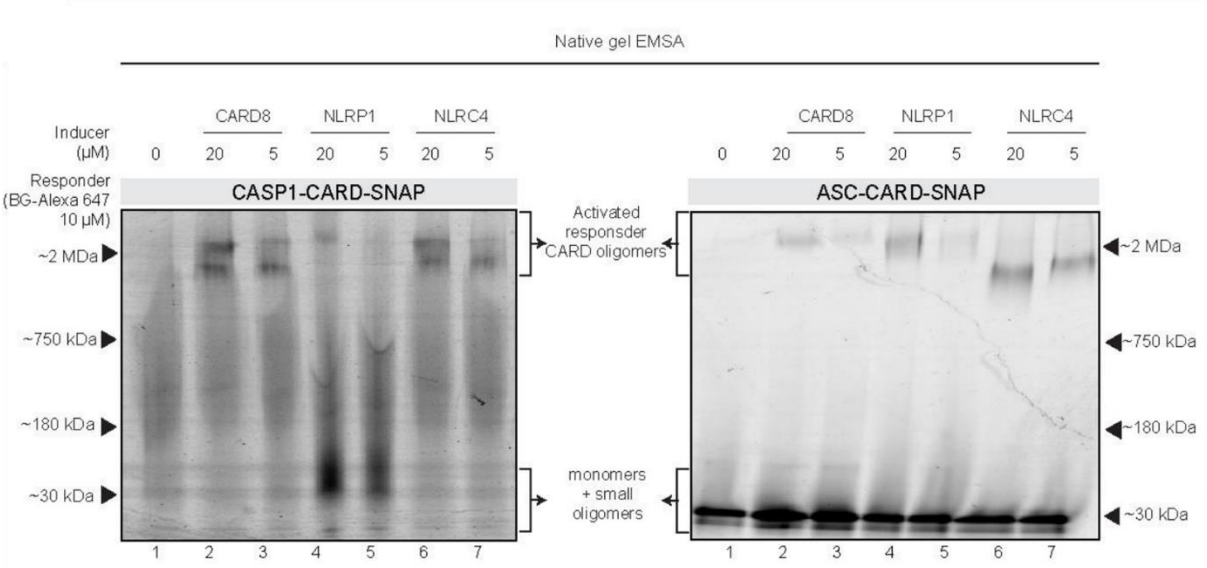

Figure 3.2. 13. Left panel shows caspase-1 activation through different seeds. Right panel demonstrate ASC activation through different seeds.

3.3 Cryo-EM structure of native UMOD filament

3.3.1 UMOD filament owns unique helical parameters

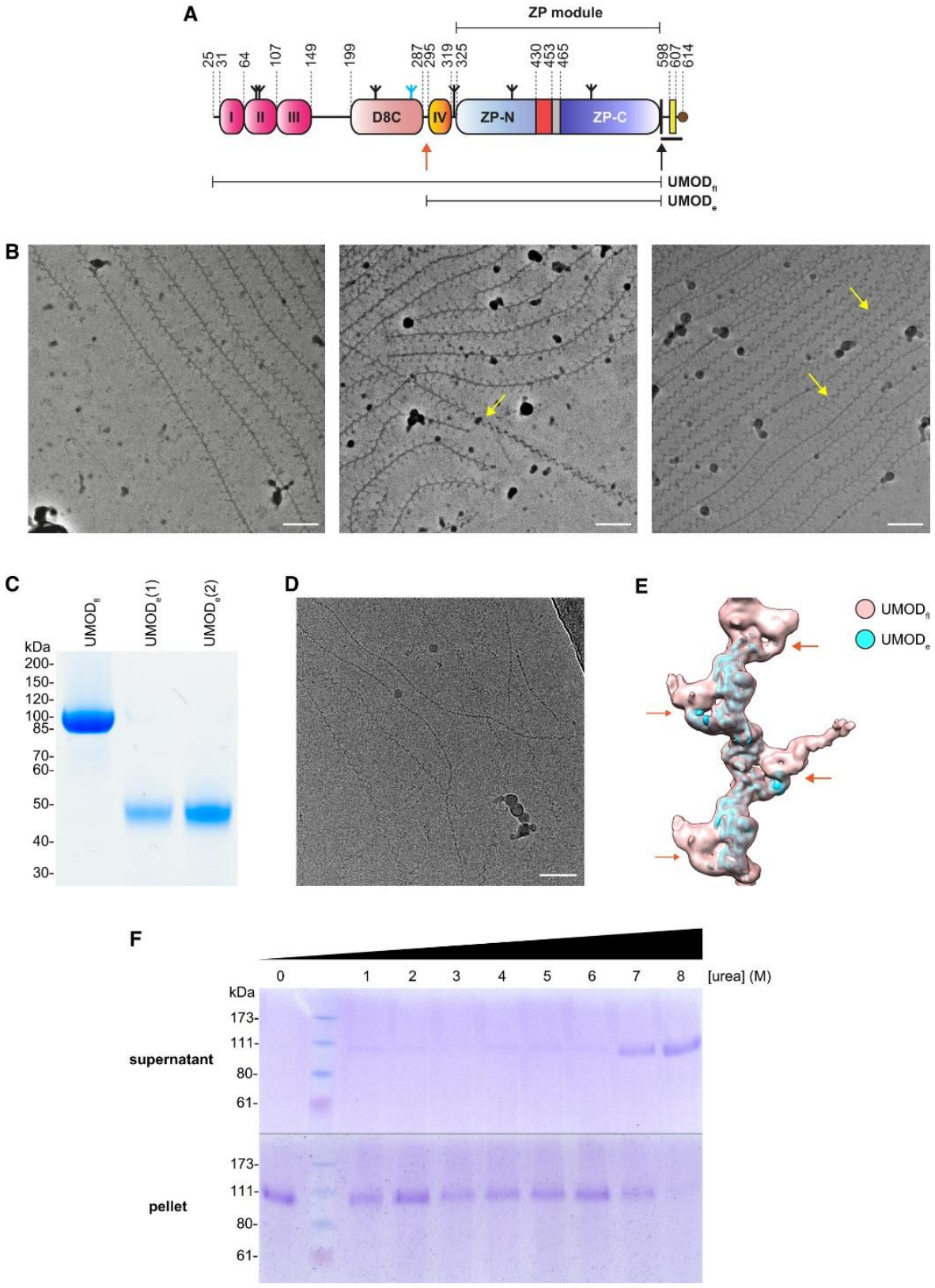

Figure 3.3. 1 (A) Domain architecture of secretive human uromodulin (UMOD) precursor. EGF I II III domains are labeled with magenta; D8C domain is labeled with salmon; EGF IV domain is labeled with orange; ZP-N domain is labeled with light blue; ZP-C domain is labeled

with dark blue; ZP-N/ZP-C linker is labeled with red; IHP domain is labeled with gray; CCS domain is labeled with black; EHP domain is labeled with yellow; CTP domain is marked with thick black horizontal line; CGI anchor attaching point is marked with a brown circle. N-glycans is marked with Inverted tripods, in which the N275 residue located at D8C is attached to a high-mannose chain and is labeled with cyan. Hepsin (F587 and R588) is marked with a black arrow, and elastase digestion sites (S291 and S292) is marked with orange arrows. UMODfl and UMODe is indicated with different bars. (B) Imaging of native UMODfl filament under phase plate Cryo-EM. Both tree-like structures and zig-zag structures appear in the side views of the filament. The transient points between the three-like structure and zig-zag structure are pointed out with yellow arrows. Scale bars: 50nm. (C) SDS-PAGE result with Reducing Coomassie-stained. Lane 1: UMODfl (6 µg). Lanes 2: UMODe (3 µg). Lanes 3: UMODe (5 µg). These specimens were used for Cryo-EM structure determination later. (D) Imaging of UMODe filament, where the branches are absent due to elastase digestion. Scale bar: 50 nm. (E) Density overlapping analysis for UMODfl (labeled with salmon) and UMODe (labeled with cyan) shows only UMODfl density demonstrate the globular domain extending from the central core. These results suggest the approximate position of the elastase digestion site in the UMODfl structure (pointed out with orange arrows), which correspond to the UMODe N-terminus. (F) SDS-PAGE result of gradient urea concentration shows no significant phenomenon of oligomers digestion when urea concentration is below 7 M.

In order to obtain information on UMOD oligomer, firstly, we screen processed human urine samples through Cryo-EM. The images show that the supramolecular structure of UMOD exhibits filamentous and sheet-like shape. The pairing features are similar to the projection, from previous reported double-helical structure (Jovine, Qi et al. 2002) (Figure 3.3.2A). UMODfl (full-length native UMOD) demonstrates a filament form

under Cryo-EM. The majority of the filaments exhibit tree-like substructure, with consistent alternation of branches with a length of around 12 nm extending from either side of the central core at an angle of roughly 55 degrees (Figure 3.3.2B)—as for other filaments, they adopted a zig-zag structure, which is consistent with previous negative staining EM study for UMOD (Bayer 1964) (Figure 3.3.2C). These two types of filament structure can occasionally exhibit interconversion within individual filaments (Figure 3.3.1B).

Consequently, diverse distinct conformation appears during helical reconstructions, which corresponds to different side views of a single type filament, resulting from its unusual helical parameter; 180° rotation twist and 62.5 Å axial rise. Structure determination is severely affected by the flexibility and the thinness (~35 Å) of the central core filament. Apart from that, the 180° rotation twist is another critical problem to handle, which is more extreme than the helical twist of F-actin (−166.6°) (Dominguez and Holmes 2011). During structure reconstruction, we select the most homogenous helical segments with a total of 288,403 particles. Cryo-EM map derives from these particles reaches an overall resolution of 3.8 Å, as well as a filament core density of 3.4 Å (Figure 3.3.2F–J; Figure 3.3.3; Figure 3.3.12).

To help model building, we also try to solve the structure of UMODe (native UMOD digested with elastase). Elastase is a protease that is able to remove UMOD N-terminal region (EGG I-III and D8C), with a digestion site just before EGF IV (Jovine, Qi et al. 2002) (Figure 3.3.1A and C). This digestion results in the loss of UMOD filament branches (Figure 3.3.2D and E; Figure 3.3.1D). With an overlapping comparison between UMODfl and UMODe, we can identify the position for the N-

terminus region of EGF IV in the maps (Figure 3.3.2F and K; Figure 3.3.1E). As UMOD ZP-N and EGF IV domain is solved through crystallography (Bokhove, Nishimura et al. 2016), we could unambiguously dock these models into our UMODfl map and then fit the model of ZP-C (Bokhove, Nishimura et al. 2016). The corresponding strong density supported the docking results to the ZP-N N396 and its attached N-glycans (Figure 3.3.2G) and ZP-C N513 (Figure 3.3.3B, right panel). Subsequently, a prominent and continuous stretch of unexpected density linking both domains was further identified as ZP-C/ZP-N linker from a third UMOD subunit (UMOD 3), which embraces the placed ZP-C (UMOD 2) and ZP-N (UMOD 4), an adjacent subunit of UMOD 3 (Figure 3.3.2H and I; Figure 3.3.4). This finding revealed that the arrangement of UMOD for filamentous complex is different from its homodimeric precursor reported before (Bokhove, Nishimura et al. 2016). Consequently, the distance between the mass centers of ZP-C and ZP-N β-sandwiches increases from 41 Å to 91 Å during polymerization (Figure 3.3.5A). In agreement with a roughly 120 Å axial periodicity (Jovine, Qi et al. 2002), the head-to-tail (ZP-N to ZP-C) interaction conformation is achieved through this ZP domain conformation, which shows one and two-half per turn (Figure 3.3.2F; Figure 3.3.4).

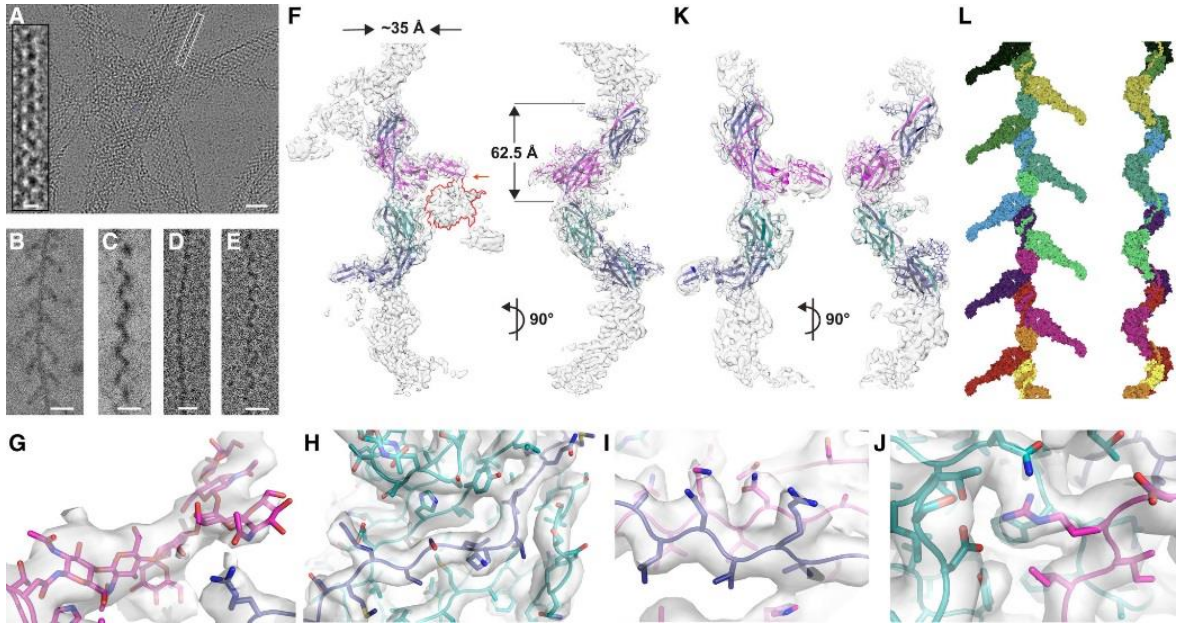

Figure 3.3. 2. Demonstration of native human UMOD filamentous complex. (A) Imaging of unstained UMOD filamentous sheets purified from human urine under the electron microscope. The zoom-in figure demonstrates a double helix structure resulting from two adjacent UMOD filament positioning in parallel. Scale bars: 50 nm and 10 nm (zoom-in). (B, C) Representative imaging for side views from UMOD filamentous complex with Volta phase plate. Tree-like side view in (B) and "zig-zag" side view in (C). Scale bars: 10 nm. (D, E) Two distinct views are taken from UMODe filamentous complex, which shows the pattern of branches' absence. Scale bars: 10 nm. (F) The polished UMODfl electronic map demonstrates a resolution of 3.8 Å. Placing to similar orientation to (B, C) correspondingly. The density map together with fitted coordinates model including a complete EGF IV domain and ZP domain (labeled with blue, chain A), a ZP-C domain from another molecule (labeled with teal, chain B), and EGF IV and ZP-N domain from the third molecule (labeled with magenta, chain C). (G–J) Zoom-in structural detail from UMODfl map shows in (F). (G) Demonstrate N396 located at ZP-N and labeled with glycan. (H) Demonstrate the ZP-N/ZP-C linker α1 β. (I) Demonstrate β1 strand. (J) Demonstrate the intermolecular interface of ZP-C αEF β/ZP-N βF'. (K) Polished electronic map of UMODe with resolution 4.0 Å with orientation the same as shown in (D and

E). By comparing these two maps, the extra part, which results from digestion, is labeled with orange (F). (L) UMODfl filamentous model is shown with Goodsell-style, and model from different subunit is labeled with a different color.

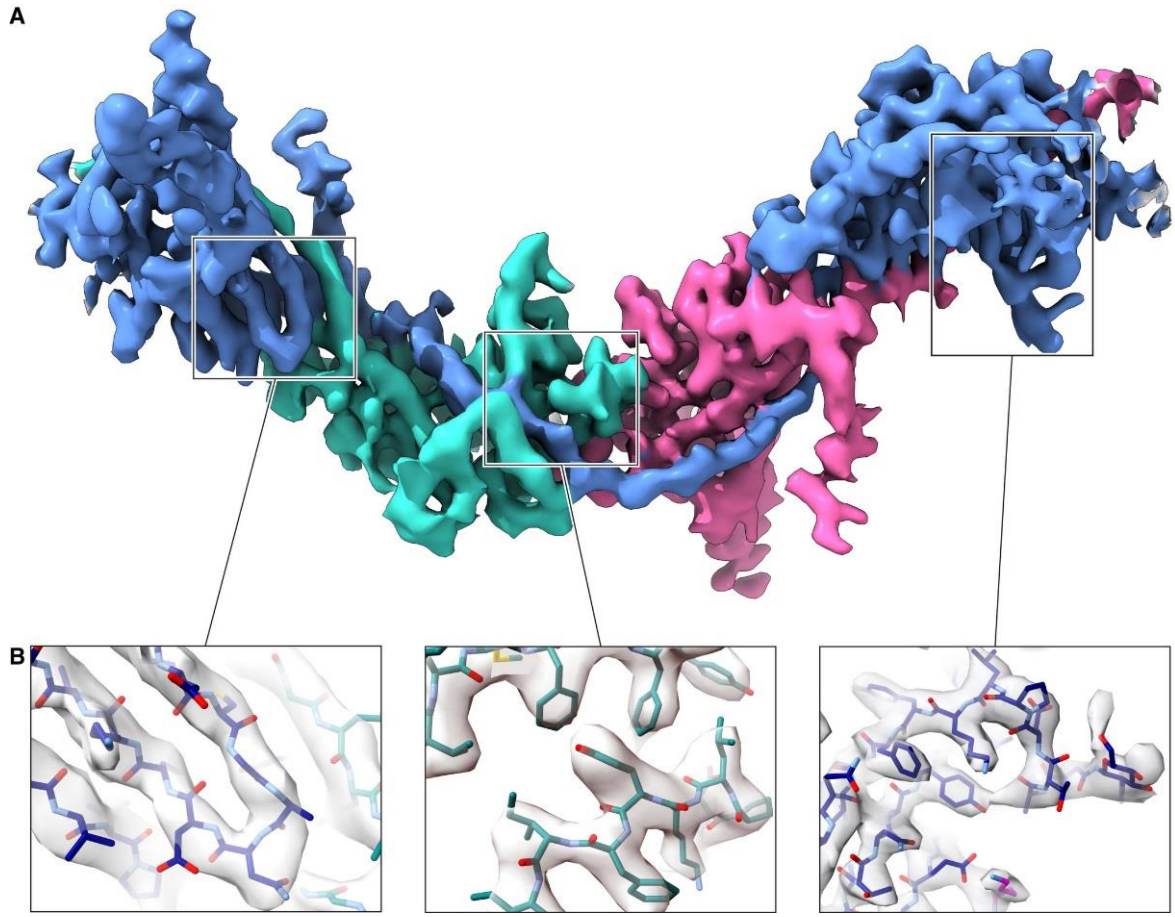

Figure 3.3. 3. Polished density map of core UMODfl filament with resolution 3.4 Å. (A) Overall architecture of core UMODfl filamentous complex. Here, a complete UMOD unit molecule is labeled with blue, which is surrounded by teal ZP-C domain and magenta ZP-N+EGF IV domains belonging to adjacent UMOD subunits. (B) Zoom-in views for some vital region of this map. The left panel demonstrates the isolation of β-strands. The middle and right panels show the side chain detail of the density map, reflecting the quality of our map. The coordinates model is fitted into this UMOD map and label with the same color as corresponding chains in panel A

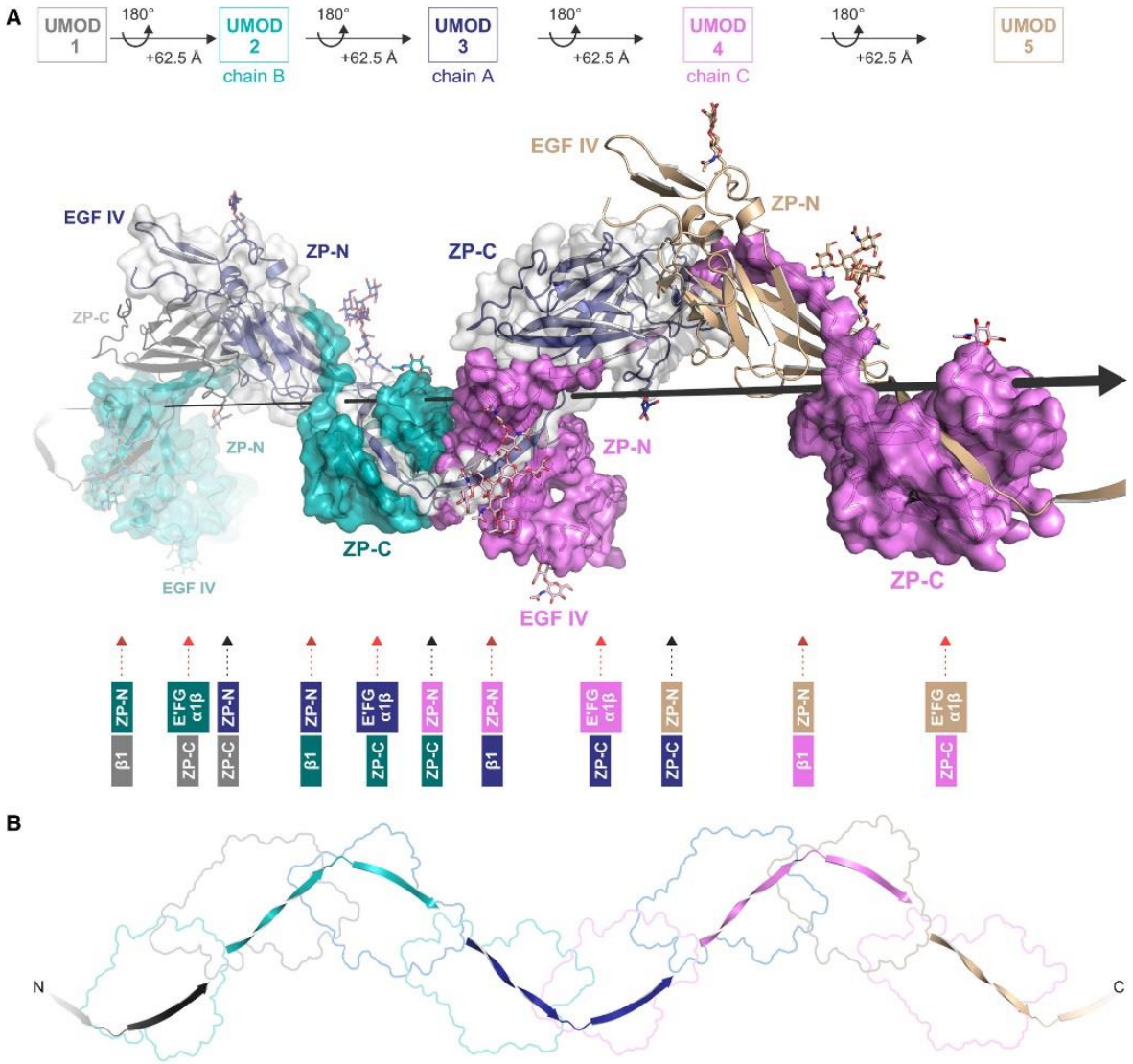

Figure 3.3. 4. Illustration of how one particular UMOD molecule interacts with the other four adjacent UMOD subunits through its ZP domain. (A) The top part demonstrates five continuous UMOD subunits (labeled from UMOD1-5), which compose one segment of the UMOD filamentous complex through helical symmetry repeat. The middle part shows a tilted view of the UMOD filament, and the central helical axis is shown as the giant black arrow. Here, UMOD-1, UMOD-3, and UMOD-5 are shown with cartoons, and the surface of UMOD-2, UMOD-3, and UMOD-4 are exhibit with a different color to highlight inter-protein interaction. UMOD-5 ZP-C and UMOD-1 ZP-N are omitted. The architecture of the linker between ZP-N and ZP-C can be described as that, linker of UMOD-3 surrounds the UMOD-2

ZP-C domain that is preceded to UMOD-3 and the UMOD-4 ZP-N domain that is behind UMOD-3. Furthermore, the ZP-C domain of one particular molecular is able to interact with ZP-N domains and ZP-C domains from the following it; for example, UMOD-1 ZP-C interacts with UMOD-2 ZP-N and ZP-C domains. And the ZP-N domain, one particular molecule is able to interact with ZP-N domains and ZP-C domains that precede it; for example, UMOD-5 ZP-N interacts with UMOD-4 ZP-N and ZP-C domains. The bottom part of (A) summary the position of important domains and motifs in the filament. The interaction that promotes one particular UMOD subunit interacting with the other four adjacent subunits involves six interfaces, which belong to three distinct kinds. These three types of interface is pointed out with different color arrows; ZP-N/ZP-C pointed out with a black arrow, ZP-N/β1 pointed out with a dark red arrow, E'FG and α1β/ZP-C pointed out with a light red arrow. UMOD subunits UMOD-3, UMOD-2, UMOD-4 here response to (A), (B), (C) in Figure 3.3.5, respectively. (B) linkers demonstrated with different color cartoons from UMOD-1 to UMOD-5. This side view is 40° tilted around the Y-axis, compared with that of the panel (A).

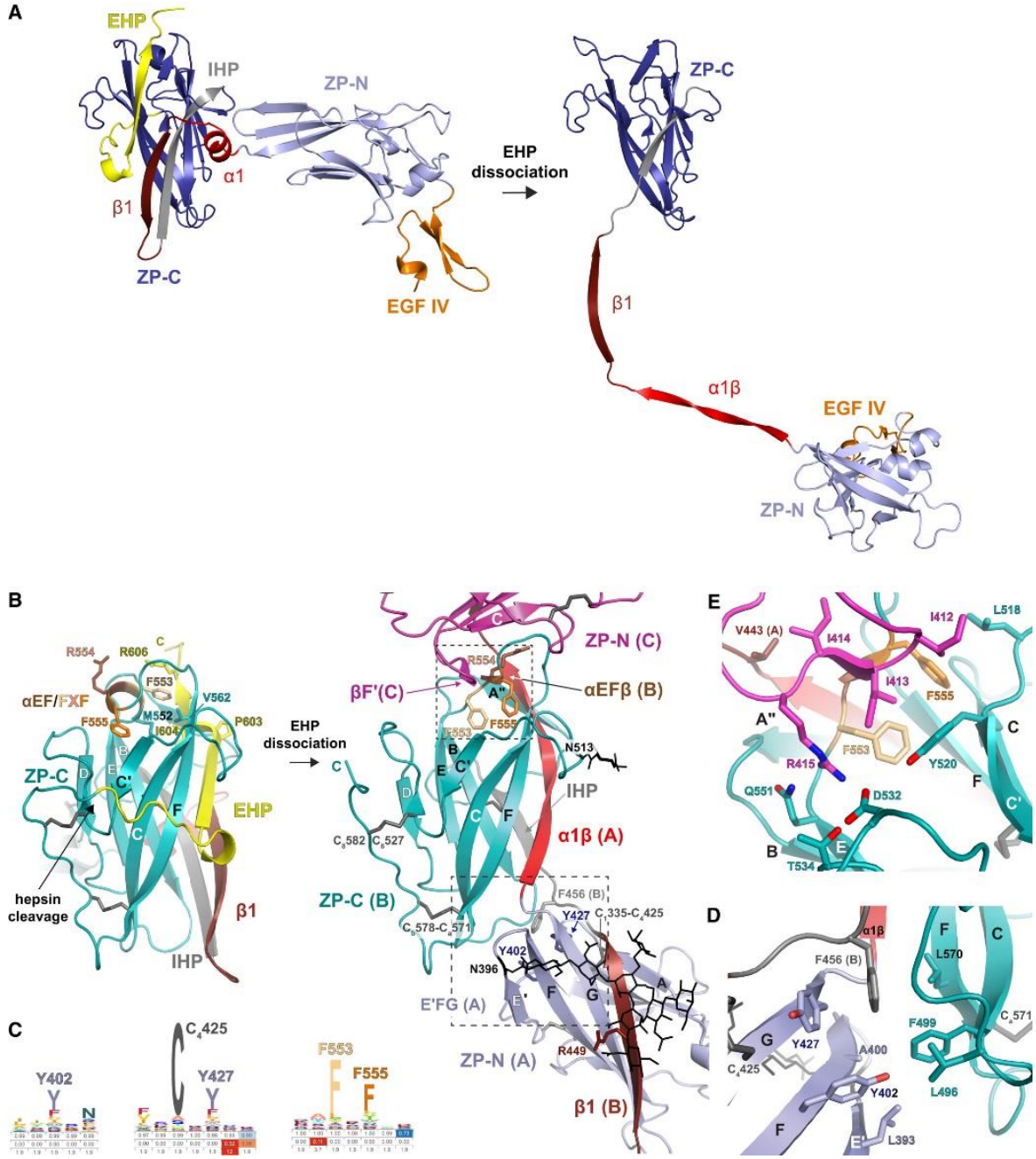

Figure 3.3. 5. significant conformation changes happen from UMOD precursor to UMOD filamentous complex. (A) By comparing the precursor and oligomerized structure of UMOD, it is suggested that EHP disassociation triggers the significant conformation change in ZP domains. For the interdomain linker, where remarkable rearrangement happens, this conformational change leads to total separation from the ZP-C domain and transfers the secondary structure from α1 that is in the precursor to α1β that is in the filament. The model is illustrated with a cartoon, which only shows one monomer belonging to the precursor dimer

structure. Structure motifs are labeled with the same color as that is shown in Figure 3.3.1A. the N-terminal and C-terminal of the linker between ZP-N and ZP-C is a label with bright red and dark red, corresponding. (B) As shown in the left part, among the ZP-C domain of UMOD precursor structure, the EHP β-strand, which work to prevent polymerization, binding to αEF (a short α-helix) that wrap FXF motif with hydrophobic force. Moreover, in the right panel, relief of EHP is mediated through Hepsin, which digest the CCS of this subunit (UMOD-2 chain B). Furthermore, the interdomain linker α1β from the other UMOD subunit (UMOD-3 chain A) replaces its position. This conformation change promotes the formation of intermolecular β-sheet by FXF motif (αEFβ/βF';), shown with the dashed box in the upper part and the fg loop of ZP-N domain, which belongs to the another incoming UMOD subunit (UMOD-4 chain C, label with magenta). Furthermore, CCS digestion also leads to that the mature C-terminus of UMOD-2 can interact with D8C of itself, not showing here. Disulfide bonds here is marked with thick dark gray and glycan residues is labeled with thin black sticks, corresponding. In addition, β-strands are illustrated with brown color. (C) Modeling the amino acid sequence with HMM (Hidden Markov model), the results highlight the conserved residues that are labeled in panels (B), (D) and (E). (D) The hydrophobic interaction helps stabilize the interface between the ZP-N domain E'FG extension of chain A and the chain B ZP-C domain, which corresponds to the dotted box in the lower right part (B). (E) Here, a different view of chain B ZP-C domain and chain C ZP-N domain (boxed region of panel B) show more residue contact details.

Model building for D8C is complex, because of the limited resolution in the corresponding map region, which is apart from the central filamentous complex and absence of homologous structures. However, from previous findings of the D8C predicted structure, which was derived from multiple protein structure prediction server,

suggest that there exist multiple intramolecular disulfides (Hamlin and Fish 1977, Yang, Wu et al. 2004). Therefore, this domain forms a compact conformation with similar dimensions, which matches the spherical electronic density extending from EGF IV (Figure 3.3.6A and B). A notable finding is that the density for hepsin-processed UMOD C-terminal tail merges with D8C because the flexibility of the tail is restricted by the last disulfide ($C_6$527-$C_8$582) of UMOD (Figure 3.3.6B). This result suggests that digestion of UMOD only promotes its ZP-C domain for oligomerization and favors it to interact with D8C, which in turn guides the N-terminal of the protein towards the core filament position. Apart from the addiction conformation constrain, the N-terminal region is notably mobile and flexible relative to the central core part, which results in smeared densities in 2D averages. Multi-body refinement is applied to refine the branch region separately, including the whole N-terminal part of UMOD (Figure 3.3.6C). When compared with the previous location of D8C, this locally refined map at higher resolution is consistent with the dimension of EGF I-III from a homology model (Figure 3.3.6C and D). With this information and our refined central region (EGF IV + ZP) coordinates, a full filament model is built, consistent with both tree-like and zig-zag side views from UMODfl (Figure 3.3.2L; Figure 3.3.6D).

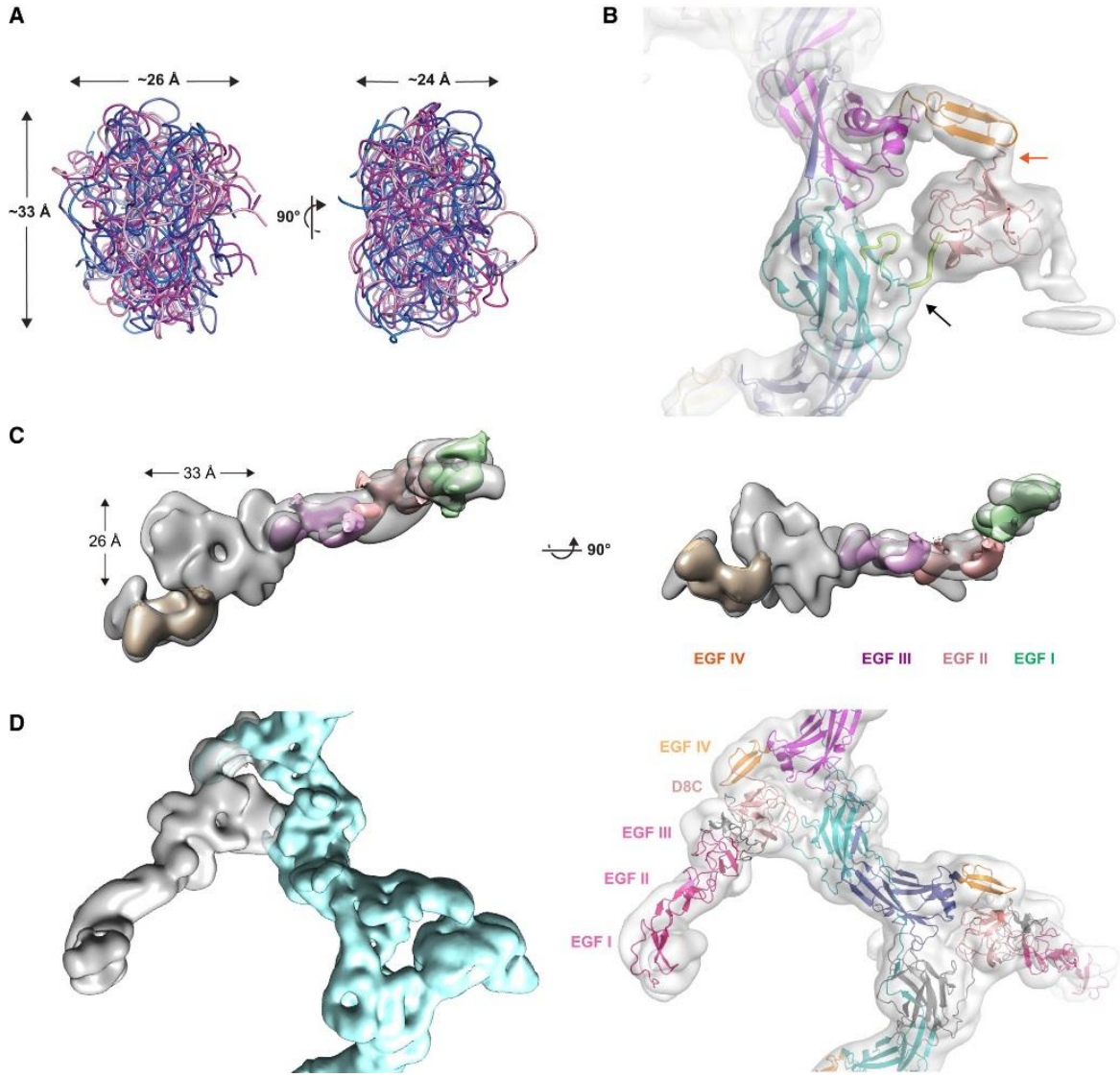

Figure 3.3. 6. Model fitting result of D8C and EGF I, EGF II, and EGF III to electronic density map of UMODfl filament. (A) Prediction models from two different protein structure prediction servers (I-TASSER labeled with blue shades and Robetta labeled with magenta shades) demonstrate similar dimensions. (B) In agreement with the position of the elastase digestion site (labeled with orange arrow) immediately preceding the EGF IV domain (labeled with orange), the model with the highest TM-score of D8C generated from server Robetta (labeled with salmon) can be directly fitted to the spherical density extending from the rigid core region of the UMOD filament. The before-polishing electronic map of UMODfl is demonstrated. Here the black arrow points to the disulfide of C6527-C8582, which tilts the mature UMOD C-terminal tail (labeled with thick lemon tube) to the D8C domain. D8C also wraps around the

loop (display with fine lemon tube) connecting βD and C6527. (C) To deal with flexibility of the extending region of UMODfl construct, Relion multibody refinement was used to perform locally and focused refinement on different regions, respectively. With this focused refined density map, the potential position of EGF I, EGF II and EGF III appears, which is helpful for the model building of UMODfl. (D) The density label with gray represents the extending N-terminus branch, considered as body 1, while the density labeled with cyan demonstrates the remaining core regions, considered as body 2. These two focused and locally refined maps merged to generate a complete map of UMODfl as a reference for protein coordinates building and as a subunit to model the matrix structure shown in Figure 3.3.11B.

3.3.2 Filamentous complex assembly involves a significant conformational rearrangement in the ZP module's inter-domain linker

A significant conformation change happens on the ZP-N/ZP-C linker when UMOD precursor rearranges and polymerizes into a filamentous form (Figure 3.3.5A). In the precursor, the inter-domain linker is composed of two parts: α1 (α-helix) and β1 (β-strand), which pack with ZP-C β-strand A and interact with EHP (external hydrophobic patch) (Bokhove, Nishimura et al. 2016). ZP-C β-strand A here is an essential element involved in oligomerization and is known as IHP (internal hydrophobic patch) (Jovine, Qi et al. 2004). As for polymerized UMOD, α1 and its following amino acids transform into a twisted β-strand of 13 residue. (α1β, which is composed of two different β strands: α1β' (D430-S434) and α1β'' (A438-M442), connected with a three-reside linker) (Figure 3.3.2H).

Consequently, the EHP from the previous subunit is replaced by hydrogen bonding to its A'' strands, F strands and facing βA/IHP (internal hydrophobic patch) (Figure 3.3.4A; Figure 3.3.5B), with linker residue L435 bind to the conserved hydrophobic pocket which is formed by IHP M460 and L462, βF L564 and βB V487 (Figure 3.3.7A and B). Another interesting finding is that an exact same copy of the DMKVSL sequence, which includes α1β', composes the UMOD ZP-N β-strand B (residues 339-344). Interdomain linker antiparallel replacing EHP/β G resembles the donor-strand exchange (DSE) reaction between bacterial pili subunits (Waksman 2017); however, as for the ZP module, the β-strand exchange is further established through a parallel pairing of the β1 region linker to ZP-N domain βG of the following subunit (Figure 3.3.2I). In agreement with the evolutionary conservation analysis on the ZP-N domain AG face (Monne, Han et al. 2008), this expands the E'FG β-sheet of the latter, exposing a surface, against which the well-resolved carbohydrate chain extended to N396 packs (Figure 3.3.4A; Figure 3.3.5B). Notably, the position where glycan located is conserved among many ZP domain proteins, and its loss-of-function mutation in inner ear α-tectorin is related to human deafness (Sagong, Park et al. 2010, Bokhove, Nishimura et al. 2016). This position is stabilized by the β1 linker itself, which binds to lactosamine GlcNac amide group through the side chain of R449 (Figure 3.3.2G; Figure 3.3.5B). Interacting with both ZP-N and ZP-C domains, this interdomain linker plays a molecular belt role in connecting three adjacent UMOD units, hiding a complete accessible surface region of 1488 Å2 (Figure 3.3.4).

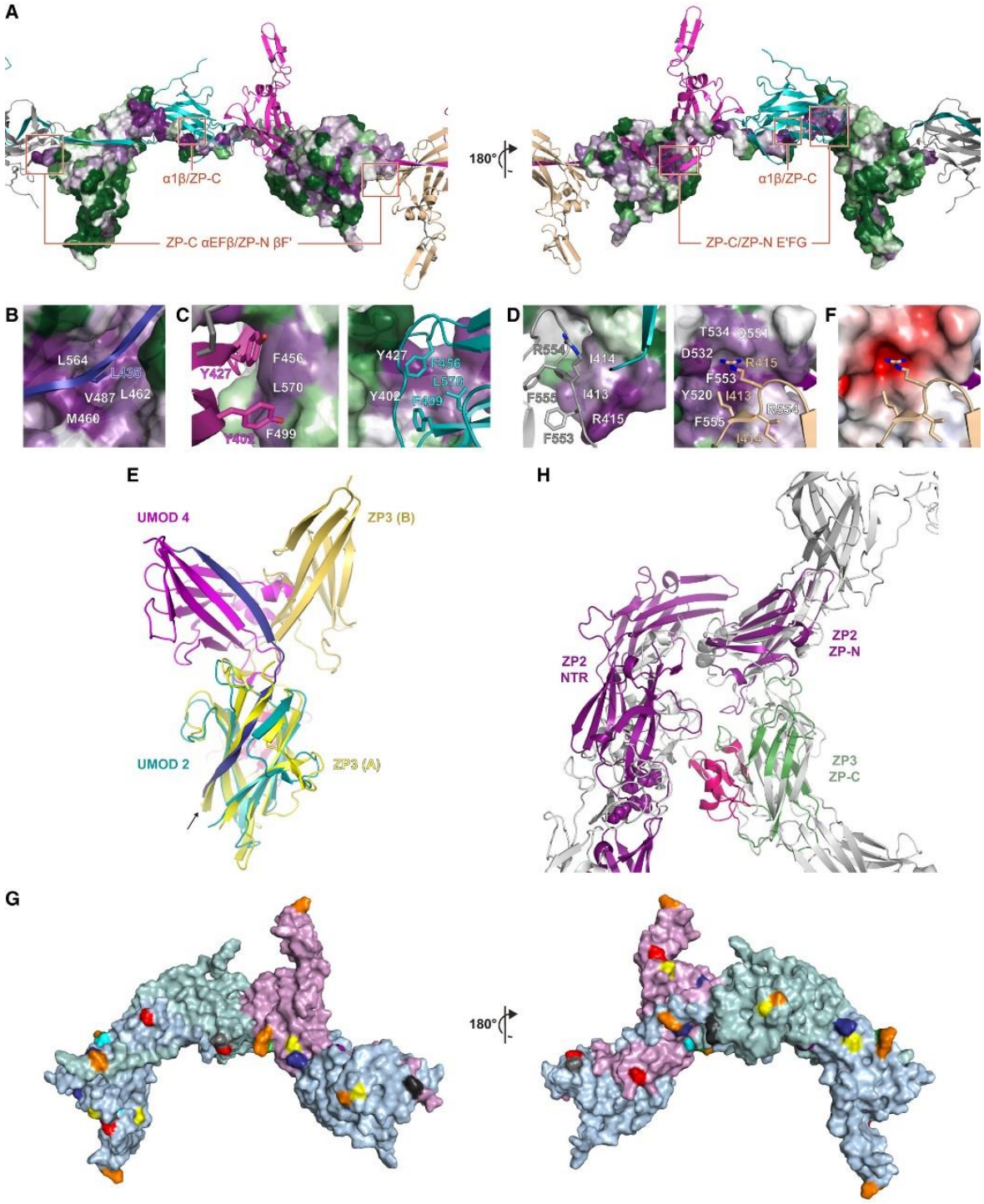

Figure 3.3. 7. UMOD inter-molecule interaction analysis and comparison of cZP3 dimeric precursor with UMOD filament. Highlight the conserved features among ZP domain protein. (A) UMOD-3 is represented by its molecule surface, and marking resides with color from green to violet based on the ConSurf. Figure 3.3.4 shows the other subunits within the filament in cartoon form, with the interaction region in the B-D section represented as a red rectangle. (B)

The linker between ZP-N and ZP-C of a UMOD subunit forms interface with the ZP-C domain of precedes UMOD subunit. Important residues that play critical roles in the interface is labeled in white. (C) A UMOD ZP-C domain forms a closed-up interface with the following UMOD subunit E'FG extension, which is located at the ZP-N domain. (D) UMOD subunits that are involved in the ZP-C αEFβ/ZP-N βF' interface are placed in exciting manners. That is to say, this interface is achieved through UMOD-(n) subunit and UMOD-(n+2) subunit. (E) Structure alignment analysis of the polymeric UMOD ZP-C domain (UMOD-2 of Figure 3.3.4; labeled with teal) and monomer of the cZP3 dimer precursor (PDB 3NK3 A chain; labeled with yellow, and ZP3-specific subdomain is labeled with hot pink). The ZP-N counterparts (colored with magenta and orange-yellow, individually) interacting with these structural domains are placed in different orientations relative to their ZP-C, respectively, although the two interaction interfaces are formed through nearly identical elements. The arrows point out how the overlapped ZP-C structural domain brings the ZP3 EHP to the same position as α1β located at the inter-domain linker of the UMOD monomer that follows UMOD-2 (UMOD-3, which is labeled with blue). (F) The identical side view of αEFβ/ZP-N βF' interface located at ZP-C demonstrate on panel D right half. Moreover, surface of ZP-C is colored based on coulombic distribution; the color gradient is based on the mapping; red (−5 kT/e), blue (+ 5 kT/e) and white (0 kT/e). (G) The ZP domain interaction face that involves in UMOD filamentous complex formation agrees with the expected N-glycosylation and O-glycosylation sites located at other ZP domain proteins, which is exposing to solvent. Based on sequence and structure alignment analysis, the modeled N-glycosylation sites of human glycoprotein-2 (labeled with orange), human α-carrier and β-carrier proteins (labeled with yellow and blue, individually), human ZP2 (labeled with green), human ZP3 (labeled with gray), human ZP4 (labeled with cyan), chicken ZPD (labeled with purple), and chicken ZP3 O-glycosylation site-1 (labeled black) were projected to the surfaces of three continue UMOD

subunit (subunit A is labeled with light blue; subunit B is labeled with light teal; subunit C is labeled with light magenta). Furthermore, The UMOD N-glycosylation site is specially labeled with red. (H) The homologous protein coordinates of NTR (N-terminal repeat region) of mouse ZP2 and its ZP-N domain and mouse ZP3 ZP-C domain (labeled with green) are respectively aligned with on the UMOD ZP-N and ZP-C domains of two adjacent subunits in UMOD filamentous complex, which is labeled with gray. Furthermore, the ZP2 NTR is placed in the same direction as the one used for alignment between the UMOD N-terminal branch and its ZP-N/ZP-N within the same UMOD subunit. The alignment modeling results show that, similar to UMOD N-terminus region (EGF I, II, III and D8C), the ZP-N domain repeats region preceding the ZP2 ZP module can protrude from the core of the eggshell filamentous complex without disturbing the oligomerization interface of the subunit. In the same way, the ZP3-specific C-terminus region (labeled with hot pink) is modeled to localize laterally to the egg coat filamentous complex body and may face to the ovastacin digestion site in the 2nd ZP2 N-terminal repeat, which is spheres.

3.3.3 UMOD polymerization relies on multiple domain-domain interactions

Because of the significant conformation change happens on this inter-domain linker, UMOD ZP-C domain from one particular subunit (e.g., UMOD-2) is exposure to interact with two ZP-N domains from adjacent UMOD subunit (UMOD-3 and UMOD-4) in a distinct way (Figure 3.3.4A).

In the first kind of contact, hydrophobic amino acids on ZP-C (UMOD-2), especially conserved F499 of βC and L570 of βF, interact with amino acids located at E'FG

extension of ZP-N (UMOD-3). These residues include conserved L393, the featured Tyr (Y402) on the ZP-N domain, and the other near-invariant Tyr (Y427) located at the E'FG extension that stacks with highly conserved F456 on the ZP-C IHP (Figure 3.3.5B – D and Figure 3.3.7A and C). Different from other Ig-like domains (Immunoglobulin-like Domain), ZP-N owns its featured E'FG extension, which is also highly conserved (Monne, Han et al. 2008); together with α1 strand and the IHP, the signature Tyr work as a critical component in ZP domain oligomerization (Jovine, Qi et al. 2004, Monne, Han et al. 2008, Schaeffer, Santambrogio et al. 2009).

As for the second kind of contact, which happens around the C-terminal region of α1β strand that inserts into the ZP-C domain, it involves the latter ef loop and fg loop on the ZP-N domain, which belongs to the second to next UMOD subunit (e.g., UMOD 2 ef/UMOD 4 fg) (Figure 3.3.4A; Figure 3.3.5B and E). Notably, ef loop of the ZP-C domain contains a highly conserved FXF motif in the ZP domain (553-FRF-555; Figure 3.3.5C; Figure 3.3.7A and D). In a previous study, the homodimeric structure of cZP3 (chicken ZP3) precursor requires the FXF motif to stabilize the conformation through binding ZP-N fg loop of another molecular. This leads to the formation of a β-sheet, working as an intermolecular linker. It involves the FXF motif itself and ZP3s' hydrophobic sequence located on the fg loop (139-VII-141) of ZP-N (Han, Monne et al. 2010). Resulting from the fg loop flexibility and different oligomerization states, UMOD and ZP3 own different relative domain orientations. However, the UMOD ZP-C FXF motif, which is an α-helix (αEF) in the protein precursor, generates a similar interface, through a short β-strand (αEFβ). This equivalent interface pairs with a hydrophobic sequence of the ZP-N fg loop βF' (412-III-414) (Figure 3.3.5B). Highly conserved R415 follows it, which also sticks into the negatively charged pocked

formed through Y520, D532, T534, and Q551 of ZP-C (Figure 3.3.2J; Figure 3.3.5E; Figure 3.3.7A, D, and F). UMOD R415 is the corresponding conserved residue of ZP3 ZP-N R142, which binds to ZP-C D254 and Y243 of ZP3 (corresponding to UMOD Y532 and D520 respectively) and is necessary for cZP3 homodimerization and secretion (Han, Monne et al. 2010). Notably, in the precursor form UMOD, the FXF motif helical conformation is stabilized through the EHP, which immediately gives insight into how hepsin-dependent dissociation of EHP can promote the conformation change that stimulates ZP-C to interact with ZP-N of another molecule (Figure 3.3.5B).

3.3.4 UMOD oligomerization involves a head-to-tail mechanism

In order to supplement our structural data and functionally study the molecular mechanism of filament assembly, firstly we expressed UMODfl (full-length) construct, which carries mutations on interface of ZP-N fg loop and ZP-C ef loop, in MDCK (Madin-Darby Canine Kidney) cells, which conduct UMOD secretion and polymerization (Schaeffer, Santambrogio et al. 2009). Consistent with our above structural study, single mutation of ZP-N R415A or a two-residue deletion (ΔF555-A556, ΔFA), which affect the ZP-C FXF motif, although it does not alter UMOD expression and secretion, completely abolish its oligomerization (Figure 3.3.8A; Figure 3.3.9A and B). Notably, co-transfection experiments show that neither mutant affect wild-type UMOD (wt) secretion (Figure 3.3.9C); however, ΔFA (but not R415A) has a significantly negative effect on its capacity to assemble filamentous UMOD (Figure 3.3.8B–D). A similar result was observed when wild-type UMOD is co-expressed with a protein variant (4A variant), which cannot polymerize due to its CCS stays in an inactive state (Schaeffer, Santambrogio et al. 2009) (Figure 3.3.8E; Figure 3.3.9C).

With the phenomenon that, although ΔFA and 4A mutations dominant-negatively affect filament formation, their effect is suppressed in double mutants R415A/ΔFA and R415A/4A (Figure 3.3.8F and G). We conclude that UMOD filament is polarized, and its oligomerization relies on the recruitment from an activated ZP-C domain to the ZP-N end of the incoming UMOD subunit, which agrees with the structural information (Figure 3.3.4)

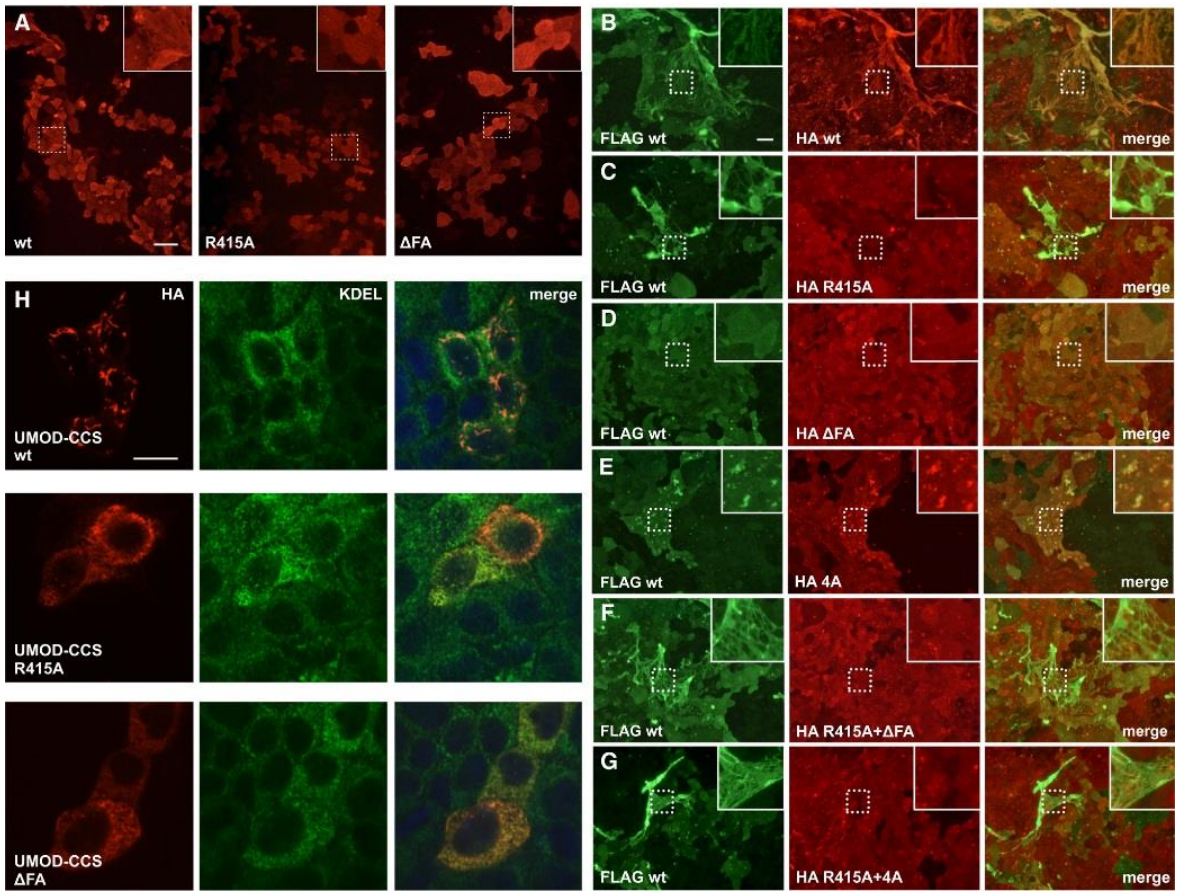

Figure 3.3. 8. Critical residue located at oligomerization interface and effect of their mutations on UMOD filamentous complex formation. (A) Immunofluorescence imaging of MDCK cell, which is unpermeabilized but stably transfected with these mutants. Compared with the UMODfl construct, ZP-C (ΔFA) or ZP-N (R415A) mutants on UMOD lead to the unsuccessful assembly of UMOD filament. Scale bar: 50 μm. (B–G) Immunofluorescence imaging of cells which is unpermeabilized co-express UMOD (tagged with FLAG, shown with green) and

desirable HA-labeled isoforms (shown with red). The results demonstrate that UMOD mutation R415A does not involve oligomers that only own wild-type protein (C). The highly negative effect of UMOD ΔFA on wild-type UMOD oligomerization (D), however double mutation located on ZP-N and ZP-C construct is able to rescue polymerization capacity (F). Identically, although mutation (4A) on CCS is also lead to an extremely negative effect on the dissociation of EHP (E), R415A mutation located at the 4A isoform is able to suppress this effect (G). Scale bar: 50 µm. (H) Immunofluorescence imaging of cells which is permeabilized and express UMOD-CCS a soluble isoform which is taken segment before EHP, demonstrates that wild-type forms intra-cellular oligomers, however constructs with mutations on interface do not. And these intra-cellular oligomers locate in ER (endoplasmic reticulum), as demonstrated through co-staining with the KDEL sequence which is used as an endoplasmic reticulum marker. Scale bar: 10 µm.

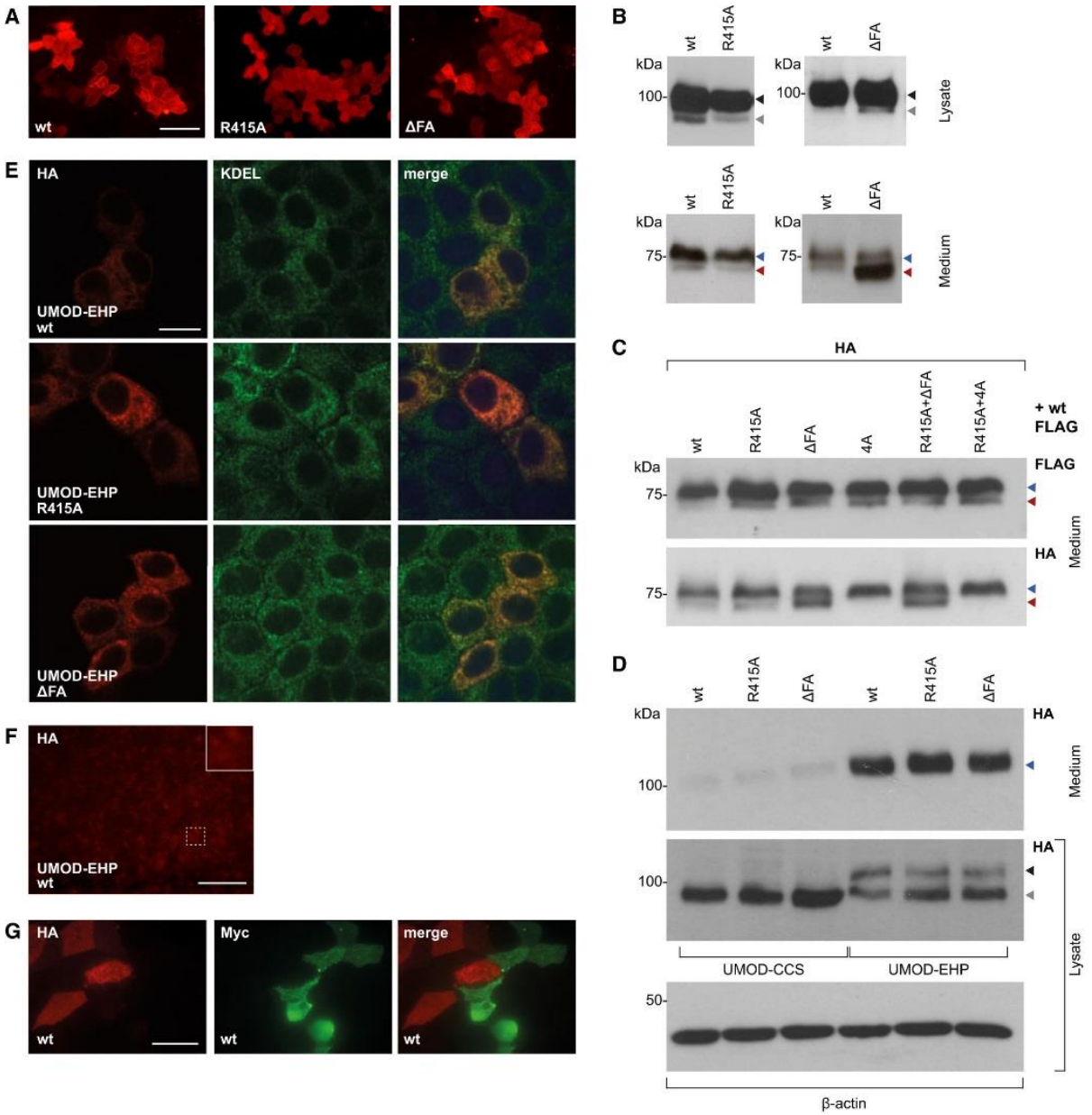

Figure 3.3. 9. Mutagenesis analysis for the effect of UMOD mutants on UMDO protein expression, secretion, and filamentous complex assembly. (A) Immunofluorescence imaging of MDCK cell which is unpermeabilized and express wild-type UMODfl, R415A mutation construct and ΔFA construct. No intra-cellular polymerization phenomenon is captured. Scale bar: 50 μm. (B) MDCK cells used in this western blot assay, which express UMODfl construct with mutation of the R415A or ΔFA. Western blot result of total cell lysates (in the top part) shows that both wild-type UMOD and constructs with the polymeric mutation are predominantly present as completely glycosylated isoforms (higher part, label with black

arrows). Furthermore, there exist ER glycosylated samples (lower part, labeled with gray arrows). Western blot assay analysis of the PNGase F deglycosylated protein, secreted through MDCK cells (bottom part), demonstrates that constructs that own ether R415A mutation or ΔFA mutation do not influence UMOD secretion in culture media. Blue arrows point to protein digested inside the joint membrane part among EHP and GPI. Red arrows point to proteins that process at CCS region. Results above indicate comparable intra-cellular protein trafficking and secretion of wild-type and isoforms with mutations. Notably, ΔFA mutation promotes the number of processed proteins processed at CCS, suggesting that altered FXF motif influences accessibility close to digestion site (Figure 3.3.5B, left part). However, filamentous complex is totally absent with this mutation (Figure 3.3.8A, right part), which then suggests the specific influence of the mutation for UMOD oligomerization. (C) Western blot analysis for deglycosylated proteins generated in the MDCK cells culture media stably co-express wilt-type UMOD (tagged with FLAG) and wilt-type or isoforms with mutations. Co-expressed with UMOD construct with mutation does not affect the digestion of wild-type protein (tagged with FLAG), indicating that the highly negative influence of 4A and 4A mutants does not derive from unusual processing of wilt-type UMOD. (D) Western blot analysis of UMOD in cell lysis solution and culture media from MDCK cells, which transiently transfected with suggested isoforms (tagged with HA). The process protein existing from the endoplasmic reticulum requires EHP motif, which is indicated by comparing the intra-cellular levels and secretion between UMOD-EHP and UMOD-CCS. (E) MDCK cells are transiently transfected and express UMOD-EHP construct with R415A located on ZP-N and ΔFA located on ZP-C mutations. Immunofluorescence imaging from the permeabilized cells demonstrates that intra-cellar oligomers are absent for both wild-type and mutant construct. Scale bar: 10 μm. (F) Immunofluorescence imaging results for unpermeabilized MDCK cells, which express UMOD-EHP. Without a membrane-anchoring point, UMOD localization and oligomerization are

affected in the plasma membrane. Scale bar: 50 µm. (G) MDCK cells, which stable express wild-type UMOD (tagged with HA, red) or UMOD (tagged with Myc, green), are co-cultured. Filamentous complex observed exhibit uniformly colored status, indicating that oligomerization relies on the participating membrane-bound subunit rather than digested subunit that is relieved in the culture media. Scale bar: 25 µm.

Another set of experiments was performed, which uses truncated constructs after UMOD-CCS and UMOD-EHP. The result shows that, compared with UMOD-CCS that is not secreted but exhibit intracellular oligomer-like structures (Schaeffer, Santambrogio et al. 2009), ΔFA or UMOD-CCS R415A are impaired in both polymerization and secretion. From another aspect, wild-type UMOD EHP and its ΔFA or R415A variant can be secreted normally and do not exhibit intracellular oligomers (Figure 3.3.8H; Figure 3.3.9D and E). Remarkably, these findings further validate the functional importance of the ZP-C motif and ZP-N fg loop. At the same time, they reveal that, although the interdomain linker α1β replaces the EHP part entirely in the context of UMOD filament (Figure 3.3.5B), the EHP plays a critical role in UMOD secretion even in the presence of mutations, which recovers polymerization. UMOD-EHP cannot polymerize extracellularly due to the absence of CCS cleavage (Figure 3.3.9D and F), and wild-type UMOD can only incorporate filaments attached to the secretion cell (Figure 3.3.9G). Besides, these data also support the hypothesis that EHP dissociation and filament growing with head-to-tail incorporation are coupled processes, which happened at the plasma membrane.

3.3.5 The architecture of UMOD homogenous polymer is similar to that of heteropolymer egg coat filamentous complex

As homologous proteins, UMOD filament is very similar to those of other ZP domain proteins; UMOD polymer structure is able to give insight into how others ZP domain proteins assemble. Mouse egg ZP filament comprises heterodimers ZP2 and ZP3 subunits with around 140 Å structural repeat (Wassarman and Mortillo 1991), which is very similar to the helical pitch of UMODfl filament (Figure 3.3.2F). Moreover, the UMOD filament structure agrees with the expected solvent exposure of the many glycosylation sites, which is diversely distributed among the other ZP domains, including ZP2 and ZP3 (Figure 3.3.7G). Very similar, the superposition, which is composed of ZP-N and ZP-C, suggests that although the C-terminal subdomain of ZP3 and the repeat region of ZP2 are very important for sperm binding (Wassarman and Litscher 2018), they are not necessary for ZP filament growing (Jovine, Qi et al. 2002). This is compatible with the structural basis of UMOD filament (Figure 3.3.7H).

In order to gain more information on how heteromeric ZP domain protein forms filament, soluble native complexes are prepared through digesting the UFE (unfertilized egg coats) of medaka fish with LCE and HCE (low and high choriolytic hatching enzymes) (Yasumasu, Kawaguchi et al. 2010). SDS–PAGE and MS (mass spectrometry) analysis of the two fractions of the sample separated through SEC (size-exclusion chromatography) (F1-2; Figure 3.3.10A) demonstrate that each of the part contains polypeptide with a size of approximately 37 kDa encompassing the ZI-3 ZP module, and fragments of size 16 kDa and 18 kDa, which correspond to the ZI-1 ZP-N and ZI-2 ZP-C (the corresponding subunit of ZP2 in the fish egg coat) (Figure 3.3.10B, D and E). In agreement with their distinct native-PAGE profiles (Figure

3.3.10C), the native mass spectrum of chemically cross-linked F1 and F2 shows that the latter is a heterogeneous complex including three aforementioned species (Figure 3.3.10F), and as for the former, it contains ZI-3 monomers, dimers, and trimers bound to ZI-1,2 ZP-N/ZP-C with one or two separate copies (Figure 3.3.10G). This result agrees with a previous study that LCE is able to cleave the interdomain between the ZP domain of ZI-1,2 (however not that of ZI-3) (Yasumasu, Kawaguchi et al. 2010), creating a site that corresponds to the short linker that connects UMOD α1β and β1 (Figure 3.3.10H). Moreover, it is also consistent with MALS (SEC-multi angle light scattering) results that precursor of ZI-1,2 and ZI-3 exhibit monomeric and monomeric/dimeric forms, respectively. Another finding shows that the polymeric core of medaka UFE comprises units with interspace around 65 Å (Figure 3.3.10I). Take these finding together; all evidence strongly suggest that filament of vertebrate egg coat should share the same overall architecture with UMOD.

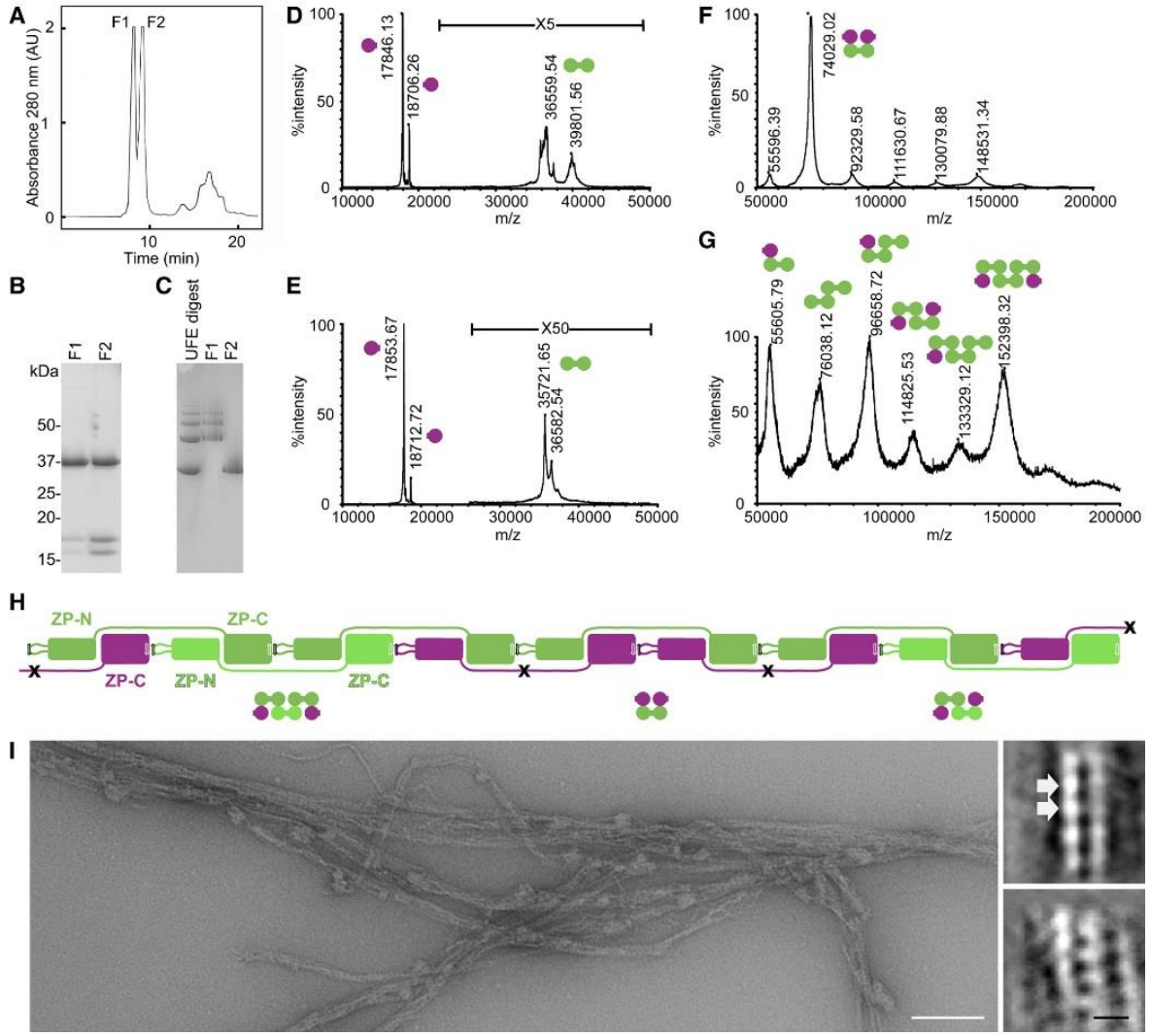

Figure 3.3. 10. UMOD filament shares similar architecture with that of egg coat filamentous complex. (A) Size-exclusion chromatography analysis for unfertilized egg coats, processed with high and low choriolytic hatching enzymes, generates two peaks. (B, C) Analysis of reducing SDS-PAGE (B) and native gel (C) of unfertilized egg coats purified with size-exclusion chromatography, along with molecular weight marker. The difference between size-exclusion chromatography elution volumes for F1 and F2 suggests the distinct digestion levels for egg coat filamentous complex by choriolytic hatching enzymes. (D, E) Time of Flight mass spectrum spectrograms analysis for samples from F1 (E) and F2 (D). Spherical dot refers to ZI-3 ZP module (labeled with green), which contains two moieties incomplete form and ZI-1 ZP-N domain and ZI-2 ZP-C domain (labeled with purple), which is processed with low

choriolytic hatching enzymes. (F, G) Time of Flight mass spectrum spectrograms analysis for cross-linked sample of from F1 (G) and F2 (F). The spherical dot here represents the modeled architecture of cross-linked samples. (H) Simple illustration for architecture for medaka egg coat filamentous complex and its interaction model. ZI-1 and ZI-2 (labeled with purple) are in monomer form before activation and can participate in filamentous complex assembly when activated. In addition, as for ZI-3, it owns diverse polymers state. Both the ZI-3 dimer (labeled with green) and ZI-3 monomer (labeled with light green) is able to participate in filament formation. After digestion processing of low choriolytic hatching enzymes (cleavage site is between ZI-1 and ZI2 inter-domain linker, pointed out with black crosses), the filamentous complex digested to soluble species, which correspond for various oligomers that are identified in mass spectrum spectrograms (F) and mass spectrum spectrograms (G). (I) Imaging of medaka egg coat filamentous complex with negative staining electron microscopy. The right part demonstrates the best 2D average of the sample, which suggests an inter pitch distance of around 65 Å in the complex—scale bars: 100 nm for the left panel and 10 nm for the bottom right panel.

3.3.6 Supramolecular organization of UMOD filament sheets

When incubating purified UMOD with NaCl solution, which has a similar concentration with that of urine, sheet-like complex forms (Figure 3.3.11A, top) that is similar to what shows in Figure 3.3.2A. An antiparallel filament arrangement was suggested from the negative staining sample, which exhibits a two-fold axis perpendicular against the plane of the sheets (Figure 3.3.11A, bottom). UMOD sheet matrix model (Figure 3.3.11B) can be built by combining UMODfl (Figure 3.3.6D) and the prior sheet conformation information. This sheet model explains the unique double-helical patterns of the native protein (Figure 3.3.2A, inset). Furthermore, the location of the

N-glycans of the ZP domain is able to explain the diffuse density shown in the filament interspace (Figure 3.3.11C).

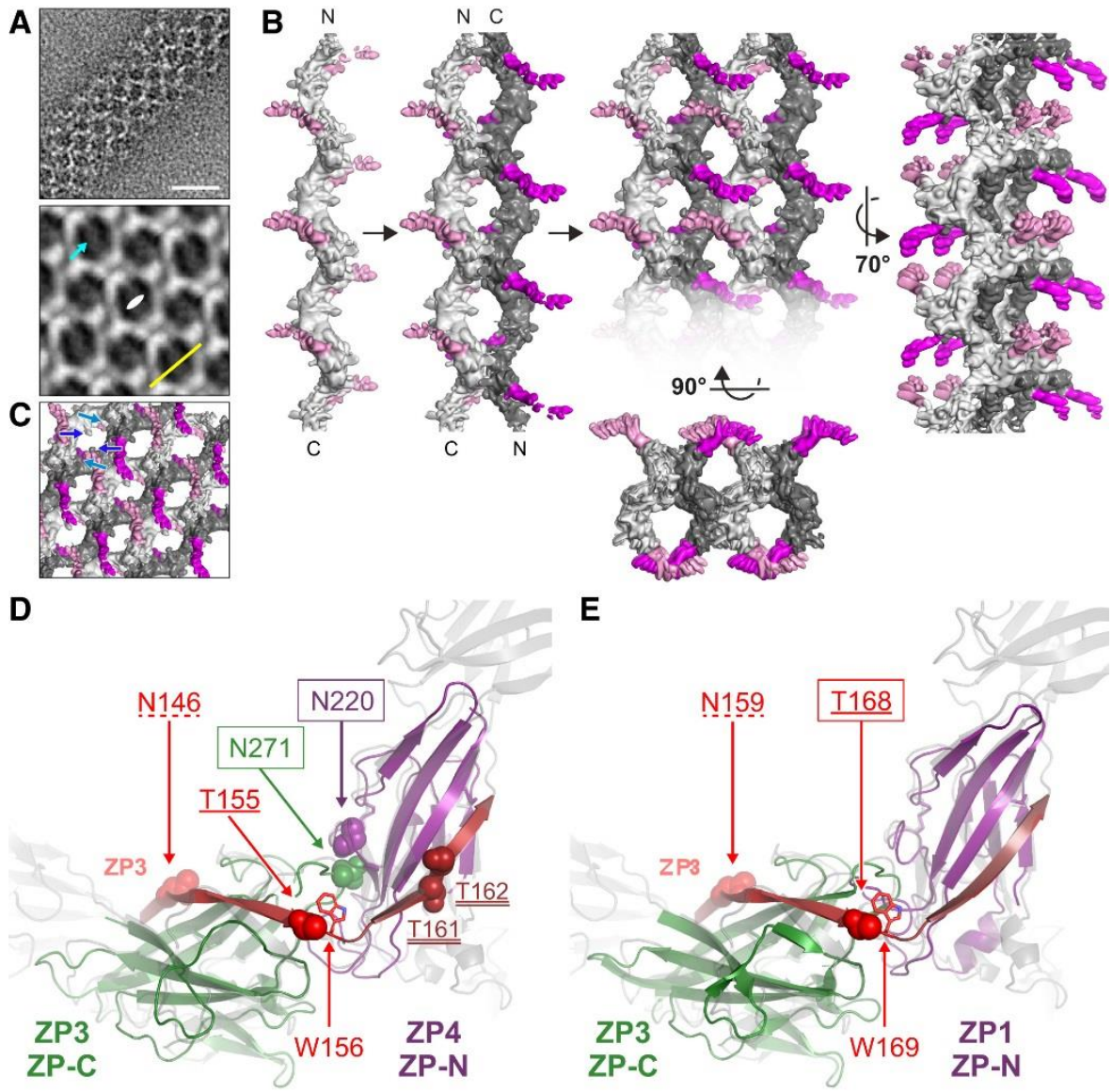

Figure 3.3. 11. The illustration for a predicted model of sheet structure composed of UMOD and ZP domain filamentous complex binding sites for sperm and uropathogenic E.coli. (A) Imaging of sheet matrix composed of purified UMODfl filamentous complex in 100 mM NaCl buffer, with negative staining electron microscopy (top panel). Alignment and averaging analysis for the sample suggests that a repeating feature, which owns double-fold symmetry axis upright to the sheet plane (labeled with white symbol). Scale bars: 250 Å for the top part

and 125 Å for the bottom part. (B) Sheet architecture model that represents the matrix structure of the sample in (A). The sheet structure is formed through antiparallel packing of multiple UMODfl merged density map; alignment is guided by the pattern shown in (A). Lacking of direction information in the space right up to the sheets plane, the precise relative position of two adjacent filamentous complexes is hard to model. Consequently, the predicted model is based on the hypothesis that the central regon of filaments compose the matrix on an identical plane. Distinct filamentous complex is then labeled with different colors; light gray and dark gray. Moreover, the N-terminus edge region is labeled with pink and magenta individually. The double-helix-like view of the projection from the sheet matrix can be explained by the zig-zag side view of two antiparallel UMOD filament (see inset of Figure 3.3.2A), and the extending edge of UMOD forms a binding surface to capture uropathogenic E.coli. (C) Side-by-side comparison of 2D average pattern and the UMOD sheet matrix modeled in (B), and the results reveal that the smeared density that extending into the space generated from two close filaments (labeled with cyan arrow) is generated by the glycan(s) of UMOD N396 and/or UMOD N513 residues (labeled with light and dark blue arrows, individually). (D) Protein structural detail of porcine ZP filamentous complex, modeled with superposition from homology structure coordinates of neighboring ZP3 subunit and ZP4 subunit (labeled with green and purple, individually) to the model of UMODfl (labeled with semi-transparent gray). The corresponding residues of the ZP3 inter-domain linker onto the ZP-N/ZP-C linker of UMOD (labeled with red). Glycosylation residue sites, which are experimentally verified, are demonstrated in the model with cartoon. It shows the close distance between N271 of ZP3 and N220 of ZP4, whose N- glycans regulate porcine sperm specific binding. Notably, these amino acids are very close to conserved ZP3 N-glycosylation (labeled with dashed underline) and C-glycosylation (site-1 is labeled with single underline and site-2 is labeled with a double underline) located at inter-domain linker. The conserved Trp of ZP3, which follows site-1, is

demonstrated with stick at the ZP-3/ZP-4 interface. (E) Structural details of interface involved in ZP-3 and ZP-1, which compose avian egg coat filamentous complex, polymerized and demonstrated the same as described in (D). The rectangle emphasizes the position of site-1 O-glycosylation, which is of vital importance for chicken sperm and ZP3 binding.

3.3.7 Near atomic resolution structure of native UMOD filament

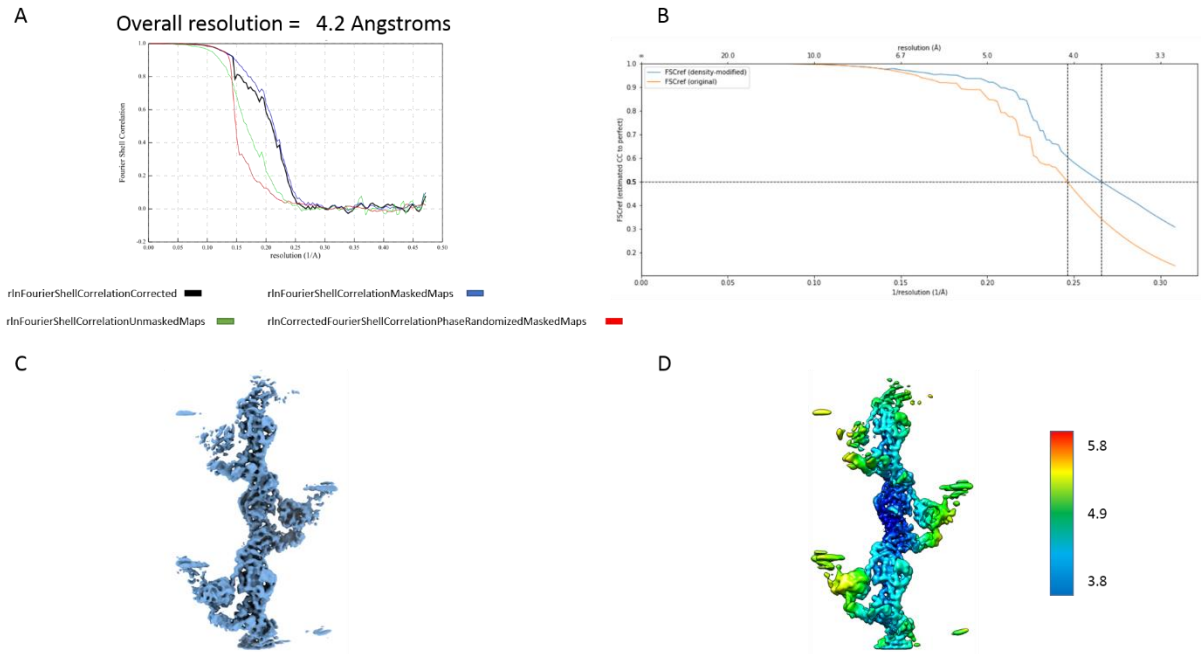

Figure 3.3. 12. (A) FSC plot of half-maps. (B) Density Modification plot with Phenix ResolveCryoEM. (C) Polished UMOD filament. (D) local resolution distribution of UMOD filament Native UMOD filament structure is solved with resolution 4.2 Å. From the local resolution heat map, we can see that the central part resolution is much higher than the edge part.

Table 3.3. 1. Data collection, image processing, and coordinates refinement statistic for UMOD

Data Collection	
Microscope	Titan Krios
Voltage (kV)	300
Detector	K2
Pixel size (Å)	1.06
Defocus range (μm)	-1.5 to -3.5
Electron Dose (e-/A2)	39.6
Movie Frames	40
Helical Reconstruction	
Software	RELION 3.0.4
Particles	288403
Helical rise (Å)	62.5
Helical rotation (°)	180
Resolution (Å)	4.2
Coordinate Refinement	
Software	Phenix
Resolution Cut Off	3
Model	
Oligomer	center dimer
Number of residues	593
B-factor overall	115.88
R.M.S. deviation	
Bond length (Å)	0.003
Bond angle (°)	0.547
Validation	
Molprobity clashscore	4.16
Rotamer Outliers (%)	0
Cβ deviation (%)	0
Ramachandran plot	
Fovored (%)	96.9
Allowed (%)	3.1
Outliers (%)	0
Deposit	
PDB code	6TQK
EMDB code	10553

3.3.8 Discussion and Summary

As the first detailed native structure of ZP domain protein, UMOD filament reveals this unique architecture, whose features were not reported in previously reported biological oligomers. In particular, this arrangement of UMOD subunit reported in this study

clearly demonstrates that distinct from the hypothesis proposed from the egg coat proteins (Jovine, Janssen et al. 2006, Egge, Muthusubramanian et al. 2015, Okumura, Sato et al. 2015, Louros, Chrysina et al. 2016), UMOD monomer neither assembles by contacts, which involve the ZP-C or ZP-N domain or polymerizes through forming an amyloid cross-β structure. On the contrary, an unprecedented linker, which achieves interlocked configuration of ZP-C and ZP-N domains, determents the basis of UMOD filament core. Significantly, our data also point out the fault of UMOD architecture suggested from a recent Cryo-ET analysis, where the filament is assembled through simply stacking of ZP domains in a zig-zag manner. Furthermore, due to the limited resolution, the subtle conformational change of the ZP-N/ZP-C linker and separation of ZP-C and ZP-N resulted from it cannot be observed from the Cryo-ET map (Weiss, Stanisich et al. 2020). On the other hand, another similar study of the UMOD core filament structure result (Stanisich, Zyla et al. 2020) is consistent with our high-resolution model.

Along with the mutagenesis and biochemical assay results, our Cryo-EM structure of UMOD filament has long-ranging implications to different aspects of ZP domain matrix function and assembly mechanism function (Figure 3.3.11; Figure 3.3.13).

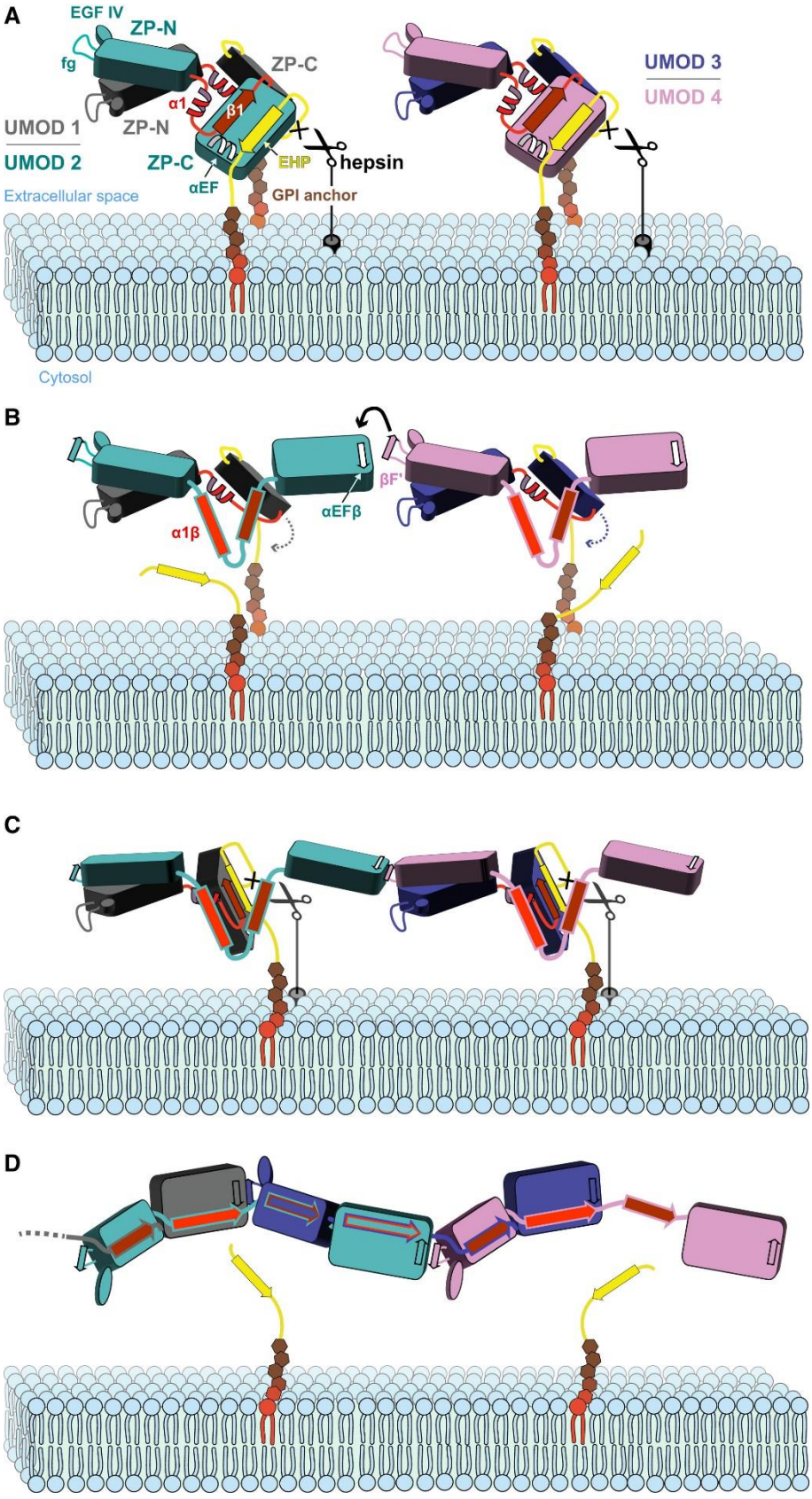

Figure 3.3. 13. The hypothesized mechanism of UMOD oligomerization. (A) Firstly, in the UMOD dimer, the GPI-anchored UMOD CCS sequence that is close to the membrane is

digested by hepsin, and then induces separation of EHP from the correspondent ZP-C domain. To simply the views, the flexible edge part of UMOD is not shown here. (B) EHP disassociation then triggers the ZP-C domain for oligomerization through promoting the formation of an inter-molecular β-sheet between EHP and the ZP-N domain of incoming homodimer UMOD. (C) Conformation change of the following ZP-C domain promotes the hepsin digestion, which relieves it from the membrane and participates in filament growing. (D) EHP displacement also triggers the transformation from the inter-domain linker precursor structure to mature inter-domain linker conformation. A significant rearrangement happened, including UMOD precursor UMOD precursor rearrangement initializing DSE to make up the absence of EHP. All of these steps complete filament assembly.

Benefiting from the unique interlocked architecture of UMOD filament, for each UMOD subunit, 25% of the surfaces are buried in the interface with four adjacent subunits. This structure achieves a remarkably resistance for chemical denaturing factors (Figure 3.3.1F) and ensures that UMOD filament is highly stable to proteolysis (Figure 3.3.1C–E). These properties, which derive from the great reinforcement of UMOD subunit–subunit interactions through semi-continuous wrapping of interdomain linkers surrounding the filament core (Figure 3.3.4), can be further strengthened through covalent interaction with other ZP domain proteins, as reported before, for example, ZP domain proteins in the egg coat from fish to human (Greve and Wassarman 1985, Yasumasu, Kawaguchi et al. 2010, Nishimura, Dioguardi et al. 2019) or insoluble cuticlin components from the nematode larval alae (Sapio, Hilliard et al. 2005). These considerations give insight on why ZP domain is evolutionarily selected for working as

protective matrices in a variety of organisms. With UMOD filament as a reference, further structural study of other proteins, which belong to this large protein family of mosaic extracellular molecules, will clarify how the interaction between a conserved ZP domain and N-terminal divers-functional domain mediates a diverse range of critical biological functions.

3.4 Fully automatic picking for filamentous sample

3.4.1 Naïve pipeline of direct Binarization for negative staining images

Images generated with electronic microscopy contain more noise compared with other biomedical images. Consequently, tools should be selected appropriately to deal with it. Thanks to the fast development of computer vision in past decades, many powerful image-processing libraries are available now. Among them, OpenCV is one the most popular and widely used libraries, which is powerful and easy to use. Our method is mainly implemented with OpenCV.

Although Cryo-EM dominates electronic microscopy structure determination, negative staining is also widely used in sample quality screening. A negative staining micrograph usually owns high contrast. Therefore, simple adaptive thresholding and morphing can separate samples out of the background. This simple picking pipeline can be described with four steps. Firstly, some preprocess denoising operation was done. In the second step, image adaptive local thresholding was processed. Here, based on contrast change, the outline of filament was detected. In the third step, noise

inside the object was removed through morphology erosion and dilution, which enhances the signal. The solid filament region was separated from the background. In the final step, the Coordinate of filament particles was represented by the central pixel of filament, which is generated by skeletonizing erosion (Figure 3.4.1).

Skeletonizing erosion algorism is shown below (Equation 2). A is the original image, and B is the structuring element, n indicates the iteration of erosion. N is the number of the last iteration when A is empty, And S here is the erosion function.

$$S_n(A) = (A \ominus nB) - (A \ominus nB) \circ B$$

$$S(A) = \bigcup_{n=0}^{N} S_n(A)$$

Equation 2. Formula of Skeletonizing

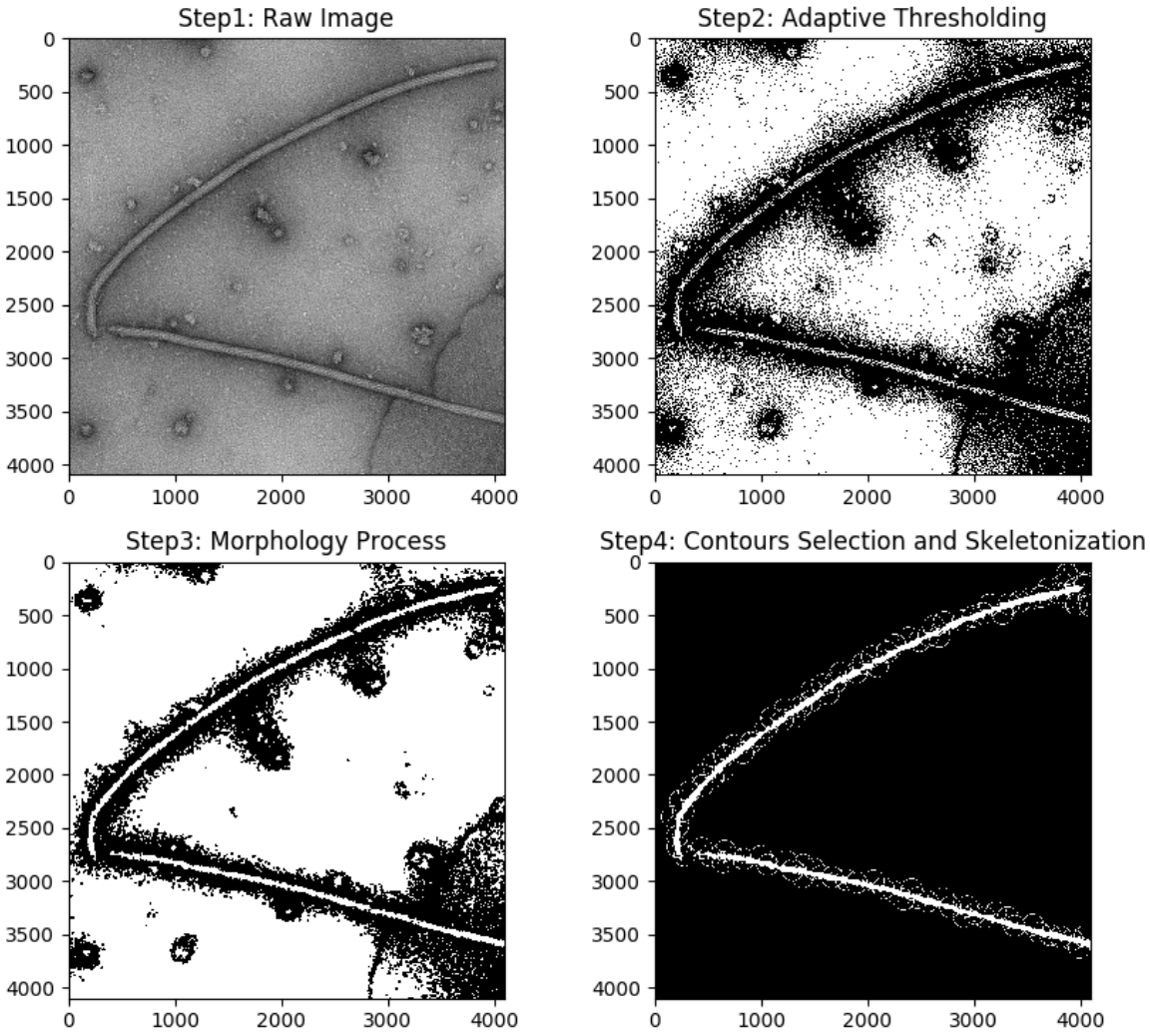

Figure 3.4. 1. Negative straining micrograph auto-picking pipeline illustration includes de-nosing, adaptive thresholding, morphology processing, and contours generation.

However, it is not wise to apply this pipeline on Cryo-EM micrographs directly. Image contrast of Cryo-EM micrographs is not as strong as negative staining. Identify high contrast region is more challenging than negative staining. In addition, high contrast contamination is common in Cryo-Grid preparation. As we can see, too much undesired region is picked with the simple pipeline (Figure 3.4.2).

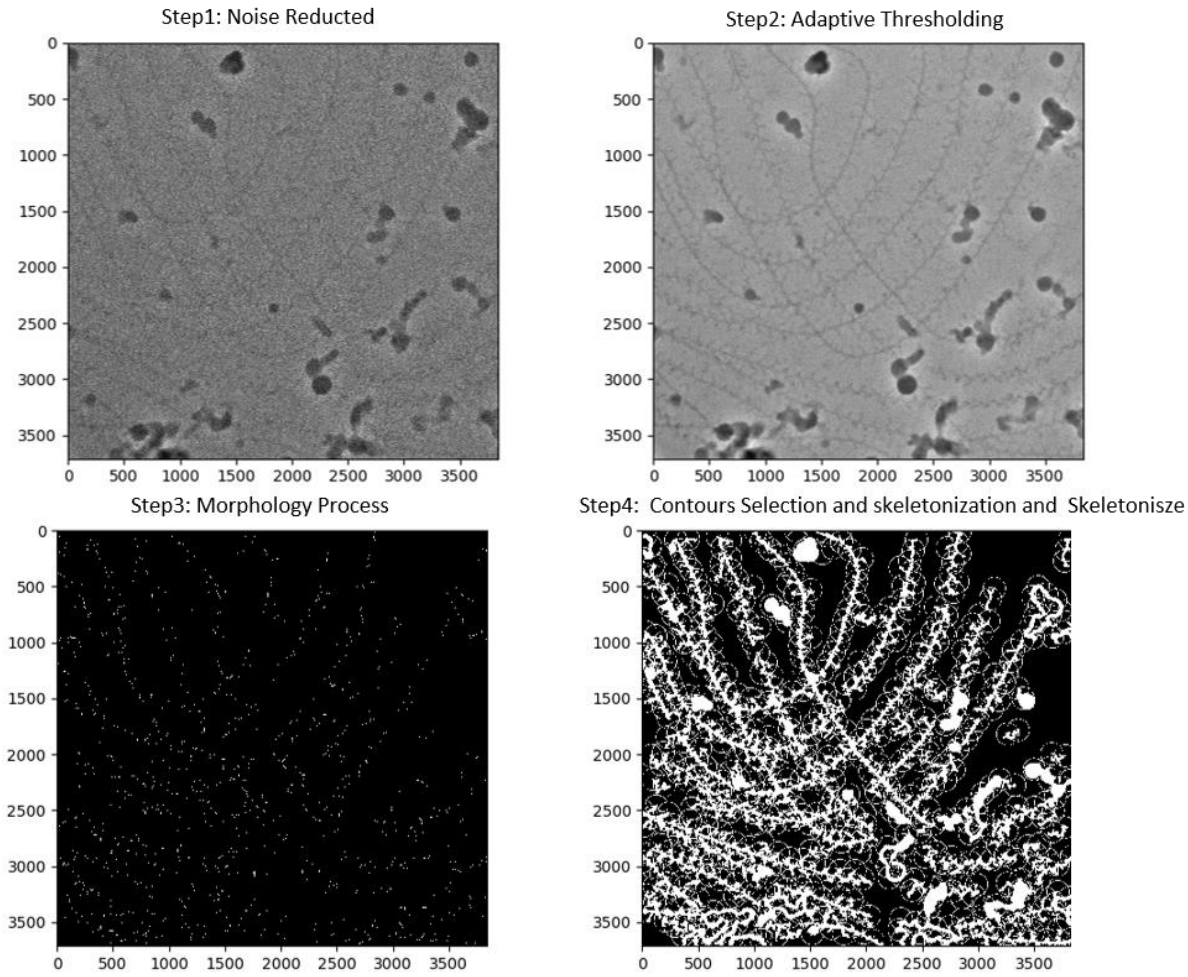

Figure 3.4. 2. Cryo-EM micrograph auto-picking pipeline illustration, which owns similar steps with negative staining image processing, however, exhibit the worse result

Besides, helical information like filament labels and Euler Angles is necessary for downstream helical reconstruction tasks. Euler Angles estimation is of vital importance during 3D classification and 3D alignment. As for single particles analysis, each particle distributes on Cryo-Grid independently, and the initial orientation is regarded as random. Single particle 3D reconstruction aims to find the correct orientation through iterative alignment. However, helical particles are not distributed independently. Particles derive from the same filament should be labeled, and their Euler Angles need to be calculated. During helical 3D reconstruction, the central axis

is fixed, requiring helical label and Euler Angles of each helical particle (Figure 3.4.3). This highlights the importance of helical particles' prior information.

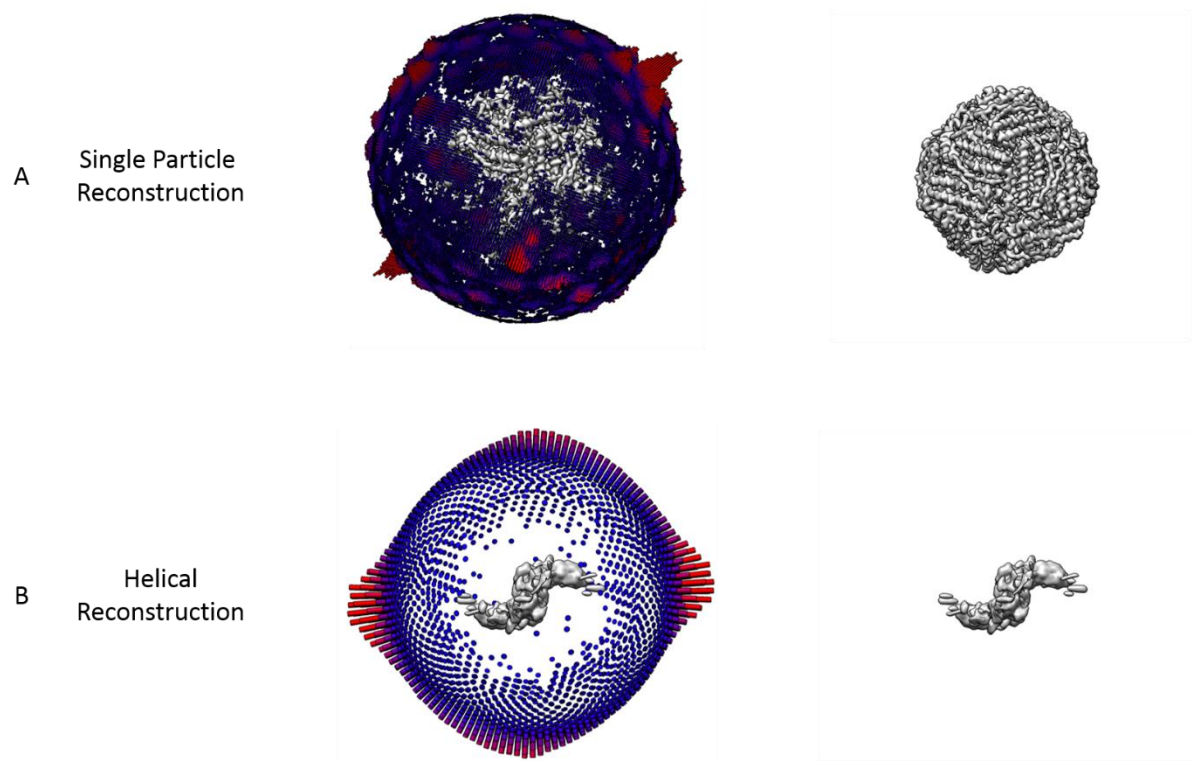

Figure 3.4. 3, orientation distribution difference between Single particle reconstruction and Helical reconstruction. (A) Apoferritin, (B) Uromodulin.

In summary, there are two disadvantages of the simple pipeline: introducing high contrast noise and lacking helical label information in coordinates. To decrease the influence of these factors and retain helical information, we propose a general-template matching method.

3.4.2 Feature-based with a general template

In order to separate filament more robustly, we design a general template, which follows below 2D Gaussian distribution to match filament in images (Equation 3).

$$\mathcal{N}(x;u,\Sigma) = \frac{1}{(2\pi)^{d/2}}|\Sigma|^{-1/2} exp\left\{-\frac{1}{2}(x-u)\Sigma^{-1}(x-u)^{T}\right\}$$

$$u = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 0.5 & 0.4995 \\ 0.4995 & 0.5 \end{pmatrix}$$

Equation 3. 2D Gaussian Distribution of the template

Some similar preprocessing operation is used in the first stage. We filter the image and shrink it to the proper size. With this preprocessing, contrast is enhanced, and receive field is increased. Moreover, the shrunk image increases efficiency, as the size of the raw micrograph is quite large, using a full-size picture is time-consuming.

In order to capture the location of filament, we need a metric to measure the confidence of each region on micrographs. Here, we use *TM_CCOEFF* algorism to measure the similarity between the template image and queried image (Equation 4). R (x, y) is the similarity between the template image and inquiry image at position (x, y). (x, y) are pixels coordinates in the template image. And (x', y') are pixels coordinates in inquiry image, w and h here are width and height of template image.

$$R(x,y) = \sum_{x'y'} \left(T'(x',y') \cdot I'(x+x', y+y')\right)$$

$$T'(x',y') = T(x',y') - 1 \Big/ (w \cdot h) \cdot \sum_{x''y''} T(x'',y'')$$

$$I'(x+x', y+y') = I(x+x', y+y') - 1 \Big/ (w \cdot h) \cdot \sum_{x''y''} I(x+x'', y+y'')$$

Equation 4. The formula of TM_CCOEFF algorism

Furthermore, a template with one single orientation is not enough to capture the full feature of filaments. There may be many filaments in one image, and their orientation varies. For some flexible filament, its orientation will change as it extends. Consequently, we need multiple rotated templates to track different orientated filaments. Considering the trade-off between efficiency and performance, we generate 36 rotated feature maps and aggregate all features through layer-wise Max-Pooling operation (Figure 3.4.4).

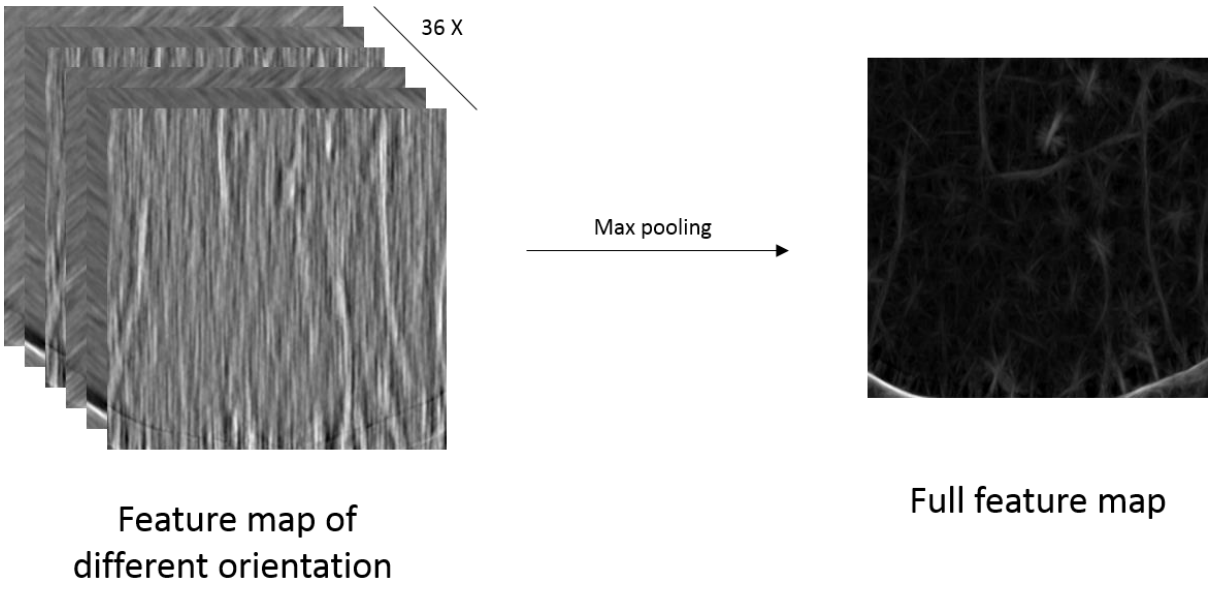

Figure 3.4. 4, illustration of multiple channel feature map and pooling operation to capture information from different orientations

The whole picture of our non-template picking method workflow is shown below (Figure 3.4.5). In summary, firstly, a filament image following Gaussian distribution is used as a generalized template. To capture orientations of all filament existing in images, rotation operation is applied on this template and generated. From there, we obtain 36 rotated filaments and a 36-layer feature map. Max pooling operation is applied on the 36-layer feature map to generate the final likelihood map. Based on the likelihood map, we apply adaptive binary thresholding. We will use this filament trace map in other filament fitting and coordinates generating steps.

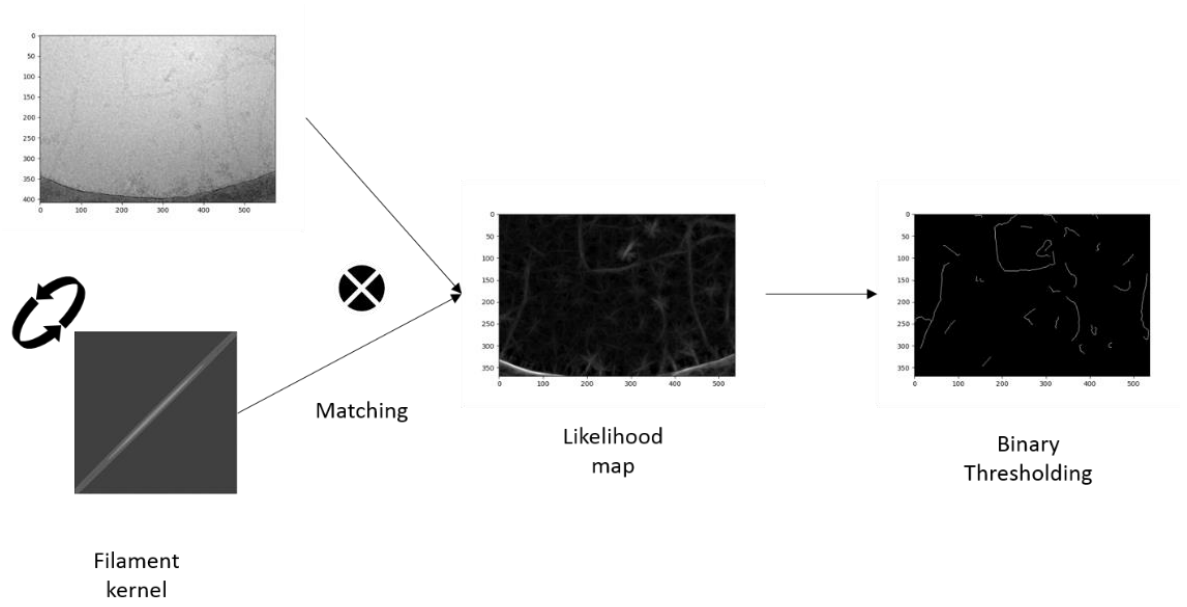

Figure 3.4. 5. The new workflow of filament contours identification, which relies on a filament kernel that follows Gaussian distribution.

In order to retain relative spatial information of helical segments, we design a strategy for filament fitting. For rigid and straight filament specimens, we only need to find its start/terminal ends. Then the line between these two ends can represent this filament. Filament ID and Euler angle between each helical segment can be generated through

helical segment extraction. As for flexible filament, the Euler angle between each helical segment may change as the orientation of filament changes. Therefore, we need to use multiple lines to fit one flexible filament; we hope these fitting lines are as long as possible and the number of lines is as few as possible with some constraints. Finally, we will get multiple lines to represent the flexible filament.

With end-to-end coordinates, the helical segment can be extracted from the representative line with customized interval distance. This distance is determined by the helical rise and step size of the filament, which the user should input. The below figure demonstrates the performance of our method (Figure 3.4.6). Our method shows robust performance for both flexible samples and rigid samples, which achieve high precision and recall ratio.

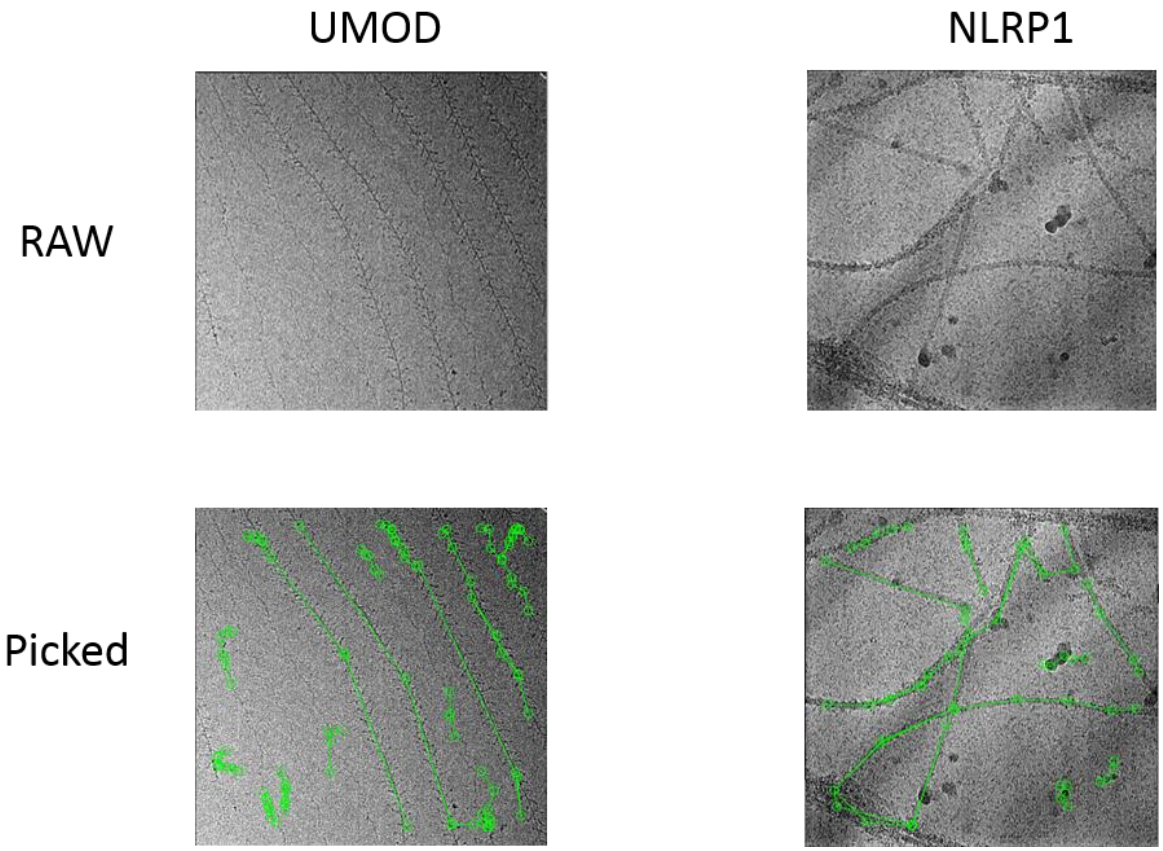

Figure 3.4. 6, Auto-picking performance illustration of non-mask filament picking method, on different samples.

However, the high contrast contamination problem is not solved entirely so far. Indeed, we can set some feature-based rules to separate contaminations from real sample. For example, when we get proposed contours from the likelihood map. We can filter contours based on some features, which are predefined. For example, the area/perimeter ratio is a good indicator for filament-contamination discrimination. We can set a threshold to filter contamination-like contours. However, this strategy still cannot remove contamination with other shapes. As contaminations across Cryo-EM images are pretty similar. A deep learning based method is a good choice for contamination labeling.

Pixel-level labeling usually refers to semantic segmentation, which is dominated by fully convolutional networks (FCN). Compared with simple discrimination convolutional networks, FCN replaces the fully-connected layer with the reverse convolution layer, which serves up sampling function. U-net belongs to FCN and is a popular architecture for biological images segmentation (Figure 3.4.7).

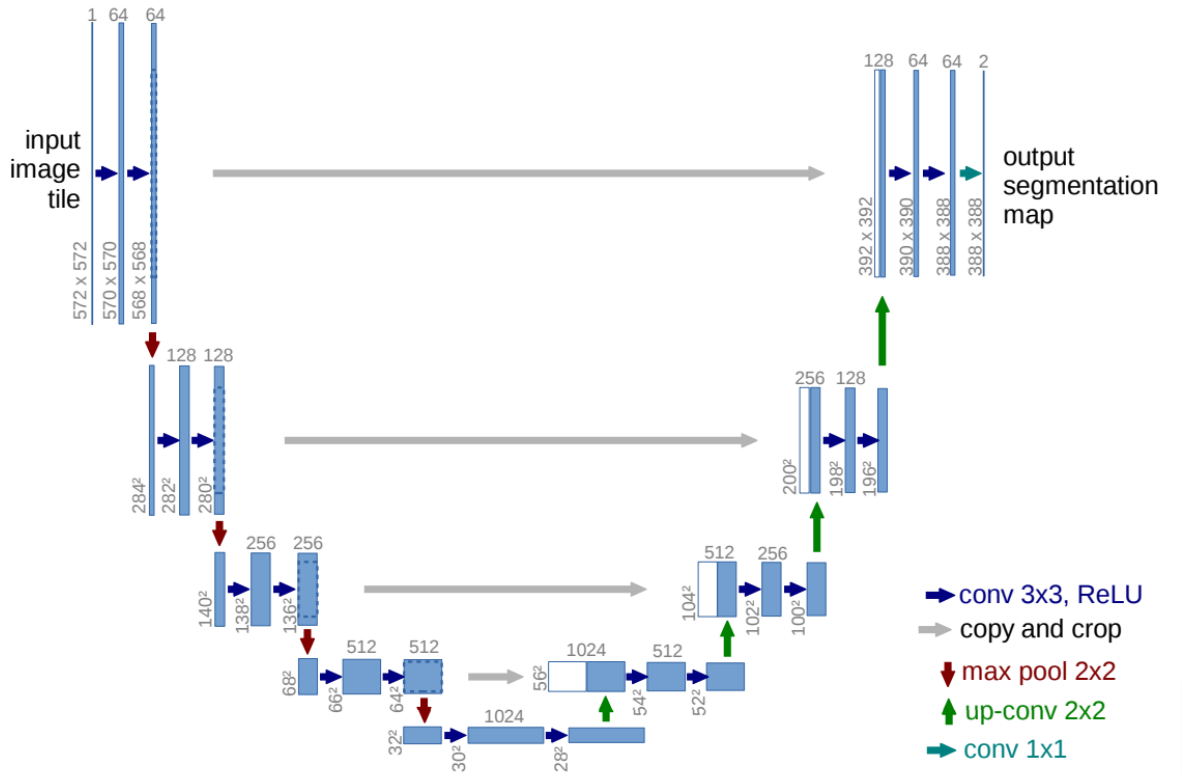

Figure 3.4. 7, U-net a fully connected convolution network for biomedical image segmentation tasks. Adopted from (Ronneberger, Fischer et al. 2015)

With this architecture, accurate contamination masking becomes possible. The contaminations across different Cryo-EM grids share similar patterns. Consequently, micrographs from different research groups can be used to train this network. As a classification task, pixel-level labeling is very laborious and time-consuming. The good news is that Micrographcleaner has finished data gathering, label marking, and network training (Sanchez-Garcia, Segura et al. 2020). In addition, their pre-trained network is ready to use or fine-tune with customized data. We ensemble this network as an option preprocessing step in our method. This dust mask can be applied in many steps of filament contours generation. Our experiment shows that applying the mask

on the likelihood map achieves the best performance—demonstrated by the picking performance show below.

Below, with-and-without comparing results, indicate the effect of the dust mask. Figure 3.4.8 below compares results with and without the effect of the dust mask. The hole edge of Cryo-Grid is one of the most common artifacts. Without a dust mask, contaminations make up 20% of the picked filament and influence downstream tasks significantly. Ensemble with the dust mask, artifacts, and some ice contamination is removed clearly.

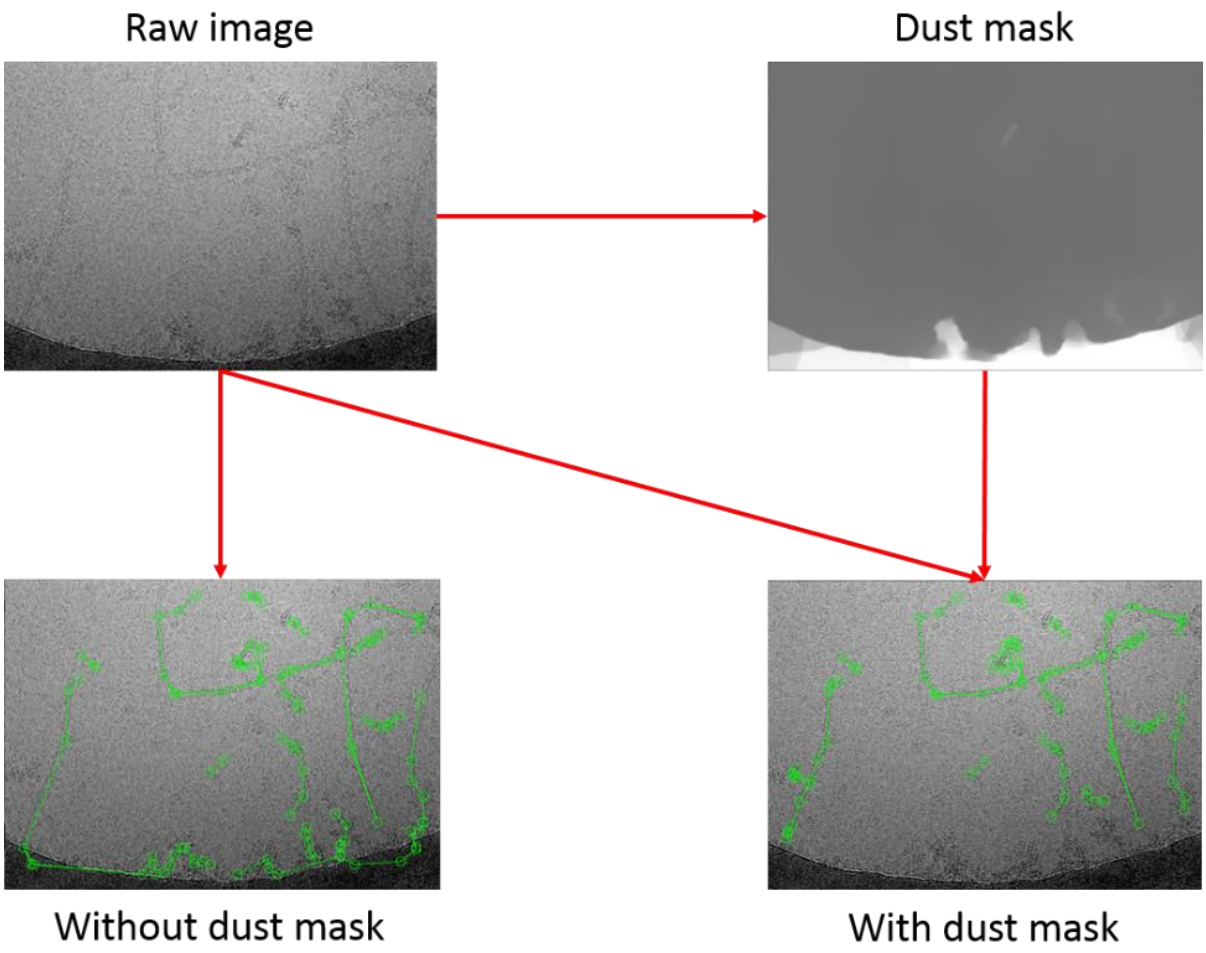

Figure 3.4. 8. Auto-picking performance illustration of method integrated with contamination mask shows better performance than non-mask method.

Although, there are many exiting filament picking algorithms available. Manually picked coordinate for training or template generation is necessary. Besides, these methods are hyper-parameters sensitive. The picking results depend on fine-tuning of hyper-parameters, which is unfriendly for users without relative prior knowledge. Compared with existing methods, our methods show an out-of-box property. Our method support reference-free mode, which does not rely on picked coordinates. Furthermore, our method is nearly parameters-free. Excluding basic information: pixel size and minimal filament length, our method does not require other compulsory parameters, which is achieved through an adaptive mechanism. The default output format of our method is Relion STAR. This makes sure that our picked coordinate is ready to use immediately after picking.

Figure 3.4.9 shows comparing results between our method and existing methods. The sample used for demonstration is UMOD filament, which is very flexible with diverse side views. SPHIRE-cyYOLO is a deep learning based method, and Relion-Helical-Autopicking is template based method. We can see these two methods can only detect rigid and straight UMOD filament. Our method captures most of the sample regions, including both rigid and flexible parts. In the lines fitting aspect, these two methods do not represent the filaments well. Our method achieves higher precision and recall ratio compared with the current method.

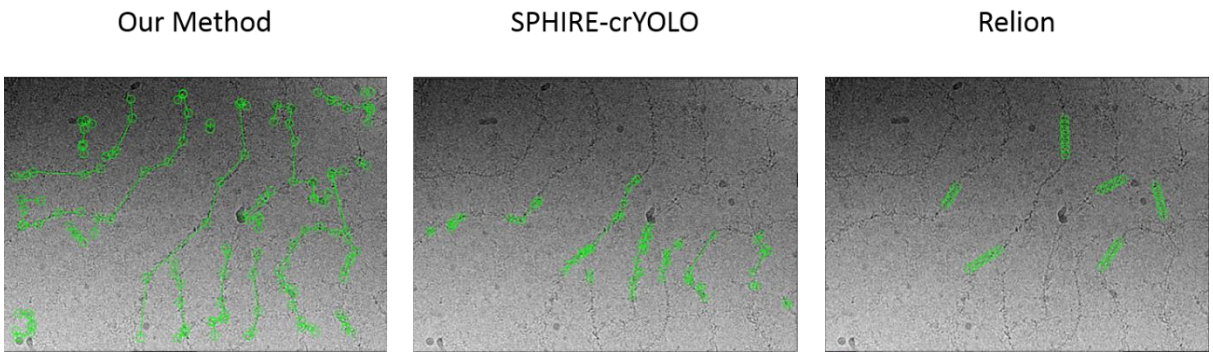

Figure 3.4. 9. picking performance comparison between different picking methods, including SHIRE-cyYOLO picking module and Relion helical picking module.

Finally, to evaluate the performance of our picking method, we feed the picked particles into 2D and 3D classification (Figure 3.4.10). As for both rigid filament and flexible filament. 2D classification generates clear and information-filled 2D averages, which indicates good picking precision. We run helical 3D reconstruction with helical symmetry and generates good 3D filament density. That means the helical information is well preserved during picking.

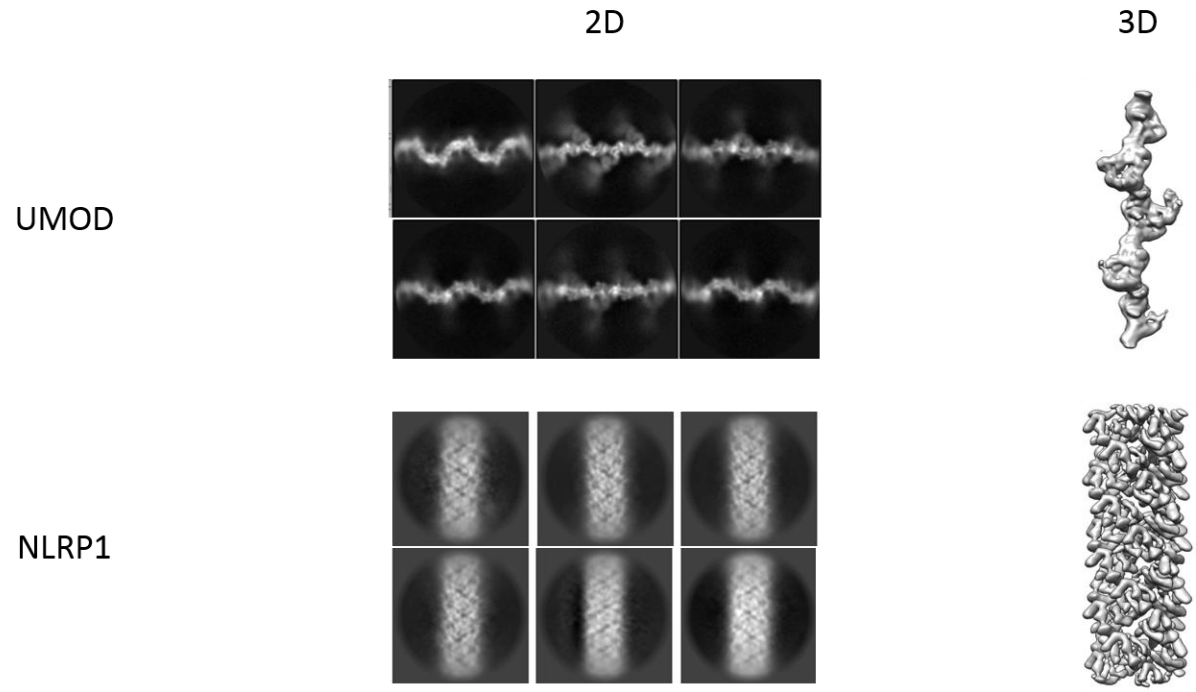

Figure 3.4. 10. performance demonstration for the downstream task, including 2D and 3D tasks, using our method's particles.

3.4.3 Conclusion and Discussion

Croy-EM image processing is quite different from other image process tasks. Low SNR (signal-to-noise ratio) and weak contrast limit the direct application of mature image processing algorithms on Cryo-EM data. And among Cryo-EM images processing pipeline, particle picking is the bottleneck step. Especially, filament picking is more challenging than single particle picking. The filament auto-picking method should be able to track filament orientation and retain the helical information. The existing method is either hard-to-use or requiring a prior in. The big data era is coming, as the dataset size is growing larger and larger. A user-friendly and powerful filament auto-picking is necessary. Compared with the current method, our method shows better

performance. Our method is easier to use and demonstrates higher accuracy and recall ratio at the same time; furthermore, no prior picking or reference is necessary.

3.5 CARD-CARD interaction prediction and in silico screening for potential filament nucleator

CARD domain protein is well studied in the past 20 years and recorded CARD-CARD interaction, and mutagenesis assay data is enough to train an AI model, which helps identify new CARD-CARD interaction pairs. Here, Known CARD-CARD interaction data and mutagenesis data was used as training dataset in the experiment.

3.5.1 The pipeline of modeling CARD-CARD interaction

The pipeline includes Four parts: data preprocessing, feature engineering, modeling training, and analysis.

Data used for training includes two parts, amino acids sequence of CARD domain protein and interaction annotations. Specifically, annotation here includes CARD-CARD interactions reported in previous research and loss-of-function mutations data. We can construct CARD-CARD interaction pairs with these data, and these pairs will be used for feature engineering.

The main idea in the feature extraction step is counting the frequency of specific residue patterns. Firstly, 20 amino acids were clustered into nine classes based on their physical properties to capture these patterns (Table 3.5.1). Then a window of three residues slides through the whole amino acid sequence, and the three-reside pattern is counted. Finally, the counting result is then transferred into the frequency. This feature engineering relies on the fact that nearly all CARD domains share a typical architecture: containing six alpha helixes. Because of this property, the exposing interface is quite similar, and key resides patterns will determine CARD-CARD interaction.

Table 3.5. 1. Amino acid category encoding

Amino Acid	Category
R, K	1
H	2
A, V	3
F, W, Y	4
I, L, M	5
G, P	6
N, Q	7
C, S, T	8
D, E	9

Random forest (RF) is one of the most influential classic machine learning, achieving good representative learning and generalization ability. Prepared features are fed into the RF model to fit the patterns. Regulation of hyper-parameters optimization is performed through grid search strategy, which prevents the model from overfitting.

The validation result is also generated during hyper-parameters searching. With 5-fold cross-validation and grid search, the ROC figure shows as above. Sensitivity is also called true positive rate, and 1- Specificity is called false positive rate. ROC figure demonstrates the value change of TPR and FPR under different thresholds. The large the area under the curve, the better the classifier performance. Our curve completely covers the red line, representing a random model generating positive and negative with both 50% probability. This result demonstrates that the hyper-parameters used to train the model will not result in overfitting (Figure 3.5.1). Due to limited data points, this curve is not smooth enough. Although this model does not achieve very high AUC value, as a proposal model, it is able to give insight on potential CARD-CARD pair.

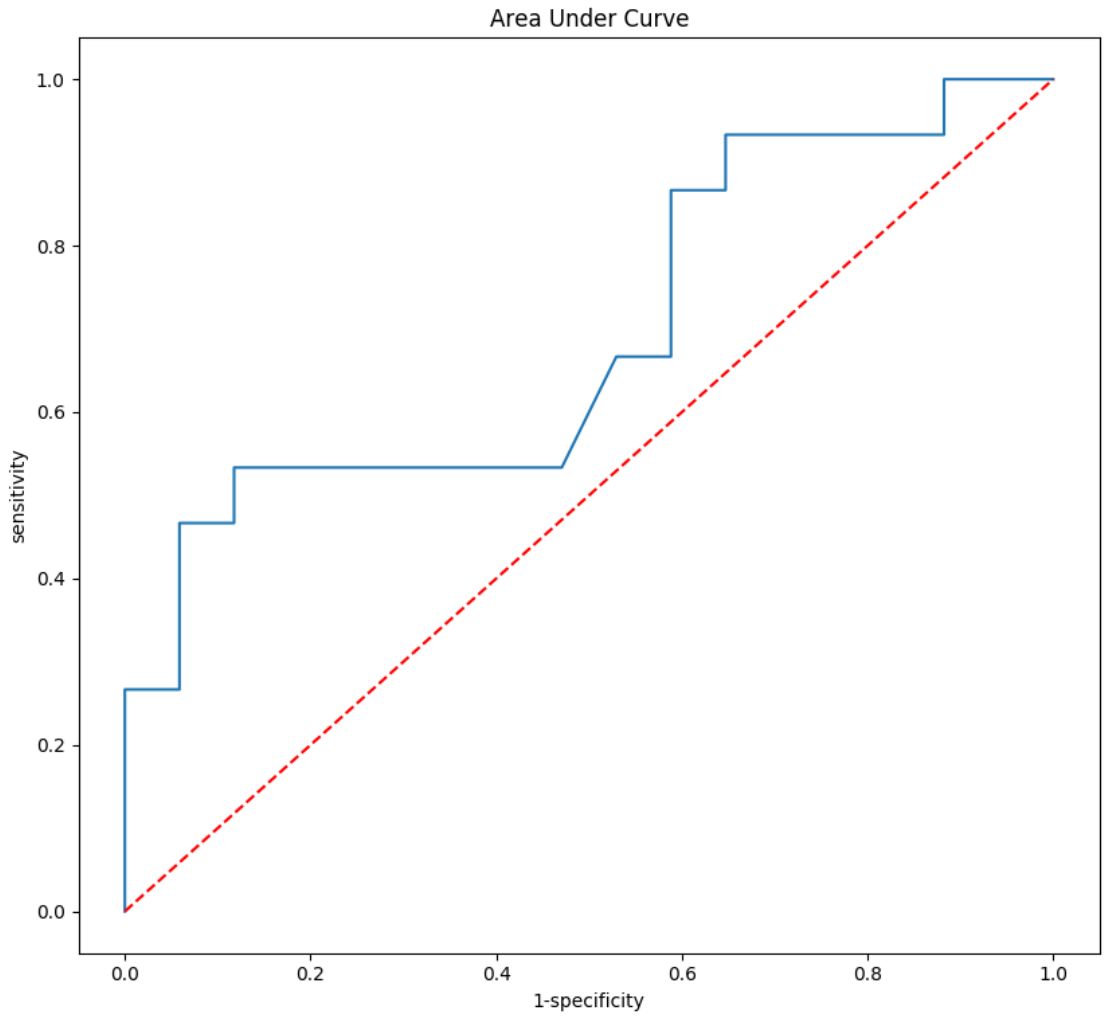

Figure 3.5. 1. Performance demonstration through AUC metric.

3.5.2 Analysis for the prediction results

The output of this CARD-CARD interaction prediction algorithm represents the probability of interaction happens. Exclude from the training set; a test set comprised of 38 proteins was used in this prediction. We get a matrix of interaction probability of 1369 CARD-CARD interaction pairs. Probability between these CARD-CARD interaction pairs constitutes a network of CARD-CARD affinity. This information in the network can be used to analyze the distance between different CARD containing proteins. With a hierarchical clustering map, the hierarchical tree illustrates the distance between different CARD containing proteins. A protein cluster with a similar property was observed. Notably, one cluster contains NLRP1, CASPASE2, MDA5, RIG-I, second CARD, RIP2. All of them are reported to form filament oligomers in previous studies. CARD8, CARD9, CARD10, CARD11, NLRC4, RIG-I first CARD were classified into one cluster. In this cluster, NLRC4, RIG-I, and CARD9 were reported to form filament. Besides, CARD10 and CARD11 are reported to participate in others protein filament assemblies (Holliday, Witt et al. 2018, Schlauderer, Seeholzer et al. 2018, Chiriaco, Di Matteo et al. 2019). In a previous study, CARD8 are known to inhibit NLRP3 signaling oligomer assembly. However, the mechanism is not precise. Besides, as the figures demonstrate, CARD8-CARD8 interaction was predicted as positive, which means it tends to form oligomers through self-assembly. (Figure 3.5.2)

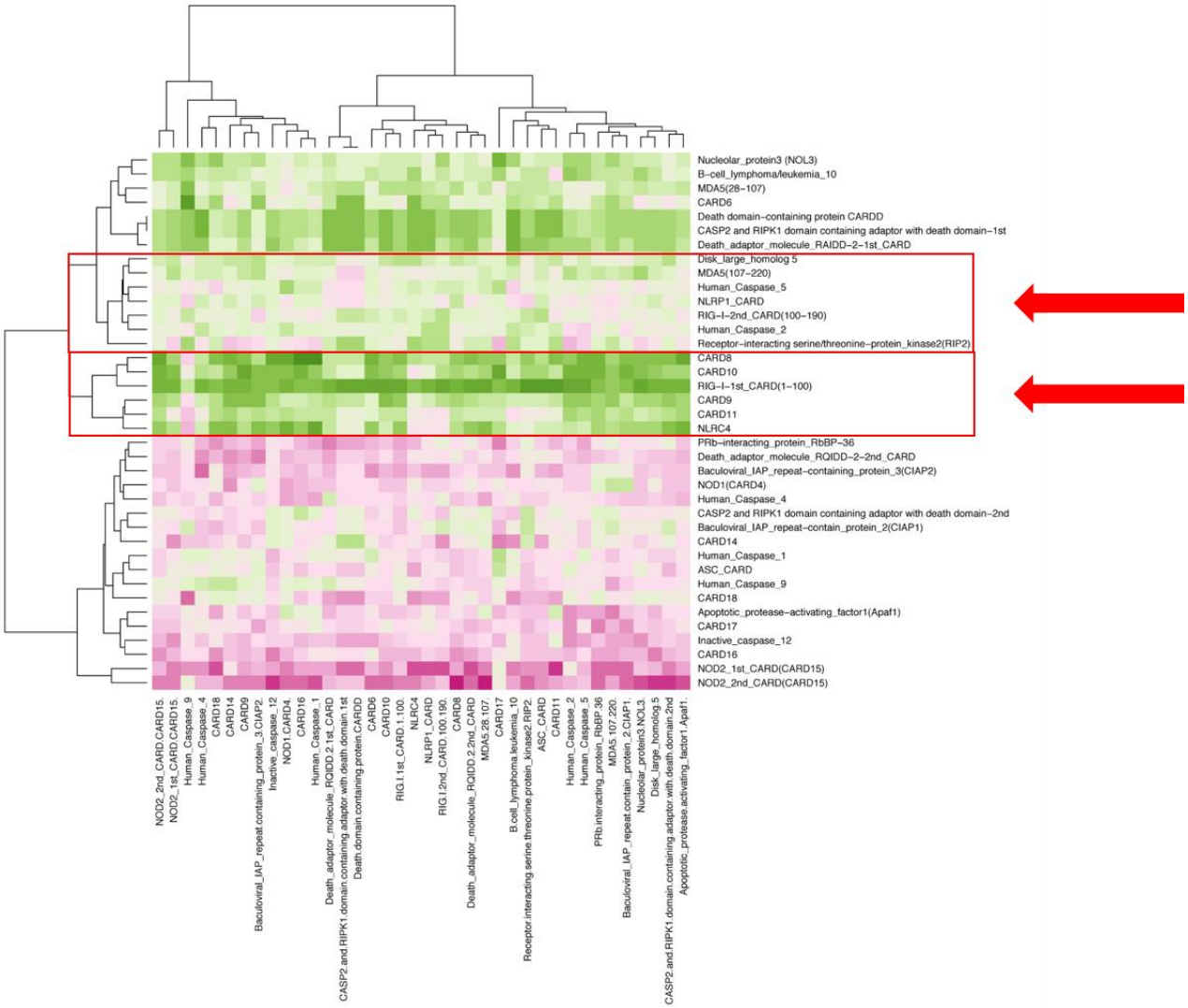

Figure 3.5. 2. Hierarchical tree clustering for CARD-CARD interaction pairs based on the predicted results

To validate this prediction result. The CARD8-SNAP fusion protein was constructed. Filamentous sample preparation is similar to that of NLRP1, which is accomplished through boosting protein concentration. With negative staining and T12 electron microscope, CARD8-CARD filament oligomer was observed (Figure 3.5.3). This work

guides us to consider CARD8 for the oligomerization study. We talk in more detail about this filament in Chapter 3.2 part.

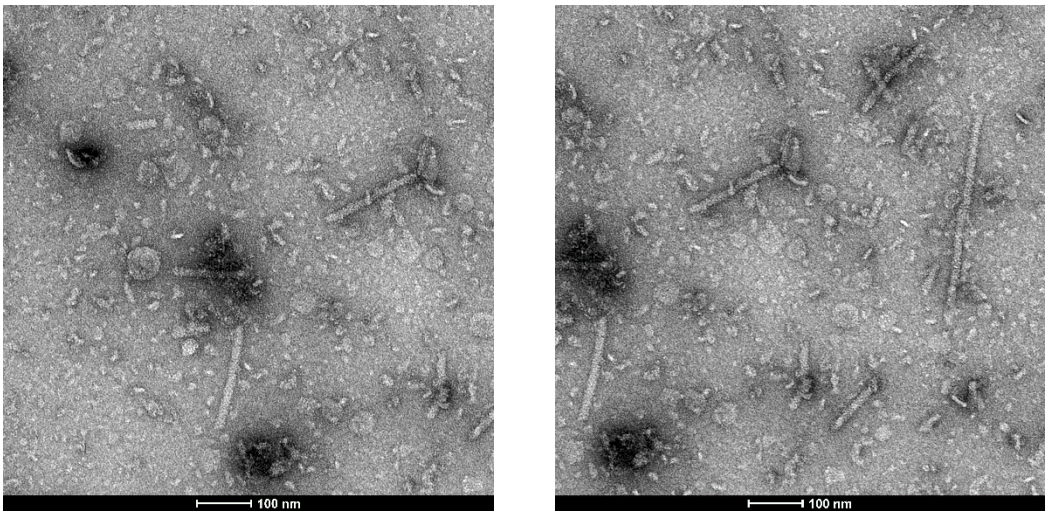

Figure 3.5. 3. Imaging of CARD8-SNAP under negative staining microscopy,

3.5.3 Conclusion and Discussion

Compared with general protein-protein interactions, CARD-CARD interaction is recalcitrant to be predicted. For general protein-protein interaction prediction task, PPI dataset is used to train, which is giant dataset covers various type of protein-protein interaction. However, training with this dataset does not work appropriately on CARD-CARD interaction because of the sequence similarity of CARDs. Consequently, the training set and model should be carefully selected. For the limited number of known CARD-CARD interactions, mutagenesis data should also be added to the training set. Classifiers should be able to identify tiny feature changes caused by single mutation. Based on this property, we develop a new feature extraction method and choose Random Forest as the classifier. Random Forest is a powerful and classic machine learning model; it is sensitive enough to identify slight differences in features and owns

many parameters that can be adjusted to avoid overfitting. With cross-validation, the result shows that for CARD-CARD interaction prediction, and the accuracy reaches high accuracy around 85% with our model. Random Forest model with best-optimized hyper-parameters was used in the following experiment, optimized through parameters grid searching and cross-validation. The model output the probability; if interaction will happen between two CARD domain containing protein.

Based on the probability matrix of CARDs pair, we get the hierarchical tree of CARD-CARD interaction, which show a cluster of protein with similar property. Based on positive marker protein RIG-I, CARD9, and NLRC4. We identify the protein cluster, members of which trends to form filament oligomers. Here CARD8 is also a member of this cluster. The previous study only shows that CARD8 is involved in caspase1 activation without a precise mechanism. However, our computation method shows that it trends to form signaling oligomers in the innate immune signaling pathway. Our EM study validates this hypothesis. We purified this protein and observed a filament complex under negative stain EM. This pipeline shows its potential to discover possible interaction in other protein interaction systems where protein secondary structures are high conserved and interaction mechanism is complicated.

Chapter 4. Conclusion and Future directions

4.1 Conclusions

Protein filament is a special kind of sizeable molecular complex, which plays diverse roles in biological activity. 3D structure of native filaments provides insight into drug discovery or therapy development. Our protein filament preparing protocol, Cryo-EM helical reconstruction pipeline, and new filament auto-picking algorism accelerate filament structure determination significantly. Here, I demonstrate filament electron density maps of NLRC4-, NLRP1- and CARD8-CARD, which are obtained with these methods. The high-resolution structure reveals detail of filament oligomer assembly mechanism and demonstrates essential residues contact between interfaces. Notably, Cryo-EM image processing efficiency increased with our fully automatically filament picking software, which is less time-consuming and more user-friendly.

In addition, CARD-CARD interaction is the basis for CARD-base filament assembly. Consequently, the identification of unknown CARD-CARD interaction is also of vital importance. Experimental assay with exclusive grid search for CARD-CARD interaction is not possible, due to limited resources and labor, highlighting the importance of CARD cross-talk prediction in silico. Based on the property of CARD-CARD interface, protein interaction prediction algorithm specific for CARD-containing protein was developed, which accurately predicts interactions and protein clusters with similar interaction pattern. Our prediction algorithms rank the CARD-CARD interaction pairs by happening probability, which can then guide unknown signaling pathway discovery. With this pipeline, we successfully identified CARD8, which shows a similar interaction pattern with NLRP1 and NLRC4 in hierarchical clusters. Furthermore, this

method can be applied to other protein families, whose members share similar interaction patterns.

These high-resolution filament structures explain many relevant biochemistry mechanisms and reveal Novo protein structure conformations. In the innate immunity system, especially PRRs, signaling filament aggregation is the bottleneck step for completing inflammasome formation and signaling transduction. NAIP/NLRC4 is one of the canonical signaling pathways, and NLRC4 inflammasome is well studied previously. Here, our NLRC4 CARD filament serves as the last piece of the puzzle. It explains the assembly mechanism of NLRC4 CARD and how NLRC4 recruits ASC and Caspase-1, which is meaningful for inhibitor or agonist discovery in NLRC4 related disease therapy. Compared with NLRC4, NLRP1 and CARD8 are exceptional cases in NOD-like receptors. The functional domain, which is involved in filament formation, usually locates at N-terminal. However, the functional CARD domain exists at the C-terminal of NLRP1 and CARD8. From structure analysis of our NLRP1 and CARD8 filament model, NLRP1 and CARD8 are similar in the sequence alignment and share similar helical rise and rotation angle when forming filament. The interface analysis demonstrates that NLRP1 and CARD8 filament assemble relies on Type-I interaction, compared with MAVS and RIP2 filament. These findings improve our understanding of the assembly mechanism for NLRs members.

For another critical filamentous protein UMOD, its high-resolution structure reveals a unique inter-domain linker, explaining why UMOD is so resistant to proteolysis, and other denaturing chemical factors. Besides, as the first solved native filament of the

ZP domain family, it provides a model to describe the egg coat matrix, which is vital for sperm detection and fertilization.

Supramolecular complex usually works as an organizer for signal transduction. In order to achieve flexibility and efficiency in signal transduction, signal receiving function and activation function should be assigned to different domains in a single molecular. For example, in NLRs, CARD/PYD and NATCH/LRR work as activators and sensors, respectively. This architecture ensures that the inflammasome is able to respond to diverse antigens and mediate pyroptosis signals dynamically. This fact suggests that to handle complex signaling networks, other supramolecular signaling organizers may also own this function separation architecture.

4.2 Future work - Representative learning for protein amino acid sequence and 3D structure

Amino acid sequence of a protein determines its property and functions. Although calculating the property from amino acid sequence with the first principal method is theoretically possible, current computation resource does not meet the requirements. Therefore, many approximation methods are developed. These traditional methods rely on domain knowledge and manual feature engineering. This paradigm demonstrates some good results but cannot achieve the most expressive feature learning.

Deep learning is growing fast in the past decade, and it plays more and more critical roles in diverse domains. For sequence data processing, it also defines a new

paradigm, including RNN (Recurrent neural network) and self-attention mechanism. These methods achieve sort-of-art performance on NLP (natural language processing). For protein, if we only consider its amino acid sequence, we may find the intrinsic similarity between protein and natural language. Consequently, successful language models like the self-attention mechanism and BERT are able to model protein sequence as well. The revolution already happens in the protein modeling area. Recently, publication dedicates that these deep learning based methods achieve better performance than traditional methods on many tasks, including protein-protein interaction, protein-ligand interaction, and critical contact site prediction. It even solves one of the biggest challenges in biology, protein folding prediction. Alphafold2 achieves high precision protein 3D structure prediction from amino acid sequence only data. This accelerates protein-related research dramatically.

However, protein is more than amino acid sequence. Protein 3D structure contains more abundant information than plain amino acid sequence. A deep learning method, which embeds protein 3D structure, is missing. Graph neuron network, which is designed for non-Euclidean data, is suitable to take protein 3D structure as input. Designing a protein 3D structure embedding method should achieve better expressive learning and predict the dynamic and function of protein.

References

Abate, E., R. Blomgran, D. Verma, M. Lerm, M. Fredrikson, M. Belayneh, P. Soderkvist, O. Stendahl and T. Schon (2019). "Polymorphisms in CARD8 and NLRP3 are associated with extrapulmonary TB and poor clinical outcome in active TB in Ethiopia." Sci Rep **9**(1): 3126.

Akira, S. and K. Takeda (2004). "Toll-like receptor signalling." Nat Rev Immunol **4**(7): 499-511.

Allen, I. C., J. E. Wilson, M. Schneider, J. D. Lich, R. A. Roberts, J. C. Arthur, R. M. Woodford, B. K. Davis, J. M. Uronis, H. H. Herfarth, C. Jobin, A. B. Rogers and J. P. Ting (2012). "NLRP12 suppresses colon inflammation and tumorigenesis through the negative regulation of noncanonical NF-kappaB signaling." Immunity **36**(5): 742-754.

Baldwin, P. R. and P. A. Penczek (2007). "The Transform Class in SPARX and EMAN2." J Struct Biol **157**(1): 250-261.

Ball, D. P., C. Y. Taabazuing, A. R. Griswold, E. L. Orth, S. D. Rao, I. B. Kotliar, L. E. Vostal, D. C. Johnson and D. A. Bachovchin (2020). "Caspase-1 interdomain linker cleavage is required for pyroptosis." Life Sci Alliance **3**(3).

Banchereau, J. and R. M. Steinman (1998). "Dendritic cells and the control of immunity." Nature **392**(6673): 245-252.

Barclay, A. N. (1999). "Ig-like domains: evolution from simple interaction molecules to sophisticated antigen recognition." Proc Natl Acad Sci U S A **96**(26): 14672-14674.

Bayer, M. E. (1964). "An Electron Microscope Examination of Urinary Mucoprotein and Its Interaction with Influenza Virus." J Cell Biol **21**: 265-274.

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.

Bepler, T., A. Morin, M. Rapp, J. Brasch, L. Shapiro, A. J. Noble and B. Berger (2019). "Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs." Nat Methods **16**(11): 1153-1160.

Bernabeu, C., J. M. Lopez-Novoa and M. Quintanilla (2009). "The emerging role of TGF-beta superfamily coreceptors in cancer." Biochim Biophys Acta **1792**(10): 954-973.

Berndsen, Z., C. Bowman, H. Jang and A. B. Ward (2017). "EMHP: an accurate automated hole masking algorithm for single-particle cryo-EM image processing." Bioinformatics **33**(23): 3824-3826.

Bokhove, M. and L. Jovine (2018). "Structure of Zona Pellucida Module Proteins." Curr Top Dev Biol **130**: 413-442.

Bokhove, M., K. Nishimura, M. Brunati, L. Han, D. de Sanctis, L. Rampoldi and L. Jovine (2016). "A structured interdomain linker directs self-polymerization of human uromodulin." Proc Natl Acad Sci U S A **113**(6): 1552-1557.

Bokhove, M., H. Sadat Al Hosseini, T. Saito, E. Dioguardi, K. Gegenschatz-Schmid, K. Nishimura, I. Raj, D. de Sanctis, L. Han and L. Jovine (2016). "Easy mammalian expression and crystallography of maltose-binding protein-fused human proteins." J Struct Biol **194**(1): 1-7.

Boyden, E. D. and W. F. Dietrich (2006). "Nalp1b controls mouse macrophage susceptibility to anthrax lethal toxin." Nat Genet **38**(2): 240-244.

Broz, P., J. von Moltke, J. W. Jones, R. E. Vance and D. M. Monack (2010). "Differential requirement for Caspase-1 autoproteolysis in pathogen-induced cell death and cytokine processing." Cell Host Microbe **8**(6): 471-483.

Brunati, M., S. Perucca, L. Han, A. Cattaneo, F. Consolato, A. Andolfo, C. Schaeffer, E. Olinger, J. Peng, S. Santambrogio, R. Perrier, S. Li, M. Bokhove, A. Bachi, E. Hummler, O. Devuyst, Q. Wu, L. Jovine and L. Rampoldi (2015). "The serine protease hepsin mediates urinary secretion and polymerisation of Zona Pellucida domain protein uromodulin." Elife **4**: e08887.

Canna, S. W., A. A. de Jesus, S. Gouni, S. R. Brooks, B. Marrero, Y. Liu, M. A. DiMattia, K. J. Zaal, G. A. Sanchez, H. Kim, D. Chapelle, N. Plass, Y. Huang, A. V. Villarino, A. Biancotto, T. A. Fleisher, J. A. Duncan, J. J. O'Shea, S. Benseler, A. Grom, Z. Deng, R. M. Laxer and R. Goldbach-Mansky (2014). "An activating NLRC4 inflammasome mutation causes autoinflammation with recurrent macrophage activation syndrome." Nat Genet **46**(10): 1140-1146.

Carneiro, L. A., L. H. Travassos and S. E. Girardin (2007). "Nod-like receptors in innate immunity and inflammatory diseases." Ann Med **39**(8): 581-593.

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

Chavarria-Smith, J. and R. E. Vance (2015). "The NLRP1 inflammasomes." Immunol Rev **265**(1): 22-34.

Chen, G., M. H. Shaw, Y. G. Kim and G. Nunez (2009). "NOD-like receptors: role in innate immunity and inflammatory disease." Annu Rev Pathol **4**: 365-398.

Cheng, T. C., C. Hong, I. V. Akey, S. Yuan and C. W. Akey (2016). "A near atomic structure of the active human apoptosome." Elife **5**.

Chiriaco, M., G. Di Matteo, F. Conti, D. Petricone, M. De Luca, S. Di Cesare, C. Cifaldi, R. De Vito, M. Zoccolillo, J. Serafinelli, N. Poerio, M. Fraziano, I. Brigida, F. Cardinale, P. Rossi, A. Aiuti, C. Cancrini and A. Finocchi (2019). "First Case of Patient With Two Homozygous Mutations in MYD88 and CARD9 Genes Presenting With Pyogenic Bacterial Infections, Elevated IgE, and Persistent EBV Viremia." Frontiers in Immunology **10**.

D'Osualdo, A., C. X. Weichenberger, R. N. Wagner, A. Godzik, J. Wooley and J. C. Reed (2011). "CARD8 and NLRP1 undergo autoproteolytic processing through a ZU5-like domain." PLoS One **6**(11): e27396.

Darie, C. C., W. G. Janssen, E. S. Litscher and P. M. Wassarman (2008). "Purified trout egg vitelline envelope proteins VEbeta and VEgamma polymerize into homomeric fibrils from dimers in vitro." Biochim Biophys Acta **1784**(2): 385-392.

David, L., Y. Li, J. Ma, E. Garner, X. Zhang and H. Wu (2018). "Assembly mechanism of the CARMA1-BCL10-MALT1-TRAF6 signalosome." Proc Natl Acad Sci U S A **115**(7): 1499-1504.

De Rosier, D. J. and A. Klug (1968). "Reconstruction of three dimensional structures from electron micrographs." Nature **217**(5124): 130-134.

Demkowicz, W. E., Jr., R. A. Littaua, J. Wang and F. A. Ennis (1996). "Human cytotoxic T-cell memory: long-lived responses to vaccinia virus." J Virol **70**(4): 2627-2631.

Devuyst, O., E. Olinger and L. Rampoldi (2017). "Uromodulin: from physiology to rare and complex kidney disorders." Nat Rev Nephrol **13**(9): 525-544.

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.

Diebolder, C. A., E. F. Halff, A. J. Koster, E. G. Huizinga and R. I. Koning (2015). "Cryoelectron Tomography of the NAIP5/NLRC4 Inflammasome: Implications for NLR Activation." Structure **23**(12): 2349-2357.

Dieude, P., M. Guedj, J. Wipff, B. Ruiz, G. Riemekasten, P. Airo, I. Melchers, E. Hachulla, M. M. Cerinic, E. Diot, N. Hunzelmann, P. Caramaschi, J. Sibilia, K. Tiev, L. Mouthon, V. Riccieri, J. L. Cracowski, P. H. Carpentier, J. Distler, Z. Amoura, I. Tarner, J. Avouac, O. Meyer, A. Kahan, C. Boileau and Y. Allanore (2011). "NLRP1 influences the systemic sclerosis phenotype: a new clue for the contribution of innate immunity in systemic sclerosis-related fibrosing alveolitis pathogenesis." Ann Rheum Dis **70**(4): 668-674.

Dominguez, R. and K. C. Holmes (2011). "Actin structure and function." Annu Rev Biophys **40**: 169-186.

Dranoff, G. (2004). "Cytokines in cancer pathogenesis and cancer therapy." Nat Rev Cancer **4**(1): 11-22.

Dudley, D. J. (1992). "The immune system in health and disease." Baillieres Clin Obstet Gynaecol **6**(3): 393-416.

Duncan, J. A. and S. W. Canna (2018). "The NLRC4 Inflammasome." Immunol Rev **281**(1): 115-123.

Earl, L. A., V. Falconieri, J. L. Milne and S. Subramaniam (2017). "Cryo-EM: beyond the microscope." Curr Opin Struct Biol **46**: 71-78.

Egge, N., A. Muthusubramanian and G. A. Cornwall (2015). "Amyloid properties of the mouse egg zona pellucida." PLoS One **10**(6): e0129907.

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.

Fernandes, R., C. Tsuda, A. L. Perumalsamy, T. Naranian, J. Chong, B. M. Acton, Z. B. Tong, L. M. Nelson and A. Jurisicova (2012). "NLRP5 mediates mitochondrial function in mouse oocytes and embryos." Biol Reprod **86**(5): 138, 131-110.

Finger, J. N., J. D. Lich, L. C. Dare, M. N. Cook, K. K. Brown, C. Duraiswami, J. Bertin and P. J. Gough (2012). "Autolytic proteolysis within the function to find domain (FIIND) is required for NLRP1 inflammasome activity." J Biol Chem **287**(30): 25030-25037.

Fontalba, A., O. Gutierrez and J. L. Fernandez-Luna (2007). "NLRP2, an inhibitor of the NF-kappaB pathway, is transcriptionally activated by NF-kappaB and exhibits a nonfunctional allelic variant." J Immunol **179**(12): 8519-8524.

Franchi, L., N. Warner, K. Viani and G. Nunez (2009). "Function of Nod-like receptors in microbial recognition and host defense." Immunol Rev **227**(1): 106-128.

Friedman, L. M., A. A. Dror and K. B. Avraham (2007). "Mouse models to study inner ear development and hereditary hearing loss." Int J Dev Biol **51**(6-7): 609-631.

Ghosal, D. and J. Lowe (2015). "Collaborative protein filaments." EMBO J **34**(18): 2312-2320.

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.

Gong, Y. N. and F. Shao (2012). "Sensing bacterial infections by NAIP receptors in NLRC4 inflammasome activation." Protein Cell **3**(2): 98-105.

Grandemange, S., E. Sanchez, P. Louis-Plence, F. Tran Mau-Them, D. Bessis, C. Coubes, E. Frouin, M. Seyger, M. Girard, J. Puechberty, V. Costes, M. Rodiere, A. Carbasse, E. Jeziorski, P. Portales, G. Sarrabay, M. Mondain, C. Jorgensen, F. Apparailly, E. Hoppenreijs, I. Touitou and D. Genevieve (2017). "A new autoinflammatory and autoimmune syndrome associated with NLRP1 mutations: NAIAD (NLRP1-associated autoinflammation with arthritis and dyskeratosis)." Ann Rheum Dis **76**(7): 1191-1198.

Grant, T., A. Rohou and N. Grigorieff (2018). "cisTEM, user-friendly software for single-particle image processing." Elife **7**.

Green, D. P. (1997). "Three-dimensional structure of the zona pellucida." Rev Reprod **2**(3): 147-156.

Greve, J. M. and P. M. Wassarman (1985). "Mouse egg extracellular coat is a matrix of interconnected filaments possessing a structural repeat." J Mol Biol **181**(2): 253-264.

Griswold, A. R., D. P. Ball, A. Bhattacharjee, A. J. Chui, S. D. Rao, C. Y. Taabazuing and D. A. Bachovchin (2019). "DPP9's Enzymatic Activity and Not Its Binding to CARD8 Inhibits Inflammasome Activation." ACS Chem Biol **14**(11): 2424-2429.

Hamlin, L. M. and W. W. Fish (1977). "Physical properties of Tamm-Horsfall glycoprotein and its glycopolypeptide." Int J Pept Protein Res **10**(4): 270-276.

Han, L., M. Monne, H. Okumura, T. Schwend, A. L. Cherry, D. Flot, T. Matsuda and L. Jovine (2010). "Insights into egg coat assembly and egg-sperm interaction from the X-ray structure of full-length ZP3." Cell **143**(3): 404-415.

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

He, X., S. Mekasha, N. Mavrogiorgos, K. A. Fitzgerald, E. Lien and R. R. Ingalls (2010). "Inflammation and fibrosis during Chlamydia pneumoniae infection is regulated by IL-1 and the NLRP3/ASC inflammasome." J Immunol **184**(10): 5743-5754.

Hibino, T., M. Loza-Coll, C. Messier, A. J. Majeske, A. H. Cohen, D. P. Terwilliger, K. M. Buckley, V. Brockton, S. V. Nair, K. Berney, S. D. Fugmann, M. K. Anderson, Z. Pancer, R. A. Cameron, L. C. Smith

and J. P. Rast (2006). "The immune gene repertoire encoded in the purple sea urchin genome." Dev Biol **300**(1): 349-365.

Hoffman, H. M., J. L. Mueller, D. H. Broide, A. A. Wanderer and R. D. Kolodner (2001). "Mutation of a new gene encoding a putative pyrin-like protein causes familial cold autoinflammatory syndrome and Muckle-Wells syndrome." Nat Genet **29**(3): 301-305.

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

Holliday, M., A. Witt, A. Estevez, A. Rohou, E. Dueber and W. Fairbrother (2018). "The Structural Basis for CARD9 and CARD11 Autoinhibition." Protein Science **27**: 112-112.

Holliday, M. J., A. Witt, A. Rodriguez Gama, B. T. Walters, C. P. Arthur, R. Halfmann, A. Rohou, E. C. Dueber and W. J. Fairbrother (2019). "Structures of autoinhibited and polymerized forms of CARD9 reveal mechanisms of CARD9 and CARD11 activation." Nat Commun **10**(1): 3070.

Hollingsworth, L. R., H. Sharif, A. R. Griswold, P. Fontana, J. Mintseris, K. B. Dagbay, J. A. Paulo, S. P. Gygi, D. A. Bachovchin and H. Wu (2021). "DPP9 sequesters the C terminus of NLRP1 to repress inflammasome activation." Nature **592**(7856): 778-783.

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.

Hu, Z., C. Yan, P. Liu, Z. Huang, R. Ma, C. Zhang, R. Wang, Y. Zhang, F. Martinon, D. Miao, H. Deng, J. Wang, J. Chang and J. Chai (2013). "Crystal structure of NLRC4 reveals its autoinhibition mechanism." Science **341**(6142): 172-175.

Huang, M., X. Zhang, G. A. Toh, Q. Gong, J. Wang, Z. Han, B. Wu, F. Zhong and J. Chai (2021). "Structural and biochemical mechanisms of NLRP1 inhibition by DPP9." Nature **592**(7856): 773-777.

Huber, S. T., T. Kuhm and C. Sachse (2018). "Automated tracing of helical assemblies from electron cryo-micrographs." J Struct Biol **202**(1): 1-12.

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

Janeway, C. A., Jr. (1989). "Approaching the asymptote? Evolution and revolution in immunology." Cold Spring Harb Symp Quant Biol **54 Pt 1**: 1-13.

Janeway, C. A., Jr. (1992). "The immune system evolved to discriminate infectious nonself from noninfectious self." Immunol Today **13**(1): 11-16.

Janssens, S. and A. Tinel (2012). "The PIDDosome, DNA-damage-induced apoptosis and beyond." Cell Death Differ **19**(1): 13-20.

Jin, T., J. Curry, P. Smith, J. Jiang and T. S. Xiao (2013). "Structure of the NLRP1 caspase recruitment domain suggests potential mechanisms for its association with procaspase-1." Proteins **81**(7): 1266-1270.

Jin, T., M. Huang, P. Smith, J. Jiang and T. S. Xiao (2013). "The structure of the CARD8 caspase-recruitment domain suggests its association with the FIIND domain and procaspases through adjacent surfaces." Acta Crystallogr Sect F Struct Biol Cryst Commun **69**(Pt 5): 482-487.

Johnson, D. C., M. C. Okondo, E. L. Orth, S. D. Rao, H. C. Huang, D. P. Ball and D. A. Bachovchin (2020). "DPP8/9 inhibitors activate the CARD8 inflammasome in resting lymphocytes." Cell Death Dis **11**(8): 628.

Johnson, D. C., C. Y. Taabazuing, M. C. Okondo, A. J. Chui, S. D. Rao, F. C. Brown, C. Reed, E. Peguero, E. de Stanchina, A. Kentsis and D. A. Bachovchin (2018). "DPP8/DPP9 inhibitor-induced pyroptosis for treatment of acute myeloid leukemia." Nat Med **24**(8): 1151-1156.

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

Jounai, N., K. Kobiyama, M. Shiina, K. Ogata, K. J. Ishii and F. Takeshita (2011). "NLRP4 negatively regulates autophagic processes through an association with beclin1." J Immunol **186**(3): 1646-1655.

Jovine, L., C. C. Darie, E. S. Litscher and P. M. Wassarman (2005). "Zona pellucida domain proteins." Annu Rev Biochem **74**: 83-114.

Jovine, L., W. G. Janssen, E. S. Litscher and P. M. Wassarman (2006). "The PLAC1-homology region of the ZP domain is sufficient for protein polymerisation." BMC Biochem **7**: 11.

Jovine, L., H. Qi, Z. Williams, E. Litscher and P. M. Wassarman (2002). "The ZP domain is a conserved module for polymerization of extracellular proteins." Nat Cell Biol **4**(6): 457-461.

Jovine, L., H. Qi, Z. Williams, E. S. Litscher and P. M. Wassarman (2004). "A duplicated motif controls assembly of zona pellucida domain proteins." Proc Natl Acad Sci U S A **101**(16): 5922-5927.

Kagan, J. C., V. G. Magupalli and H. Wu (2014). "SMOCs: supramolecular organizing centres that control innate immunity." Nat Rev Immunol **14**(12): 821-826.

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.

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.

Kitamura, A., Y. Sasaki, T. Abe, H. Kano and K. Yasutomo (2014). "An inherited mutation in NLRC4 causes autoinflammation in human and mice." J Exp Med **211**(12): 2385-2396.

Kofoed, E. M. and R. E. Vance (2011). "Innate immune recognition of bacterial ligands by NAIPs determines inflammasome specificity." Nature **477**(7366): 592-595.

LaFavers, K. A., E. Macedo, P. S. Garimella, C. Lima, S. Khan, J. Myslinski, J. McClintick, F. A. Witzmann, S. Winfree, C. L. Phillips, T. Hato, P. C. Dagher, X. R. Wu, T. M. El-Achkar and R. Micanovic (2019). "Circulating uromodulin inhibits systemic oxidative stress by inactivating the TRPM2 channel." Sci Transl Med **11**(512).

Lander, G. C., S. M. Stagg, N. R. Voss, A. Cheng, D. Fellmann, J. Pulokas, C. Yoshioka, C. Irving, A. Mulder, P. W. Lau, D. Lyumkis, C. S. Potter and B. Carragher (2009). "Appion: an integrated, database-driven pipeline to facilitate EM image processing." J Struct Biol **166**(1): 95-102.

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.

Levandowski, C. B., C. M. Mailloux, T. M. Ferrara, K. Gowan, S. Ben, Y. Jin, K. K. McFann, P. J. Holland, P. R. Fain, C. A. Dinarello and R. A. Spritz (2013). "NLRP1 haplotypes associated with vitiligo and autoimmunity increase interleukin-1beta processing via the NLRP1 inflammasome." Proc Natl Acad Sci U S A **110**(8): 2952-2956.

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.

Li, Y., M. Zhou, Q. Hu, X. C. Bai, W. Huang, S. H. Scheres and Y. Shi (2017). "Mechanistic insights into caspase-9 activation by the structure of the apoptosome holoenzyme." Proc Natl Acad Sci U S A **114**(7): 1542-1547.

Ligtenberg, A. J., E. C. Veerman, A. V. Nieuw Amerongen and J. Mollenhauer (2007). "Salivary agglutinin/glycoprotein-340/DMBT1: a single molecule with variable composition and with different functions in infection, inflammation and cancer." Biol Chem **388**(12): 1275-1289.

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.

Linder, A., S. Bauernfried, Y. Cheng, M. Albanese, C. Jung, O. T. Keppler and V. Hornung (2020). "CARD8 inflammasome activation triggers pyroptosis in human T cells." EMBO J **39**(19): e105071.

Liu, Z. and T. S. Xiao (2015). "STRUCTURAL BIOLOGY. Assembling the wheel of death." Science **350**(6259): 376-377.

Louros, N. N., E. D. Chrysina, G. E. Baltatzis, E. S. Patsouris, S. J. Hamodrakas and V. A. Iconomidou (2016). "A common 'aggregation-prone' interface possibly participates in the self-assembly of human zona pellucida proteins." FEBS Lett **590**(5): 619-630.

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.

Lu, A., V. G. Magupalli, J. Ruan, Q. Yin, M. K. Atianand, M. R. Vos, G. F. Schroder, K. A. Fitzgerald, H. Wu and E. H. Egelman (2014). "Unified polymerization mechanism for the assembly of ASC-dependent inflammasomes." Cell **156**(6): 1193-1206.

Ludtke, S. J., P. R. Baldwin and W. Chiu (1999). "EMAN: Semiautomated software for high-resolution single-particle reconstructions." Journal of Structural Biology **128**(1): 82-97.

Mao, L., A. Kitani, M. Similuk, A. J. Oler, L. Albenberg, J. Kelsen, A. Aktay, M. Quezado, M. Yao, K. Montgomery-Recht, I. J. Fuss and W. Strober (2018). "Loss-of-function CARD8 mutation causes NLRP3 inflammasome activation and Crohn's disease." J Clin Invest **128**(5): 1793-1806.

Martinon, F., K. Burns and J. Tschopp (2002). "The inflammasome: a molecular platform triggering activation of inflammatory caspases and processing of proIL-beta." Mol Cell **10**(2): 417-426.

Massague, J. and Y. G. Chen (2000). "Controlling TGF-beta signaling." Genes Dev **14**(6): 627-644.

Matyszewski, M., W. Zheng, J. Lueck, B. Antiochos, E. H. Egelman and J. Sohn (2018). "Cryo-EM structure of the NLRC4(CARD) filament provides insights into how symmetric and asymmetric supramolecular structures drive inflammasome assembly." J Biol Chem **293**(52): 20240-20248.

McMullan, G., A. R. Faruqi and R. Henderson (2016). "Direct Electron Detectors." Methods Enzymol **579**: 1-17.

Medzhitov, R. and C. Janeway, Jr. (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.

Miao, E. A., C. M. Alpuche-Aranda, M. Dors, A. E. Clark, M. W. Bader, S. I. Miller and A. Aderem (2006). "Cytoplasmic flagellin activates caspase-1 and secretion of interleukin 1beta via Ipaf." Nat Immunol **7**(6): 569-575.

Miao, E. A., D. P. Mao, N. Yudkovsky, R. Bonneau, C. G. Lorang, S. E. Warren, I. A. Leaf and A. Aderem (2010). "Innate immune detection of the type III secretion apparatus through the NLRC4 inflammasome." Proc Natl Acad Sci U S A **107**(7): 3076-3080.

Miao, E. A. and S. E. Warren (2010). "Innate immune detection of bacterial virulence factors via the NLRC4 inflammasome." J Clin Immunol **30**(4): 502-506.

Mitchell, P. S., A. Sandstrom and R. E. Vance (2019). "The NLRP1 inflammasome: new mechanistic insights and unresolved mysteries." Curr Opin Immunol **60**: 37-45.

Mo, L., X. H. Zhu, H. Y. Huang, E. Shapiro, D. L. Hasty and X. R. Wu (2004). "Ablation of the Tamm-Horsfall protein gene increases susceptibility of mice to bladder colonization by type 1-fimbriated Escherichia coli." Am J Physiol Renal Physiol **286**(4): F795-802.

Monne, M., L. Han, T. Schwend, S. Burendahl and L. Jovine (2008). "Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats." Nature **456**(7222): 653-657.

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

Nishimura, K., E. Dioguardi, S. Nishio, A. Villa, L. Han, T. Matsuda and L. Jovine (2019). "Molecular basis of egg coat cross-linking sheds light on ZP1-associated female infertility." Nat Commun **10**(1): 3086.

Ogura, Y., N. Inohara, A. Benito, F. F. Chen, S. Yamaoka and G. Nunez (2001). "Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF-kappaB." J Biol Chem **276**(7): 4812-4818.

Ogura, Y., F. S. Sutterwala and R. A. Flavell (2006). "The inflammasome: first line of the immune response to cell stress." Cell **126**(4): 659-662.

Okumura, H., T. Sato, R. Sakuma, H. Fukushima, T. Matsuda and M. Ujita (2015). "Identification of distinctive interdomain interactions among ZP-N, ZP-C and other domains of zona pellucida glycoproteins underlying association of chicken egg-coat matrix." FEBS Open Bio **5**: 454-465.

Park, C. K. and N. C. Horton (2019). "Structures, functions, and mechanisms of filament forming enzymes: a renaissance of enzyme filamentation." Biophys Rev **11**(6): 927-994.

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.

Passmore, L. A. and C. J. Russo (2016). "Specimen Preparation for High-Resolution Cryo-EM." Methods Enzymol **579**: 51-86.

Peisley, A., B. Wu, H. Xu, Z. J. Chen and S. Hur (2014). "Structural basis for ubiquitin-mediated antiviral signal activation by RIG-I." Nature **509**(7498): 110-114.

Pellegrini, E., A. Desfosses, A. Wallmann, W. M. Schulze, K. Rehbein, P. Mas, L. Signor, S. Gaudon, G. Zenkeviciute, M. Hons, H. Malet, I. Gutsche, C. Sachse, G. Schoehn, H. Oschkinat and S. Cusack (2018). "RIP2 filament formation is required for NOD2 dependent NF-kappaB signalling." Nat Commun **9**(1): 4043.

Peng, H., B. Chang, C. Lu, J. Su, Y. Wu, P. Lv, Y. Wang, J. Liu, B. Zhang, F. Quan, Z. Guo and Y. Zhang (2012). "Nlrp2, a maternal effect gene required for early embryonic development in the mouse." PLoS One **7**(1): e30344.

Petit, C. (2006). "From deafness genes to hearing mechanisms: harmony and counterpoint." Trends Mol Med **12**(2): 57-64.

Poyet, J. L., S. M. Srinivasula, M. Tnani, M. Razmara, T. Fernandes-Alnemri and E. S. Alnemri (2001). "Identification of Ipaf, a human caspase-1-activating protein related to Apaf-1." J Biol Chem **276**(30): 28309-28313.

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.

Punjani, A., J. L. Rubinstein, D. J. Fleet and M. A. Brubaker (2017). "cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination." Nat Methods **14**(3): 290-296.

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.

Qu, Y., S. Misaghi, A. Izrael-Tomasevic, K. Newton, L. L. Gilmour, M. Lamkanfi, S. Louie, N. Kayagaki, J. Liu, L. Komuves, J. E. Cupp, D. Arnott, D. Monack and V. M. Dixit (2012). "Phosphorylation of NLRC4 is critical for inflammasome activation." Nature **490**(7421): 539-542.

Rampoldi, L., F. Scolari, A. Amoroso, G. Ghiggeri and O. Devuyst (2011). "The rediscovery of uromodulin (Tamm-Horsfall protein): from tubulointerstitial nephropathy to chronic kidney disease." Kidney Int **80**(4): 338-347.

Rathinam, V. A., S. K. Vanaja and K. A. Fitzgerald (2012). "Regulation of inflammasome signaling." Nat Immunol **13**(4): 333-342.

Robinson, K. S., D. E. T. Teo, K. S. Tan, G. A. Toh, H. H. Ong, C. K. Lim, K. Lay, B. V. Au, T. S. Lew, J. J. H. Chu, V. T. K. Chow, Y. Wang, F. L. Zhong and B. Reversade (2020). "Enteroviral 3C protease activates the human NLRP1 inflammasome in airway epithelia." Science **370**(6521).

Roh, J. S. and D. H. Sohn (2018). "Damage-Associated Molecular Patterns in Inflammatory Diseases." Immune Netw **18**(4): e27.

Romberg, N., K. Al Moussawi, C. Nelson-Williams, A. L. Stiegler, E. Loring, M. Choi, J. Overton, E. Meffre, M. K. Khokha, A. J. Huttner, B. West, N. A. Podoltsev, T. J. Boggon, B. I. Kazmierczak and R. P. Lifton (2014). "Mutation of NLRC4 causes a syndrome of enterocolitis and autoinflammation." Nat Genet **46**(10): 1135-1139.

Ronneberger, O., P. Fischer and T. Brox (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation, Cham, Springer International Publishing.

Rovida, C. (1873). "Conclusione degli studi intorno all'origine istologica dei cilindri dell'urina." Riv Clin Bologna **2**: 303-306.

Sagong, B., R. Park, Y. H. Kim, K. Y. Lee, J. I. Baek, H. J. Cho, I. J. Cho, U. K. Kim and S. H. Lee (2010). "Two novel missense mutations in the TECTA gene in Korean families with autosomal dominant nonsyndromic hearing loss." Ann Clin Lab Sci **40**(4): 380-385.

Sanchez-Garcia, R., J. Segura, D. Maluenda, C. O. S. Sorzano and J. M. Carazo (2020). "MicrographCleaner: A python package for cryo-EM micrograph cleaning using deep learning." J Struct Biol **210**(3): 107498.

Sandstrom, A., P. S. Mitchell, L. Goers, E. W. Mu, C. F. Lesser and R. E. Vance (2019). "Functional degradation: A mechanism of NLRP1 inflammasome activation by diverse pathogen enzymes." Science **364**(6435).

Sapio, M. R., M. A. Hilliard, M. Cermola, R. Favre and P. Bazzicalupo (2005). "The Zona Pellucida domain containing proteins, CUT-1, CUT-3 and CUT-5, play essential roles in the development of the larval alae in Caenorhabditis elegans." Dev Biol **282**(1): 231-245.

Saxena, M. and G. Yeretssian (2014). "NOD-Like Receptors: Master Regulators of Inflammation and Cancer." Front Immunol **5**: 327.

Schaeffer, C., S. Santambrogio, S. Perucca, G. Casari and L. Rampoldi (2009). "Analysis of uromodulin polymerization provides new insights into the mechanisms regulating ZP domain-mediated protein assembly." Mol Biol Cell **20**(2): 589-599.

Scheres, S. H. W. (2012). "RELION: Implementation of a Bayesian approach to cryo-EM structure determination." Journal of Structural Biology **180**(3): 519-530.

Schlauderer, F., T. Seeholzer, A. Desfosses, T. Gehring, M. Strauss, K. P. Hopfner, I. Gutsche, D. Krappmann and K. Lammens (2018). "Molecular architecture and regulation of BCL10-MALT1 filaments." Nat Commun **9**(1): 4041.

Schneider, M., A. G. Zimmermann, R. A. Roberts, L. Zhang, K. V. Swanson, H. Wen, B. K. Davis, I. C. Allen, E. K. Holl, Z. Ye, A. H. Rahman, B. J. Conti, T. K. Eitas, B. H. Koller and J. P. Ting (2012). "The innate immune sensor NLRC3 attenuates Toll-like receptor signaling via modification of the signaling adaptor TRAF6 and transcription factor NF-kappaB." Nat Immunol **13**(9): 823-831.

Serafini-Cessi, F., N. Malagolini and D. Cavallone (2003). "Tamm-Horsfall glycoprotein: biology and clinical relevance." Am J Kidney Dis **42**(4): 658-676.

Sharif, H., L. R. Hollingsworth, A. R. Griswold, J. C. Hsiao, Q. Wang, D. A. Bachovchin and H. Wu (2021). "Dipeptidyl peptidase 9 sets a threshold for CARD8 inflammasome formation by sequestering its active C-terminal fragment." Immunity.

Shen, C., A. Lu, W. J. Xie, J. Ruan, R. Negro, E. H. Egelman, T. M. Fu and H. Wu (2019). "Molecular mechanism for NLRP6 inflammasome assembly and activation." Proc Natl Acad Sci U S A **116**(6): 2052-2057.

Srinivasula, S. M., J. L. Poyet, M. Razmara, P. Datta, Z. Zhang and E. S. Alnemri (2002). "The PYRIN-CARD protein ASC is an activating adaptor for caspase-1." J Biol Chem **277**(24): 21119-21122.

Stanisich, J. J., D. S. Zyla, P. Afanasyev, J. Xu, A. Kipp, E. Olinger, O. Devuyst, M. Pilhofer, D. Boehringer and R. Glockshuber (2020). "The cryo-EM structure of the human uromodulin filament core reveals a unique assembly mechanism." Elife **9**.

Steubl, D., P. Buzkova, P. S. Garimella, J. H. Ix, P. Devarajan, M. R. Bennett, P. H. M. Chaves, M. G. Shlipak, N. Bansal and M. J. Sarnak (2020). "Association of serum uromodulin with mortality and cardiovascular disease in the elderly-the Cardiovascular Health Study." Nephrol Dial Transplant **35**(8): 1399-1405.

Sun, T. L., B. Zhou, L. H. Lai and J. F. Pei (2017). "Sequence-based prediction of protein protein interaction using a deep-learning algorithm." Bmc Bioinformatics **18**.

Sutterwala, F. S., L. A. Mijares, L. Li, Y. Ogura, B. I. Kazmierczak and R. A. Flavell (2007). "Immune recognition of Pseudomonas aeruginosa mediated by the IPAF/NLRC4 inflammasome." J Exp Med **204**(13): 3235-3245.

Suzuki, S., L. Franchi, Y. He, R. Munoz-Planillo, H. Mimuro, T. Suzuki, C. Sasakawa and G. Nunez (2014). "Shigella type III secretion protein MxiI is recognized by Naip2 to induce Nlrc4 inflammasome activation independently of Pkcdelta." PLoS Pathog **10**(2): e1003926.

Tamm, I. and F. L. Horsfall, Jr. (1950). "Characterization and separation of an inhibitor of viral hemagglutination present in urine." Proc Soc Exp Biol Med **74**(1): 106-108.

Tenthorey, J. L., N. Haloupek, J. R. Lopez-Blanco, P. Grob, E. Adamson, E. Hartenian, N. A. Lind, N. M. Bourgeois, P. Chacon, E. Nogales and R. E. Vance (2017). "The structural basis of flagellin detection by NAIP5: A strategy to limit pathogen immune evasion." Science **358**(6365): 888-893.

Thurber, K. R., Y. Yin and R. Tycko (2021). "Automated picking of amyloid fibrils from cryo-EM images for helical reconstruction with RELION." J Struct Biol **213**(2): 107736.

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.

Vance, R. E. (2015). "The NAIP/NLRC4 inflammasomes." Curr Opin Immunol **32**: 84-89.

Veesler, D., M. G. Campbell, A. Cheng, C. Y. Fu, Z. Murez, J. E. Johnson, C. S. Potter and B. Carragher (2013). "Maximizing the potential of electron cryomicroscopy data collected using direct detectors." J Struct Biol **184**(2): 193-202.

Wagner, T., L. Lusnig, S. Pospich, M. Stabrin, F. Schonfeld and S. Raunser (2020). "Two particle-picking procedures for filamentous proteins: SPHIRE-crYOLO filament mode and SPHIRE-STRIPER." Acta Crystallogr D Struct Biol **76**(Pt 7): 613-620.

Waksman, G. (2017). "Structural and Molecular Biology of a Protein-Polymerizing Nanomachine for Pilus Biogenesis." J Mol Biol **429**(17): 2654-2666.

Wang, L., Q. Qiao and H. Wu (2017). "Understanding CARD Tricks in Apoptosomes." Structure **25**(4): 575-577.

Wang, L., H. Sharif, S. M. Vora, Y. Zheng and H. Wu (2021). "Structures and functions of the inflammasome engine." J Allergy Clin Immunol **147**(6): 2021-2029.

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.

Wang, Q., H. Gao, K. M. Clark, C. S. Mugisha, K. Davis, J. P. Tang, G. H. Harlan, C. J. DeSelm, R. M. Presti, S. B. Kutluay and L. Shan (2021). "CARD8 is an inflammasome sensor for HIV-1 protease activity." Science **371**(6535).

Wang, R., Z. Wei, H. Jin, H. Wu, C. Yu, W. Wen, L. N. Chan, Z. Wen and M. Zhang (2009). "Autoinhibition of UNC5b revealed by the cytoplasmic domain structure of the receptor." Mol Cell **33**(6): 692-703.

Wassarman, P. M. and E. S. Litscher (2018). "The Mouse Egg's Zona Pellucida." Curr Top Dev Biol **130**: 331-356.

Weichhart, T., M. Haidinger, W. H. Horl and M. D. Saemann (2008). "Current concepts of molecular defence mechanisms operative during urinary tract infection." Eur J Clin Invest **38 Suppl 2**: 29-38.

Weiss, G. L., J. J. Stanisich, M. M. Sauer, C. W. Lin, J. Eras, D. S. Zyla, J. Truck, O. Devuyst, M. Aebi, M. Pilhofer and R. Glockshuber (2020). "Architecture and function of human uromodulin filaments in urinary tract infections." Science **369**(6506): 1005-1010.

Wen, H., E. A. Miao and J. P. Ting (2013). "Mechanisms of NOD-like receptor-associated inflammasome activation." Immunity **39**(3): 432-441.

Wong, S. H., L. Hamel, S. Chevalier and A. Philip (2000). "Endoglin expression on human microvascular endothelial cells association with betaglycan and formation of higher order complexes with TGF-beta signalling receptors." Eur J Biochem **267**(17): 5550-5560.

Wu, B., A. Peisley, C. Richards, H. Yao, X. Zeng, C. Lin, F. Chu, T. Walz and S. Hur (2013). "Structural basis for dsRNA recognition, filament formation, and antiviral signal activation by MDA5." Cell **152**(1-2): 276-289.

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.

Yang, H., C. Wu, S. Zhao and J. Guo (2004). "Identification and characterization of D8C, a novel domain present in liver-specific LZP, uromodulin and glycoprotein 2, mutated in familial juvenile hyperuricaemic nephropathy." FEBS Lett **578**(3): 236-238.

Yang, J., Y. Zhao, J. Shi and F. Shao (2013). "Human NAIP and mouse NAIP1 recognize bacterial type III secretion needle protein for inflammasome activation." Proc Natl Acad Sci U S A **110**(35): 14408-14413.

Yasumasu, S., M. Kawaguchi, S. Ouchi, K. Sano, K. Murata, H. Sugiyama, T. Akema and I. Iuchi (2010). "Mechanism of egg envelope digestion by hatching enzymes, HCE and LCE in medaka, Oryzias latipes." J Biochem **148**(4): 439-448.

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, Q. C., D. Petrey, L. Deng, L. Qiang, Y. Shi, C. A. Thu, B. Bisikirska, C. Lefebvre, D. Accili, T. Hunter, T. Maniatis, A. Califano and B. Honig (2012). "Structure-based prediction of protein-protein interactions on a genome-wide scale." Nature **490**(7421): 556-+.

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.

Zheng, S. Q., E. Palovcak, J. P. Armache, K. A. Verba, Y. F. Cheng and D. A. Agard (2017). "MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy." Nature Methods **14**(4): 331-332.

Zhong, F. L., O. Mamai, L. Sborgi, L. Boussofara, R. Hopkins, K. Robinson, I. Szeverenyi, T. Takeichi, R. Balaji, A. Lau, H. Tye, K. Roy, C. Bonnard, P. J. Ahl, L. A. Jones, P. J. Baker, L. Lacina, A. Otsuka, P. R. Fournie, F. Malecaze, E. B. Lane, M. Akiyama, K. Kabashima, J. E. Connolly, S. L. Masters, V. J. Soler, S. S. Omar, J. A. McGrath, R. Nedelcu, M. Gribaa, M. Denguezli, A. Saad, S. Hiller and B. Reversade (2016). "Germline NLRP1 Mutations Cause Skin Inflammatory and Cancer Susceptibility Syndromes via Inflammasome Activation." Cell **167**(1): 187-202 e117.

Zhong, F. L., K. Robinson, D. E. T. Teo, K. Y. Tan, C. Lim, C. R. Harapas, C. H. Yu, W. H. Xie, R. M. Sobota, V. B. Au, R. Hopkins, A. D'Osualdo, J. C. Reed, J. E. Connolly, S. L. Masters and B. Reversade (2018). "Human DPP9 represses NLRP1 inflammasome and protects against autoinflammatory diseases via both peptidase activity and FIIND domain binding." J Biol Chem **293**(49): 18864-18878.

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