THE STRUCTURAL BASIS OF TIGHT NUCLEOSOME TETHERING AND INACTIVATION OF CGAS

A Dissertation

by

PENGBIAO XU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee, Junjie Zhang
Committee Members, Pingwei Li
Jae-Hyun Cho
Robert Watson
Head of Department, Josh Wand

May 2021

Major Subject: Biochemistry

ABSTRACT

Cells defend against bacteria and viruses' invasion through their immune system. Innate immunity employs a wide spectrum of germ-line encoded pattern-recognition receptors (PRRs) to sense microbial infection and initiate a protective immune response. The detection of foreign nucleic acids is a central strategy in the innate immune system. The innate immune system has evolved to distinguish pathogenic RNA by secondary structure or specific modifications. However, self and non-self DNA are very similar in structure and therefore DNA recognition has more complicated systems. Because the cellular DNA normally presents in the nuclei and mitochondria, the accumulation of cytosolic DNA is a danger signal, either from pathogens or from the host itself.

Cyclic GMP-AMP synthase (cGAS) is a cytosolic DNA sensor that catalyzes the synthesis of a cyclic dinucleotide second messenger, c[G(2',5')pA(3',5')p] cGAMP. 2'3'-cGAMP binds to the adaptor STING (Stimulator of interferon genes protein), which forms puncta on the ER membrane and mediates the recruitment and activation of the protein kinase TBK1 and transcription factor IRF-3. Phosphorylated IRF-3 translocates to the nucleus and induces the expression of type I interferons.

Recently studies also identify that endogenous cGAS is predominantly a nuclear protein. cGAS locating inside the nuclei tethers tightly with chromatin and maintains the inactive state. In this work, a mechanism of cGAS being inhibited by nucleosomes is studied. A hybrid structure model of mouse cGAS in complex with human nucleosome is established. The binding of cGAS to the acidic patch on the nucleosome is discovered and

several key residues are identified for both cGAS and nucleosome for the interaction. This study determines that although in the existence of nucleosome DNA, the cGAS in the complex is in its inactive state and the complex clashes in space for binding extra dsDNA.

Another part of this thesis is dedicated to understanding the molecular mechanisms governing how cGAMP activates STING at the ER membrane. The N-terminal transmembrane domain of STING is essential for STING oligomerization and signaling. Moreover, cGAMP binding induces the oligomerization of full-length STING both in vitro and in vivo. The central hypothesis is that cGAMP binding makes the conformational changes of STING dimer which is located discretely on the ER membrane. cGAMP also increases the affinity of STING dimers to form oligomer, which facilitates the activation of kinase TBK1. These comprehensive structural and functional studies provide new insights into the mechanism of STING activation.

Taken together, this work presents an advanced understanding of the cGAS-STING pathway. These findings provide a structural basis for the development of cGAS and STING agonists and antagonists for the treatment of cancer and autoimmune disorders.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. Pingwei Li for his patient training, kind guidance, and unconditional support throughout my entire graduate studies. I would also like to thank all my committee members: Dr. Junjie Zhang, Dr. Robert Watson, and Dr. Jae-Hyun Cho for their valuable and constructive ideas and suggestions on my thesis.

I also want to thank all my lab members for their generous help with my research. Baoyu Zhao, a post-doc in my lab, taught me many techniques including expression and purification of proteins in various expression systems and surface plasmon resonance. Tao Jing, a graduate student, talked with me about my project and gave me sincere support. I want to thank graduate students Zhicheng Cui and Mengqiu Jiang in Dr.Junjie Zhang's group. They maintained our local TF20 microscope in very good condition and taught me the skills in data collection and processing of Cryo-EM.

In addition, I want to send special thanks to all the faculty and staff in the Department of Biochemistry and Biophysics. I took many useful courses taught by the faculty members, which broadened my horizons. Whenever I had any concerns about my study or life, the staff members were always willing to help me in a timely manner.

I also want to extend my gratitude to The UT Health Center in Houston for collecting the X-Ray diffraction data, the Welch Foundation, and the National Institute of Health for funding my research.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a thesis (or) dissertation committee consisting of Professor Junjie Zhang, Pingwei Li, and Jae-Hyun Cho of the Department of Biochemistry and Biophysics and Professor(s) Robert Watson of the Department of Microbial Pathogenesis & Immunology.

The Western blot and Fluorescent Microscopy analyzed for Chapter 2 was provided by Professor Phillip West. The analyses depicted in Chapters 2 and 3 were conducted in part by Dr. Baoyu Zhao of the Department of Biochemistry and were published in 2019 and 2020.

All other work conducted for the thesis (or) dissertation was completed by the student independently.

Funding Sources

Graduate study was supported in part by NIH under Grant Number AI145287. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

NOMENCLATURE

A	Alanine
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
ATP	Adenosine-5’-triphosphate
BLI	Biolayer interferometry
C	Cysteine
CBP	CREB-binding protein
cDNA	Complementary deoxyribonucleic acid
Caspase	Cysteine-aspartic proteases
CARD	Caspase recruitment domains
cGAMP	Cyclic guanosine monophosphate–adenosine monophosphate
cGAS	Cyclic GMP-AMP synthase
Cryo-EM	Cryogenic Electron Microscopy
CTD	Carboxyl-terminal domain
CTT	Carboxyl-terminal tail
D	Aspartic acid
DAPI	4'-6-diamidino-2-phenylindole
DDM	n-dodecyl-β-D-maltopyranoside
DTT	Dithiothreitol

dsDNA	Double-stranded deoxyribonucleic acid
E	Glutamic acid
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
F	Phenylalanine
FBS	Fetal bovine serum
FL	Full Length
Gln	Glutamine
Glu	Glutamic acid
H	Histidine
HEK	Human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
I	Isoleucine
IDT	Integrated DNA technologies
IFN	Interferon
IFN-I	Type-I interferons
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRF	Interferon regulatory factor
K	Lysine
KD	Kinase domain
L	Leucine

LB	Luria Bertani broth
LPS	Lipopolysaccharide
Lys	Lysine
MAVS	Mitochondrial antiviral signaling
MR	Molecular replacement
MS	Mass spectrometry
NF-kB	Nuclear factor kB
NLRP	NOD-, LRR- and pyrin domain-containing protein
NP	Nucleoprotein
OAS	Oligoadenylate synthase-like
P	Proline
PAGE	Polyacrylamide gel electrophoresis
PAMP	Pathogen-associated molecular pattern
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
PRR	Pattern recognition receptor
R	Arginine
RIG-I	Retinoic acid-inducible gene I
RLR	RIG-I-like receptor

r.m.s.d	Root mean square deviation
RNA	Ribonucleic acid
S	Serine
SDS	Sodium dodecyl sulphate
SEC	Size-exclusion chromatography
Ser	Serine
Sf9	Spodoptera frugiperda 9
SPR	Surface plasmon resonance
STING	Stimulator of interferon genes
SUMO	Small ubiquitin-like modifiers
TBD	Tank-binding domain
TBK1	TANK-binding kinase 1
Thr	Threonine
TLR	Toll-like receptor
TRIF	TIR domain-containing adaptor inducing IFN-β
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
V	Valine
W	Tryptophan
WT	Wild type
X	Any amino acid residue
Y	Tyrosine

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Innate immunity system

The immune system contains innate immunity and adaptive immunity. The adaptive immunity system generates a random and highly diverse repertoire of antigen receptors, followed by clonal selection and expansion of receptors with relevant specificities. Therefore, it generates immunological memory. The limitations of adaptive immunity are that it cannot tell the sources of the antigen and it usually takes more than 4 days to function [1-4]. The basic strategy of innate immunity recognition is to detect the constitutive and conserved products of microbes [1]. These products are usually the metabolic outputs that are unique to bacteria or viruses. For example, lipopolysaccharide (LPS) is a molecule from bacteria but not from eukaryotic cells. Therefore, the host eukaryotic cells consider the LPS as a signal from microbial invaders and initiate the response after sense it. Usually, the innate immunity responds to the common structure of the microbial product. The lipid-A portion of LPS is a quite conserved part for LPS from different species in bacteria. And the lipid-A is sufficient to activate the LPS sensitive pathway. The targets of the innate immunity recognition are conserved, which are called pathogen-associated molecular patterns (PAMPs). And the receptors of the innate immune system that recognizes PAMPs are called pattern-recognition receptors (PRR).

The innate immunity dates back to at least the 1908 Nobel Prize for Ilya Mechnikov and it undergoes rapid development in this century. Innate immunity is performed by both hematopoietic and nonhematopoietic origin [5]. It includes

macrophages, dendritic cells, mast cells, neutrophils, eosinophils, natural killer (NK) cells and NK T cells. At least three strategies of innate immunity are used, which contains detecting microbial non-self PAMPs, detecting common metabolic consequences of cell infection or injury, and detecting the missing self-components. Detecting microbial PAMP contains the receptors family-like Toll-like receptors (TLRs), NOD-like receptors, and collection family. The consequences of cell infections or injuries are detected by NOD-like receptors like NLRP3 and Receptor of advanced glycation and product (RAGE) family. The damage-associated molecular patterns (DAMPs) are an immunologic danger signal, including high mobility group box 1 protein and other endogenous alarmins, heat shock protein, and uric acid. The last one, to detect the missing self, MHC class I-specific inhibitory receptors are responsible for ensuring NK cells to attack infected cells. Some molecular is only expressed in normal healthy cells. Cells expressing these molecules will be marked as healthy and have an inhibited signal to immune response. Upon detecting the missing molecule, the inhibitory effect will diminish and NK cells would be recruited. KIR and CD94-NKG2A heterodimer are two examples for self MHC class I inhibitory signal [5-10].

Pattern recognition receptors

The pattern recognition receptors (PRR) mediates the initial sensing of the infection. The PRR includes Toll-like receptors, NOD-like receptors, RIG-I-like receptors, and C-type lectin receptors. After the sensing of the PAMP ligands, the PRRs initiate the cascade of the intracellular signaling to start the transcription of

proinflammatory cytokines, type I interferons (IFN), chemokines, and antimicrobial proteins to eliminate the invaded pathogens [11].

The TLRs are a class of the PRRs that senses a broad range of PAMPs [12]. The typical TLRs are type I transmembrane protein and contain three domains: a leucine-rich repeats (LRRs) motif, a single helix transmembrane domain, and a cytoplasmic Toll/IL-1 receptor (TIR) domain [12]. The LRRs are responsible for detecting the PAMPs, and the TIR domain works for interacting with the signaling adaptors to initiate the immunity. There are 10 TLRs been identified in human (TLR1-TLR10) and 13 in mouse (TLR1-TLR13). Different TLRs have specific ligand for detecting, for example, the TLR1/TLR2 forming a heterodimer which detects the triacylated lipopeptides, while the TLR6/TLR2 forming a heterodimer to detect the diacylated lipopeptides. The TLR5 homodimers sense the flagellin, and TLR4 homodimer, together with CD14 and MD2, senses the LPS (Figure 1). Other than the mentioned TLRs located on cytoplasm, there are some TLRs located on endosome membranes. The TLR3/7/8/9 detects nucleic acids and initiate the cascade to type I IFNs. After the TLRs sense the ligands, they form the activated dimer and recruit the adaptors such as MyD88 or TRIF. The MyD88 dependent response is shared by almost

all TLRs as signaling cascade, except for the TLR3, which utilizes the TRIF-dependent pathway.

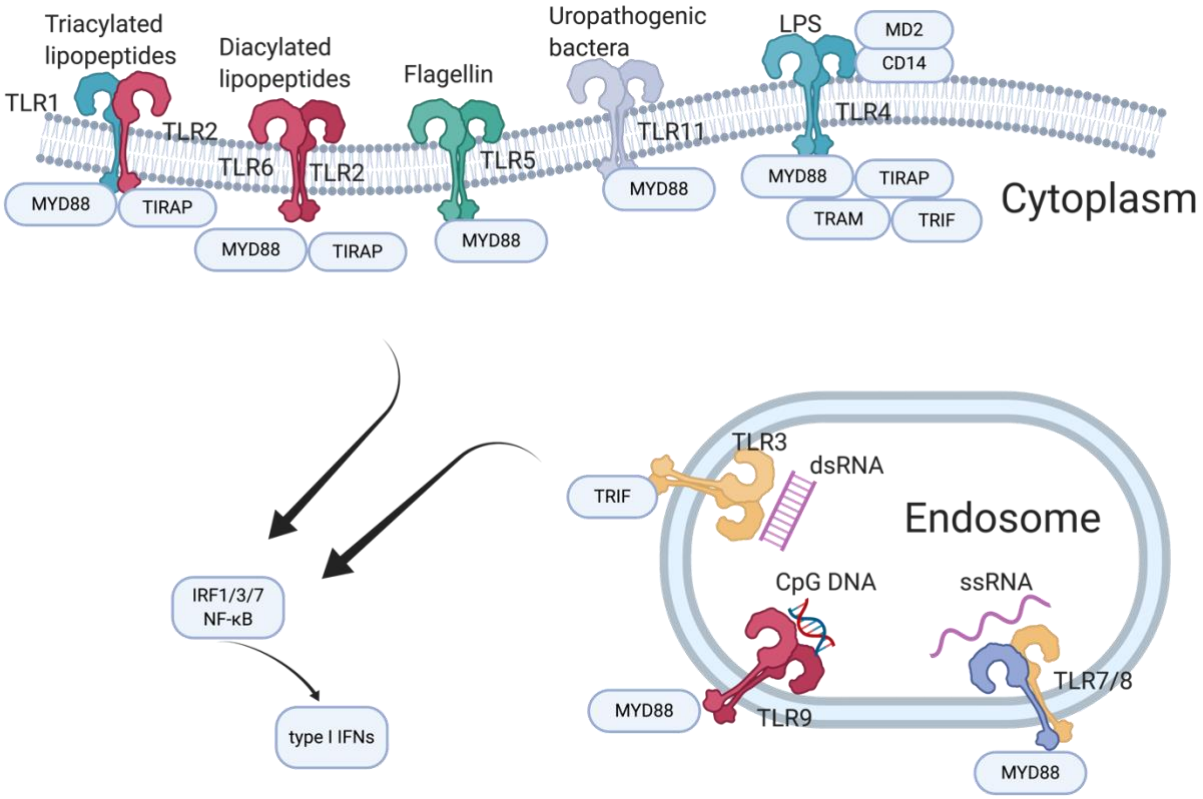


Figure 1. Schematic of TLRs

Nucleic acids sensors

Toll-like receptors

Among all the TLRs, TLR3, TLR7, TLR8, and TLR9 recognize nucleic acids from the bacteria and virus [11]. These TLRs activate the expression of type I interferons and cytokines. TLR3 detects the viral dsRNA in the endolysosome. It also detects the

polyinosinic polycytidylic acid (poly I: C), which is an analogy of dsRNA. The TLR7/8 recognizes ssRNA. TLR9 recognizes the unmethylated DNA with CpG motifs. CpG motifs are considered pathogenic products because they are abundant in the microbial genome but rare in the vertebrate genome [13-17]. The TLR7 and TLR9 are highly expressed in plasmacytoid DCs, producing type I IFNs in response to infection.

AIM2 and NOD-like receptors

The NOD-like receptors (NLR) are a large family consisting of 23 reported members in human. NLRs contain a C-terminal leucine-rich repeat, a nucleotide-binding oligomerization (NOD) domain, and various N-terminal protein binding motifs such as caspase activation and recruitment domain (CARD), Pyrin domain (PYD), or baculoviral IAP repeat (BIR) domain [18, 19]. The well-studied NLRs includes NOD1, NOD2, and NLR containing inflammasome. NOD1 detects the component of bacterial wall peptidoglycan and NOD2 senses muramyl dipeptide in both Gram-positive and Gram-negative strains of bacteria [20-24].

The inflammasome is cytosolic protein complexes formed by NLR with adaptor protein apoptosis-associated speck-like (ASC) protein containing CARD and procaspase-1. The NLR in inflammasome contains NLRP1, NLRP3, NLRP6, and NLRC4. When the stimuli or partner molecules are recognized and bound by the LRR domain, it induces a conformational change to release the autoinhibition of the NOD domain. The active NLRPs contain the clustered PYD that directs the filament formation of ASC through ASC's PYD. The ASC filament then recruits pro-caspase-1 through CARD/CARD interaction to form the active inflammasome. The formation of caspase-1 filaments

initiates proximity-driven dimerization and auto-proteolysis, and hence caspase activation. The mature and active caspase-1 has cysteine-aspartic proteases activity and cleaves pro-interleukin (IL) -1β and pro-IL-18, which secrets out of the cell [18]. NLRP1b and NLRC4 are two examples that do not have PYD and do not depend on ASC to form an inflammasome. They have CARD on the N-terminus and could recruit caspase-1 directly to form the inflammasome (Figure 2).

The inflammasome detects multiple kinds of danger and promotes the secretion of interleukin (IL) -1β and IL-18. A spectrum of seemingly unrelated stimuli via induction of K+ efflux appears to be the checkpoint to assemble the NLR inflammasome[25-30]. The NLPR3 is a well-studied NLR protein. NLRP3 inflammasome can be activated by nigericin, uric acid crystals, amyloid-β fibrils, and extracellular ATP (Figure 3). The mitotic kinase NEK7 licenses the assembly and activation of the NLRP3 inflammasome. The cryo-EM structure of the NLRP3 inflammasome with NEK7 is reported [31].

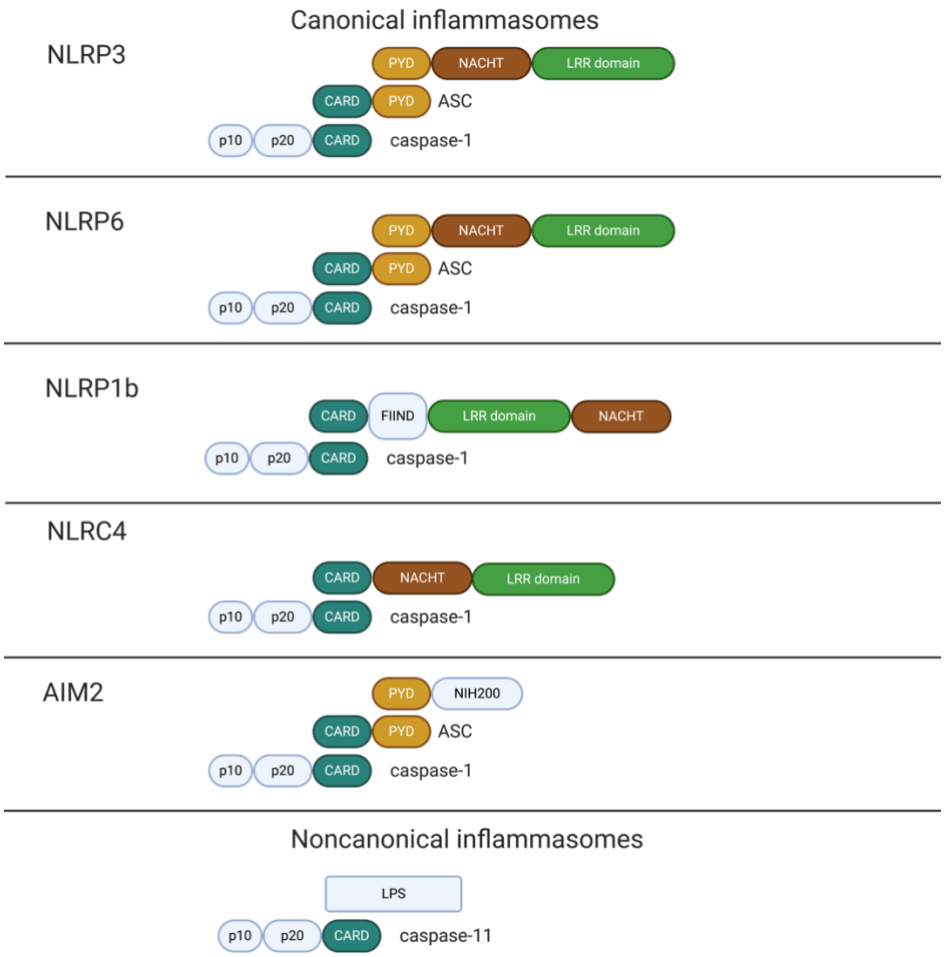


Figure 2. Inflammasome formation.

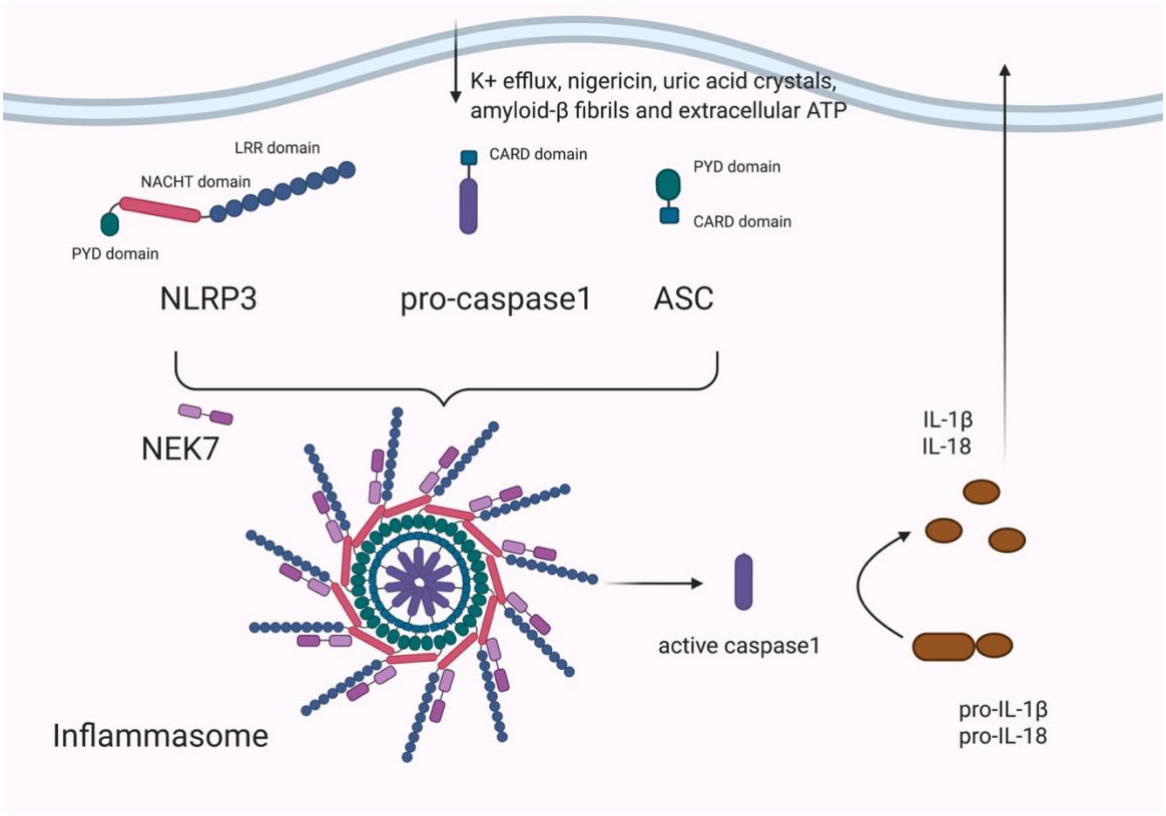


Figure 3. Mechanism of NLRP3 inflammasome.

A non-NLR PRR, absent in melanoma-2 (AIM2) is also identified as a component in inflammasome. AIM2 recognizes dsDNA in the cytosol and activates caspase-1. AIM2 is one of the hematopoietic nuclear PYHIN protein families [32-38]. AIM2 contains an N-terminal PYD and C-terminal NIH domain to recognize dsDNA in the cytosol. For long dsDNA in the cytosol, AIM2 binds it with the NIH domain and clusters the PYD. The PYD also recruits ASC's PYD and induces the filament formation of ASC. Analogously, the ASC filament recruits and pro-caspase-1 and facilitates the self-activation of caspase-1.

IFI16, an AIM2 like receptor, has a similar structure as AIM2 with PYD and NIH domain. It shuttles between the cytosol and the nucleus and functions as a nuclear pathogen sensor.

RIG-I-like receptors

RIG-I-like receptors (retinoic acid-inducible gene-I-like receptors, RLRs) are a type of PRRs that detect viral RNA in the cytoplasm. RLRs include the best-characterized receptor RIG-I and melanoma differentiation-associated 5 (MDA5), together with the laboratory of genetics and physiology 2 (LGP2). Unmodified, uncapped, full base-paired 5'-di/triohosphorylated ended RNA with a short blunt-ended double-stranded portion could activate RNA helicase RIG-I. These two features identify the viral RNA from self RNAs. The MDA5's physiological ligand has not been fully characterized. Unlike RIG-I prefers the short RNAs as ligands, the MDA5 selectively binds the long dsRNA[39-42].

OAS proteins

2'-5'-oligoadenylate synthetase is the enzyme that is encoded by the OAS gene. The OAS proteins are template-independent nucleotidyltransferases. OAS proteins sense A-form dsRNA in the cytosol and produces 2'-5' linked oligoadenylates as the second messenger. The 2'-5' linked oligoadenylates activate the enzyme RNase L to degrade host and viral RNA to limit viral propagation and induce apoptosis [43].

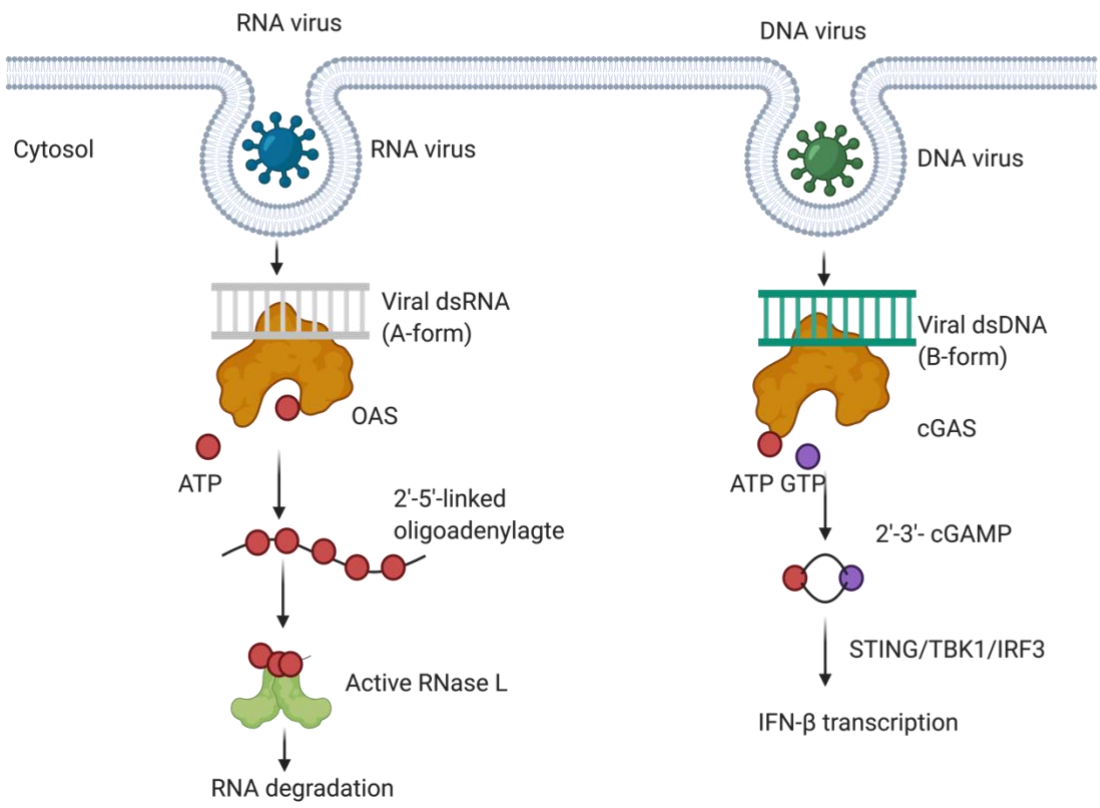


Figure 4. OAS protein and cGAS

Interferon and IFN-β

Interferon (IFN) mainly includes three classes: Type I IFN, Type II IFN, and Type III IFN. Type I IFN consists of many types, including IFN-α, IFN-β, and IFN-ω. Type I IFN all binds to a common receptor on the cell surface. Type II IFN has only one type, IFN-γ, which binds to a distinct cell surface receptor, type II IFN receptor. Type III IFN, which includes four members, IFN-λ1, IFN-λ2, IFN-λ3 and IFN-λ4, all bind to the type III receptor [44, 45]. The type I IFN receptor has two subunits, IFNAR1 and IFNAR2. The Janus activated kinases (JAKs) tyrosine kinase 2 (TYK2) and JAK1 are in the signal

cascade. TYK2 and JAK1 both bind to the IFN I receptor. Activation of the JAKs that are associated with the type I IFN receptor results in tyrosine phosphorylation of STAT2 (signal transducer and activator of transcription 2) and STAT1. It results in the formation of STAT1–STAT2–IRF9 (IFN-regulatory factor 9) complexes, which are known as ISGF3 (IFN-stimulated gene (ISG) factor 3) complexes. These complexes translocate to the nucleus, and bind IFN-stimulated response elements (ISREs) in DNA in order to initiate gene transcription. Both type I and type II IFNs also induce the formation of STAT1–STAT1 homodimers that translocate to the nucleus and bind GAS (IFN-γ-activated site) elements that are present in the promoter of certain ISGs, thereby initiating the transcription of these genes [46-51].

Under healthy conditions, a limited number of IFNs are produced [52]. Upon the viral infection, IFN-β is upregulated through TLR, RIG-I, and cGAS-STING pathway. This leads to the secretion of IFN-β out of the cell and binds to the IFN receptor, initiating the immune pathway for the infected and neighboring cells [53].

cGAS-STING pathway

Cells can defend against bacteria and virus's invasion through their immune system [11, 54-62]. Innate immunity employs a wide spectrum of germ-line encoded pattern-recognition receptors (PRRs) to sense microbial infection and initiate protective immune response[63-66]. Previous studies identified cyclic GMP-AMP synthase (cGAS) is a key dsDNA sensor in the cytosol [67-69]. cGAS is activated by dsDNA and catalyzes the synthesis of a noncanonical cyclic dinucleotide cGAMP containing mixed 2′,5′ and 3′,5′ phosphodiester linkages [70-75]. Unlike bacterial cyclic dinucleotides (CDNs) with

canonical 3'-5' linkages, 2'3'-cGAMP is a unique metazoan CDN that comprises both 3'-5'and a non-canonical 2'-5' phosphodiester linkages connecting adenosine with guanosine and guanosine with adenosine, respectively.

cGAMP, behaving as a second messenger, stimulates the induction of type I interferons (IFNs), such as IFN-α and β, through the adaptor STING located on the ER membrane [64, 68, 71, 76-78]. cGAMP binding induces conformational change and oligomerization of STING, leading to the recruitment and activation of TBK1 [68, 69, 74, 79]. The transcription factor IRF-3 is then recruited to the signaling complex and activated by TBK1 [80-82]. Mechanistically, STING employs a conserved pLxIS (p, hydrophilic residue, x, any residue, S, phosphorylation site) motif to recruit IRF-3 upon phosphorylation [83]. Phosphorylated IRF-3 oligomerizes, translocates to the nucleus [84, 85], and initiates the expression of type I IFNs [86-89]. The transcription factor NF-κB is also activated upon the stimulation of STING and is likely involved in the transcriptional regulation of IFN-β expression together with IRF-3 [54, 78, 90, 91]. Aberrant activation of innate immune responses by self nucleic acids causes autoimmune disorders such as systemic lupus erythematosus (SLE) and polyarthritis [92-99]. Several mutations of human STING cause autoimmune disorders [100, 101]. In addition to its roles in innate immunity, cGAMP and its phosphorothioate derivatives have potent antitumor activity and have applications in cancer immunotherapy [97, 102-104].

MAVS, STING, and TRIF as adaptors for IRF3

TBK1 is activated by multiple pathways. However, IRF3 phosphorylation is only observed in the IFN producing pathways including RIG-I-MAVS, cGAS-STING, and

TLR3/4-TRIF. It indicates the specificity of the mechanism of IRF3 phosphorylation, where the STING, MAVS, and TRIF work as adaptor proteins to recruit both IRF3 and TBK1 [83]. STING recruits IRF3 with the motif pLxIS (p, hydrophilic residue; x, any residue; S, phosphorylation site). Strikingly, the pLxIS motif is also involved in IRF-3 recruitment by MAVS and TRIF in the RIG-I like receptor (RLR) and Toll-like receptor (TLR) signaling pathways that sense viral RNA and bacterial lipopolysaccharides (LPS), respectively [11, 42, 83, 105]. The molecular mechanisms by which these adaptors recruit IRF3 upon phosphorylation are determined by structural biology [106]. The pLxIS motif of phosphorylated STING, MAVS, and TRIF binds to IRF3 in a similar manner, whereas residues upstream of the motif confer specificity.

Structure basis of cGAS-STING pathway

Recent structural and functional studies provide an overall picture and some molecular details of the cGAS-STING pathway [27, 30, 31, 34-38]. The determination of the structures of STING ligand-binding domain in isolation and in complex with c-di-GMP and cGAMP show that cGAMP binding induces a conformational change of the STING dimer. These comprehensive structural studies significantly advanced the understanding of DNA sensing through the cGAS-STING pathway. However, there are still a number of major gaps in our understanding of the cGAS-STING pathway [23, 24]. For example, the exact mechanism of how cGAMP activates STING at the ER membrane remains poorly understood. The recently reported structure of chicken STING in its activated form reaches 7Å, which is still not high resolution enough to validate the STING activation mechanism. It is also not clear how cGAMP binding by STING mediates the

recruitment and activation of TBK1. The mutagenesis study has found that several mutants on transmembrane domain block the IFN signaling but have no effect on cGAMP binding, indicating the N-terminal transmembrane is also vital for STING function. We will address these problems through comprehensive structural and functional studies. To elucidate the molecular basis of STING activation by cGAMP, we will determine the structures of cGAMP bound to full-length human STING (FL-STING) by a hybrid approach of X-ray crystallography and cryo-electron microscopy (cryo-EM). The long-term goal of this research is to elucidate how cGAMP and STING mediate the activation of TBK1 and IRF-3 in the cGAS-STING pathway. These rigorous and comprehensive structural and functional studies will enlighten the molecular mechanism of STING activation by cGAMP and also significantly advance our understanding of the mechanisms of innate immunity towards viral infection, providing a structural framework for conceptually new approaches to block undesired immune responses in autoimmune disorders or manipulate the innate immune responses to fight against cancer.

CHAPTER II

THE STRUCTURAL BASIS OF TIGHT NUCLEOSOME TETHERING AND INACTIVATION OF CGAS[1]

Introduction

The presence of billions of DNA in the mammalian genome raises the question of how the cGAS activity is inhibited to self-DNA. The prevalent explanation of this is that the cGAS localizes in the cytosol and have no access to the self-genome DNA in normal and healthy cells [58, 67, 107].

[1] Material from: Baoyu Zhao, Pengbiao Xu, Chesley M Rowlett, Tao Jing, Omkar Shinde, Yuanjiu Lei, A Phillip West, Wenshe Ray Liu, Pingwei Li. The molecular basis of tight nuclear tethering and inactivation of cGAS, Nature, published 2020.

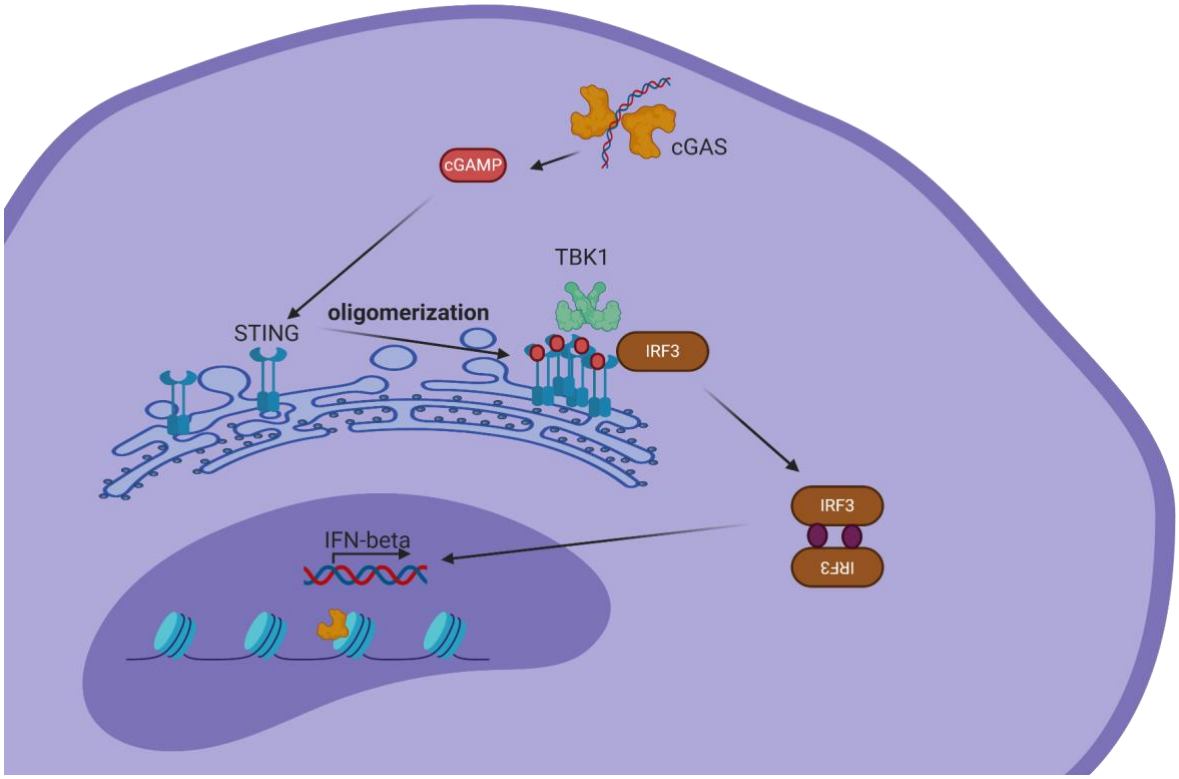


Figure 5. Schematic of the cGAS-STING pathway.

This idea is challenged by the recent studies that identified the overexpressed cGAS is located in micronuclei [108-112]. And further research supported that endogenous cGAS is also predominate the main nucleus protein [107, 113]. cGAS locating inside the nuclei tethers tightly with chromatin and maintains in the inactive state. The cGAS-chromatin tight interaction prevents cGAS autoreactivity against self-DNA. Further study also determines that cGAS has direct interaction with H2A&H2B dimer in vitro [114]. And the interaction between cGAS with nucleosomes is much higher than that of cGAS to free double-strand DNA. All of the experiments indicate that rather than

directly binding to nucleosomal DNA, there is a novel interaction for cGAS binding to nucleosomes.

cGAS binds tightly to individual nucleosomes

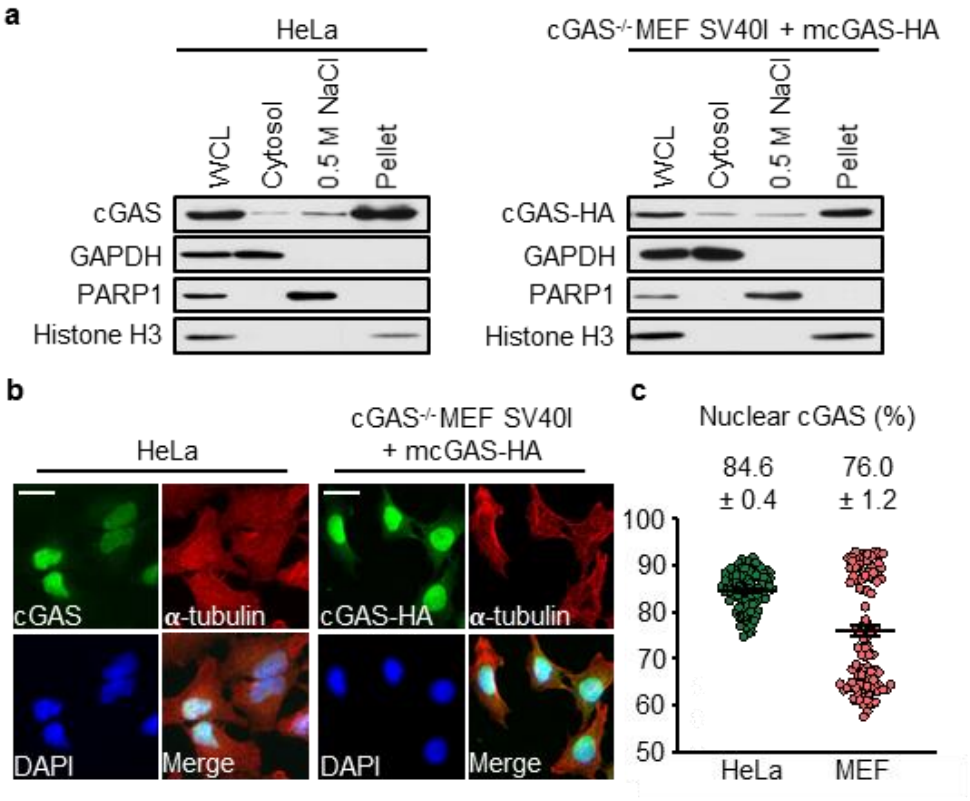


Figure 6. cGAS localizes in the nucleus of HeLa and MEF cells.

a. Western blot analyses of cGAS expression in 1% SDS whole cell lysates (WCL), 0.2% NP-40 cytosolic fractions, 0.5 M NaCl nuclear extracts, and 1% SDS nuclear lysates of HeLa and cGAS$^{-/-}$ MEF SV40I + mcGAS-HA cells. b. Immunofluorescence microscopy images of HeLa and cGAS$^{-/-}$ MEF SV40I + mcGAS-HA cells stained with anti-cGAS or –HA, and -α-tubulin antibodies and DAPI. The scale bars denote 50 µm. c. The percentage of nuclear- to whole cell-cGAS ratio quantified by microscopy. All data are presented as mean ± SEM.

Nucleosome binding inhibits the activity of cGAS

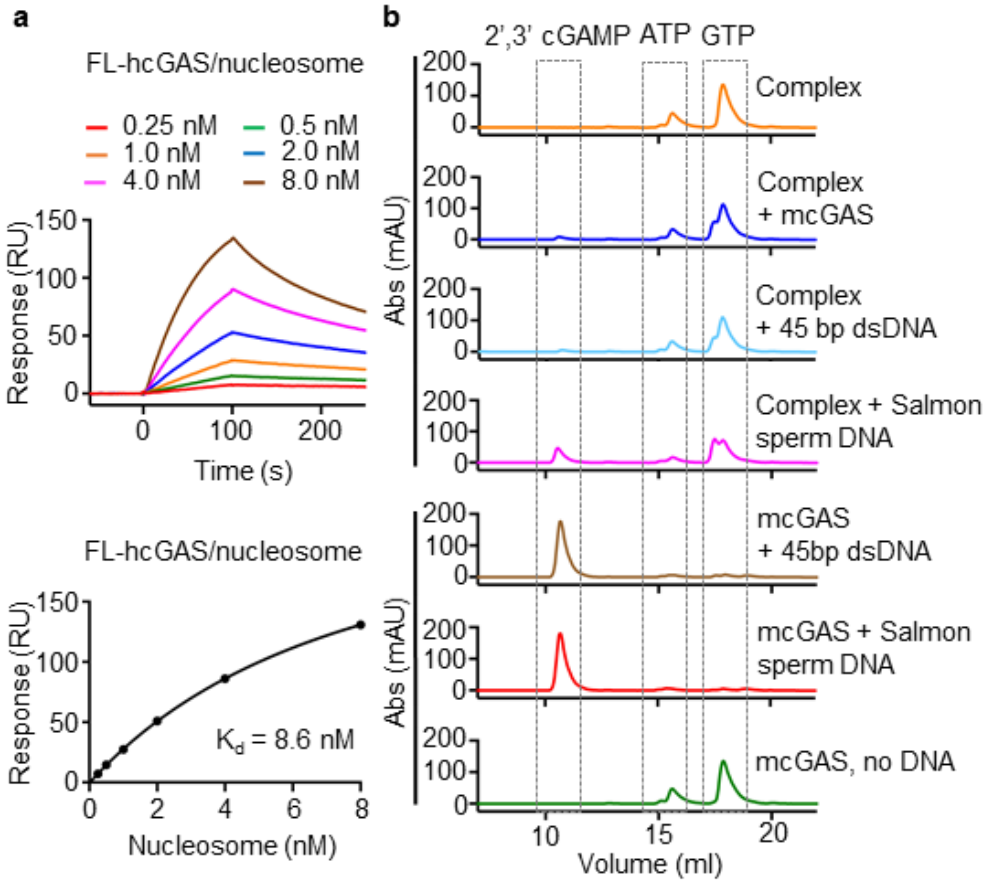


Figure 7. Nucleosome tethering inactivates cGAS.

a. Surface plasmon resonance (SPR) shows that full-length human cGAS binds to in vitro reconstituted human nucleosome with nanomolar affinity. The binding affinity (Kd) was determined by fitting the binding data to a simple one-to-one binding model. b. Enzyme activity assays by ion-exchange chromatography show that nucleosome binding potently inhibits the catalytic activity of cGAS.

According to the published results and our data, it was convinced that cGAS actually located in the nucleus as well as in the cytoplasm. Postdoc Dr. Baoyu Zhao did the preliminary work to test if cGAS binds to the nucleosome, which is the basic module of chromatin. Mono nucleosome was purified from HEK293T cells. And indeed, human full-length cGAS bound to nucleosome with a rather high, 8.6 nM, affinity. It was also interesting whether this tight binding would have any effects on the enzymatic activities of cGAS. Using ion-exchange chromatography, the binding of nucleosome was found to inactivate cGAS enzymatic function.

In the activity assay, different combinations of cGAS, dsDNA and nucleosomes were tested. The cGAS/nucleosome complex had no activities any more. Adding more free cGAS to the complex did not activate the synthase of cGAMP, indicating the cGAS cannot be activated by the dsDNA which wrapped around the nucleosome. In addition, adding more free dsDNA would not activate cGAS, indicating the cGAS in the complex is not able to bind dsDNA or to induce conformational changes to activate itself after binding (Figure 7).

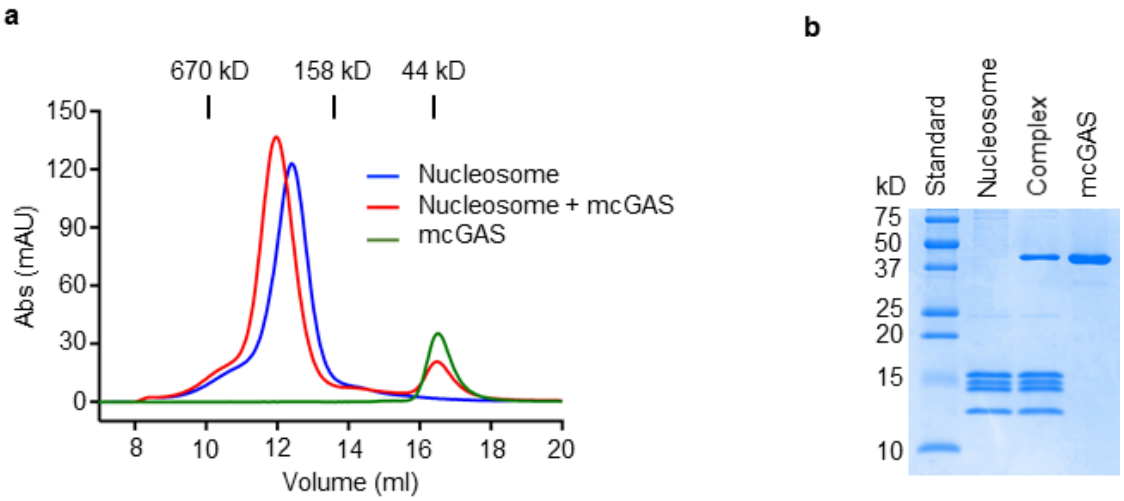


Figure 8. cGAS binds to mono-nucleosome in-vitro.
a. Size Exclusion Chromatography of Nucleosome, mouse cGAS catalytic domain, and the nucleosome-cGAS complex. b. The SDS-PAGE analyses of Nucleosome, mouse cGAS catalytic domain, and the nucleosome-cGAS complex.

After that cGAS binding tightly with nucleosome in vitro was validated, Size Exclusion Chromatography was used to purify the complex. By adding extra cGAS, the peak of the nucleosome left-shifted. The peak of the complex was collected and analyzed by SDS-PAGE, confirming it was the peak of the cGAS/Nucleosome complex. The complex was approximately 300kD on the SEC (Figure 8). The symmetric peak indicated a homogenous specimen, which was good to do structural study.

Structure determination of cGAS-nucleosome by cryo-EM

The preliminary study of the nucleosome by negative stain

The nucleosomes purified from HEK293T were tested by Electron Microscope with negative stain. When the specimen was diluted to 0.0045mg/ml, the white dots in the

figure was predicted to be the intact nucleosome. The black extended lines in the image might be the dissociated dsDNA from nucleosomes. When the specimen was diluted the 0.045mg/ml, the intact nucleosome was increased a lot. This was a good sign of good specimen in cryo-EM (Figure 9, Figure 10).

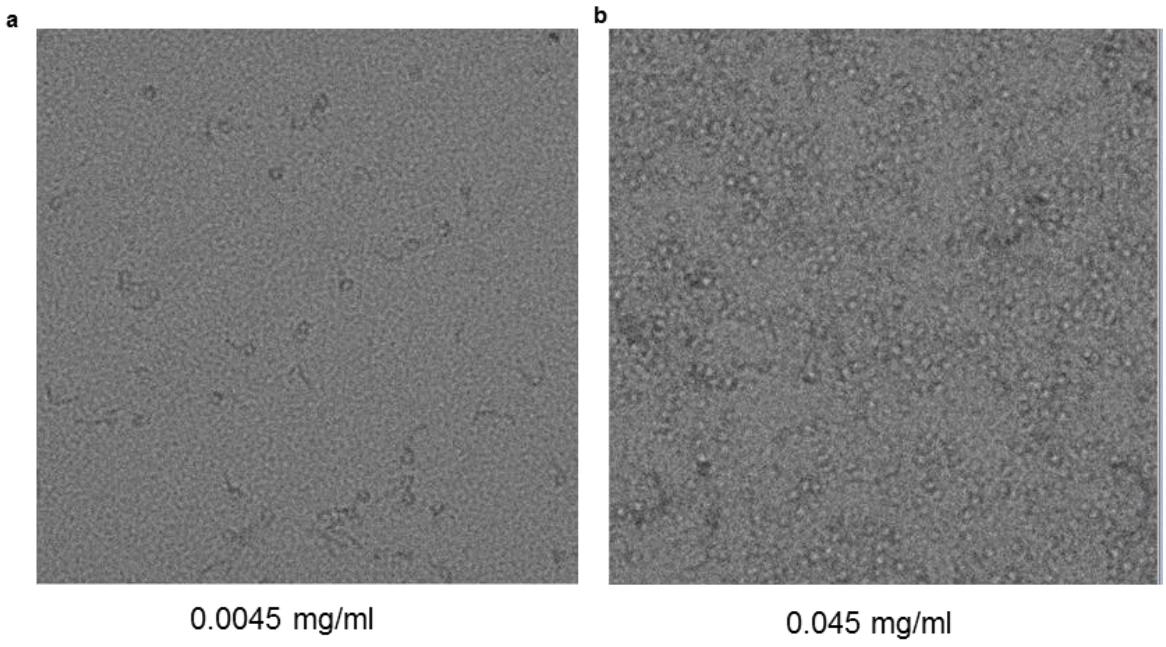


Figure 9. The negative stain representative micrographs of mono-nucleosome purified from HEK293T cells.

a. The negative stain representative micrographs of mono-nucleosome purified from HEK293T cells at a concentration of 0.0045mg/ml. b. The negative stain representative micrographs of mono-nucleosome purified from HEK293T cells at a concentration of 0.045mg/ml.

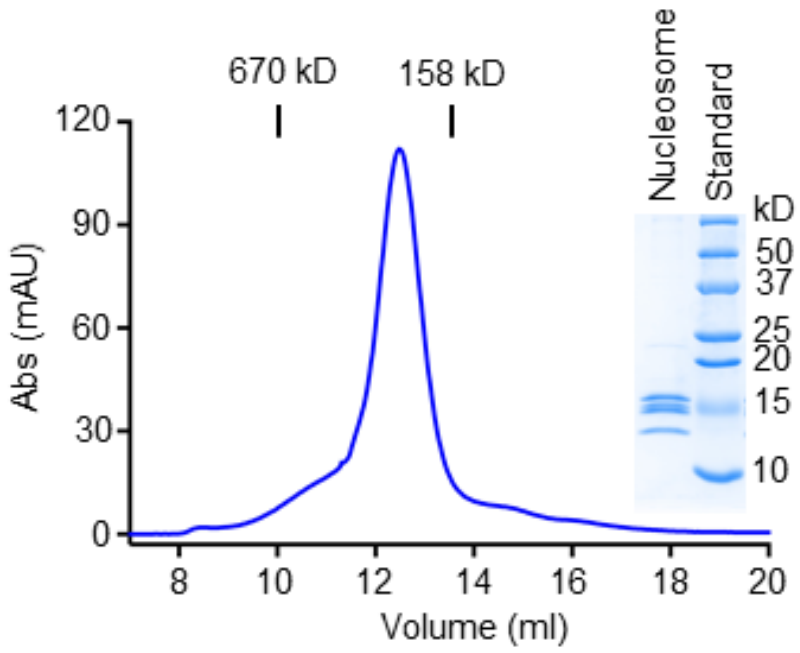


Figure 10. The Size Exclusion Chromatography of mono-nucleosome purified from HEK293T cells. The peak is analyzed by SDS-PAGE.

The cryo-EM study with a local TF20 microscope

After testing the nucleosome specimen, I tried cryo-EM data collection of cGAS/nucleosome complex with TF20 microscope in the campus. The particles in the cryo-EM had lower contrast than those in negative stain, but there were no aggregates or dissociation (Figure 11).

Around 400 micrographs were collected for the mcGAS/nucleosome complex specimen. Here, the nucleosome was still purified from HEK293T cells. Then Relion, cryosparc, and cisTEM were used individually to process the data. The picked particles

were around 80k for each software. All of the software produced 2D classifications with extra densities on the top of the nucleosome, which were directed by the blue arrows (Figure 12).

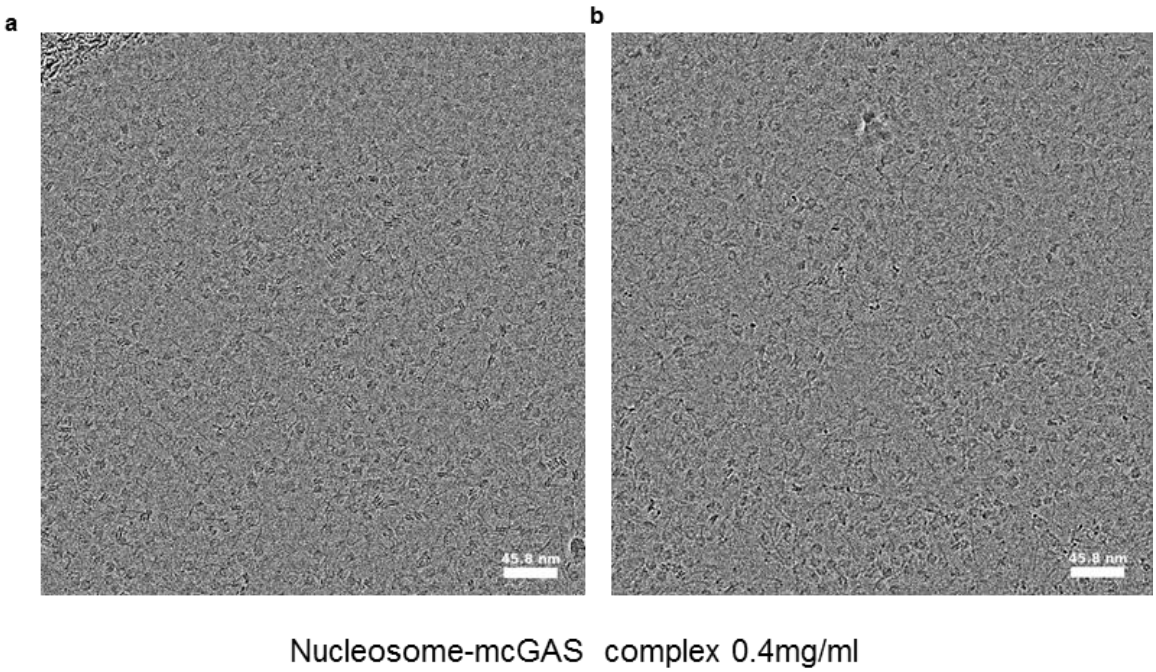


Figure 11. The Cryo-EM representative micrographs of mono-nucleosome purified from HEK293T cells with mouse cGAS at the complex concentration at 0.4 mg/ml.

a. and b. are images taken from a different position in the same grid.

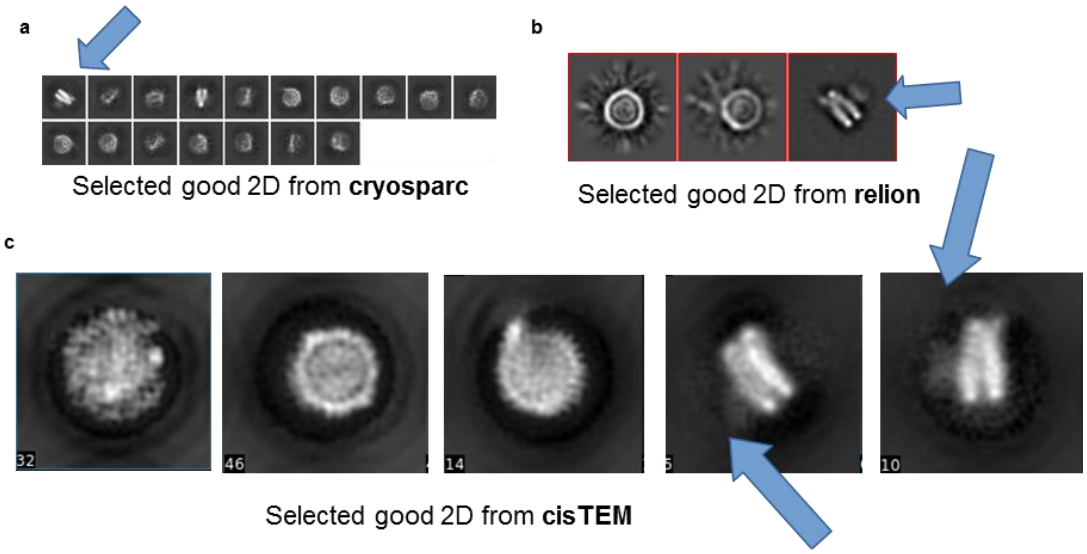


Figure 12. The representative 2D classification from different image processing software.

a. 2D classification of the nucleosome-cGAS complex from cryosparc. b. 2D classification of the nucleosome-cGAS complex from relion. c. 2D classification of the nucleosome-cGAS complex from cisTEM.

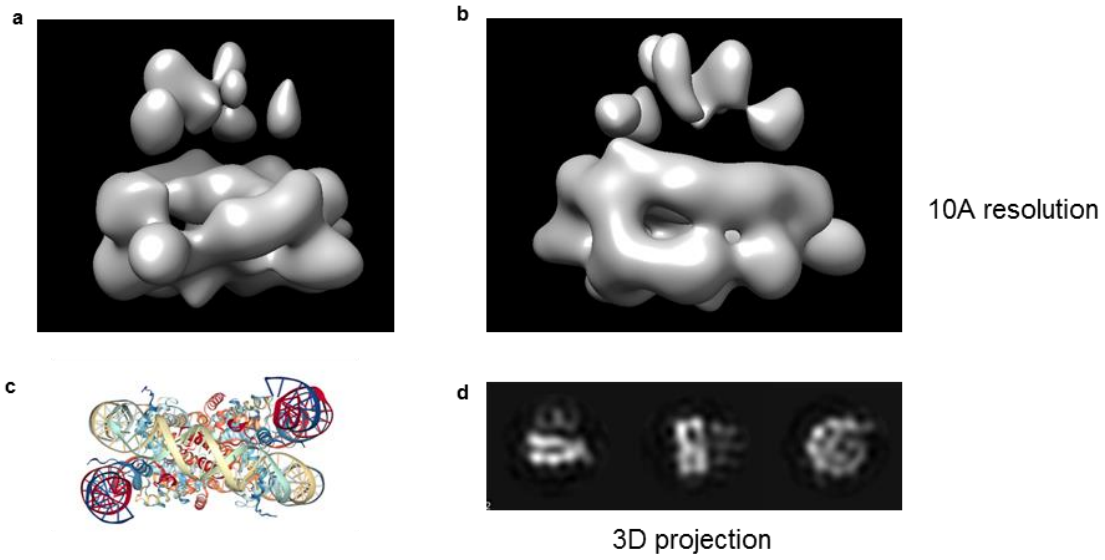


Figure 13. The 3D reconstruction of the nucleosome-cGAS complex from cisTEM.

a. and b. Two different orientations of the reconstructed structure. b. Nucleosome with the natural handedness. d. The back-projection of the reconstructed structure.

In generating a 3D reconstruction, Relion and cryosparc failed due to the low signal-to-noise ratio and limited particle amounts. cisTEM produced a 10Å resolution structure, reflecting there were some densities on the protein surface of nucleosomes. The 3D reconstruction actually had a wrong handedness compared with the real nucleosome. It was common in 3D reconstruction and could be easily post-processed to the correct handedness. The back-projection of this reconstruction also fitted well with the 2D averages, indicating a reasonable data processing (Figure 13).

At that stage, we were worried about the natural nucleosomes which may have different modifications and result in small conformational changes. It might be a problem for high resolution reconstruction. We found a collaboration with Prof. Wenshe Liu. Chesley in the Liu group produced reconstituted nucleosomes with histones purified from E. coli and 601 Widom DNA. This kind of nucleosomes have no modifications and should be more homogeneous (Figure 14).

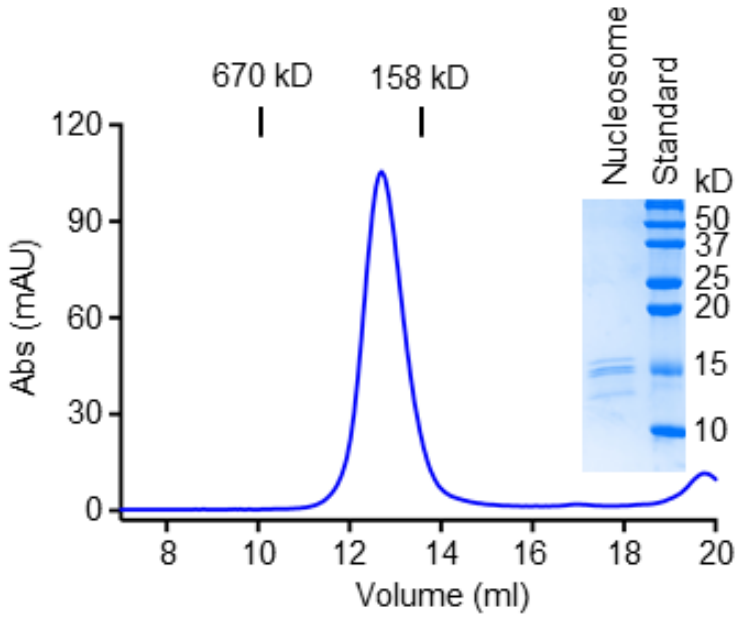


Figure 14. The Size Exclusion Chromatography of the reconstituted nucleosome.

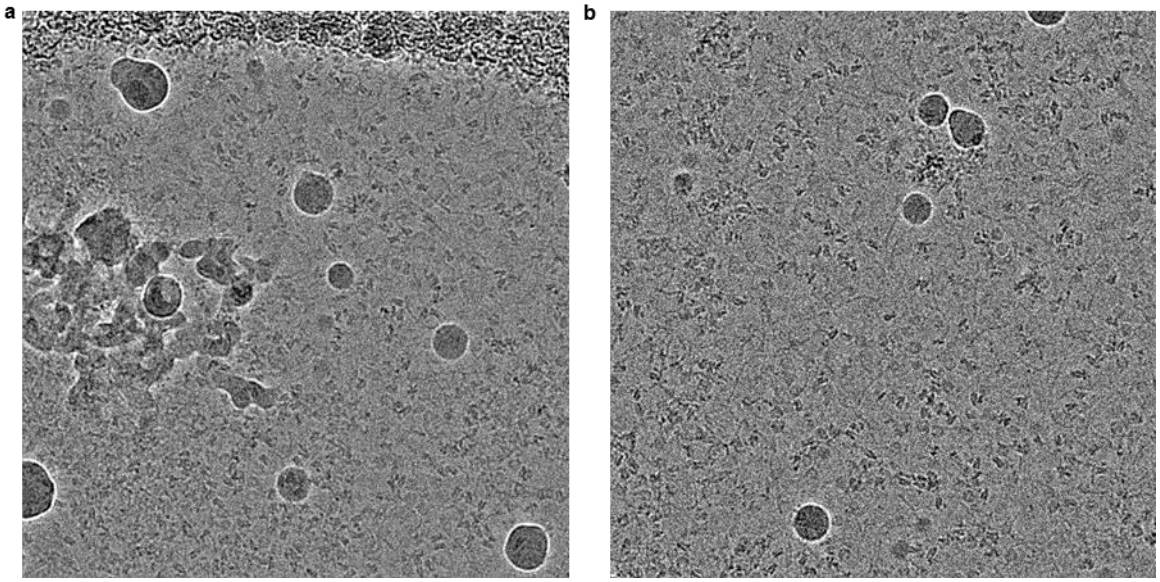


Figure 15. The Cryo-EM representative micrographs of reconstituted mono-nucleosome with mouse cGAS at the complex concentration at 0.4 mg/ml.
a. and b. are images taken from a different position in the same grid.

Thereafter, I made the cGAS/nucleosome complex via the same method on SEC for the reconstituted nucleosome. About 400 micrographs were collected for this specimen with our local microscope (Figure 15).

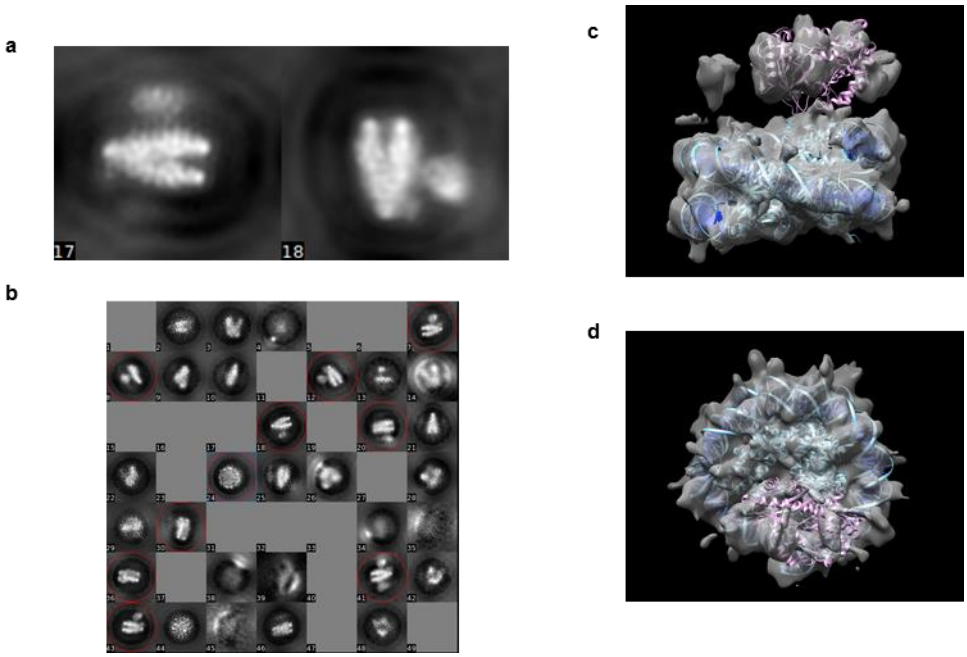


Figure 16. Data process of the reconstituted nucleosome with mcGAS from cisTEM.

a. The representative 2D class averages of the specimen. b. The overall 2D class average in the data process. c and d. The reconstructed structure in two different orientations.

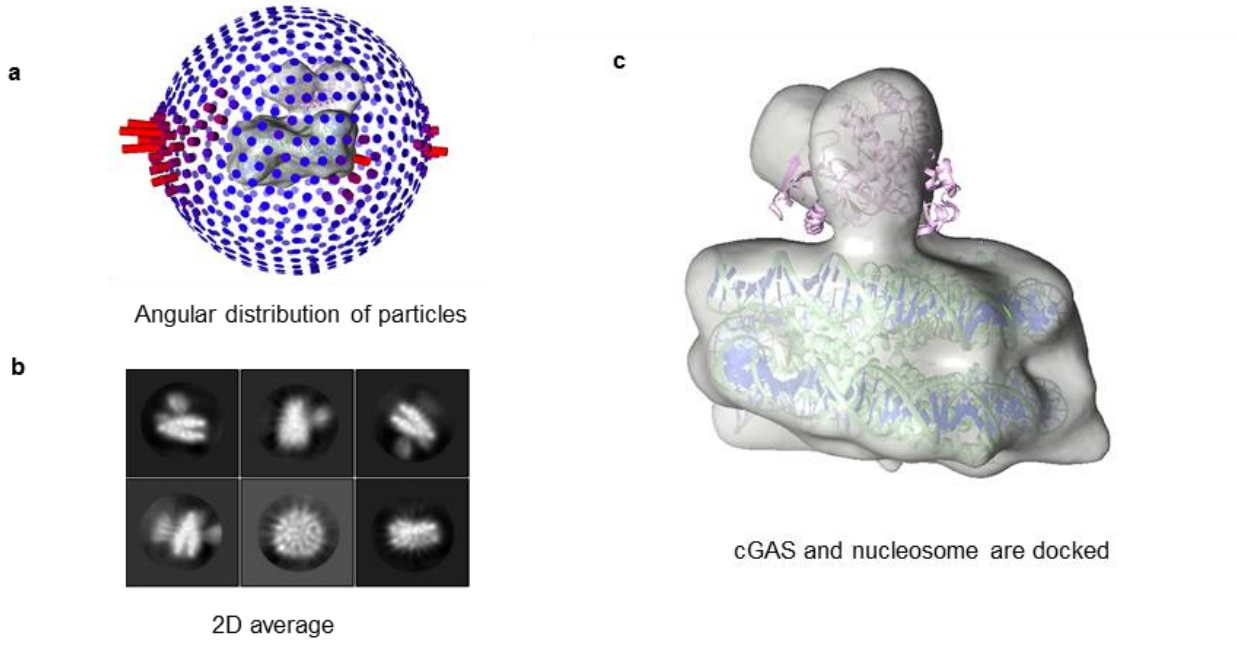


Figure 17. Data process of the reconstituted nucleosome with mcGAS from relion.

a. The particle distribution of the final particles used in the reconstruction. b. The representative 2D averages of the specimen. c. The reconstructed structure docked with the nucleosome and mouse cGAS domain.

For the reconstituted nucleosome/cGAS complex in cryo-EM, it contained more particles with side views, compared with the natural complex which preferred to be shown as top view. More side views benefited the data analyses. Both cisTEM (Figure 16) and Relion (Figure 17) could do a successful reconstruction this time. Typical 2D averages of the complex presented very clear protein densities on the nucleosome. It indicated the cGAS had a fixed binding site on the nucleosome. I could dock the cGAS monomer model into both the 3D reconstructions. The model fitted very well to the map. The angular

distribution of particles from Relion processing also delineated around six major particles' views, indicating that this specimen did not suffer from the preferred orientation.

The cryo-EM data collection and processing of Titan Krios microscope

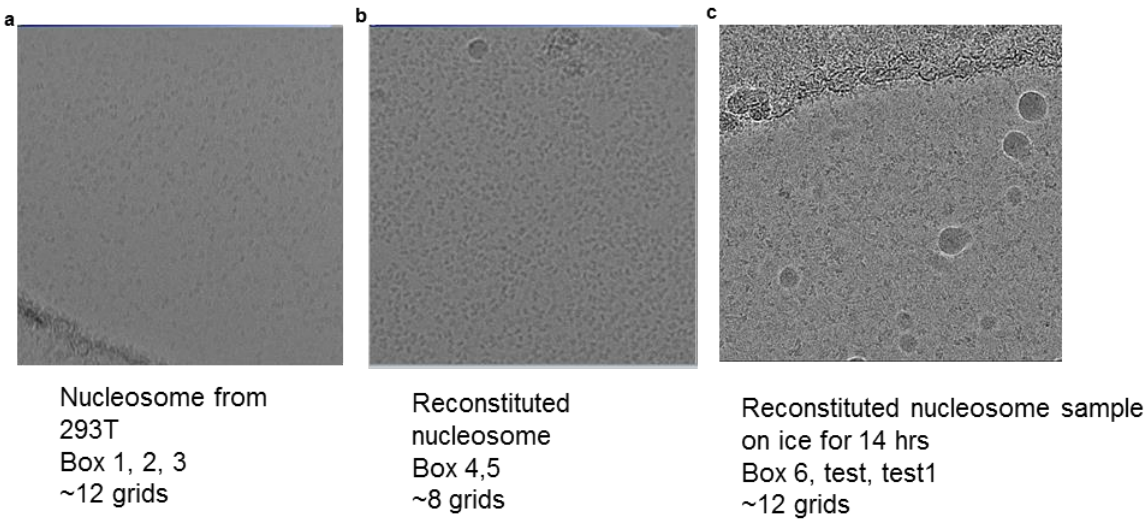


Figure 18. The representative micrographs of nucleosome-cGAS specimens were used for data collection with Titan Krios.

a. The representative micrograph of the nucleosome-mcGAS complex. The nucleosome is purified from HEK293T cells. b. The representative micrograph of the nucleosome-mcGAS complex. The nucleosome is reconstituted with Histones expressed in E. coli. c. The representative micrograph of the nucleosome-mcGAS complex. The nucleosome is reconstituted with Histones expressed in E. coli. The complex is kept on ice for 14 hours before vitrification.

To reconstruct the map to high resolution, a higher-end microscope is required. I applied for the data collection session at UT health center in Houston. 12 grids were prepared with HEK293T-purified nucleosome/mcGAS specimen. I also made 8 grids in

two boxes with the reconstituted nucleosome/mcGAS specimen. The next day, after 14 hours, I made another 12 grids in three boxes, considering that grids might not enough for the reconstituted one. All of the grid groups were tested by selected one to two grids with our local microscope (Figure 18). The particle distribution and ice thickness optimization were verified. All grids showed a monodispersed particle distribution and good ice thickness to make sure enough area can be collected.

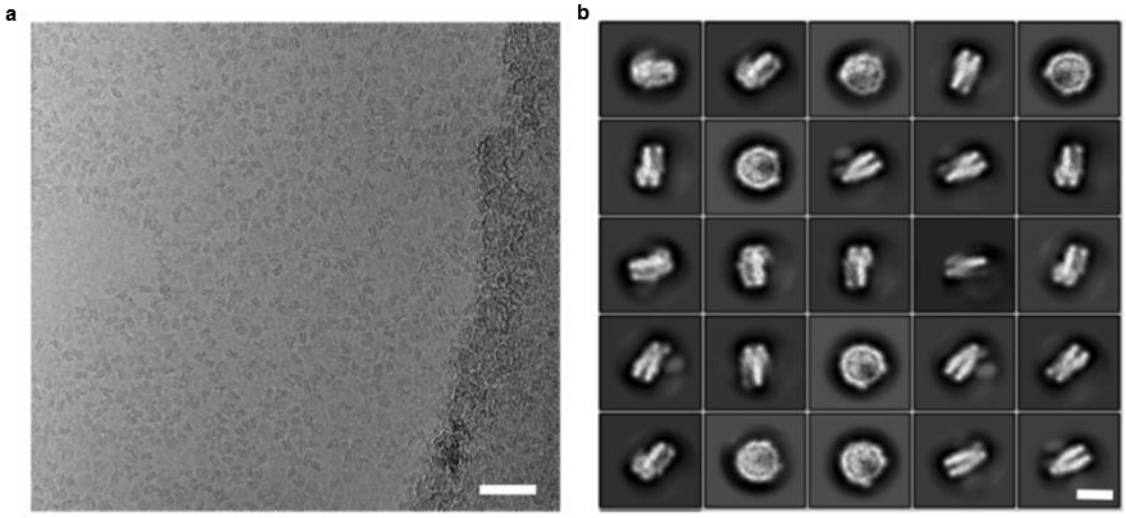


Figure 19. The data collection of the reconstituted nucleosome-mcGAS complex.

a. Representative micrograph of the mcGAS-nucleosome complex in vitrified ice. The scale bar denotes 50 nm. b. 2D class averages of mcGAS-nucleosome complex particles. The scale bar denotes 10 nm.

With the help of UT Health center in Houston, we collected 5354 movies with Titan Krios operated at 300kV and equipped with a K2 direct electron camera (Figure 19).

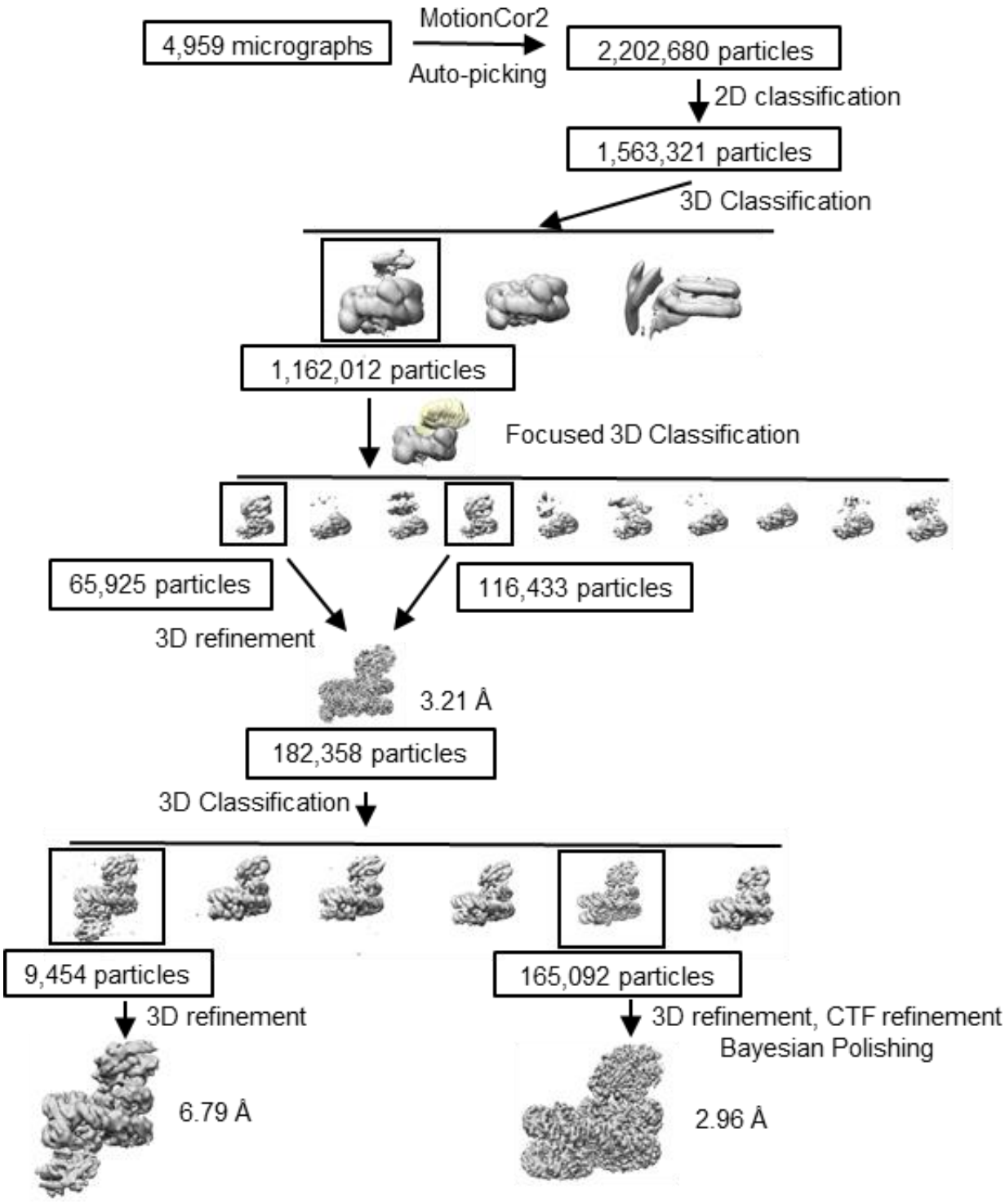


Figure 20. Flowchart of data processing; see Methods for details.

The collected stack movies were aligned by Motioncor2. After CTF estimation by GCTF, the micrographs were screened by max signal resolution higher than 5Å, leaving 4959 good micrographs. Particle picking was performed by Relion auto-picking with templates. The templates were generated with 100 manually picked particles from the same datasets in advance. After several rounds of 2D classification to remove the false picked particles, crystalline ice, and grid edges, the particles were submitted to 3D classification to identify all the nucleosome/cGAS complexes (Figure 20).

From the collected datasets, the nucleosome/cGAS (1:1) was reconstructed to 2.96Å. The generated map presented clearly that the one cGAS bound to the acidic patch of the nucleosome. The question raised that why there was no cGAS on the other side of the nucleosome. It might because the cGAS binding changed the conformation for the nucleosome so that it closed the opposite binding site. However, when I flipped and aligned the map to compare the empty acidic patch and the cGAS bound acidic patch, the two acidic patches were exactly the same.

This comparison presented the idea that cGAS binding did not make conformation changes on the nucleosome. And it also indicated that the other nucleosome acidic patch could bind another cGAS.

One more iteration of 3D classification was done and succeeded to produce a class showing two cGAS bound to the nucleosome. This class contained around 10k particles and was refined to 6.79Å. The helices of both cGAS could be determined in the map (Figure 21).

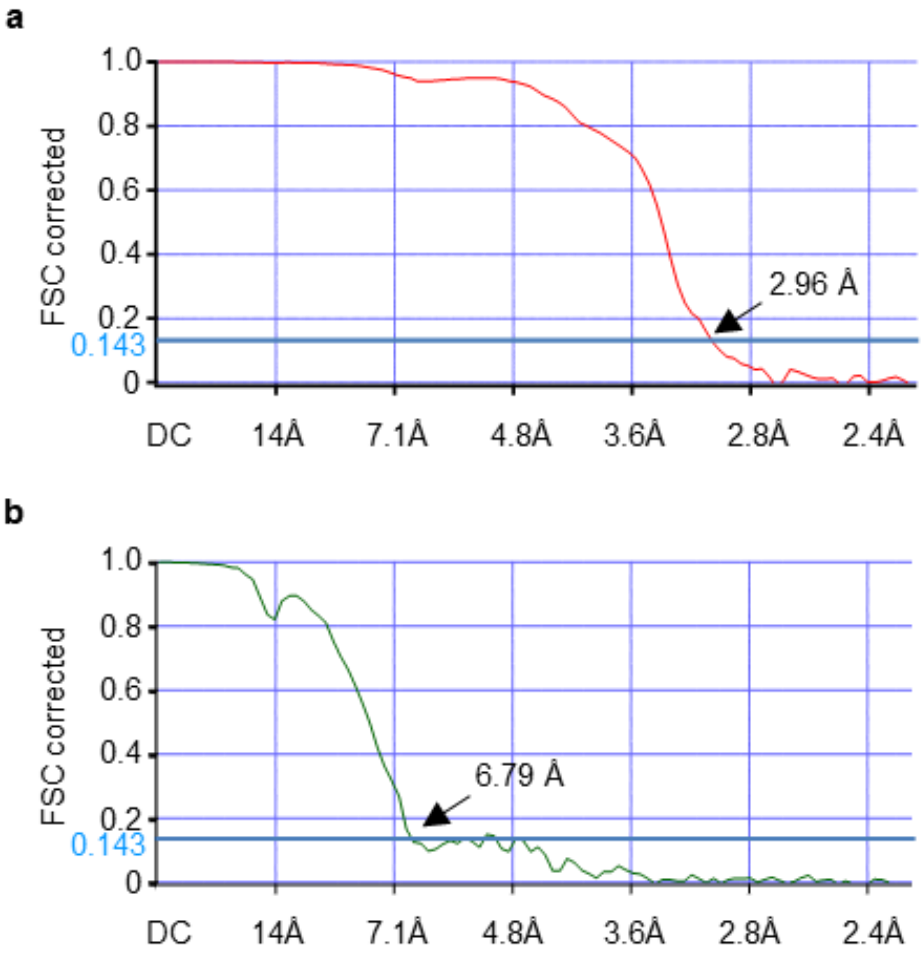


Figure 21. The corrected FSC curve of the data process.
a. The corrected FSC for the structure of nucleosome-cGAS (1:1). b. The corrected FSC for the structure of nucleosome-cGAS (1:2).

The structural have relatively good particle distributions with more than 6 major particle views and almost no missing views. And overall speaking, the resolution is high in the core part of nucleosome and cGAS/nucleosome interaction regions, but it goes lower

to the outer side, mainly because the cGAS itself has some flexibilities in structure (Figure 22, Figure 23, Figure 24, Figure 25).

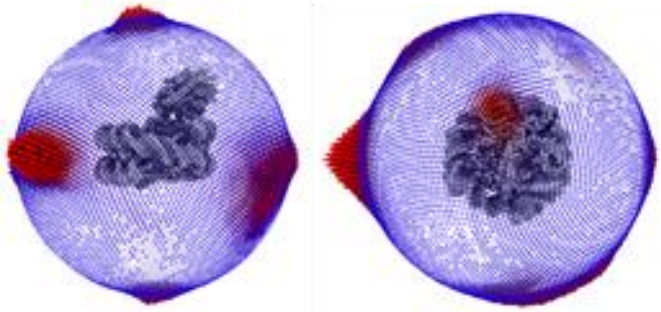


Figure 22. The angular distribution of the particles finally used in the reconstruction nucleosome-cGAS (1:1).

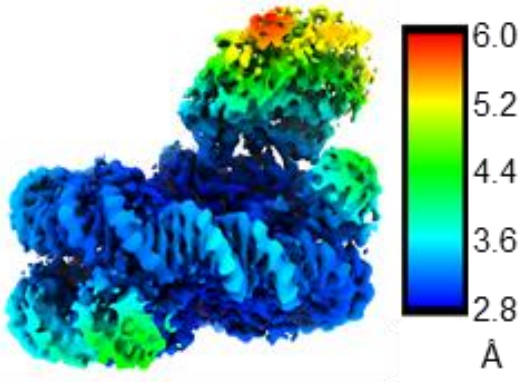


Figure 23. The local resolution of the reconstructed structure nucleosome-cGAS (1:1).

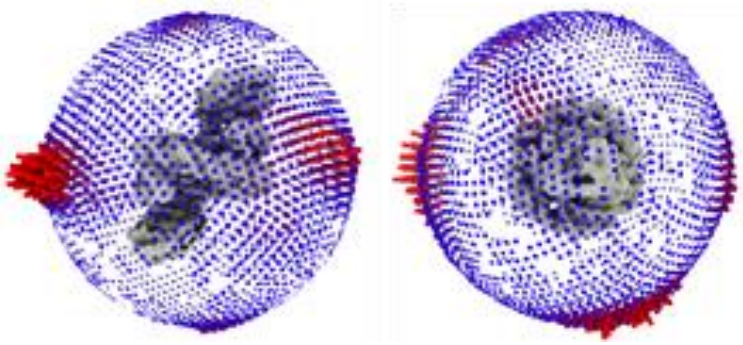


Figure 24. The angular distribution of the particles finally used in the reconstruction nucleosome-cGAS (1:2).

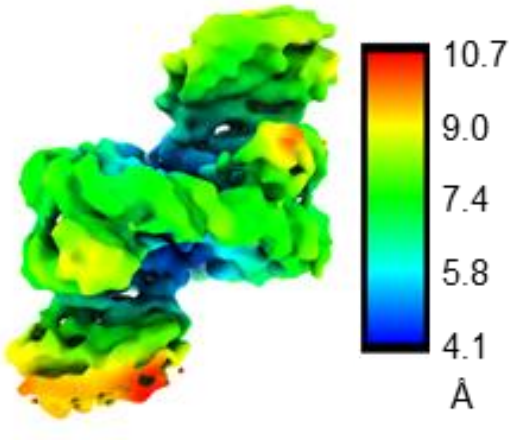


Figure 25. The local resolution of the reconstructed structure nucleosome-cGAS (1:2).

For the 1:1 map, as the reconstruction went high to 2.96Å, this atomic resolution helped to build a precise model by docking the crystal structure into the map and manually checking and adjusting the residuals. For the 2:1 map, due to the low resolution of about 7Å, I used the real-space refinement to dock the 1:1 structure model into it without any more adjusting (Figure 26, Figure 27).

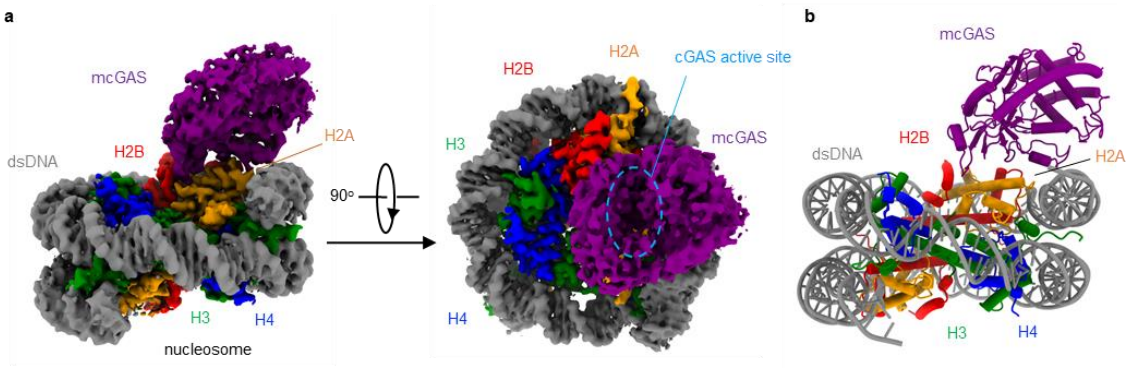


Figure 26. Cryo-EM structure of mouse cGAS catalytic domain bound to human nucleosome (1:1).

a. Cryo-EM density map of the mcGAS-nucleosome complex at 2.96 Å resolution in two different orientations contoured at 3σ. b. Ribbon representation of the mcGAS-nucleosome complex structure.

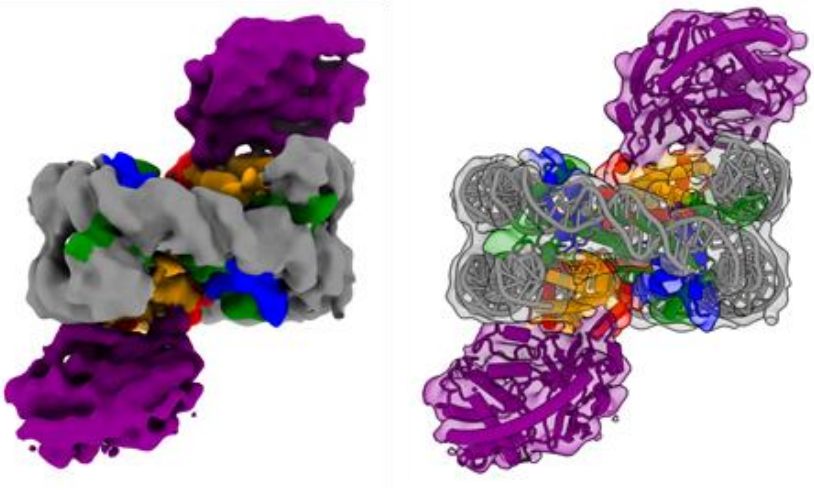


Figure 27. Cryo-EM structure of mouse cGAS catalytic domain bound to human nucleosome (1:2).

a. Cryo-EM density map of the mcGAS-nucleosome complex at 6.79 Å resolution in two different orientations contoured at 3σ. b. Ribbon representation of the mcGAS-nucleosome complex structure docked into the map.

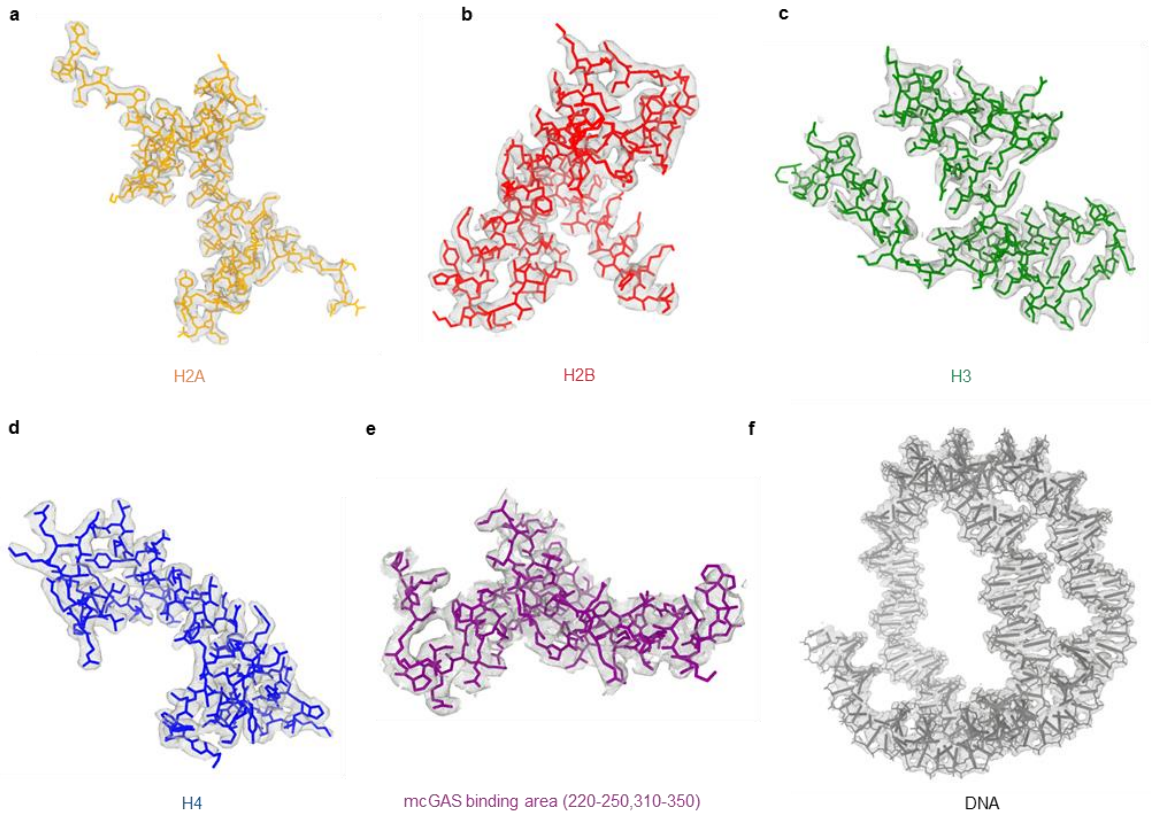


Figure 28. Density maps and structural models of cGAS-nucleosome (reconstituted, 1:1) complex.

a - f. The density maps (grey mesh) of histones H2A, H2B, H3, H4, part of mouse cGAS catalytic domain, and the Widom 601 nucleosome positioning sequence DNA contoured at 3σ. The protein and DNA structures fitted into the density map are shown by the stick models.

The cGAS was colored with purple, Histone H2A with yellow, H2B with red, H3 with green, H4 with blue, and dsDNA with grey respectively (Figure 28). By adjusting the transparency, the figures presented that both the 1:1 and 2:1 model fitted the map densities very well.

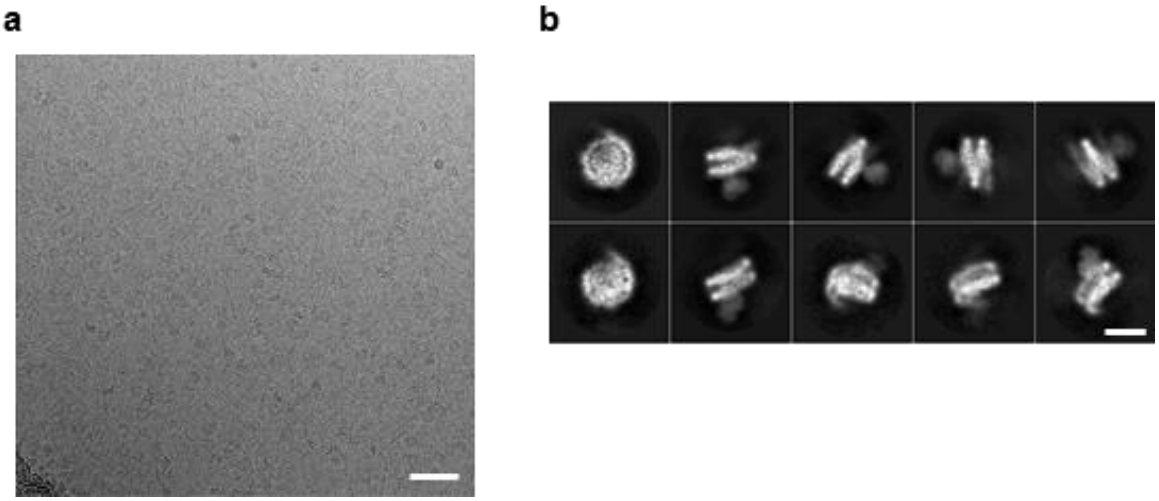


Figure 29. Cryo-EM analysis of mcGAS domain in complex with nucleosome purified from HEK 293T cells.

a. Representative micrograph of the mcGAS-nucleosome complex in vitrified ice. The scale bar denotes 50 nm. b. 2D class averages of mcGAS-nucleosome complex particles. The scale bar denotes 10 nm.

We also collected 2979 movie stacks for the 293T nucleosome/cGAS specimen. The data processing was very similar to that of the reconstituted specimen (Figure 29).

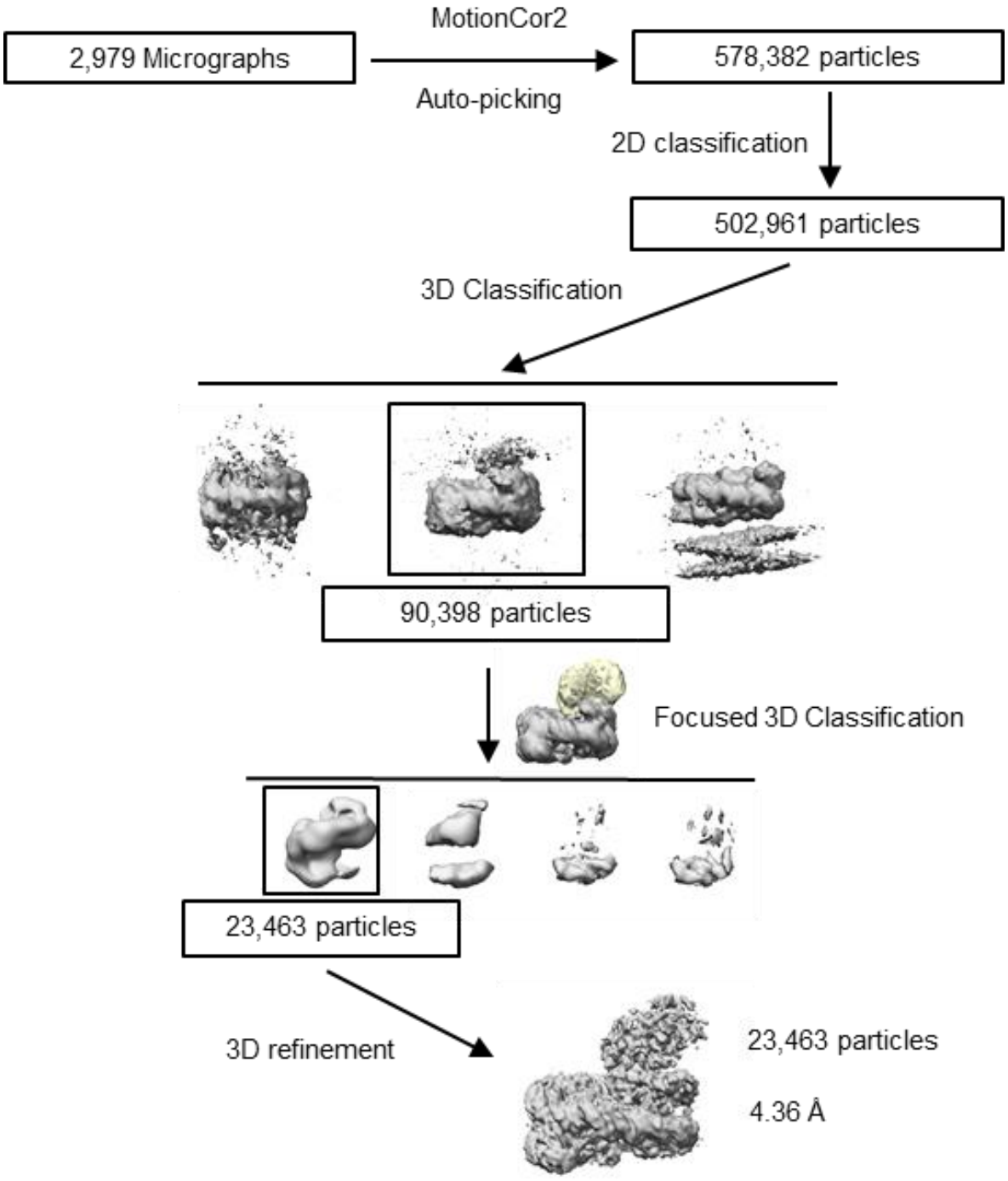


Figure 30. Flowchart of data processing; see Methods for details.

The representative 2D averages showed a clear cGAS density on the nucleosome. After several rounds of 2D classification and 3D classification, I reconstructed the structure to 4.36Å. Due to a heterogeneous DNA sequence, the DNA resolution was low to 5Å. The overall resolution was not high enough to build an atomic model, however, the cGAS can be unambiguously docked into the map. The interaction region had a resolution around 4Å, which was enough for us to conclude that cGAS bound to the natural nucleosome with the same residues and types as the reconstituted complex with higher resolution (Figure 30, Figure 31, Figure 32, Figure 33, Figure 34).

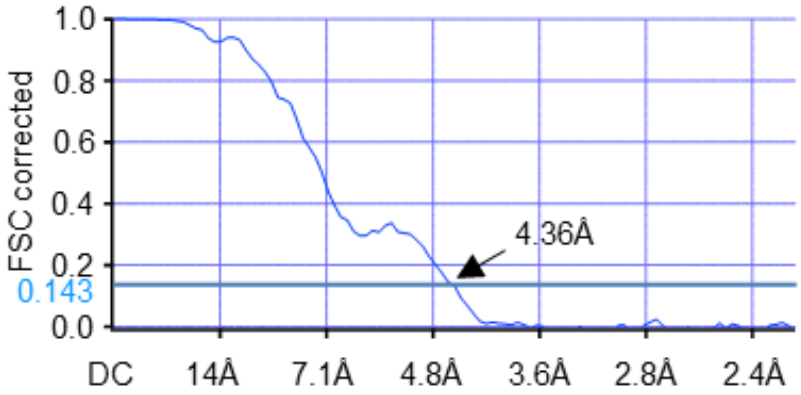


Figure 31. The corrected FSC curve of the data process.

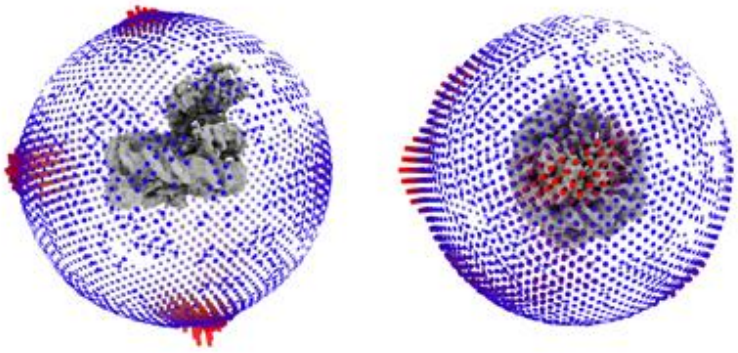


Figure 32. The angular distribution of the particles finally used in the reconstruction nucleosome-cGAS.

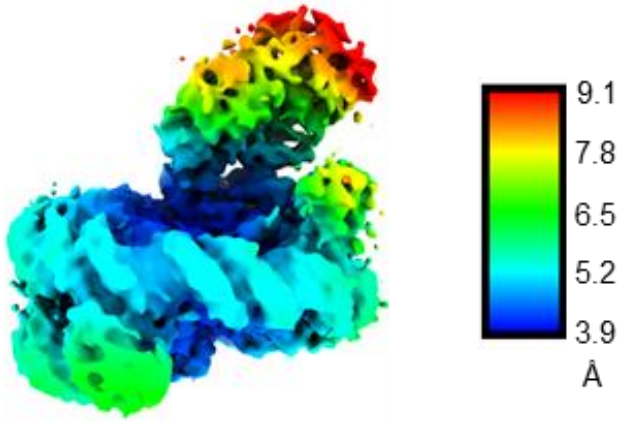


Figure 33. The local resolution of the reconstructed structure nucleosome-cGAS.

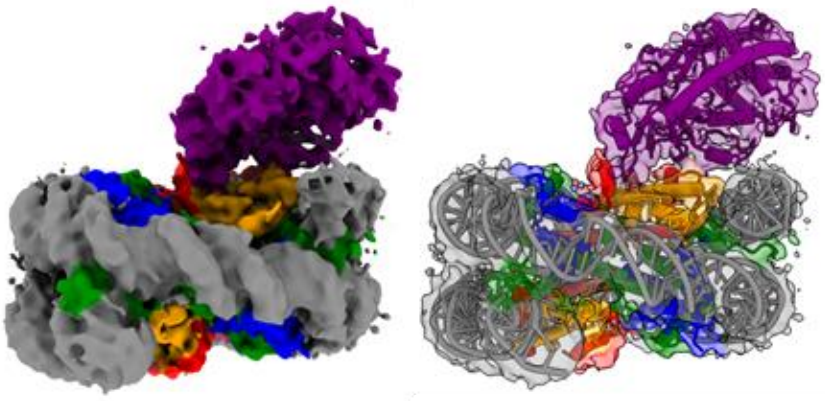


Figure 34. Cryo-EM structure of mouse cGAS catalytic domain bound to the human nucleosome.

a. map of the mcGAS-nucleosome complex at 4.36 Å resolution in two different orientations contoured at 3σ. b. Ribbon representation of the mcGAS-nucleosome complex structure docked into the map.

I reconstructed the mouse cGAS domain with the recombinant human nucleosome, or the human nucleosome directly purified from HEK 293T cells respectively. The two maps share the same binding profile that cGAS binds to the acidic patch on the nucleosome core particle (NCP) with its relatively distinct basic residues. The recombinant nucleosome-cGAS complex (~250kD) was reconstructed to a higher resolution as 3Å overall, making it possible to determine the interaction residues.

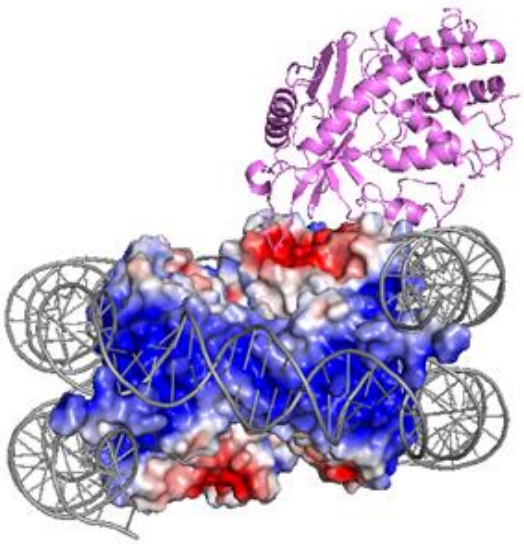


Figure 35. Histone proteins are colored by electrostatic.

The mcGAS lies on the acidic surface of H2A and H2B on NCP. By generating the electrostatic map of the histone proteins, the acidic patch is clearly presented as the red region in the figure (Figure 35). Its three loops (residues 220 to 223, 239 to 247, 314 to 345) parallelly attach to the nucleosome. mcGAS R222, K240, R241, K323, R341, and R342 have clear residual density reaching out to the H2B surface. On the NCP, E61, D90, and E92 have direct electrostatic interaction with mcGAS R241. NCP E61 and E64 interact with mcGAS R222. E64 also interacts with K240. Around nucleosomal DNA, NCP R71 reaches out to the mcGAS and interacts with the main chain of mcGAS G316 and R341. mcGAS R342 reaches out to the NCP and interacts with mcGAS T76 and K57.

The mcGAS binds to dsDNA in a 2:2 ratio and it has two surfaces for dsDNA binding, referred to as site A and site B. The crystal structure of mcGAS/DNA (pdb:4LEY) provided that site A includes K151, R158, K160, R161, S165, R180, K184, K372, and K395, while site B contains residues of K315, K323, K335, R222, K240, and R342. The mcGAS interacted residue to the NCP acidic patch overlaps with site B, including R222, K335, and R342. And the interaction is also facilitated by mcGAS R241, G316, R341, and R342 (Figure 36).

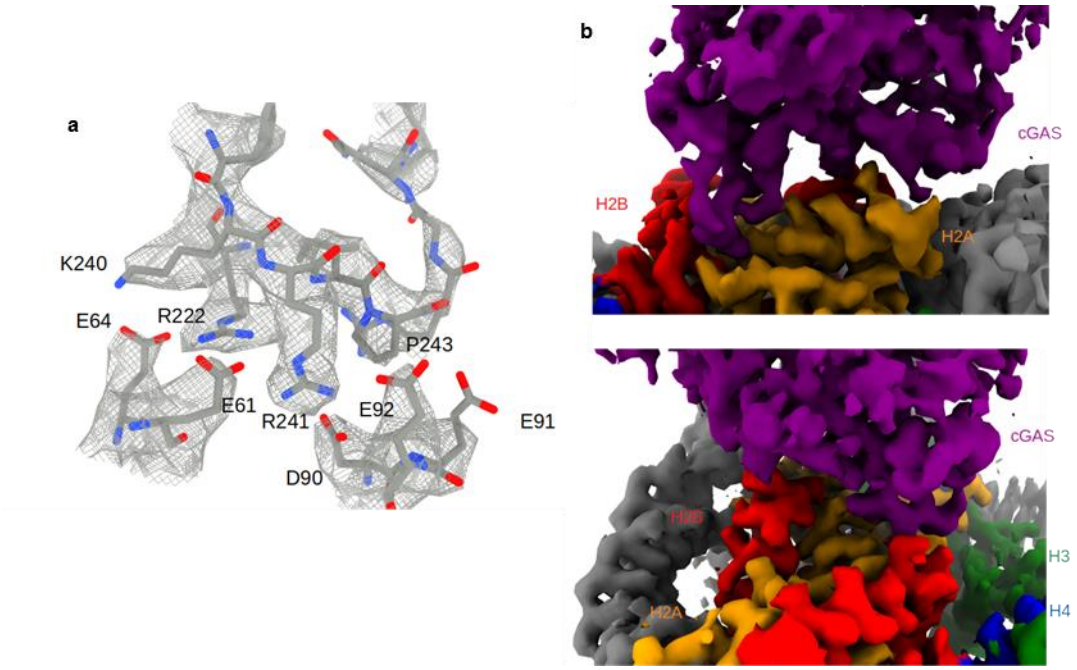


Figure 36. The interaction between cGAS and nucleosome.
a. The interaction region map is contoured by 3σ. K240, R222, and R241 of cGAS insert into the acidic patch of the nucleosome. b. Two different views of the interaction region.

The mutagenesis and in-vitro study of cGAS-nucleosome complex

The binding of cGAS to nucleosomes blocks the activation of cGAS. The activation of cGAS requires the complex formation of cGAS and dsDNA in 2:2 ratios. In the cGAS/dsDNA complex, cGAS binds to two dsDNA at both its A site and B site and undergoes the conformation in the catalytic area. However, the interaction of NCP to the cGAS B site does not activate cGAS. From the cGAS nucleosome complex EM structure, the NCP bound cGAS is in its inactive conformation. Binding to NCP also prevents the activation of cGAS. The alignment of cGAS between dsDNA bound complex and NCP bound complex clearly presents that, after cGAS binds to the nucleosome, either A site or B site of cGAS cannot bind to dsDNA because of clashes in space.

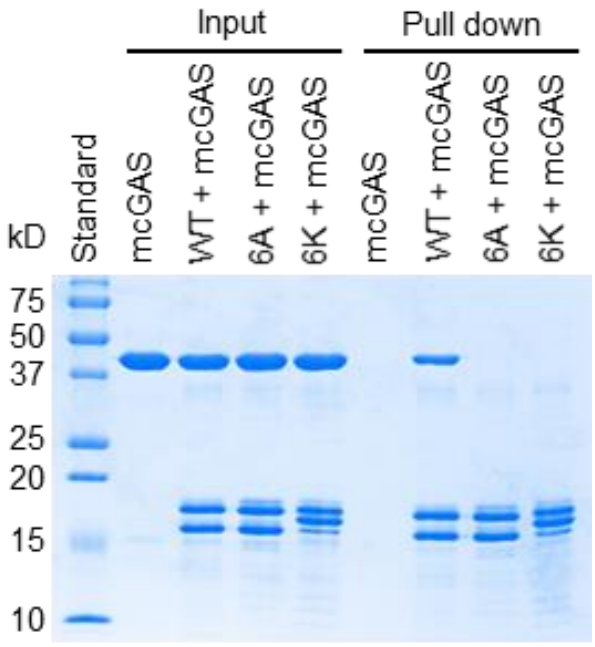


Figure 37. The pulldown assay of the H2A/H2B dimer mutants with mouse cGAS.

Moreover, some mutagenesis assays were done to validate our cryo-EM structure. After mutating the acidic patch on H2A/H2B from acidic amino acids Glu and Asp to neutral amino acid Ala or basic amino acid Lys, I did the Histidine tagged pull-down assay. The 6x His tagged wild type H2A/H2B or its 6A (H2A: E61A, E64A, D90A, E91A, E92A; H2B: D51A) and 6K (H2A: E61K, E64K, D90K, E91K, E92K; H2B: D51K) mutants were incubated with Ni-NTA resins. After washing the resins with extra buffer, mcGAS were incubated with the H2A/H2B dimers and washed with running buffer. The protein remaining on the beads were analyzed by SDS-PAGE. The pull-down assay indicated that WT H2A/H2B containing the acidic patch could bind with the mcGAS, while after mutating the acidic patch to alanine or to lysine, the interaction between H2A/H2B to mcGAS was weakened or diminished (Figure 37).

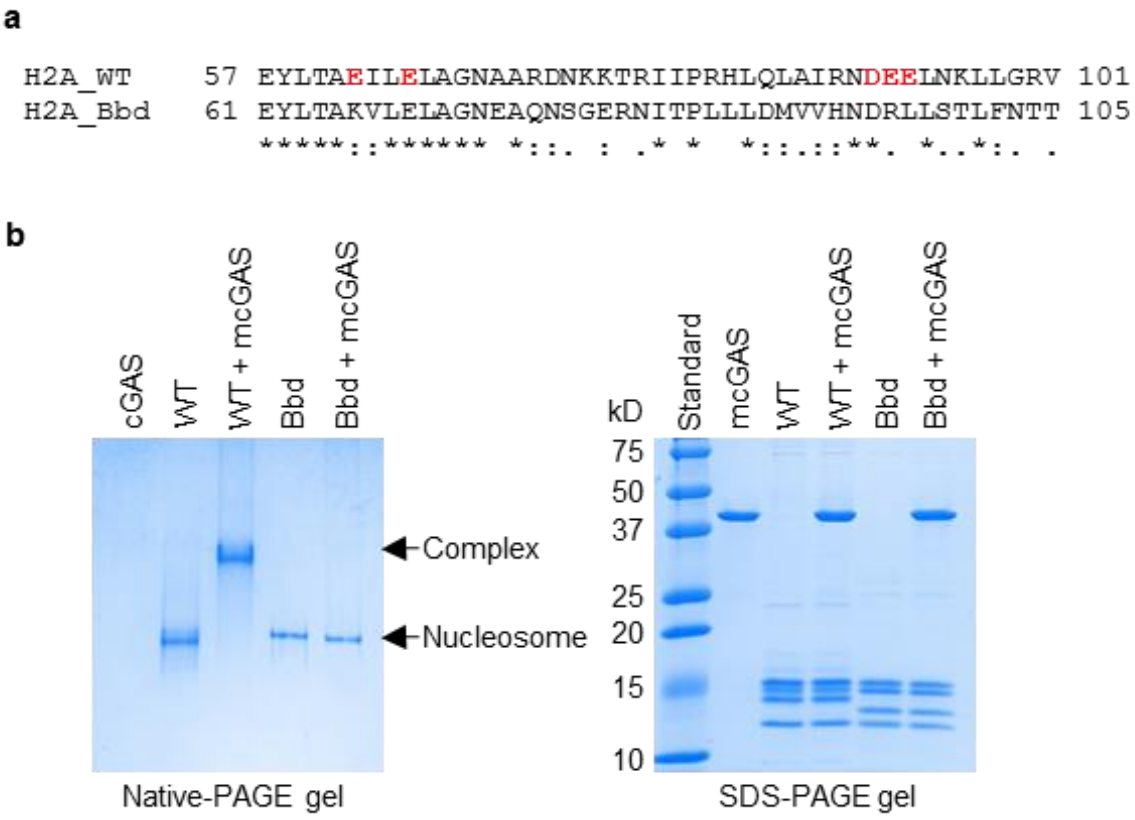


Figure 38. The native gel shift assay of nucleosome variant with mouse cGAS. a. The sequence alignment of wild type H2A and H2A variant H2A.Bbd. The acidic patch is colored red. b. The native gel shift assay of nucleosome variant with mouse cGAS. SDS-PAGE analyses of the same specimen are on the right.

To test if the acidic patch was required for the binding by another method, I also found a nucleosome formed by H2A variant call H2A.Bbd. By aligning the sequence between H2A and H2A.Bbd, it showed H2A.Bbd did not have a conserved acidic patch. There were two basic amino acids (K66, R96) in the acidic patch of H2A.Bbd. The native gel shift assay gave the result that nucleosome formed by H2A.Bbd did not have a shift

on gels after mixing with mcGAS, which supported that the acidic patch on nucleosome were the key residues to bind with cGAS. Furthermore, this experiment also supported that Histone H3, H4, and dsDNA on nucleosomes were not the binding partner for mcGAS. The sample using for native gel shift assay was analyzed by SDS-PAGE. The H2A.Bbd has a lower band compared with WT H2A, while all the other bands were the same (Figure 38).

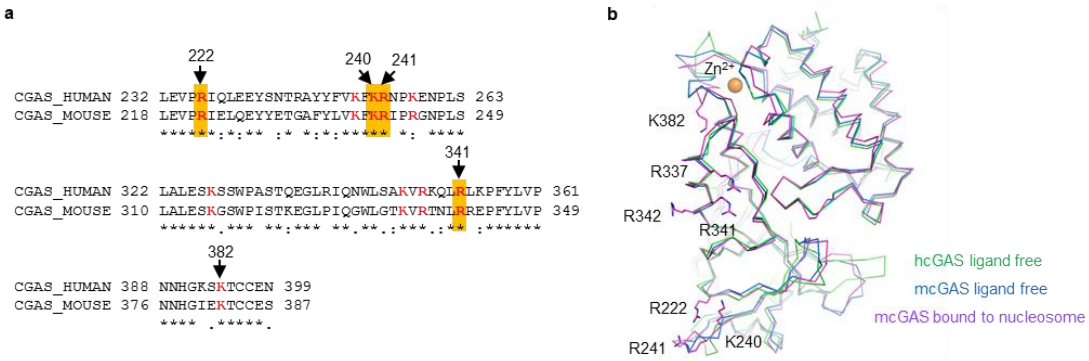


Figure 39. The sequence and structure alignment of the mouse and human cGAS.

a. Sequence alignment of human and mouse cGAS around the nucleosome binding site. The conserved basic residues around the nucleosome binding site are colored red. Residues that abolish nucleosome binding when mutated are highlighted yellow. b. Superposition for the structures of ligand-free human cGAS (PDB, 4LEV), ligand-free mouse cGAS (PDB, 4K8V), and mouse cGAS bound to the nucleosome.

After validating that the acidic patch was important binding region on the nucleosome, I worked on the cGAS side by mutagenesis assay. We did the sequence and

structural alignments of mouse cGAS and human cGAS on the binding region revealed by the cryo-EM map. The alignments showed that both the amino acid sequence and three-dimensional structure of mouse cGAS and human cGAS were very conserved. From the structure of mcGAS/nucleosome, we found four key interaction residues on cGAS and colored them orange. mcGAS R222, K240, R241, and R341 were conserved to hcGAS R236, K254, R255, and R353 (Figure 39).

We proposed the mutation of these basic residues to acidic ones would break the interaction. I also mutated the neighboring basic residues to work as controls or to test if they could have any effects on the binding.

All of the mcGAS and hcGAS mutants were generated by site-directed mutagenesis and confirmed by DNA sequencing. mcGAS mutants included R222E, K238E, K240E, R241E, R244E, K315E, K323E, K335E, R337E, R341E, R342E, and K382E. The protein was individually transfected to E. coli BL21 competent cells and induced by 0.4M IPTG in pET28a vector modified with an additional HIS-SUMO tag.

The mcGAS protein was purified by the Ni-NTA column. After the HIS-SUMO tag was cleaved by the Ulp1 enzyme, the protein runs through the Ni-NTA column again to remove the tags, and the flow-through was fractionated by Superdex 200 size exclusion column. The resulting mcGAS proteins were analyzed by SDS-PAGE (Figure 40). The single band on the SDS-PAGE indicated a very pure product and made it possible for me to do Circular Dichroism (CD). The CD spectra indicated the mutations on mcGAS did not change the protein folding. (Figure 41)

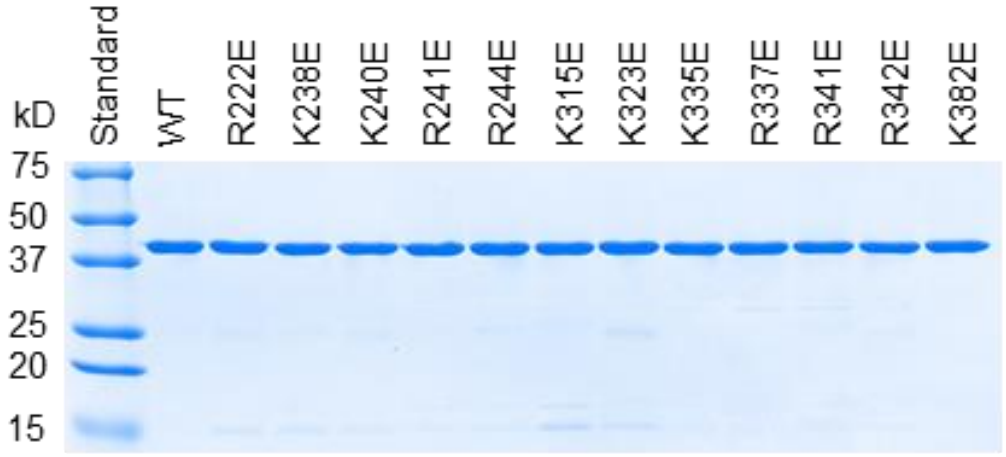


Figure 40. SDS-PAGE analysis of mouse cGAS catalytic domain mutants used for the gel shift assays and enzyme activity assays.

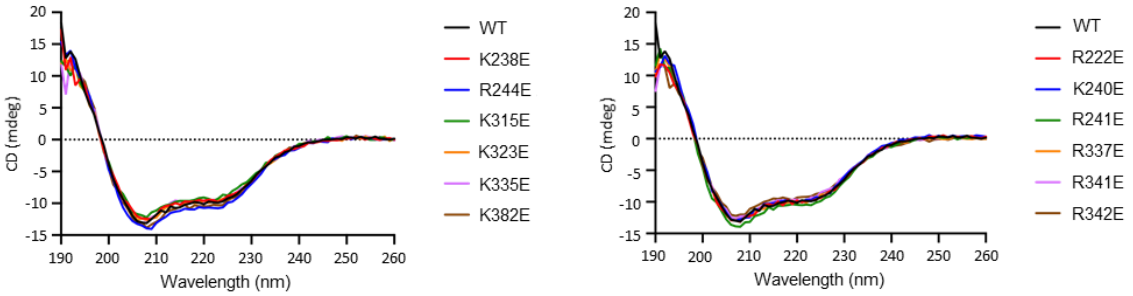


Figure 41. Circular dichroism of mouse cGAS catalytic domain and its mutants used for gel shift assays and enzyme activity assays.

cGAS mutants that have strong binding to nucleosomes are shown by the spectra on the left. cGAS mutants that have weak or no binding to nucleosomes are shown by the spectra on the right.

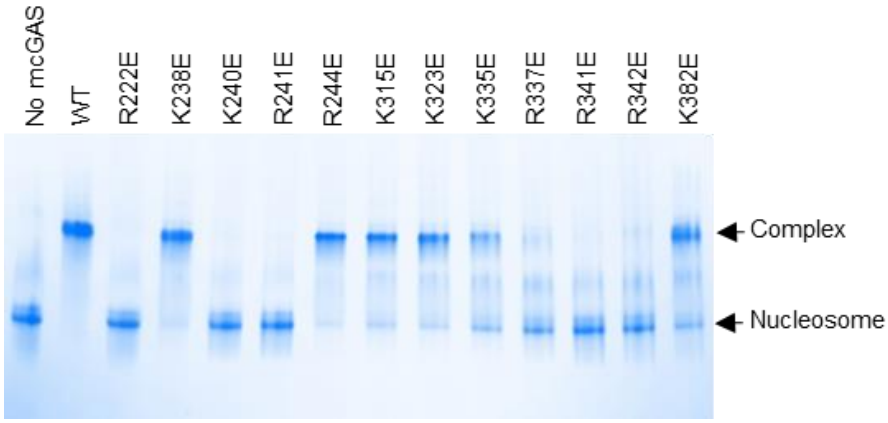


Figure 42. Polyacrylamide gel electrophoretic mobility shift assay (EMSA) shows that mutations at the cGAS-nucleosome interface affect nucleosome binding by cGAS.

In this assay, the mcGAS catalytic domain was mixed with the nucleosome at a molar ratio of 3:1.

Dr. Baoyu Zhao in our lab did the gel shift assay after we purified all the mcGAS mutants. The cryo-EM revealed that the four key residues were indeed very important for the binding. After the four key residues were mutated to acidic Glutamic acid, the native gel shift assay could no more show the complex formation (Figure 42).

I also purified the 13 human mutants (R236E, R246E, K252E, K254E, R255E, K258E, K327E, K347E, R349E, K350E, R353E, K355E, K394E) with biotin label. The biotinylated full-length cGAS and its mutants were good candidates for surface plasmon resonance (SPR) to detect the binding affinity. Because the biotinylated site was on the N-terminal to His-SUMO tag, the FL hcGAS was not cleaved. The product after the Ni-NTA

column was submitted to the Size Exclusion Column directly. The peak of the non-degraded protein was collected and concentrated.

The SDS-PAGE analyses of the FL cGAS also presented a relative pure band (Figure 43).

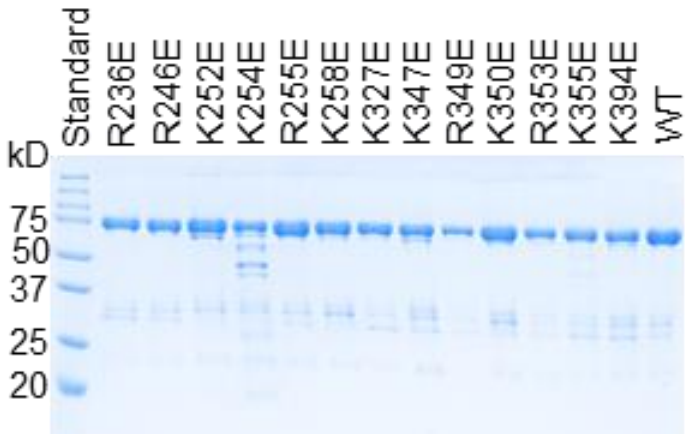


Figure 43. SDS-PAGE analysis of biotin-labeled full-length human cGAS mutants used for the SPR binding studies.

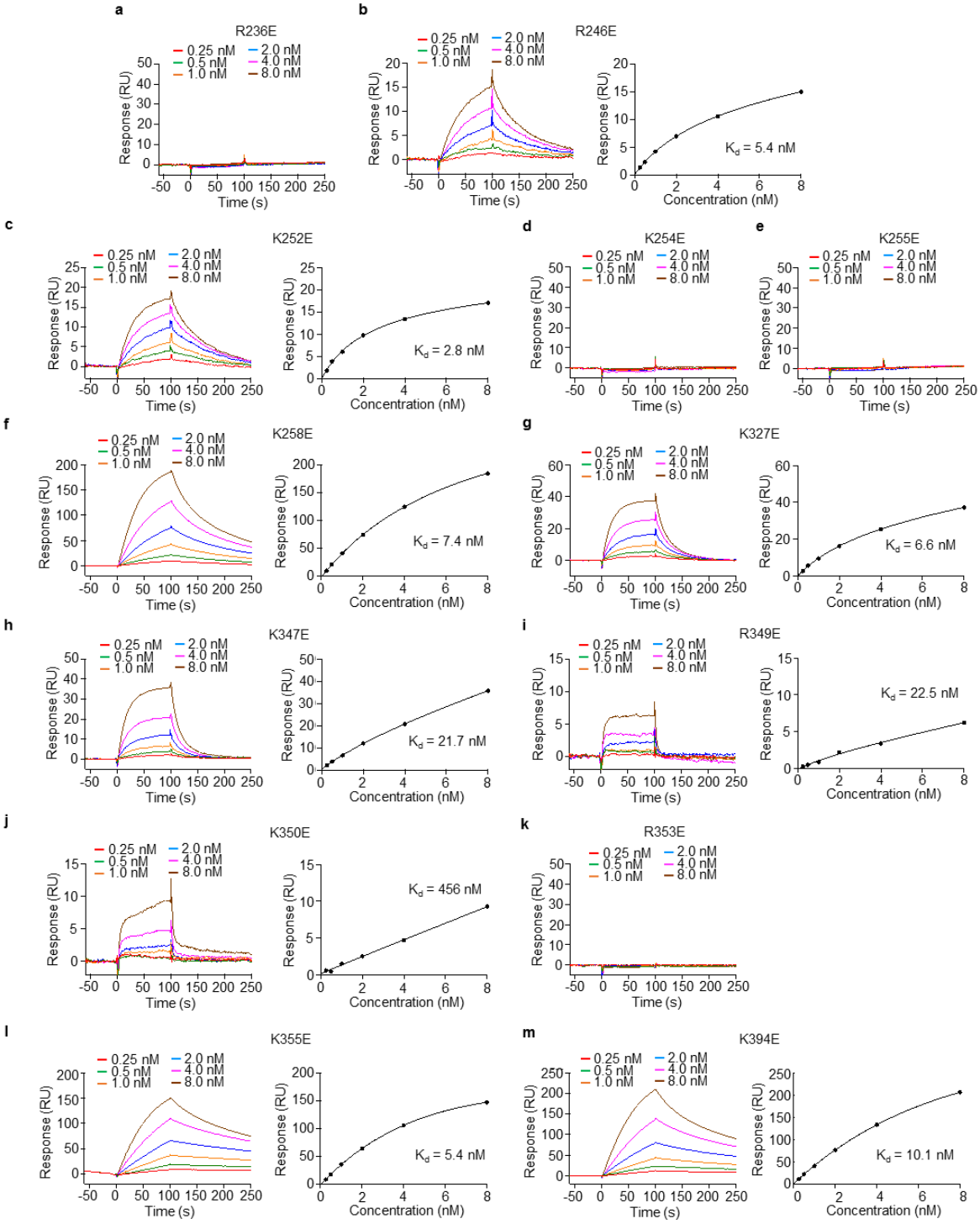


Figure 44. SPR binding studies of full-length human cGAS mutants with nucleosomes.

To do the SPR assay to determine the binding affinity, I immobilized the individual biotinylated FL hcGAS on the chip by streptavidin (CAP chip). Nucleosomes from 0.25 nM to 8 nM were injected automatically. The SPR validated that the mutations on four key residues diminished the signal, indicating no detectable binding existed. The other mutations on the neighboring basic residues did not change the interaction much when they compared with WT FL hcGAS (Figure 44).

The SPR study validated that the key residues identified from the cryo-EM structure indeed participated in the interaction. And the FL human cGAS actually had a very similar binding profile as the mouse cGAS catalytic domain. We also did SPR for FL hcGAS, FL mcGAS, hcGAS catalytic domain, and mcGAS catalytic domain. The results were very similar for these four kinds of proteins, indicating the catalytic domain of cGAS was the major part to bind nucleosome. In addition, we did SPR of FL hcGAS to both the reconstituted nucleosome and the natural nucleosome purified from 293T cells. No obvious affinity differences were found, indicating the modification on nucleosome did not affect much of nucleosome/cGAS binding (Figure 45).

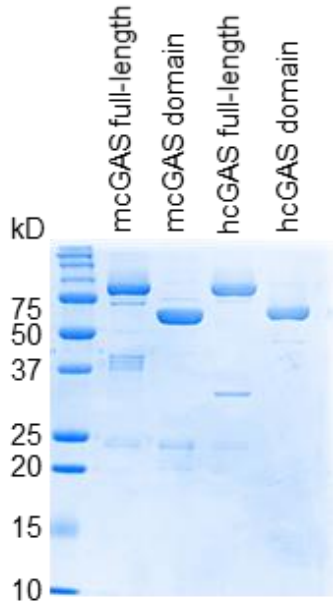


Figure 45. SDS-PAGE analyses of biotin-Avi-His6-SUMO fusion of human and mouse.

cGAS full length and catalytic domain proteins were individually used for nucleosome binding studies.

The binding of the cGAS to nucleosome core particles also potentially blocked the binding of other factors. Other than cGAS, the acidic patch on H2A&H2B also bound to RCC1 (regulator of chromosome condensation) protein, Snf5 (chromatin-remodeling complex SWI/SNF), Dot1L (Methyltransferase), PRC1 (Polycomb repressive complex 1). The cGAS was discovered to inhibit DNA repair in the nucleus and be independent of its DNA-binding ability or enzymatic activity. It is a possible reason that cGAS on

nucleosome blocks the binding of the nucleosome to chromatin-remodeling complex SWI/SNF.

Table 1. Cryo-EM data collection, refinement, and validation statistics

	mcGAS/nucleosome (reconstituted, 1:1) (EMDB-22046) (PDB 6X59)	mcGAS/nucleosome (reconstituted, 2:1) (EMDB-22206) (PDB 6XJD)	mcGAS/nucleosome (HEK 293T, 1:1) (EMDB-22047) (PDB 6X5A)
Data collection and processing			
Magnification	130,000	130,000	130,000
Voltage (kV)	300	300	300
Electron exposure (e–/Å^2)	42	42	42
Defocus range (μm)	-1.8 ~ -0.7	-1.8 ~ -0.7	-2.0 ~ -0.8
Pixel size (Å)	1.07	1.07	1.07
Symmetry imposed	C1	C1	C1
Initial particle images (no.)	2,202,680	2,202,680	578,302
Final particle images (no.)	165,092	9,454	23,463
Map resolution (Å)	2.96	6.79	4.36
FSC threshold 0.143			
Map resolution range (Å)	6.0 ~ 2.8	10.7 ~ 4.1	9.1 ~ 3.9
Refinement			
Initial model used (PDB code)	3AFA 4K8V	3AFA 4K8V	3AFA 4K8V
Model resolution (Å)	3.00	7.16	4.54
FSC threshold 0.5			
Model resolution range (Å)	32.5 ~ 2.8	39.8 ~ 6.8	35.3 ~ 4.4
Map sharpening *B* factor (Å^2)	-15	-48	N/A
Model composition			
Non-hydrogen atoms	14,963	17,957	14,963
Protein residues	1,123	1,485	1,123
Nucleotide	290	290	290
Ligands	1	2	1
B factors (Å^2)			
Protein	83.69	104.67	405.53
Nucleotide	74.19	74.19	278.45
Ligand	194.81	194.81	997.41
R.m.s. deviations			
Bond lengths (Å)	0.006	0.008	0.007
Bond angles (°)	0.834	1.533	0.887
Validation			
MolProbity score	2.64	2.73	2.23
Clashscore	10.75	10.82	13.71
Poor rotamers (%)	7.57	8.16	1.66
Ramachandran plot			
Favored (%)	93.67	91.95	93.67
Allowed (%)	6.15	8.05	6.15
Disallowed (%)	0.18	0	0.18

Discussion and future studies

Together with my paper published in *Nature* [115], there are another 4 papers published back-to-back in *Nature* and *Science* for the similar mechanism of cGAS inhibition by nucleosome [116-119]. Among all the papers, my structure is determined to the highest resolution and is the only one that has better resolution than 3Å. It allows us to investigate the most of the cGAS inhibition.

The structures determined by human cGAS in complex with nucleosomes are different from the mouse complex, in that the human cGAS has a site C which binds to the nucleosomal DNA. For this additional binding site C, human cGAS would have two sites for nucleosome binding. And it contributes to a sandwich like structure for human cGAS with nucleosome. Due to one nucleosome binds to two cGAS and one cGAS also binds to two nucleosomes, the complex will form an elongated filament like structure. The elongated filament is kind of not homogenous and not a good sample for single particle reconstruction, so that the groups using human cGAS for structural study all stuck at ~4Å overall resolution and the cGAS densities are beyond 10 Å. In our studies, other than using the mouse cGAS for structural study, we also use FL human cGAS for binding studies. We claim that the human and mouse cGAS shares the similar binding site, which is site B to bind nucleosome. The site C of human cGAS only has a weak binding to nucleosome because after we mutated site B in the FL human cGAS, we no longer detected the binding of cGAS to nucleosome any more.

The future study of this field will be focused on the N-terminal part of cGAS. Because right now, all the structural based mechanism studies are focused on the cGAS

domain around amino acids 155-522. However, the FL human cGAS, FL mouse cGAS and human cGAS domain all form aggregates with nucleosome, while the mouse cGAS domain only forms the ~300kD complex with nucleosome. From the inspiration of human cGAS domain with nucleosome, I speculate that cGAS must have more than two binding site for nucleosome, then they would form filament complex. From sequencing analyses, we found that the N-terminal parts of cGAS also contains multiple basic residues, like K and R. These basic residues also have high possibilities to bind nucleosomal acidic patch or DNA. In summary, mouse cGAS domain only has one site (site B) to bind nucleosome. Human cGAS domain has site B and site C to bind nucleosome. And FL mouse cGAS has site B and N-terminal parts to bind nucleosome, while FL human cGAS has site B, site C and N-terminal domain to bind nucleosome. This is a claim that requires more experimental support.

On the other hand, due to the potency of cGAS binding to nucleosome, the role of cGAS on nucleosome is very interesting. I speculate that, when the cell is dividing, the nuclear membrane is not intact, such strong binding will inactivate the cGAS in the cytosol and protect the cell from self-activation by genome DNA. However, what the real role of cGAS in the nucleus is still unknown. And how the cGAS is released from nucleosome or chromatin is also a challenge topic for such nanomolar binding. The role of N-terminal part of cGAS in the nuclear translocation of cGAS is another rising question.

Materials and Methods in detail

Protein expression and purification

The cDNA of cGAS full length and catalytic domain were cloned into a modified pET-28a vector with an N-terminal Avi-His6-SUMO tag. Mouse cGAS catalytic domain (residues 142 – 507) was expressed in E.coli BL21 (DE3) with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) induction overnight at 16°C and purified as described previously[120]. Biotin-Avi-His6-SUMO human and mouse cGAS full length and catalytic domains (human cGAS domain residues 157 – 522) were expressed in E.coli BL21 (DE3) cells co-transformed with the plasmids coding for cGAS and the pBirAcm plasmid coding for BirA. Protein expression was induced with 0.4 mM IPTG in the presence of 5 μg/ml biotin (Sigma-Aldrich, B4501). The proteins were first purified using a Ni2+-NTA column (Qiagen) and were further purified over a Superdex200 column (GE Healthcare Life Sciences). Biotin-Avi-His6-SUMO human and mouse cGAS full length and human cGAS catalytic domain were eluted with the buffer containing 20 mM Tris, 500 mM NaCl, pH 7.5. Biotin-Avi-His6-SUMO mouse cGAS catalytic domain was eluted with the buffer containing 20 mM Tris, 150 mM NaCl, pH 7.5. All mutants were generated using a PCR-based technique with appropriate primers and confirmed by DNA sequencing. The cGAS mutant proteins were expressed and purified the same way as the wild-type cGAS.

Nucleosome purification from HEK 293T cells

Nucleosomes were extracted and purified from HEK 293T cells using the nucleosome preparation kit (Active motif, 53504) according to the manual. Briefly, HEK 293T cells were cultured in a 10 cm tissue culture dish to 70-80% confluency and were harvested with a cell scraper. 2 × 107 cells were washed with 10 ml 1 × PBS buffer twice. Then the cells were resuspended in 1 ml ice-cold lysis buffer supplemented with 5 µl protease inhibitor cocktail and 5 µl 100 mM PMSF. After incubation on ice for 30 mins, the lysed cells were centrifuged at 2400 g for 10 mins at 4 °C to pellet the nuclei. The nuclei pellet was resuspended in 350 µl digestion buffer supplemented with 1.75 µl protease inhibitor cocktail and 1.75 µl 100 mM PMSF. After incubation at 37 °C for 5 mins, the resuspended nuclei were mixed with 17 µl diluted enzymatic shearing cocktail and incubated at 37 °C for 50 mins. To generate oligonucleosomes, the mixture was incubated at 37 °C for 15 mins. During the incubation, the mixture was vortexed approximately every 2 mins. To stop the reaction, 7 µl ice-cold 0.5 M EDTA was added into the mixture for 10 mins on ice. Then, the mixture was centrifuged at 21000 g for 10 mins. Nucleosomes or oligonucleosomes in the supernatant were purified using a Superose 6 increase 10/300 GL column (GE Healthcare Life Sciences) with the running buffer 20 mM Tris, 150 mM NaCl, pH 7.5. The purified nucleosomes and oligonucleosomes were concentrated and stored at 4 °C.

Reconstitution and purification of human nucleosome

Human histone H2A, H2B, H3-C110A, and H4 with TEV tags were cloned into a pETDuet vector and expressed in E. coli BL21 (DE3) at 37 ˚C. The cell pellets were resuspended in 50 mM Tris, 100 mM NaCl, 0.1% Triton-X100, pH 8.0 buffer and lysed by sonication. Inclusion bodies were collected and washed three times with additional lysis buffer and three times with 50 mM Tris, 100 mM NaCl, pH 8.0 buffer. Inclusion bodies were then dissolved in 6 M guanidine hydrochloride, 20 mM Tris, 250 mM NaCl, pH 8.0 buffer and purified under denatured conditions using a Ni-NTA column. The histone proteins were eluted with 6 M urea, 20 mM Tris, 500 mM NaCl, 250 mM imidazole, pH 7.8 buffer, dialyzed against 5% acetic acid, flash-frozen, and lyophilized. A plasmid containing the original Widom 601 sequence cloned in a pUR19 vector was used as a template for PCR amplification. Purified 601 DNA was diluted to 200-300 ng/μL in 20 mM Tris, 1 mM EDTA, pH 7.8 buffer, and solid NaCl was added to a final concentration of 2 M. The histone pellets were dissolved in 6 M guanidine hydrochloride, 20 mM Tris, 500 mM NaCl, pH 7.8 buffer. To prepare the H2A/H2B dimer, H2A and H2B were mixed in an equimolar ratio and diluted with 6 M guanidine hydrochloride to adjust the total protein concentration to 4 mg/mL. The denatured H2A/H2B solution was then dialyzed sequentially at 4 ˚C against a 20 mM Tris, 1 mM EDTA, pH 7.8 buffer containing 2 M, 1 M, and 0.5 M NaCl. H2A/H2B 6A and 6K dimer mutants were prepared as the wild-type H2A/H2B dimer. Refolding of the H3/H4 tetramer was done in the same manner. The refolded H2A/H2B dimer and H3/H4 tetramer were mixed in a 2:1 molar ratio and solid NaCl was added to a final concentration of 2 M to form histone octamers.

Purified 601 DNA in 2 M TE buffer was added to the histone octamer solution in a molar ratio of 0.9:1 and dialyzed against 2 M TE buffer. While stirred at room temperature, a 20 mM Tris buffer at pH 7.8 was slowly added to the 2 M salt buffer with the DNA/histone mixture until the salt concentration was reduced to below 200 mM. The reconstituted nucleosome was then dialyzed against 20 mM Tris, 20 mM NaCl, 1 mM EDTA, pH 7.8 buffer. TEV protease was then added to the nucleosome solution (TEV: nucleosome 1:30, w:w) to cleave the His-TEV tags. The reconstituted nucleosome was purified using a Superdex 200 column eluted with 20 mM Tris, 150 mM NaCl at pH 7.5.

Surface plasmon resonance (SPR)

All SPR binding studies were performed with a Biacore X100 instrument (GE Healthcare Life Sciences). The Biotin CAPture kit (GE healthcare life science, 28920233) was used to detect the binding between human cGAS and nucleosome according to the manual. Briefly, Biotin-Avi-His6-SUMO cGAS was captured on the sensor chip CAP. Serial dilutions of nucleosome were flowed through the chip at 30 µl/min in 1× HBS-EP buffer supplemented with 5 mM MgCl2. The multi-cycle kinetic/affinity protocol was used in all studies. In each cycle, the nucleosome was injected for 100 s. The sensor chip was regenerated with a buffer containing 6 M guanidine hydrochloride and 0.25 M NaOH at the end of each cycle. For full-length mouse cGAS, the sensor chip SA (GE healthcare life science, BR-1000-32) was used for the binding study, and the chip was regenerated with 2 M MgCl2 at the end of each cycle. The Biotin CAPture kit was used for the binding study of mouse cGAS catalytic domain and nucleosome with the running buffer 20 mM

Tris, 150 mM NaCl, pH 7.5. All data were analyzed using Biacore X100 evaluation software version 2.0 (GE Healthcare Life Science) and the binding affinities (Kd) were determined by fitting the data to a steady-state 1:1 binding model.

cGAS activity assay

All cGAS activity assays were performed in 20 mM HEPES, 5 mM MgCl2, 5 mM DTT, 150 mM NaCl, 2 mM ATP (Sigma Aldrich, A2383) and 2 mM GTP (Sigma Aldrich, G8877), pH 7.5. For the mouse cGAS catalytic domain, 2.5 μM protein was incubated with 0.2 mg/ml salmon sperm DNA (Invitrogen, 15632-011) for 1 h at 37 °C. For mcGAS-nucleosome complex or mcGAS-oligonucleosome complex, nucleosomes or oligonucleosomes purified from HEK 293T cells were incubated with excess mouse cGAS catalytic domain (In the mixture, the molar ratio of mcGAS: nucleosome is 10:1 and mcGAS: oligonucleosome is 20:1) for 1 h on ice. The complexes were purified by Superdex 200 10/300 GL column (GE Healthcare Life Sciences) with the running buffer 20 mM Tris, 150 mM NaCl, pH 7.5. 2.5 μM mcGAS-nucleosome complex or 1.25 μM mcGAS-oligonucleosome complex was incubated with or without 2.5 μM mcGAS catalytic domain, 0.2 mg/ml 45bp interferon stimulatory dsDNA (ISD) or 0.2 mg/ml salmon sperm DNA for 1 h at 37 °C. For human cGAS full length, 2.5 μM protein was incubated with 0.2 mg/ml salmon sperm DNA for 4 h at 37 °C. The product in the supernatant was separated from cGAS, DNA, and nucleosome by ultrafiltration and was analyzed on a Mono Q 5/50GL ion exchange column (GE Healthcare Life Sciences). The relative enzymatic activities of the mouse and human cGAS mutants were calculated by

dividing the cGAMP peak height of the cGAS mutants by the cGAMP peak height of WT cGAS.

Electrophoretic mobility shift assay (EMSA)

For nucleosome binding studies by cGAS, 500 nM nucleosome or 250 nM oligonucleosome was mixed with mouse cGAS catalytic domain at indicated molar ratios in the buffer containing 20 mM Tris, 150 mM NaCl, 5 mM DTT, pH 7.5. After incubation on ice for 1 h, the mixtures were analyzed on 4-20% precast polyacrylamide gel (Bio-rad) at a constant voltage of 100 V. After Coomassie blue staining and destaining, the gels were imaged using a ChemiDoc MP Imager (Bio-Rad). The binding study of nucleosome variant H2A.Bbd (Active motif, 31556) with the mcGAS catalytic domain was also performed as the wild-type nucleosome. For cGAS and dsDNA binding studies, 1 μM 45 bp interferon stimulatory dsDNA was mixed with WT and mutants of mouse cGAS catalytic domain at a molar ratio of 1:20. For human cGAS full-length proteins, the molar ratio of DNA and protein is 1:4. The mixtures were incubated on ice for 30 mins and then analyzed using 1% agarose gel.

Negative stain

Nucleosome was diluted to a certain concentration and was added a 5 uL droplet to the glow-discharged Carbon coated Cu grids for 10 seconds. The droplet was removed by a filter paper. Then 20 uL buffer was added to the grid to wash for 20 seconds and removed by a filter paper again. 20 uL of 0.75% Uranyl Formate was added to the grid to stain the protein. After 10 seconds, the Uranyl Formate was removed by filter paper. The

staining process was repeated twice. The grid was dried in air for 30 minutes and screened by TF-20 microscope operated at 200kV and equipped with a K2 Summit direct electron camera.

Cryo-EM data acquisition

Reconstituted nucleosomes or nucleosomes purified from HEK 293T cells were incubated with excess mouse cGAS catalytic domain for 1 h on ice. Excess cGAS was removed by Superdex 200 10/300 GL column (GE Healthcare Life Sciences). The cGAS-nucleosome complex fraction was collected and concentrated to 0.4 mg/ml. Aliquots of 3 μl nucleosome-cGAS complexes were loaded onto glow-discharged holey carbon grids (Electron Microscopy Sciences, CF312-50, CFlat, Cu, R 2/1, 300 mesh). Grids were blotted for 8 s and plunged frozen in liquid ethane using a Vitrobot at 4 °C and with 100% humidity. Grids were transferred to a Titan Krios electron microscope (Thermo fisher) operating at 300 kV equipped with a Gatan Gif Quantum energy filter (Slit width 20 eV). Micrographs were recorded by EPU through a Gatan K2 Summit detector in counting mode at a nominal magnification of ×130000 (yielding a pixel size of 1.07 Å). The dose rate on the camera was set to be 6 electrons per physical pixel per second. Exposure of 8 s was dose-fractionated into 40 moves frames, leading to a total accumulated dose of 42 electrons per Å2 on the specimen. Sample for the reconstituted nucleosome bound to mcGAS was recorded with a defocus in the range from 0.7 to 1.8 μm for a total of 4,959 micrographs. Nucleosomes purified from HEK 293T cells bound to cGAS were recorded with a defocus in the range from 0.8 to 2.0 μm for a total of 2,979 micrographs.

Cryo-EM data processing

The imaging processing is identical for the two datasets of in-vitro reconstituted and HEK 293T cells purified nucleosomes in complex with cGAS by Relion-3.023. The collected movies were subjected to MotionCor2 for whole-frame dose-weighted motion correction. 2,202,680 and 578,382 particles were auto picked by Relion for reconstituted and HEK 293T cells purified nucleosome-cGAS complexes, respectively. Multiple rounds of reference-free 2D classification were run to remove aggregates, ice contamination, and carbon edges by Relion, yielding 1,563,321 and 502,961 particles for reconstituted and HEK 293T cells purified nucleosome-cGAS complexes, respectively. Particles were processed with Relion 3D classification with ab-initio models generated in Relion. 3D class with cGAS densities were selected, yielding 1,162,012 and 90,398 particles for the reconstituted and HEK 293T cells purified nucleosome-cGAS complexes, respectively. The particles were subjected to 3D auto refinement in Relion, followed by additional cGAS focused skip-align 3D classification. 3D classes with clear cGAS densities were selected again, yielding 182,358 and 23,463 particles for reconstituted and HEK 293T cells purified nucleosome-cGAS complexes, respectively. The particles were then re-centered and re-extracted to 3D auto refinement in Relion. The HEK 293T cells purified nucleosome-cGAS complex was reconstructed to 4.36 Å. The reconstituted nucleosome-cGAS complex was reconstructed to 3.21 Å and further subjected to skip-align 3D classification. The class of two cGAS bound nucleosome contained 9,454 particles and was refined to 6.79Å. The EM map of one cGAS bound nucleosome was refined to 2.96

Å by CTF refinement and Bayesian Polishing. The reported resolutions are based on the gold-standard Fourier shell correlation (FSC) 0.143 criterion. Local resolution variations were estimated using Relion.

Cryo-EM model building and refinement

Human nucleosome model containing the 601 DNA was generated using a published nucleosome structure (PDB 3AFA). The nucleosome and mcGAS (PDB 4K8V) models were docked in the EM map in Chimera and fine-tuned by manual adjustment with Coot[121]. This model was refined against the EM map in Phenix[122]. Several loop regions of mcGAS, tails of the histone proteins, and the 601 dsDNA were manually adjusted to fit into the map using Coot. The model was refined again in Phenix and the crystal structure of mcGAS was used as a reference. This refined model was docked into the EM map derived from nucleosome purified from HEK 293T cells bound to mcGAS and the model was refined in Phenix to improve the fitting without further remodeling due to the lower resolution of the map.

Ni-NTA Pull-down assay

The H2A/H2B acidic patch residues were mutated to alanine (H2A: E61A, E64A, D90A, E91A, E92A; H2B: D51A; referred to as H2A/H2B 6A) or Lysine (H2A: E61K, E64K, D90K, E91K, E92K; H2B: D51K; referred as H2A/H2B 6K). 40 μg 6 × His tagged WT H2A/H2B, H2A/H2B 6A or H2A/H2B 6K dimer was incubated with 20 μg Ni-NTA beads in pull-down buffer (20 mM Tris, 150 mM NaCl, 5 mM DTT, pH 7.5) for 5 mins at

4°C. Beads were washed by pull-down buffer three times and mixed with 60 μg mouse cGAS domain, then incubated for 5 mins at 4°C. Excess proteins were washed off the beads using 300 uL pull-down buffer six times. 20 μl of 5 × SDS loading buffer was added to the resin and boiled for 5 mins, thereafter, the samples were centrifuged briefly. 5 μl of supernatant was analyzed by SDS-PAGE. The protein bands were visualized by Coomassie blue staining.

Circular dichroism spectroscopy

Wild type mouse cGAS catalytic domain and its mutants were buffer exchanged into 10 mM phosphate buffer containing 50 mM K2SO4 at pH 7.5. The CD spectra were measured using a Chirascan spectrometer with each protein sample at 3 μM concentration. The spectra from 280 nm to 190 nm wavelength were recorded at room temperature using a 2 mm path length quartz cuvette.

CHAPTER III

STRUCTURAL BASIS OF STING ACTIVATION BY CGAMP[2]

Introduction

Cells can defend against bacteria and virus's invasion through their immune system [11, 54-62]. Innate immunity employs a wide spectrum of germ-line encoded pattern-recognition receptors (PRRs) to sense microbial infection and initiate protective immune response[63-66]. Previous studies identified cyclic GMP-AMP synthase (cGAS) is a key dsDNA sensor in the cytosol [67-69]. cGAS is activated by dsDNA and catalyzes the synthesis of a noncanonical cyclic dinucleotide cGAMP containing mixed 2′,5′ and 3′,5′ phosphodiester linkages [70-75]. Unlike bacterial cyclic dinucleotides (CDNs) with canonical 3'-5' linkages, 2'3'-cGAMP is a unique metazoan CDN that comprises both 3'-5'and a non-canonical 2'-5' phosphodiester linkages connecting adenosine with guanosine and guanosine with adenosine, respectively.

cGAMP, behaving as a second messenger, stimulates the induction of type I interferons (IFNs), such as IFN-α and β, through the adaptor STING located on the ER membrane [64, 68, 71, 76-78]. cGAMP binding induces conformational change and oligomerization of STING, leading to the recruitment and activation of TBK1 [68, 69, 74, 79]. The transcription factor IRF-3 is then recruited to the signaling complex and activated

[2] Material from: Baoyu Zhao, Fenglei Du, Pengbiao Xu, Chang Shu, Banumathi Sankaran, Samantha L Bell, Mengmeng Liu, Yuanjiu Lei, Xinsheng Gao, Xiaofeng Fu, Fanxiu Zhu, Yang Liu, Arthur Laganowsky, Xueyun Zheng, Jun-Yuan Ji, A Phillip West, Robert O Watson, Pingwei Li. A conserved PLPLRT/SD motif of STING mediates the recruitment and activation of TBK1
, Nature, published 2019.

by TBK1 [80-82]. Mechanistically, STING employs a conserved pLxIS (p, hydrophilic residue, x, any residue, S, phosphorylation site) motif to recruit IRF-3 upon phosphorylation [83]. Phosphorylated IRF-3 oligomerizes, translocates to the nucleus [84, 85], and initiates the expression of type I IFNs [86-89]. The transcription factor NF-κB is also activated upon the stimulation of STING and is likely involved in the transcriptional regulation of IFN-β expression together with IRF-3 [54, 78, 90, 91]. Aberrant activation of innate immune responses by self nucleic acids causes autoimmune disorders such as systemic lupus erythematosus (SLE) and polyarthritis [92-99]. Several mutations of human STING cause autoimmune disorders [100, 101]. In addition to its roles in innate immunity, cGAMP and its phosphorothioate derivatives have potent antitumor activity and have applications in cancer immunotherapy [97, 102-104].

The detection of foreign nucleic acids is a central strategy in the innate immune system. Cyclic GMP-AMP synthase (cGAS) is a cytosolic DNA sensor that catalyzes the synthesis of a cyclic dinucleotide second messenger, cGAMP, that binds to the adaptor STING (Stimulator of interferon genes protein), which forms puncta on the ER membrane and mediates the recruitment and activation of the protein kinase TBK1 and transcription factor IRF-3. Phosphorylated IRF-3 translocates to the nucleus and induces the expression of type I interferons. However, the molecular mechanisms governing how cGAMP activates STING at the ER membrane remain poorly understood. Here, we report that cells have no STING mediated signaling if STING is truncated on either the N-terminal transmembrane domain or C-terminal TBK1/IRF3 binding region. Furthermore, the N-terminal transmembrane domain of STING is essential for STING oligomerization and

signaling. Moreover, cGAMP binding induces the oligomerization of full-length STING both in vitro and in vivo. The central hypothesis is that cGAMP binding makes the conformational changes of STING dimer located discretely on the ER membrane, and increases the affinity of STING dimers to form oligomer, which facilitates the activation of kinase TBK1. These comprehensive structural and functional studies provide new insights into the mechanism of STING activation.

One-step purification of GFP fusion protein with GFP-Nanobody coated streptavidin resins

Introduction of the method

The purification of target protein from the expression system is always critical before structural determination, including X-Ray crystallization and Cryo-Electron Microscopy (Cryo-EM). Purification based on conventional affinity tags is not generally satisfying for its specificity, especially when macromolecules are low-level expressed from the mammalians system. Here we have developed a new protein expression and purification tracking system, with green fluorescence protein (GFP) tag fusion and GFP-Nanobody coated streptavidin beads (GNB). This system has an advantage especially for the low-level expressed macromolecular in mammalian cells. GNB has high specificity and extremely tight affinity to the GFP tag at 8.7 picomolar. Fusing GFP tag to target protein not only provides affinity tags for GNB but also facilitates the solubility and gives easy track of the protein expression and purification by fluorescence. After GFP tagged protein is captured and purified by GNB, the GFP tag can be cleaved at the pre-inserted thrombin site and removed by size exclusion chromatography, leaving the tag-free

recombinant protein. By such a method, we purify full-length integral membrane protein STING from HEK293 Freestyle cells and its soluble truncation from *E. coils* BL21 individually. Thus, GNB offers an efficient and general method to obtain highly purified tag-free recombinant protein and is therefore suitable for the subsequent structural determination. The biotin-labeled GFP-Nanobody also works in coating the SPR chip, making it possible to immobilize GFP tagged protein for direct binding detection.

Studies of protein structure and functions typically require the purification of the target protein[123, 124]. The conventional affinity-tag purification is generally the first step. This includes but is not limited to polyhistidine tags, Glutathione S-Transferase (GST) tags, Maltose Binding Protein (MBP) tags, Streptavidin/Biotin based tags[125]. Polyhistidine tags have an advantage over other tags for their small size, which helps maintain the native function of overexpressed target proteins. The challenge to use polyhistidine tags is the nonspecific binding of the untagged protein. Even though histidine contributes to only 2% of all protein residues, some cellular proteins have two or more adjacent histidine residues, especially in mammalian expression system [126]. These proteins will have an affinity to Ni NTA resin and co-elute with the tagged protein, resulting in contamination.

The antibody is made use of to overcome the nonspecific binding in protein purification from mammalian cells. Using antibody also enhances the yield purity by higher affinity in comparison to conventional affinity tags. Such antibody systems include Flag tag (DYKDDDK), Myc tag (EQKLISEEDL), KT3 epitope (KPPTPPPEPET), V5 tag (GKPIPNPLLGLDST), and HA tag (YPYDVPDYA)[127, 128]. The antibody-based

purification is limited to a small scale, such as immunoprecipitation because the yield of antibody is low and the resins' price is accordingly high. The nanobody, also known as a single-domain antibody, is a better choice for large scale yield[129]. Nanobody has a heavy chain only and their variable region (VHH) is the smallest antigen-binding fragment found in a natural antibody[129]. With the same or even higher affinity to antigens, the expression of 12-15 kD nanobody does not require difficult refolding and secretion steps as the 150-160 kD multi-fragment antibodies[130, 131].

Here, we establish a novel one-step purification system with GFP fusion protein and GFP-Nanobody coated streptavidin beads (Figure 46). GFP tag fused to the target protein not only functions as affinity tags for GFP-Nanobody resins but also works as a fluorescence marker to track the protein expression and purification[132, 133]. The in-house prepared GFP-Nanobody is high level expressed in E. coil with avi-polyhistidine-SUMO fusion, making it easy to purify and store. GNB is prepared freshly by incubating the purified GFP-Nanobody with streptavidin resins and followed by a proper wash. The fresh-made GNB is ready to pull down the GFP fusion protein from the cell supernatant. The pull-down procedure usually takes half an hour and can be tracked by green fluorescence. With a proper wash of the GNB, SUMO protease is added to cleave out GFP fusion protein together with GFP-Nanobody from resins and additional thrombin protease would cleave off the GFP/GFP-Nanobody. The untagged target protein is fractionated on size exclusion chromatography.

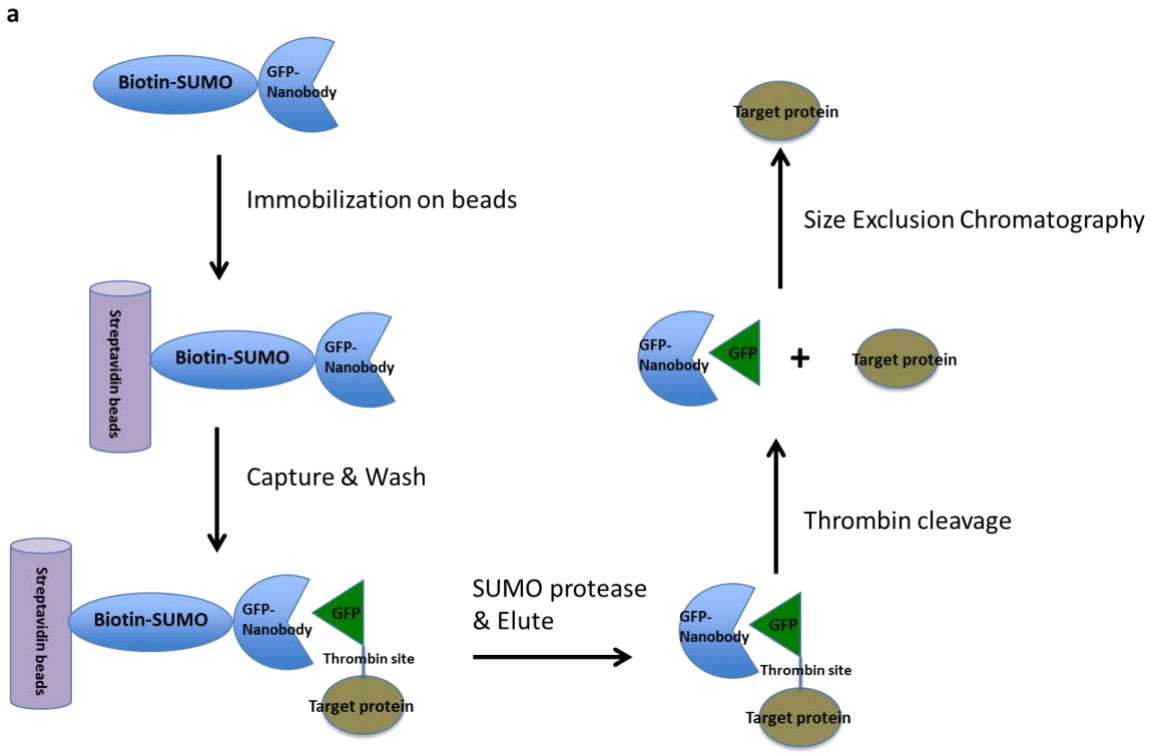


Figure 46. Schematic of the GFP-Nanobody assisted purification.

Expression and purification of biotin-labeled GFP-Nanobody

The GFP-Nanobody is fused in frame with upstream AviTag, Polyhistidine tag and SUMO tag in pET-28a vector. AviTag is for biotin label, where biotin ligase catalyzes amide linkage between the biotin and the specific lysine of the 15-aa AviTag peptide[134]. Polyhistidine tag is fused to purify the GFP-Nanobody from E. coli expression system with Ni NTA resins. The SUMO fusion is to enhance the expression and make it possible to be cleaved off the Nanobody from resins.

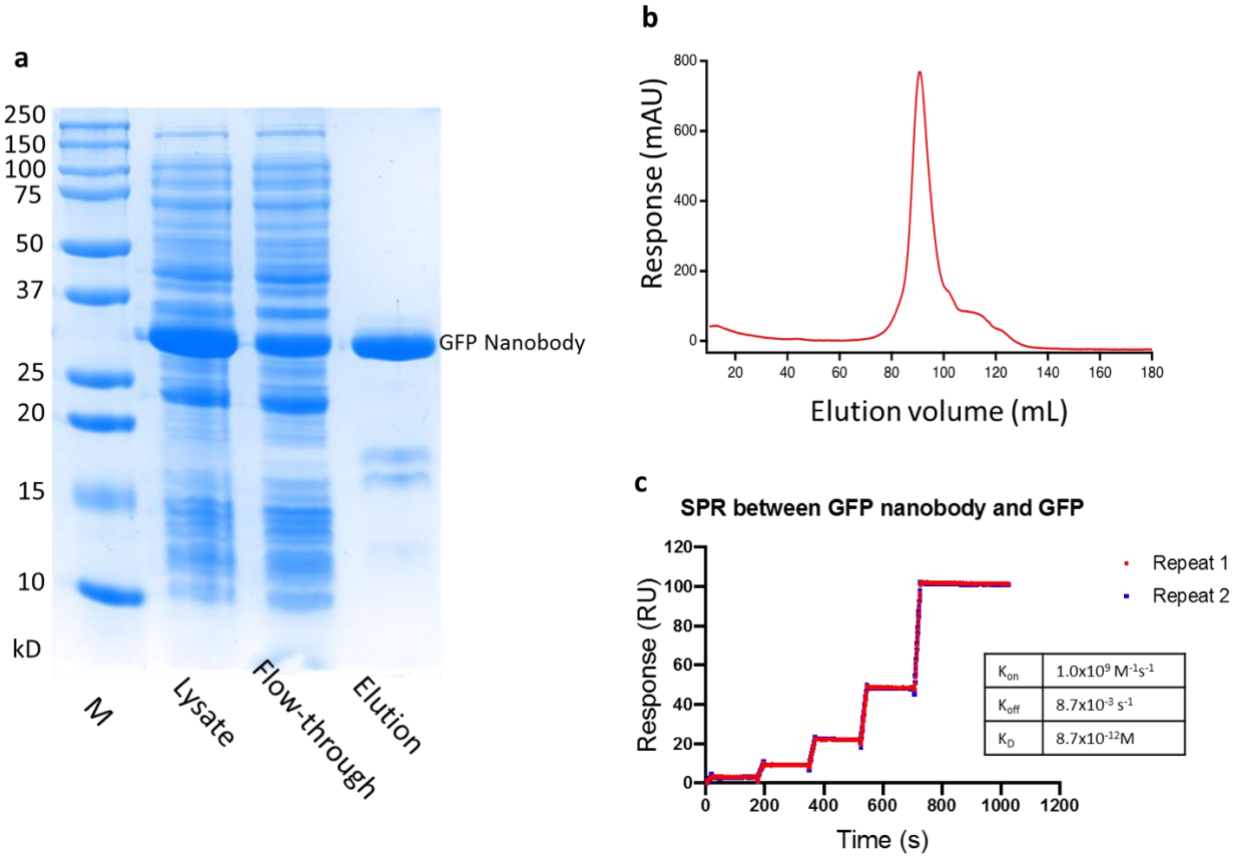


Figure 47. The purification of GFP-Nanobody and binding study.

The Avi-Polyhistidine-SUMO tagged GFP-Nanobody is highly expressed in E. coli BL21 strains. The GFP-Nanobody is purified by Ni-NTA affinity column, with subsequent size exclusion chromatography (Figure 47). The yield is high for 30 mg protein per Liter medium.

In SPR characterization, the GFP-Nanobody has an extremely high affinity to bind GFP (8.7 pM) (Figure 47). The measured K_{on} is $1.0x10^{9}$ $M^{-1}s^{-1}$ and k_{off} is $8.7x10^{-3}$ s^{-1}. It is considered that GFP Nanobody works as a bridge for SPR chip and GFP protein. It binds

to streptavidin SPR chip with its biotin label and binds to GFP with Nanobody. The SPR response is stable and horizontal with the wash, indicating the GFP-Nanobody works great to pull-down GFP proteins and maintain this tight affinity.

Purification of GFP tagged STING by GNB

Here, we test the purification of GFP fused integral membrane protein STING and its soluble truncation residues 155-379. $STING_{155-379}$ does not have transmembrane regions. We fuse $STING_{155-379}$ with an N-terminal GFP tag and express it in the E.coli system. $STING_{FL}$ is a transmembrane protein and we express it in 293F cells with a C-terminal GFP tag. The purification of GFP tagged protein can be tracked by green fluorescence. Photos of the purification of $STING_{155-379}$ are recorded (Figure 48).

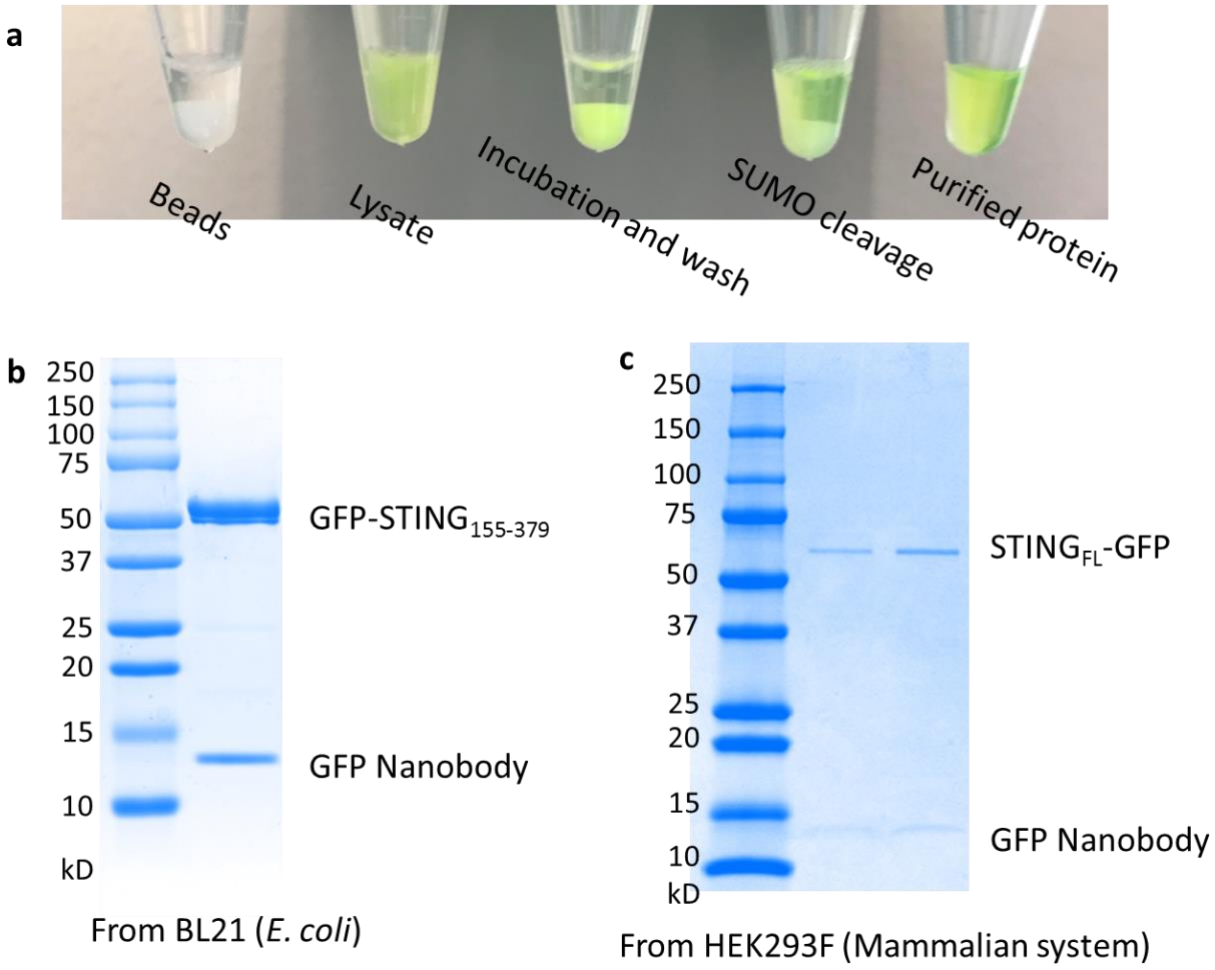


Figure 48. Purification of GFP-STING

In GFP-STING$_{155\text{-}379}$ purification, after incubating GNB with cell supernatant for 30 mins, GFP fusion protein goes to resins and leaves no green fluorescence in the supernatant, indicating all GFP fusion protein is captured by GNB. The GFP fusion protein-coupled resins are washed with Running Buffer to remove background proteins and the supernatant goes to clear. After adding SUMO protease to resins, SUMO protease cleaves after the C terminal of the SUMO tag. Leaving the SUMO tag with resins, the

GFP fusion protein coupled with GFP-Nanobody returns back to supernatant again. The GNB purified GFP $STING_{155-379}$ together with GFP-Nanobody is analyzed by SDS-PAGE (Figure 48).

In $STING_{FL}$-GFP purification, lysis buffer including 1% DDM detergent is added to cell lysate to extract full-length STING from cell membranes. After ultracentrifugation to remove cell debris, GNB is mixed with cell supernatant and standard purification protocol is followed with SUMO protease. GNB purification works effectively to produce a high-purity yield, as analyzed by SDS-PAGE (Figure 48).

Purification of untagged $STING_{FL}$

Because SUMO protease cleaves at the C-terminal of SUMO tag, GFP-Nanobody together with the GFP fusion protein is eluted from beads. As in plasmid constructs, the thrombin site is pre-inserted between the target protein and GFP tag. To purify the untagged protein, the thrombin enzyme is added to cleave between the target protein and GFP/GFP-Nanobody. In full-length STING purification, as STING is activated by cGAMP, it forms oligomers in ~400kD complex, which is much larger than 40kD GFP/GFP-Nanobody. The STING oligomers are cleaved off the GFP/GFP-Nanobody with thrombin and separate with GFP/GFP-Nanobody in a superose 6 size exclusion column (Figure 49). The purity of untagged STING is also analyzed by SDS-PAGE (Figure 49).

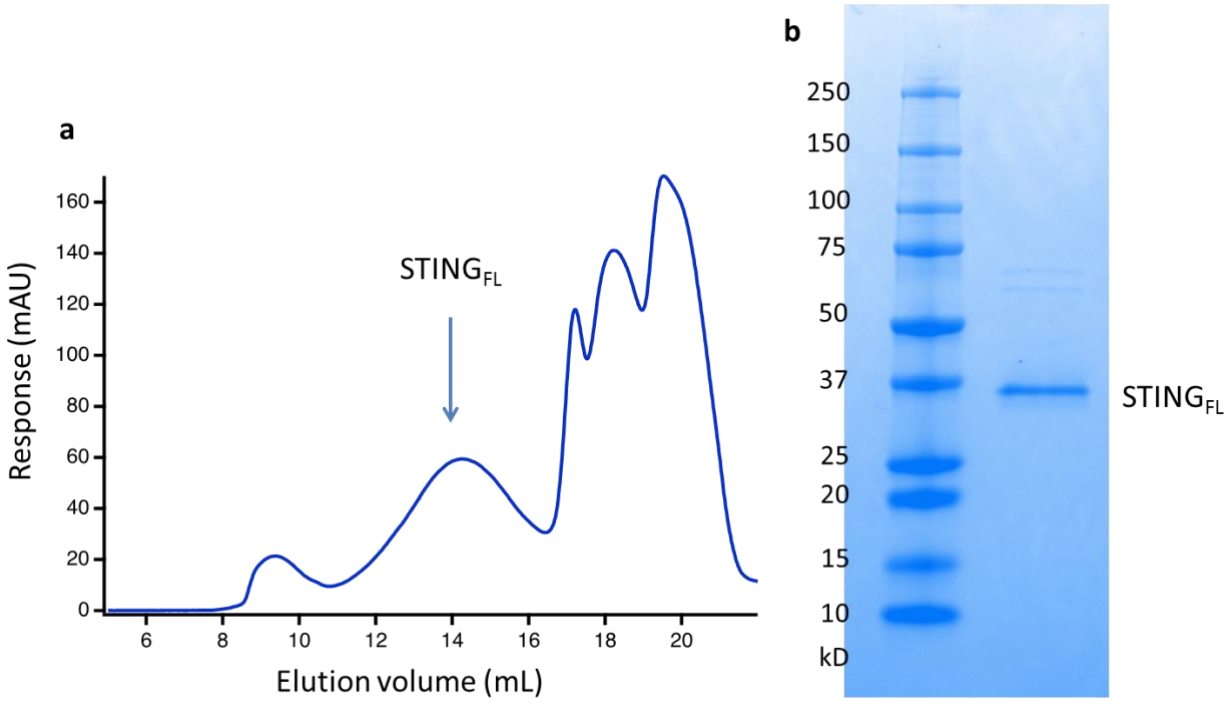


Figure 49. Purification of untagged STING

On-Chip purification and binding detection for GFP tagged protein

As for the conventional SPR experiment, immobilized protein is crosslinked to chip by amine groups. It requires the purification of this immobilized protein prior to detecting binding, or cellular protein will make contamination and interfere with the affinity detection. The purification step limits the usage of SPR because many proteins are difficult to high-level expressed or purified. On the other hand, the usage amount of the immobilized protein in SPR is very small. It only requires about 1 ug protein to coat the chip. Here we use GFP-Nanobody coated SPR chip to immobilize GFP fusion protein in cell supernatant, which does not require the purification step. The STING$_{155\text{-}379EFEEMW}$ is

a mutant of STING, which has a high affinity to $TBK1_{S172A}$ as reported[135]. We express the GFP fusion $STING_{155\text{-}379EFEEMW}$ in E. coli and flow through the cell supernatant to the GFP-Nanobody coated chip. The chip works to immobilize enough GFP-STING with only 100uL cell lysate. After washing out the background protein and getting a stable baseline, the sequentially diluted TBK1 is flowed through to detect affinity. The readout Kd of GFP-$STING_{155\text{-}379EFEEMW}$ to $TBK1_{S172A}$ is 0.5 uM (Figure 50), which is in the same order as the reported Kd 1.4 uM[135].

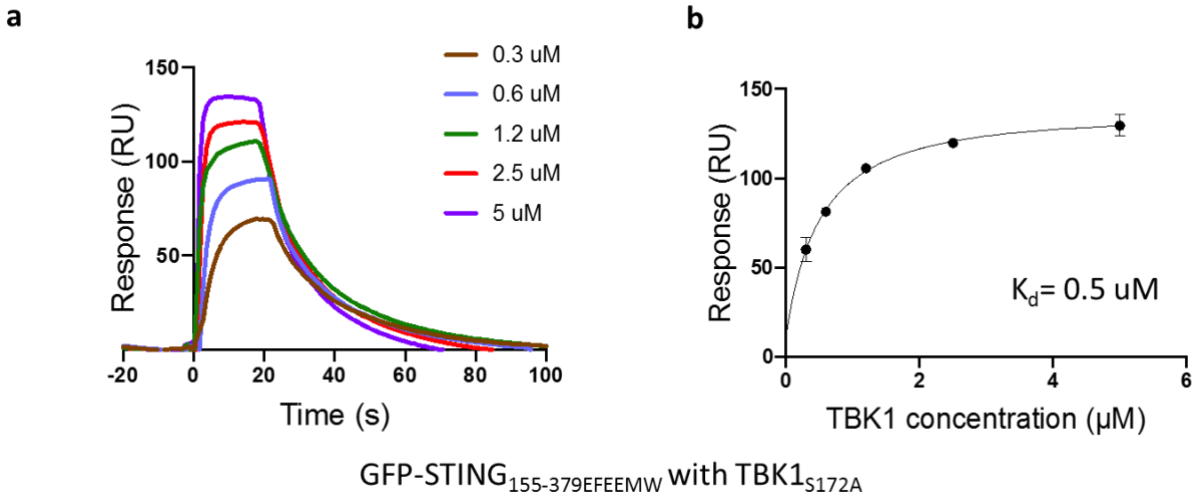


Figure 50. SPR of GFP-STING with TBK1

Discussion

The GFP-Nanobody is small (13kD) and stable[133]. With avi-polyhistidine-SUMO fusion, we purified 200mg GFP-Nanobody from 6L LB medium. At 4 mg/ml, purified GFP-Nanobody is stored in a -80°C refrigerator. To make GNB freshly, we mix the thawed GFP-Nanobody with streptavidin resins every time before purifying GFP

fusion protein. It avoids the frequent freeze and thaws for GFP-Nanobody coated streptavidin resins.

The GFP-Nanobody is biotin labeled to immobilize on the streptavidin resins. The biotin-streptavidin interaction system is the strongest noncovalent biological interaction known, having a dissociation constant, Kd, in the order of 4×10^{-14} M[136, 137]. This affinity pair is widely used due to its stability in large pH and temperature ranges. The streptavidin resin has a biotin-binding capacity of >120 nmol/ml of gel, which is calculated to bind 3 mg avi-polyhistidine-SUMO GFP-Nanobody per milliliter of resins. The extra GFP-Nanobody, which exceeds the capacity of resins, must be washed off before adding GFP fusion proteins. This washing usually takes 5 times with 5-fold volume. Washing the resins also gets rid of the contaminated proteins in GFP-Nanobody due to the defect of Ni NTA affinity purification.

The purification method is suitable when the target protein is above 150kD. For protein around 40kD, as the size is similar to the GFP/GFP-Nanobody complex, it is difficult to separate the target protein from GFP/GFP-Nanobody on size exclusion chromatography after thrombin cleavage. For such cases, when GFP fusion protein is bound to GNB, thrombin can be added directly. Thrombin does not work as efficiently as SUMO protease to cleave the protein from resins. But in difficult circumstances like this, we make use of the advantage that thrombin elutes the tag-free target protein in one step. As green fluorescence does not go together with the target protein, the time course experiment for the thrombin cleavage should be evaluated by SDS-PAGE.

In conclusion, we have developed a cleavable GFP-Nanobody system for GFP-fusion protein purification. By combining GNB with the HEK293F mammalian expression system, we have successfully expressed and purified the integral membrane protein STING-cGAMP complex. The whole expression and purification procedure are tracked by green fluorescence. The highly efficient GFP-Nanobody coated streptavidin beads preparation and simple, quick, and easy-tracking purification process will make this system generally applicable to other low-level expressed macromolecular purification.

Biophysically characterize the full-length STING-cGAMP complex

The formation of STING oligomer upon cGAMP binding is identified in cell research. Here we purify full-length STING protein out of HEK293F mammalian cells and characterize the properties of full-length STING before and after cGAMP activation.

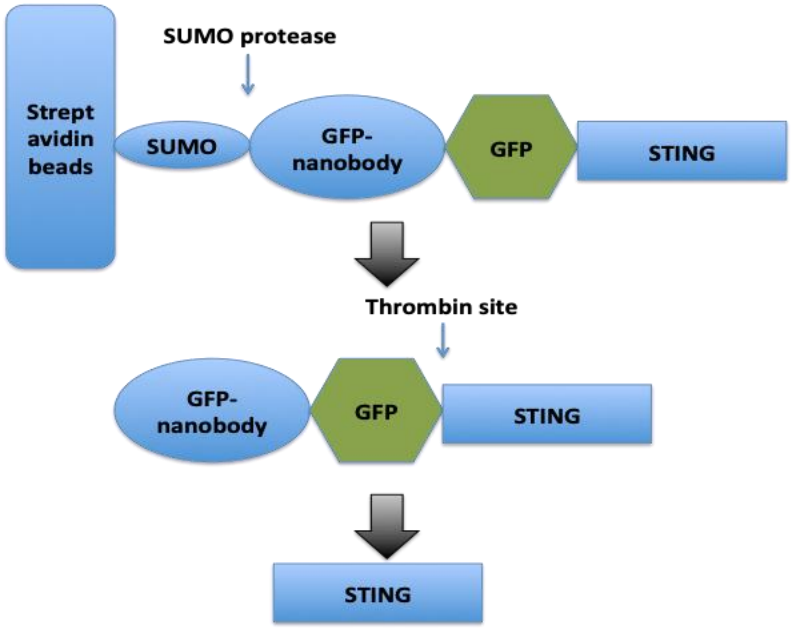


Figure 51. Schematic of the purification process of full-length STING

According to the current model of STING activation, cGAMP binding induces the oligomerization of STING, releasing its C-terminal tail (CTT) to mediate the recruitment and activation of TBK1. Next, phosphorylated STING recruits IRF-3 to the signaling complex. The crystal structures of the human STING C-terminal domain (CTD) in isolation and in complex with cGAMP have been determined. These structures reveal that STING forms a dimer and the cGAMP binding site is located at the dimer interface. Ligand binding induces a pronounced conformational change of the STING dimer. However, STING CTD will not oligomerize upon cGAMP binding in vitro and cannot be phosphorylated by TBK1 in vivo. Luciferase assay indicates that cells have no STING mediated signaling if STING is truncated on the N-terminal transmembrane domain. Here, we establish the HEK293F mammalian cell expression system, which successfully expressed the full-length STING at ~1mg per liter cells. Biophysically characterizing the cGAMP-STING complex will provide insights into how cGAMP binding induces the oligomerization of STING and activates STING mediated signaling. Other than the structural characterization of size and shape, as it is the first time that we obtained the full-length STING in vitro, we also aim to test the interaction between full-length STING and kinase TBK1 with or without cGAMP activation, which will give insights into how cGAMP activates the STING-TBK1 pathway.

Purify full-length STING with detergent

To investigate how ligand binding induces the activation of STING, we have recently established a protocol to express GFP fusion of full-length human STING in 293F cells. To purify the STING-GFP fusion, we generated biotin-labeled SUMO fusion of a

GFP nanobody (nAb) that binds GFP with high affinity (K_D ~10 pM by SPR). The GFP-nAb was immobilized on streptavidin agarose beads and used for affinity purification of the STING-GFP fusion from the cell lysate. The STING-GFP fusion was cleaved from the agarose beads with SUMO protease and purified using a Superose 6 column. The yield of the purified STING-GFP is ~1.0 mg/L of cell culture. Next, the GFP tag was cleaved with thrombin and full-length STING was purified using a Superose 6 column with 1 μM cGAMP and 0.1% DDM in the running buffer. STING oligomer elutes on chromatography at a single peak around ~400 kD (**Fig.** 2.2A). After DDM is exchanged to another membrane-stabilizing reagent Amphipol A8-35, it elutes at the same position, suggesting that cGAMP binding induces the formation of a stable STING oligomer.

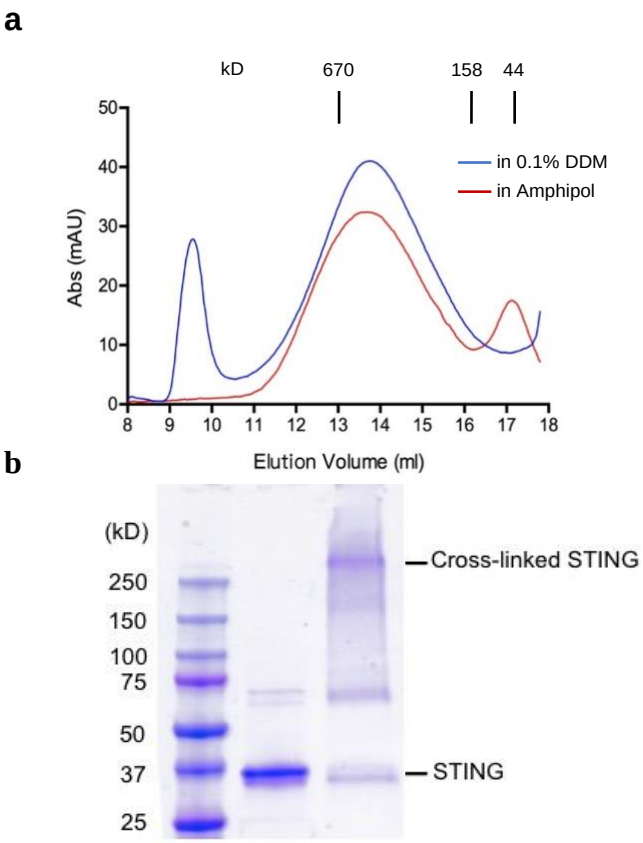


Figure 52. The crosslinking and detergent substitution of STING oligomer with cGAMP.

a. Chromatography profile of STING oligomer in different detergents b. Cross-linked STING oligomer presents a high molecular weight on SDS-PAGE.

Estimate the size of full-length STING oligomer through Crosslinking

Crosslinking reagent glutaraldehyde covalently binds amine groups in the side-chain of lysine residues, making it possible to maintain the intact structure of complex in different conditions. Due to the fact that the spacer length of glutaraldehyde is around 5Å, it usually will not make artifacts by crosslinking non-specific binding regions. Here we

use the “on-column” cross-linking strategy to stabilize the full-length STING oligomers on a Superose 6 column [55]. This method facilitated near-complete cross-linking of the preformed complexes at relatively high concentrations and simultaneously enabled the isolation of samples of different sizes. After loading the cross-linked STING oligomer to SDS-PAGE using a 4-20% gradient gel, the main protein band with a molecular mass of ~360kD was observed, (Figure 52) which agreed with the results from gel filtration chromatography. Based on these results, we estimate that there are 4~5 STING dimers present in the oligomers of the STING-cGAMP complex.

Determine the structure of the STING-cGAMP complex

Although the cytosolic domain of STING has been solved by X-ray crystallography, the truncated STING cannot form an oligomer and it cannot function to mediate TBK1 activation in cells. Since we successfully purified full-length STING, we aim to determine the oligomer structure either by Crystallography or Cryo-electron microscopy. To stabilize the transmembrane domain of STING, multiple detergents such as N-Dedecyl-beta-Maltoside (DDM) or amphipol A8-35 would be optimized in the purification and structural determination process. Basically, TEM FEI Tecnai F20 were used to collect micrographs and the data was analyzed by Relion 3.0 software to eventually generate the three-dimensional model.

Estimate homogeneity by negative stained electron microscopy

To confirm whether the full-length STING oligomer is suitable for an electron microscopy study, a negative stain screen is necessary beforehand. The full-length STING oligomer is diluted to 0.01 mg/ml and stained with 0.75% Uranyl Formate. Negative

stained specimen indicates that STING dimers likely stack against each other side by side to form oligomer. STING oligomer also presents a homogeneous formation after staining, which supports the subsequent cryo-EM study.

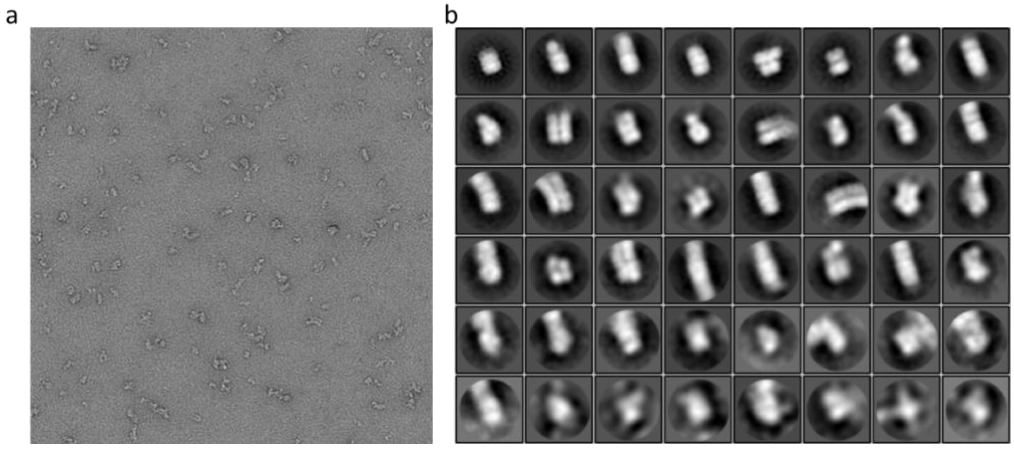


Figure 53. Negative Stain of human FL STING.
a. The raw micrograph of human FL STING oligomer with cGAMP. b. 2D averages of the particles.

Reconstruct the STING oligomer by Cryo-electron microscopy

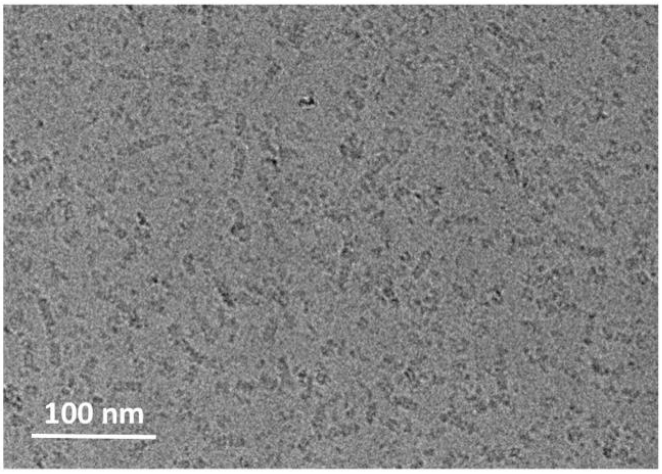
Cryo-EM is a form of transmission electron microscopy (TEM). By averaging millions of particles from diverse orientations, the single-particle reconstruction could reach an atomic resolution, revealing protein-protein interactions in the physiological state. To increase specimen homogeneity, I applied the “on column” crosslinking of STING oligomer. Then I changed detergent DDM to amphipol A8-35, which gives better particle distribution in Cryo-EM. The final protein at a concentration around 0.6 mg/ml is applied to glow discharged C-flat carbon grids and frozen by Vitrobot in liquid ethane. The raw image presents mono-dispersed layer stack particles. The subsequent 2D

classification analysis outputs a projection around 7Å resolution, indicating the particles are quite homogeneous and suitable for EM study. However, in cryo-EM 3D structure reconstruction, we encounter a preferred orientation issue, where the transmembrane domain prefers to stay in the air-water interface. In such circumstances, we cannot get structural information from other directions and the structural model is less accurate accordingly. The 3D reconstruction is reasonable because the structure of the cytosolic domain fits the bottom part and the map contains extra density where the transmembrane domain may exist. The 3D map is ~17Å, which requires improvement by solving the preferred orientation issue.

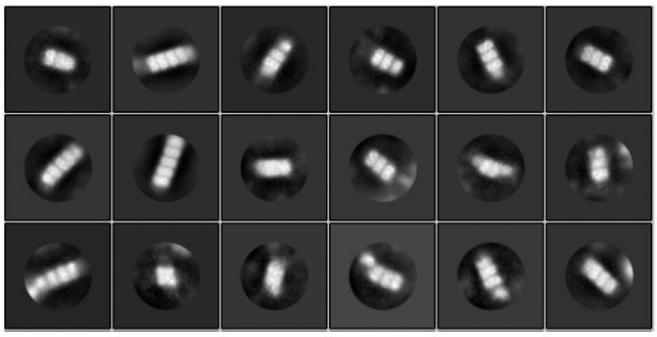
Optimize the particle distribution targeting preferred orientation

To overcome the preferred orientation issue, we need to deal with the hydrophobic property of the transmembrane domain of the STING oligomer. One of the methods is to add the detergent back to the oligomer. I try to add 0.01% DDM or Amphipol A8-35 back to change the physical properties of the water solution. In theory, DDM containing buffer will weaken hydrogen bonds that make water molecules stick to each other and decrease the surface tension of water. 0.01% DDM is below the critical micelle concentration (CMC) so that it will not form micelles that affect the cryo-EM image collection.

a



b



c

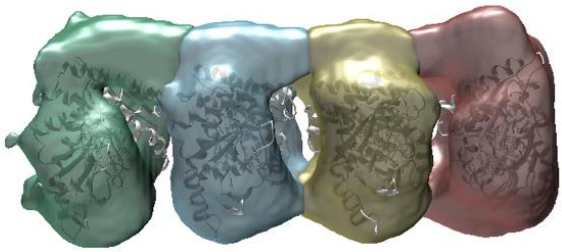


Figure 54. The cryo-EM study of STING oligomer with cGAMP.
a. Raw micrographs of STING-cGAMP complex. b. Class 2D average analysis of the particles. c. Refined 3D model at 11Å

However, adding back detergent does not critically affect the specimen property. Finally, I choose to tilt the grids in data collection, which is also a method targeting the preferred orientation. The dataset was manually collected with a TF20 microscope at 0 º, 15 º, 30 º, and 45 º tilting degrees. The data process of the tilted data indicates a significant improvement in the angular distribution of particles.

We set up the collaboration with the EM facility at Florida State University, which runs a 300kV Titan Krios microscope equipped with the K3 camera. The whole dataset collected includes 3956 micrographs with 15 º and 30 º tilting, and 880,703 particles are picked (**Fig.** 3A).

Structural determination of STING by Relion

After we solve the preferred orientation problem, I use Relion 3.0 to solve the EM structure. Specifically, the micrograph stack is subjected to software MotionCor2 for whole-frame motion correction. Particle picking is first performed manually on the summed images in EMAN2.2 to generate a reference for auto-picking. Gautomatch and Gctf are used to pick particles and estimate local CTF relatively. The generated particles are subjected to Relion 3.0 to analyze 2D classification and 3D classification. The 3D auto refinement is also run in Relion 3.0, which generates the high-resolution map. The 3D reconstruction increases resolution from 17Å to 11Å with a much-improved angular distribution (**Fig.** 3C).

Discussion and future studies

After we determine the structures of full-length STING, we can uncover the mystery of STING oligomerization via its N-terminal transmembrane domain. Then we aim to do mutagenesis studies on the newly found interactions to probe if we could block or manipulate the formation of full-length STING oligomer. Based on the structure, we are able to investigate how reported mutations of STING, such as V147L, N154S, and V155M, cause autoimmune disorders. The atomic resolution model coupled with a mutagenesis study will propose potential drug sites for subsequent research.

To elucidate the mechanism of STING activation, we determine the structure of full-length STING bound to cGAMP by cryo-EM. These structural studies provide critical insights into the mechanism of STING activation by cGAMP, which is essential for the cGAS-STING signaling pathway. As it is in biophysical characterization, the purified full-length STING with cGAMP exists as a stable oligomer of ~360 kD. The homogeneous transmembrane complex is suitable for structural determination. For cryo-EM data collection, the grids of the STING complex are first screened to optimize the sample condition. We conduct large-scale data collection from an optimized specimen on our FEI Tecnai F20 cryo-electron microscope equipped with a Gatan K2-summit electron counting direct detection camera. The cryo-EM images are analyzed using Relion 3.0 on our campus supercomputer cluster to generate the three-dimensional density map. We will dock the STING CTD model into the map and build a molecular model of transmembrane helixes using Chimera or Coot.

The innate immune response mediated by the cGAS-STING pathway is critical in host defense against viral infection. Our research represents a rigorous and comprehensive investigation into the molecular bases of key steps of the cGAS-STING pathway, especially the STING activation progress.

It's worth it to investigate how full-length STING functions to oligomerize and mediate TBK1 activation in cells. Considering the truncated cytosolic domain cannot either oligomerize or mediate pathways downstream, the transmembrane domain interaction might be the key step after cGAMP activation. After we determine the structure of the full-length STING oligomer, we build an atomic model and determine the interactions. We aim to explain the conformational changes of STING after cGAMP activation. Based on the newly found interactions, several mutations could be made accordingly. The subsequent in vivo and in vitro functional studies will help validate the structure. This structure will have more importance than that has been solved, because it is the first full-length structure contains the oligomerized transmembrane domain and it is solved by Cryo-EM, which determines a structure in physiological status. As the interferon-β is reported to repress cancer stem cells, all the proteins involved in the cGAS-STING-IRF3 pathway may work as a drug target against cancer. The structural model of full-length oligomer will give insights into the mechanism of STING activation by cGAMP, revealing the possibility to interfere with this pathway by drug development.

It's been reported that the combination of point mutations on the trans-membrane domain of STING, including the combination of E68A, E69A, Q74A, S75A, and R76A, will completely abolish IFN induction but not cGAMP binding. These residues locate

inside the transmembrane domain and are predicted to be on loops outside the membrane. We propose these mutations may affect the formation of STING oligomer. Here we split up the combination of these mutations into single point mutation, including E68A, E69A, S75A, R76A, S80A, R83A, R86A, E143A, S145A, and E149A. Each of the STING mutants is transfected into HEK293T cells, which does not express STING. After transfection, the cells are stimulated with cGAMP. The expression of the IFN-β reporter is analyzed by luciferase assay. The reporter assay indicates that some of the single point mutants (E68A, E69A) abolish IFN induction, while R76A, R83A, R86A, E143A, S145A, and E149A weaken the STING function. We will couple this result with the structural model to explain the functions of these residues.

Our studies will dramatically advance our understanding of the molecular basis of innate immunity against viral DNA. In addition, we will provide the structural basis for conceptually new approaches to treat viral diseases, cancer, and autoimmune disorders.

Materials and Methods in detail

Restriction enzymes were obtained from New England Biolabs. Bacterial expression vector pET28(a) was purchased from Novagen (EMD Millipore). Mammalian cell expression vector pEGFP-N1 was purchased from Clontech (Takara). Biotin Ligase plasmid pBirAcm was purchased from Avidity. Avi-SUMO tag sequence and GFP-Nanobody sequence were synthesized by GenScript. Human STING cDNA was purchased from Open Biosystems. pAcGHLTc baculovirus transfer vectors and Baculo-Gold bright linearized baculovirus DNA were purchased from BD Bioscience. Streptavidin agarose resin was purchased from Goldbio.

Protein expression and purification

Saccharomyces cerevisiae Ulp1 fragment $Ulp1_{403-621}$ has been shown capable to cleave a SUMO tag at its C-terminal in vitro[138, 139]. The open reading frame of $Ulp1_{403-621}$ SUMO Protease was subcloned to the pET28a vector to express a 6xHis-$Ulp1_{403-621}$ fusion protein.

pET-28(a) 6xHis-$Ulp1_{403-621}$ was transformed into Escherichia coli strain Bl21(DE3) on 50 ug/mL Kanamycin LB agar plates overnight. Several colonies from transformation plates were inoculated into 50 mL LB medium with 50 ug/mL Kanamycin and shaken at 37 °C for 4 hours. Subsequently, all of the 50 mL LB was distributed evenly into 6-liter LB medium with 50 ug/mL kanamycin, and shaken at 37 °C until the OD_{600} reached about 1.0. LB medium was cooled down by ice to 15 °C. and 0.4 mM isopropyl β-D-thiogalactoside (IPTG) was added to the medium. Cells were harvested at 16 hours post-induction and stored at -80°C refrigerator.

SUMO protease was purified through Ni NTA and Superdex 200 gel filtration columns as described[101, 140, 141]. Briefly, cell pellets were resuspended by 200 mL lysis buffer (50 mM Tris, 300mM NaCl, 1mM PMSF, pH 7.8) and applied to sonication until the lysate become non-viscous. The non-viscous lysate was centrifuged at 16000g, 4 °C for 30 min to remove the cell debris from the supernatant. The supernatant was loaded to 20 mL packed Ni NTA resin and washed with 200 mL washing buffer (50 mM Tris, 500mM NaCl, 20 mM imidazole, pH 8.0). The target protein was eluted with elution buffer (50 mM Tris, 300mM NaCl, 250 mM imidazole, pH 7.4) and concentrated to 2 mL

to Superdex 200 gel filtration column. The peak fractionation was collected and analyzed by SDS-PAGE. Pure fractionation was concentrated, freshly frozen by liquid nitrogen, and stored in a -80°C refrigerator.

pET-28(a) Avi-6xHis-SUMO GFP-Nanobody

Avi-His6-SUMO tag was cloned into the protein expression vector pET28a between NdeI and BamHI. The open reading frame of GFP nanobody was cloned into this modified avi-His6-SUMO pET28(a) at BamHI and XhoI.

pET-28(a) Avi-6xHis-SUMO GFP-Nanobody was co-transformed with pBirAcm plasmid into Escherichia coli strain Bl21(DE3) on 50 ug/mL Kanamycin, 50 ug/mL Chloramphenicol LB agar plates overnight. Several colonies from transformation plates were inoculated into 50 mL LB medium with 50 ug/mL Kanamycin, 50 ug/mL Chloramphenicol, and shaken at 37 °C for 4 hours. Subsequently, all of the 50 mL LB was distributed evenly into 6-liter LB medium with 50 ug/mL kanamycin, 50 ug/mL Chloramphenicol, and shaken at 37 °C until the OD_{600} reached about 1.0. LB medium was cooled down by ice to 15 °C. 5 mM Biotin and 0.4 mM isopropyl β-D-thiogalactoside (IPTG) was added to the medium. Cells were harvested at 16 hours post-induction and stored at -80°C refrigerator.

Biotin-labeled SUMO GFP-Nanobody was purified through Ni NTA and Superdex 200 gel filtration columns. Briefly, cell pellets were resuspended by 200 mL lysis buffer (50 mM Tris, 300mM NaCl, 1mM PMSF, pH 7.8) and applied to sonication until the lysate become non-viscous. The non-viscous lysate was centrifuged at 16000g, 4

°C for 30 min to remove the cell debris from the supernatant. The supernatant was loaded to 20 mL packed Ni NTA resin and washed with 200 mL washing buffer (50 mM Tris, 500mM NaCl, 20 mM imidazole, pH 8.0). The target protein was eluted with elution buffer (50 mM Tris, 300mM NaCl, 250 mM imidazole, pH 7.4) and concentrated to 2 mL to Superdex 200 gel filtration column. The peak fractionation was collected and analyzed by SDS-PAGE. Pure fractionation was concentrated, freshly frozen by liquid nitrogen, and stored in a -80°C refrigerator.

GFP-Nanobody coated streptavidin beads preparation

2 mL Streptavidin resin was washed with 5 mL running buffer (20 mM Tris, 150mM NaCl, pH 7.4) 5 times and incubated with 6 mg purified biotin-labeled SUMO GFP-Nanobody on ice for 20 min. The resin was then washed with 5 mL running buffer another 5 times and ready to use.

pET-28(a) GFP STING$_{155\text{-}379}$

GFP tag was cloned into protein expression vector pET28a between NdeI and BamHI. The open reading frame of human STING$_{155\text{-}379}$ was cloned into the modified GFP pET28(a) at BamHI and XhoI. Human STING$_{155\text{-}379}$ Q359E, P361F, S366E, T376E, F378M, and S379W mutation was generated by site-directed mutagenesis.

The expression of GFP STING$_{155\text{-}379}$ and its mutant is the same as the SUMO protease in BL21. In purification, 10 mL cell pellets of GFP-STING$_{155\text{-}341}$ were resuspended by 200 mL lysis buffer (50 mM Tris, 300mM NaCl, 1mM PMSF, pH 7.8)

and applied to sonication until the lysate become non-viscous. The non-viscous lysate was centrifuged at 16000g, 4 °C for 30 min to remove the cell debris from the supernatant. The supernatant was incubated with the 2 mL prepared GNB for 2 hours at 4°C with slight shaking. The beads were pelleted and washed with 30 mL running buffer for 5 times. After the final round of washing, 5 mL running buffer together with 0.1 mg SUMO protease was added to GNB to cleave the target protein from the beads. GFP-STING$_{155\text{-}379}$ was eluted after a 1-hour incubation at 4 °C.

pEGFP-N1 human STINGFL

Full-length human STING was cloned into a pEGFP-N1 plasmid with a GFP tag attached to the C terminus of STING. A thrombin cleavage site was inserted between STING and GFP.[135]

Full-length STING was expressed in FreeStyle 293-F cells. Cells were seeded into fresh FreeStyle 293 expression media with a final density of 1.0 × 106 cells ml−1 and incubated at 37 °C, 8% CO2, 130 r.p.m. After 24 h, 1 mg pEGFP-N1-STING plasmids, 2 mg of linear PEI25000 were mixed into 100 ml 1 × PBS and incubated at room temperature for 20 min. Then, the mixture was added into 1-liter FreeStyle 293-F cells. After 2 days, the cells were pelleted at 3000 rpm for 5 min and stored at -80°C refrigerator.

The pellets were resuspended in 40 ml pre-chilled lysis buffer (20 mM Tris pH 7.5, 150 mM NaCl, 1% n-dodecyl-β-d-maltoside (DDM)) containing the protease cocktail inhibitors, sonicated for 3 cycles, and shaken gently at 4°C for 2 h. After centrifugation at

16000 rpm for 30 min, the supernatant was mixed with 500 uL prepared GNB. After shaking at 4°C for another 2 h, the beads were pelleted and washed 5 times with pre-chilled washing buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.1% DDM). Finally, the beads were resuspended in 500 μl washing buffer and SUMO protease was added to cleave the target protein from the beads. The eluted $STING_{FL}$-GFP protein was incubated with thrombin at 4 °C overnight and further purified using a Superose 6 increase 10/300 GL column (GE Healthcare).

pAcGHLTc hTBK1(residues 1-657S172A)

C terminal truncated (residues 1-657) human TBK1 was cloned into the pAcGHLTc baculovirus transfer vectors[140]. The vector contains a GST tag and a 6xHis tag upstream of the multiple cloning sites. The plasmids were transfected with Baculo-Gold bright linearized baculovirus DNA into SF9 insect cells to generate recombinant baculovirus. The recombinant virus was amplified for 2 rounds before being used for large-scale expression. The insect SF9 cells were infected with the P3 virus and harvested 72 hr after infection. The cells were lysed in the lysis buffer (0.2M Tris, 150 mM NaCl, 1 mM PMSF, 1% NP-40, pH 8.0). The recombinant protein was purified by Ni NTA affinity chromatography followed by gel filtration chromatography.

Protein expression and purification

Constructs of human STING were cloned into a modified pET28(a) vector with an N-terminal Avi-His6-SUMO tag. For protein quantification, the mutation Val343Trp was introduced into the human STING C-terminal tail (residue 342–379). All the proteins were

expressed in BL21 (DE3) cells in M9 medium. The cDNA encoding human IRF-3 (residue 189–427) was cloned into another modified pET28(a) vector with an N-terminal His6-SUMO tag. IRF-3 was also expressed in BL21 (DE3) cells in regular LB medium (BD). When OD600 reached 1.0, the cells were induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and cultured overnight at 16 °C. The proteins were purified using the protocol described previously24. In brief, the proteins were first purified using a nickel-NTA column. The SUMO tag was then cleaved using SUMO protease and removed using a nickel-NTA column. The proteins were further purified by gel filtration chromatography using Superdex75 or 200 (16/60 GL) columns (GE Healthcare). The Se-methionine substituted human STING was expressed in M9 medium supplemented with selenomethionine (Acros Organics) and purified as described for the wild type protein. Human and mouse TBK1 (residue 1–657) were expressed in sf9 insect cells after infection with recombinant baculovirus and purified as described previously25. Biotin-labeled-Avi-His6-SUMO proteins and peptides were also expressed in BL21 (DE3) cells in M9 medium, with the exception of the biotin-labeled Avi-His6-SUMO-GFP nanobody, which was expressed in regular LB medium. The cells were co-transformed with the pET28(a) plasmid coding for the target proteins and the pBirAcm plasmid coding for BirA and induced with 0.4 mM IPTG in the presence of 5 μg ml−1 biotin and cultured at 16 °C overnight. The SUMO fusion proteins and peptides were first purified using a nickel-NTA column and further purified on a Superdex75 or 200 columns (GE Healthcare).

Full-length human STING was cloned into pEGFP-N1 plasmid with a green fluorescent protein (GFP) tag attached to the C terminus of STING. A thrombin cleavage

site was inserted between STING and GFP. FreeStyle 293-F cells were seeded into fresh FreeStyle 293 expression media with a final density of 1.0 × 106 cells ml−1 and incubated at 37 °C, 8% CO2, 130 r.p.m. After 24 h, 1 mg pEGFP-N1-STING plasmids, 2 mg of linear PEI25000 (Polysciences, Inc.) were mixed into 100 ml 1 × PBS and incubated at room temperature for 20 min. Then, the mixture was added into 1L FreeStyle 293-F cells. After 2 days, the cells were pelleted at 3000 rpm for 5 min. The pellets were resuspended in 40 ml pre-chilled lysis buffer (20 mM Tris pH 7.5, 150 mM NaCl, 1% n-dodecyl-β-D-maltoside (DDM)) containing the protease cocktail inhibitors (Roche), sonicated for 3 cycles, and shook gently at 4°C for 2 h. After centrifugation at 16000 rpm for 30 min, the supernatant was mixed with 500 μl streptavidin agarose beads (EMD Millipore), which have been coupled with biotin-labeled Avi-SUMO-GFP nanobody. After shaking at 4°C for another 2 h, the beads were pelleted and washed with pre-chilled washing buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.1% DDM). Finally, the beads were resuspended in 500 μl washing buffer and SUMO protease was added to cleave the target protein from the beads. The eluted protein was incubated with thrombin at 4 °C overnight and further purified using a Superose 6 increase 10/300 GL column (GE Healthcare). All mutants of STING and TBK1 were generated using the QuikChange site-directed mutagenesis kit (Agilent) or a PCR-based technique with appropriate primers. The sequences of the plasmids were confirmed by plasmid DNA sequencing.

Electron microscopy

Before cryo-EM imaging, the full-length STING sample in 0.1% DDM was exchanged into Amphipol A8-35. In brief, STING at approximately 1 mg ml−1 was

incubated for around 8 h with threefold excess by mass of A8-35. DDM was removed by incubation with Bio-Beads SM-2 (Bio-Rad) overnight at 15 mg Bio-Beads per milliliter of solution. The sample was filtered to remove Bio-Beads and purified using a Superose 6 increase 10/300 GL column eluted with a buffer containing 20 mM Tris, 150 mM NaCl and 5 μM cGAMP at pH 7.5. For cryo-EM imaging, 3 μl of STING at a concentration of 0.6 mg ml−1 stabilized with A8-35 was applied to glow-discharged C-flat holey carbon grids (1.2/1.3, 400 mesh). Girds were blotted for 8 s and plunge frozen in liquid ethane using a Vitrobot. Images were recorded by Latitude on a Titan Krios Transmission Electron Microscope operating at 300 kV. A Gatan K3 detector was used in counting mode at a nominal magnification of ×64,000 (yielding a pixel size of 1.42 Å). The dose rate on the camera was set to be 23.8 electrons per physical pixel per second. Exposure of 8.4 s was dose-fractionated into 84 movie frames, leading to a total accumulated dose of 100 electrons per Å2 on the specimen. Images were recorded with a defocus in the range from 1.0 to 4.0 μm. Movies were collected with 15° and 30° tilt to address the preferred orientation of particle distribution. A total of 3,956 movies were recorded. The cryo-EM images were subjected to MotionCor2 for whole-frame dose-weighted motion correction. Particle picking was performed automatically on the summed images in Gautomatch. A total of 880,703 particles were picked. Per-particle local CTFs were estimated by GCTF. Two rounds of 2D classification were then performed. After discarding bad class averages, 352,286 particles were re-centered and re-extracted to Relion 3D classification. Initial models for reconstruction were generated from scratch by the EMAN2 ‘e2initialmodel.py’ program using selected unsupervised 2D averages of good quality based on visual

comparison without applying any symmetry. A stack of 47,474 particles was selected to Relion 3D refinement using a 40 Å low-pass filtered reconstruction from the initial model. The final 3D reconstruction of STING oligomer with cGAMP has an overall resolution of 11.6 Å, using gold-standard Fourier shell correlation (FSC) = 0.143 criteria. The cryo-EM maps of STING oligomer were segmented using Segger. We took advantage of the availability of crystal structure of hSTING cytosolic region with cGAMP to guide the decision on the final segmentation choice. After applying the smoothing procedure for four steps to the original electron microscopy map, four regions were obtained. STING CTD structures were docked to the segmentation map using Chimera.

CHAPTER IV

SUMMARY OF STRUCTURAL INSIGHTS INTO THE MECHANISM OF HUMAN CASPASE-4 DETECTING INTRACELLULAR LPS AND FORMING ACTIVE INFLAMMASOME

Introduction

The innate immune system detects the pathogen-associated molecular patterns (PAMPs), such as nucleic acids and LPS, through pattern recognition receptors (PRRs), which triggers the induction of a variety of cytokines including type-I interferons (IFN-I)

to initiate host defense against pathogens [63, 142]. A set of TLRs (Toll-like receptor) have been characterized as PRRs for bacterial and viral products [143].

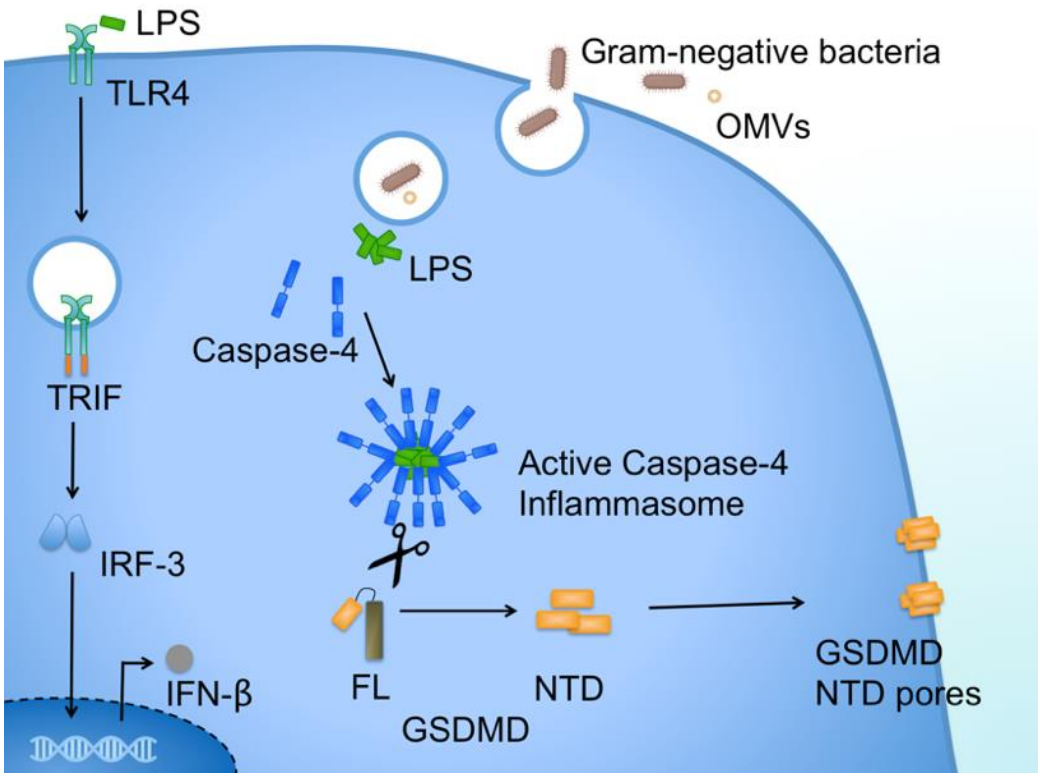


Figure 55. Schematic of the caspase4 inflammasome pathway

LPS (lipopolysaccharide) is one of the most studied immunostimulatory components of bacteria and can induce strong systemic inflammation and sepsis. LPS is an essential and abundant component of the outer membrane of Gram-negative bacteria and some Gram-positive bacteria. LPS contains three parts, the lipidA, core

polysaccharide and O-antigen. LipidA is the innermost of the three regions and its hydrophobic nature allows it to anchor the LPS to the outer membrane. LipidA consists of two phosphorylated glucosamines and six acyl chains [142, 143].

LPS in the extracellular area is detected through a series of interactions with proteins including LPS binding protein (LBP), CD14, MD-2 and TLR-4 [144-147]. After LPS recognition, plasma membrane located TLR-4 undergoes oligomerization and recruits its downstream adaptors including TRIF (TIR domain-containing adaptor inducing IFN-β), which mediates the recruitment and activation of TBK1 (TANK-Binding Kinase 1) and IRF-3 [148, 149]. IRF-3 is activated upon phosphorylation by TBK1, and translocated to the nucleus and initiates the transcription of IFN-β [83]. Thereafter, IFN-β induces the expression of a large set of anti-bacterial proteins, including IRGB10 and GBPs [150, 151].

The intracellular recognition of LPS features the binding of caspase-4 to LPS [152]. Recent structural and functional studies provide an overall picture and some molecular details of the caspase-4 mediated pyroptosis. The apo caspase-4 firstly oligomerizes with intracellular LPS with its CARD, resulting in autocatalysis, self-cleavage and activation to the matured inflammasomes. The matured inflammasome in turn cleaves the substrate GSDMD between the C-terminal (CTD) and N-terminal (NTD) [153, 154]. Full-length apo GSDMD has been revealed autoinhibition that a hydrophobic pocket in the CTD binds to NTD, as shown in the crystal structure [155]. After the linker is cleaved by active inflammasome, the NTD of GSDMD is released and forms pores on the membrane by its hydrophobic nature. The pore contains a 27-28 folds of single

GSDMD NTD and has an inner diameter of 18 nm [156-158]. The LipidA region of LPS is required and sufficient to bind Caspase-4 to form the active inflammasome. The mutagenesis study of caspase-4 reveals the N-terminal CARD 59 amino acids are required for the binding to LPS or LipidA. The substitution of repeated lysine blocks this binding, which might explain the binding is related to the acidic phosphorylated glucosamine on LipidA. These studies have significantly advanced our understanding of intracellular sensing of LPS. However, the structures of the apo caspase-4 CARD or LPS bound caspase-4 CARD have not been determined, neither has the active caspase-4 LPS oligo inflammasome.

To elucidate exactly the binding profile of CARD to LPS and shed light on the mechanism of the prerequisite of LPS binding and oligomer formation for capase4 self-cleavage to activation, the structure of the human hcaspase4 LPS complex should be determined in a hybrid approach of X-ray crystallography and cryo-electron microscopy. The structural studies are needed to confirm the model and define the features, such as size, number of monomers and conformational changes that induce caspase-4 autoactivation. To this end, we have expressed full length caspase-4 or CARD with cleavable SUMO tag. Our size-exclusion chromatography and analytical ultracentrifugation reveal that the purified caspase-4 binds to and disaggregates LPS or LipidA in vitro.

In addition, with more caspase-4 added to the mixture system, the caspase-4 LPS complex goes to smaller size and becomes more homogeneous. To determine the molecular mechanism of caspase-4 autocatalysis and self-activation, it is a good idea to

crystalize the catalytical domain of caspase-4 for both states of inactivate and active. The long-term goal of this research is to elucidate the detailed molecular basis of intracellular LPS recognition by caspase-4. These results will significantly advance our understanding of the molecular mechanism of innate immunity towards microbial infection and provide a structural framework for conceptual new approaches to block undesired innate immune response. Furthermore, Chemotherapeutic and immunotherapeutic agonists can be developed by mediating the caspase-4 GSDMD related pryoptosis in the absence of LPS.

Determine the structural basis of caspase-4 attaching to LPS through CARD domain

Expression, purification and characterization of human caspase4

The WT caspase4 autocatalyzes the cleavage at Asp289 of itself and induces autoactivation, killing the E. coli strains. We have made a SUMO-fused full-length human caspase4 (C258A) with a thrombin site downstream of CARD domain. This mutated protein expressed in E. coli lacks autocatalysis. The fusion protein was firstly purified using Ni-NTA agarose, followed by the cleavage of the SUMO tag. The caspase-4 protein was then purified by gel filtration chromatography. The resultant protein was further cleaved through the thrombin site to generate CARD which was purified through gel filtration chromatography. It is clearly observed that caspase-4 with the SUMO fusion can be highly expressed in E. coli as a monomer which was not affected by LPS present in E. coli.

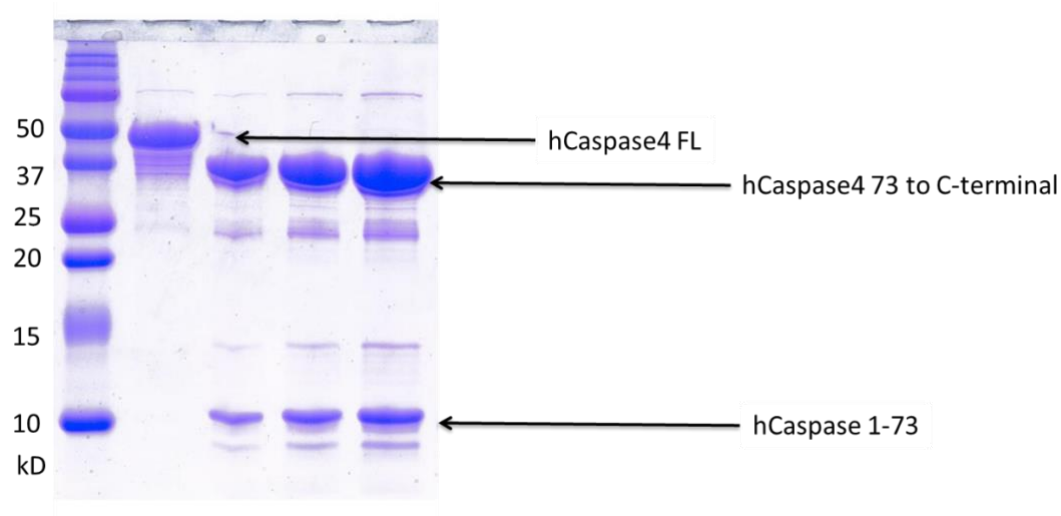


Figure 56. The purification of full-length casp4 and CARD domain

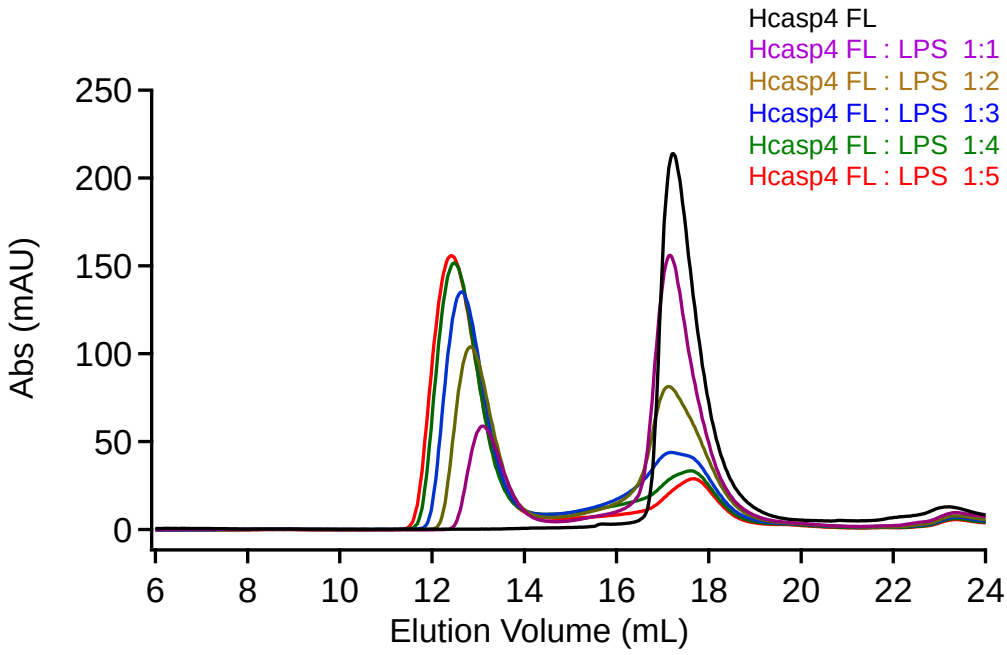


Figure 57. Size Exclusion Chromatography study of caspase4-LPS complex.

The binding profile of caspase-4 to LPS

The tag-cleaved caspase-4 or its CARD domain binds with LPS to form larger complexes. We have mixed the purified caspase-4 with LPS in different molar ratios, incubated them for 1 hour and purified by superose 6 size exclusion chromatography. In the experiment, constant LPS concentration was used. When increasing concentration of apo caspase-4 was added, the complex peak was shifted right, indicating the complex became smaller in size (Figure 57). This trend was also observed in the complex of caspase-4 CARD binding with LPS or LipidA. When excess caspase-4 was added, the

peak of the complex in SEC became sharper, suggesting the complex was more homogeneous.

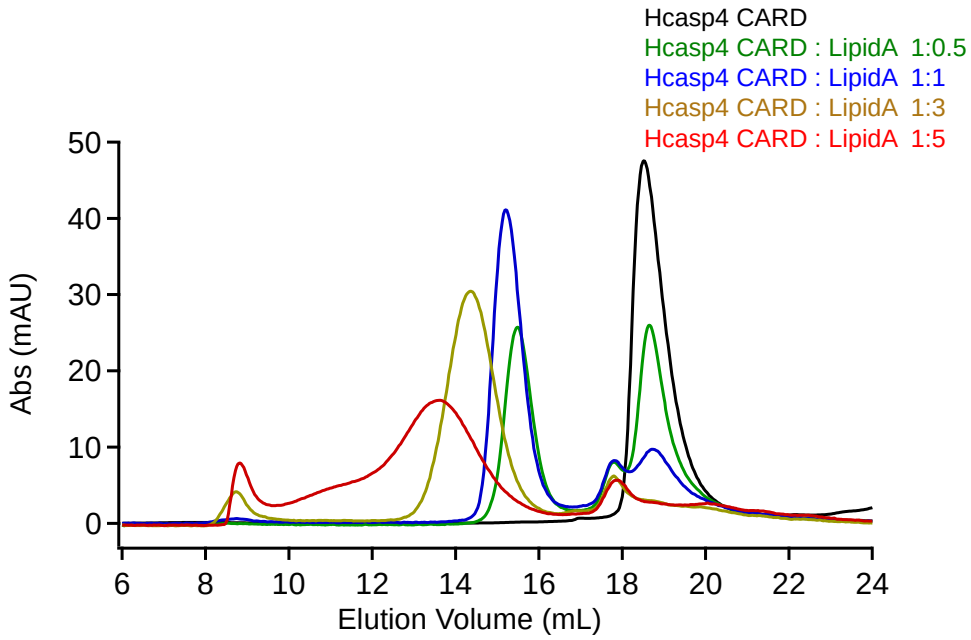


Figure 58. SEC study of the caspase4-LipidA complex.

Cryo-Electron microscopy of CARD/LPS complex

To elucidate the structural basis of caspase-4 oligomer induced by LPS, we have purified SUMO fused full-length caspase-4 with a thrombin site inserted downstream of CARD. The full-length protein was cleaved with SUMO and thrombin proteases followed by gel filtration chromatography to generate CARD domain. The purified CARD was incubated with LPS at a 1:3 molar ratio. In Cryo-EM single particle reconstruction, the

CARD/LPS complex presents a football-like structure, with a bunch of spikes inserted. Due to the heterogeneity of this complex, the structure determination is stuck at a low resolution.

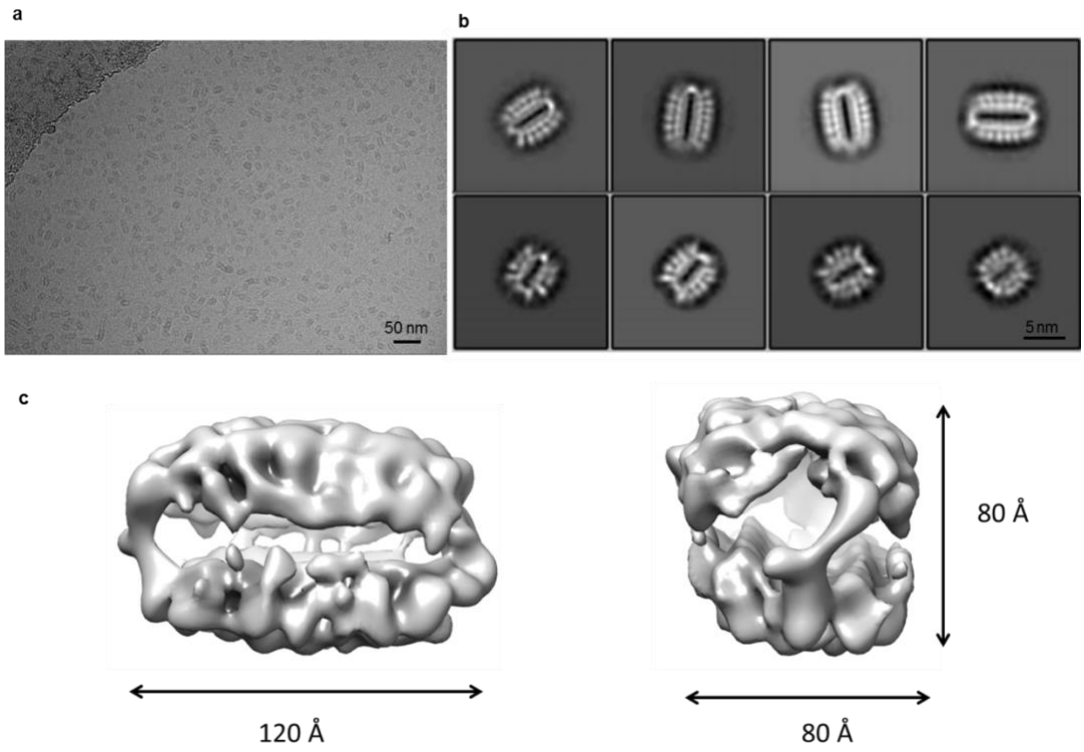


Figure 59. Cryo-EM study of the caspase4-LPS complex.

To examine how heterogeneous the specimen is, the caspase-4/LPS complex was analyzed by analytical ultracentrifugation. The AUC experiment demonstrated that the peak of the complex CARD/LPS was broad and expanding in x-axis at a 1:8 molar ratio.

By increasing CARD protein concentration, the peak shifted to the left and became sharper and more symmetric in AUC, indicating the complex was smaller and more homogenous.

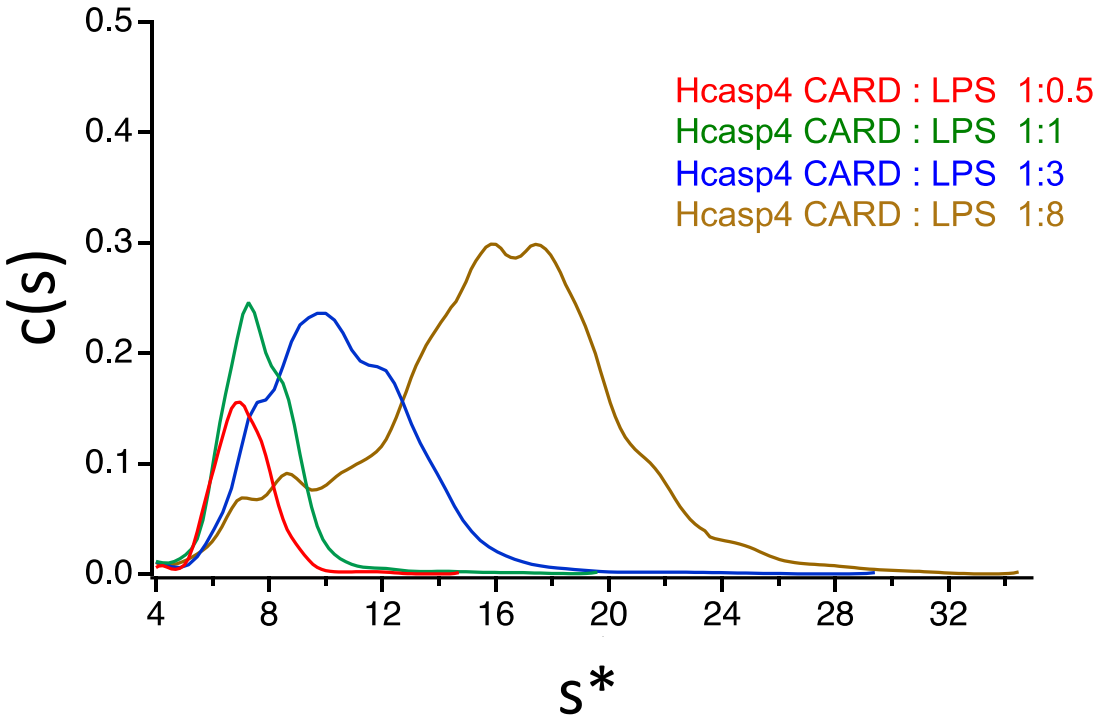


Figure 60. AUC study of the caspase4-LPS complex

Since the excess caspase-4 protein added to the complex continuously reduces the complex size, the addition of excess LPS to the complex has also been investigated. Compared to the continuously sharper and constantly more right shifted peak due to addition of CARD protein, the peak was not significantly shifted to the left or became broader with LPS added to the formed complex. It indicates that the caspase-4 does not

dissociate from the formed complex to bind with other LPS molecular.

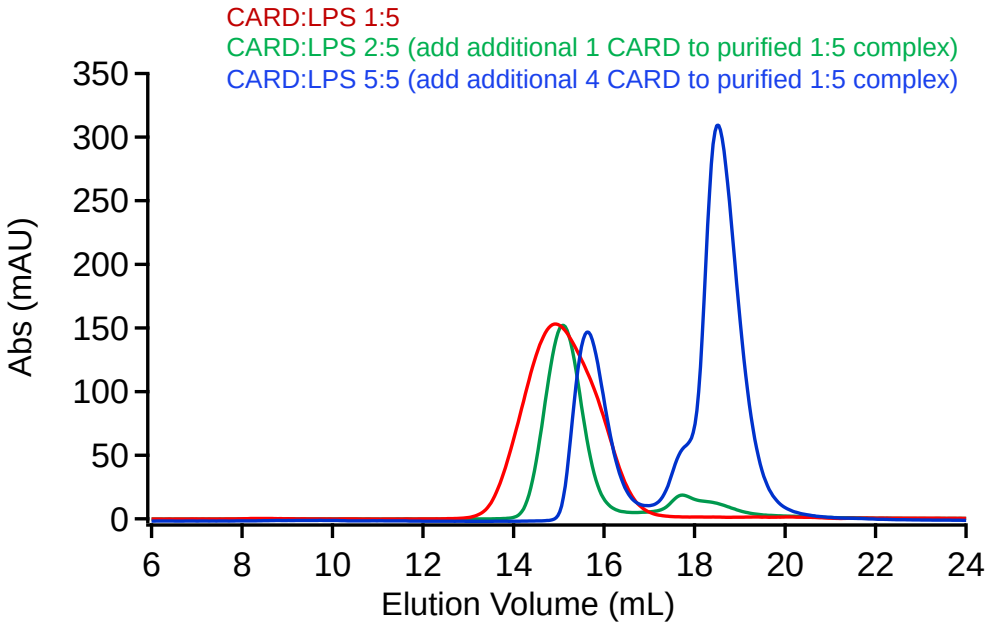


Figure 61. The addition of excess caspase4 reduces the complex size.

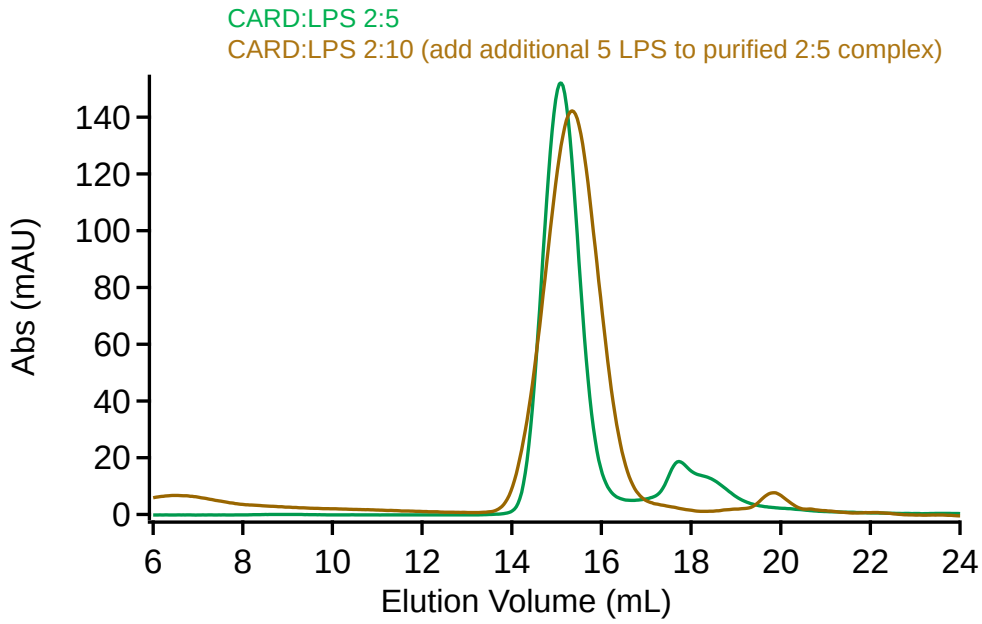


Figure 62. The addition of excess LPS does not affect the complex size.

Homogenization of the caspase-4/LPS complex

Obtaining a stably homogenous oligomer of caspase-4/LPS complex is critical for the single particle studies in Cryo-EM. Even though much effort has been contributed to make the caspase-4/LPS complex homogenous, the specimen cannot be resolved to high resolution yet. When excessive CARD protein was mixed with LPS at a total molar ratio of 1:1, generating the sharpest peak of the mixture in SEC , we collected the complex peak and applied it to the Liquid chromatography coupled Small-angle X-ray scattering(SEC-SAXS), which showed that the radius of gyration reflected a sloping trend within the SEC

peak, indicating the SEC peak contained multiple species. To make the complex homogeneous, we substituted LPS with kdo2-LipidA, which has a more defined structure. The overnight incubation of caspase-4 with LipidA at 37°C also produced more homogeneous complex compared to ice incubation, possibly due to the high viscosity of lipid at low temperatures. The caspase-4/LipidA complex, or the CARD/LipidA complex, with excess proteins in the mixture to minimize the component heterogeneity, will be used in cryoEM.

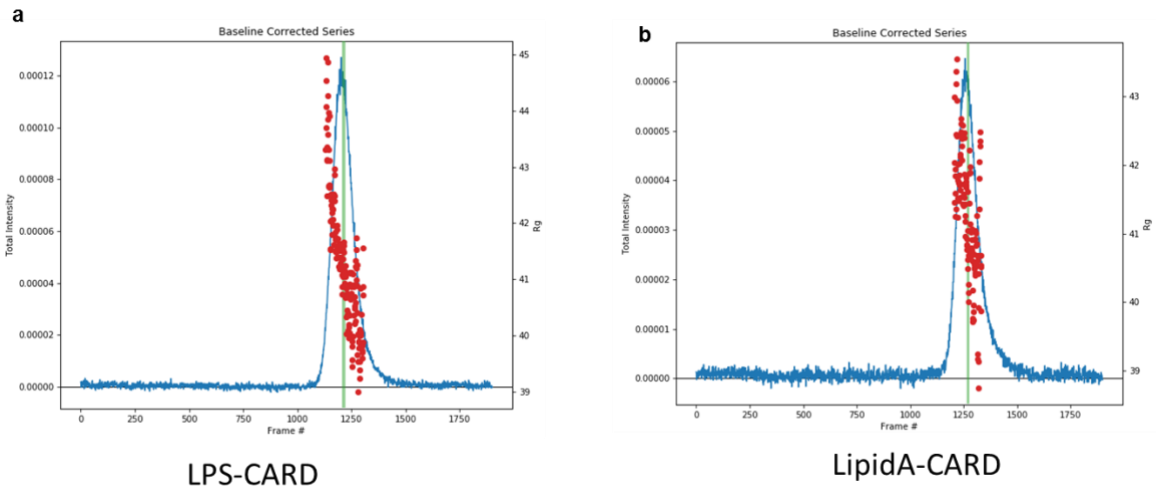


Figure 63. SAXS study of the caspase4-LPS complex.

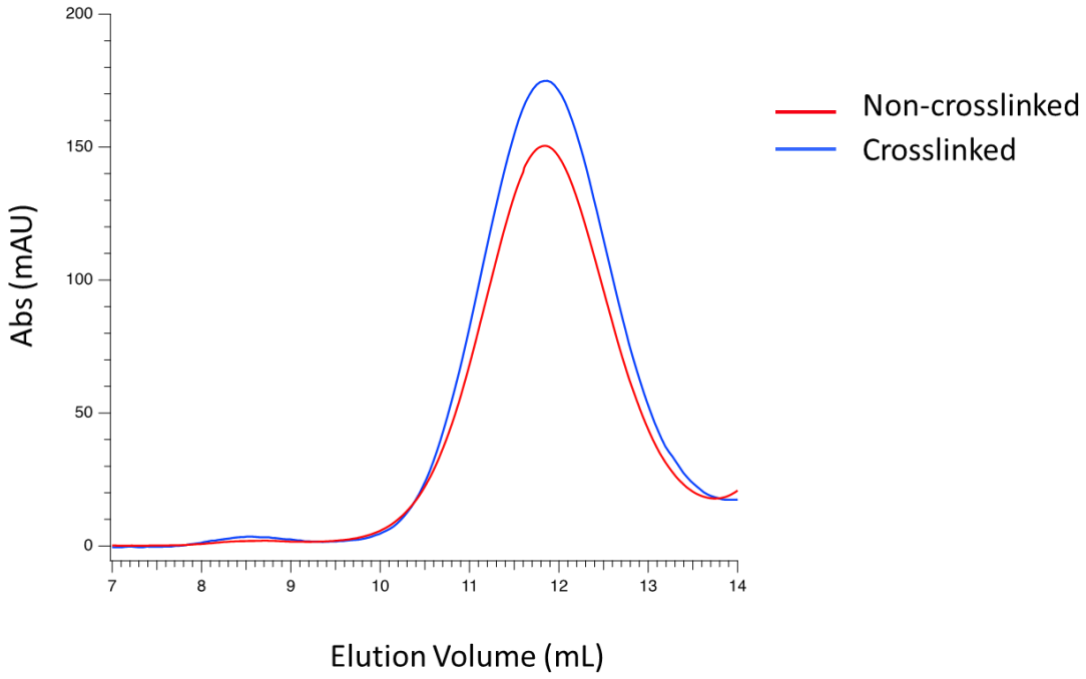


Figure 64. The on-column crosslinking of casp4-LPS complex.

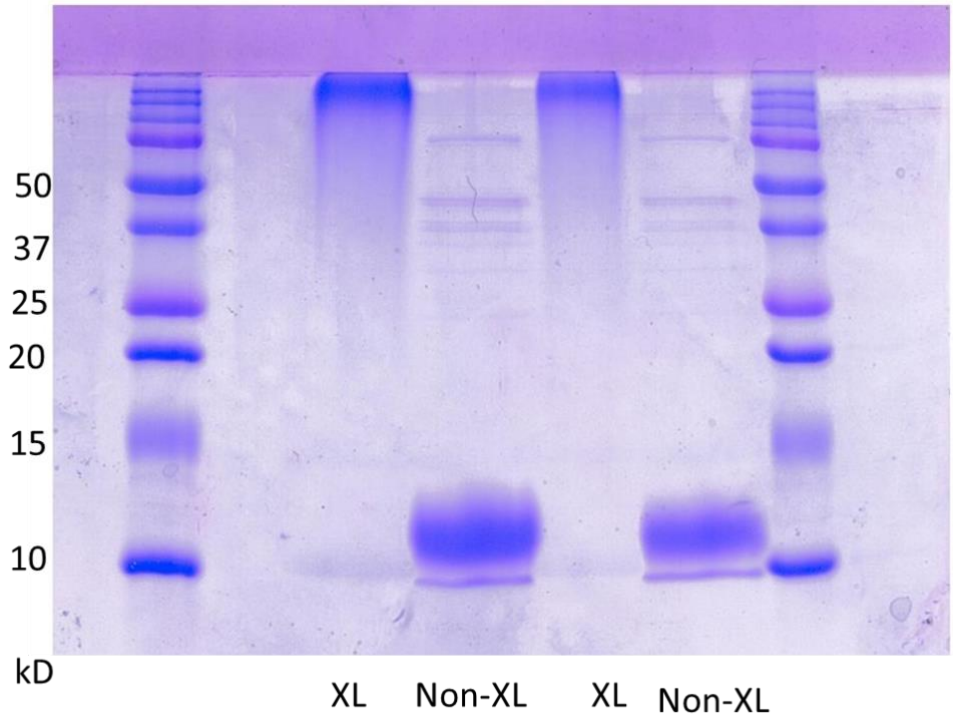


Figure 65. SDS-PAGE of the crosslinked and non-crosslinked specimen.

Crosslinking study of caspase4-LPS complex

To understand and evaluate heterogeneity of the complex, on-column crosslinking method is used. The glutaraldehyde crosslinking does not affect the complex size in SEC. The crosslinked specimen in the SDS-PAGE analyses showed a single band higher than 250kD, indicating a large complex is formed.

The homogenized and crosslinked caspase-4/LPS complex is used to collect data in cryo-EM. 1,676 movies with 50 frames each are collected at magnification of 130,000. Due to the smaller particle size, the physical apix size is chosen to be 0.64 Å. The homogenized complex is mono dispersed in the vitrified ice (Figure 66).

However, the homogenized and crosslinked specimen in cryo-EM does not improve the 3D reconstruction (Figure 67).

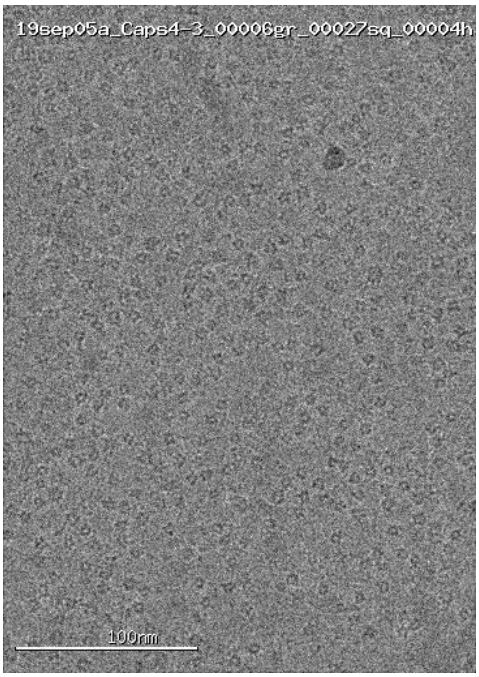


Figure 66. Raw image of homogenized caspase-4/LPS complex.

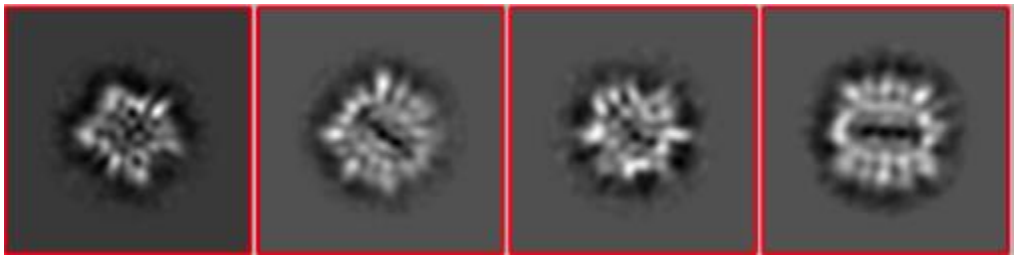


Figure 67. The 2D average of homogenized specimen.

Cryo-EM study of caspase4-Lipid A complex

To investigate how the caspase4-LipidA complex is formed, cryo-EM micrographs for the caspase-4 CARD domain in complex with LipidA are also collected and evaluated. In order to homogenize the specimen, the molar ratio of CARD protein to

LipidA was chosen to be 2:1. The LipidA/CARD complex shows similar particle distribution to LPS/CARD complex. Even though the LipidA/CARD particles show the similar shape as the LPS complex specimen, the 2D averages of the LipidA specimen still do not present a clear feature (Figure 68). Neither the caspase-4/LPS nor caspase-4/LipidA complex provides promising high-resolution 3D reconstruction no matter how we optimize the specimen preparation, indicating the complex is endogenously heterogeneous.

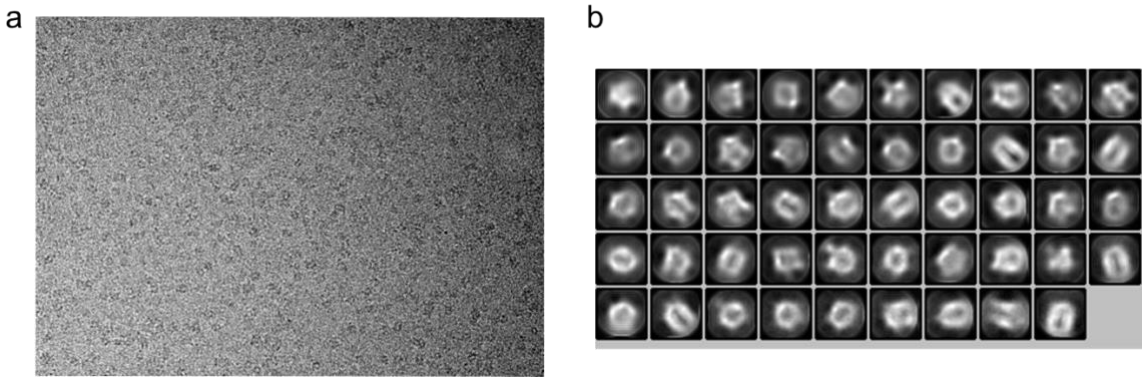


Figure 68. Raw image and 2D averages of caspase-4/LipidA complex.

Determine the structural basis of caspase-4 autocatalysis

Expression, purification and characterization of human/mouse caspase4 ΔCARD

When caspase-4 oligomerizes with LPS, it autocatalyzes the cleavage at Asp289 and induces autoactivation. The transformation of WT human caspase-4 into E. coli BL21 strains is lethal to cells, possibly due to the activity of caspase-4 inflammasome induced by cellular LPS. To harvest the active caspase-4 for functional study, CARD is truncated

off and a thrombin site is used to replace the cleavage loop at Asp289. The caspase-4 containing amino acids 100-377 was purified similar to full-length caspase-4(C258A) with SUMO fusion system. Before use, thrombin protease is added to caspase-4 100-377 to mimic the cleavage of the Asp289-containing loop and activate the caspase-4.Cryo-Electron Microscopy reconstruction of full-length caspase-4.

Cryo-Electron Microscopy reconstruction of full-length caspase-4

Even though the catalytic domain structure of caspase-4 has been determined by us with X-Ray crystallography, the full-length caspase-4 protein has not provided a promising crystal for now. To determine the structure of full-length caspase-4, Cryo EM single particle reconstruction is worth a try. The full-length human caspase-4 contains 377 amino acids and is calculated to be 43kD in size. This particular small size goes beyond the size limit of conventional cryo-EM single particle reconstruction, given that only a few 50kD proteins have been determined up to now, including streptavidin and hemoglobin dimer. However, this challenge can be overcome by those commercially available monoclonal antibodies, such as OTI4A2, 17D9 or 4B9. The antibodies are digested into Fab fragments using an IgG1 Fab preparation kit and then mixed with full-length caspase-4. The Fab fragment is believed to make the target size significantly bigger, up to 100kD. By applying the phase plate in cryo-EM, the full-length caspase-4/Fab complex will be much less challenging for single particle study. The Fab fragment might also fix the caspase-4 protein to one conformation and therefore improve the structure resolution.

Discussion and future studies

After my exhaustive optimization on complex formation of caspase-4/LPS, I still cannot determine a good sample for cryo-EM single particle reconstruction. My optimal condition right now is to use LipidA to avoid the heterogeneity of O-antigen of LPS from different sources, to use caspase-4 CARD domain instead of FL caspase-4 to avoid the heterogeneity induced by the FL caspase-4 linker, to use extremely high ratio of protein to LipidA in the mixture for homogeneity which is indicated by SEC and AUC, to use overnight incubation at 37C and to use crosslinking reagent to fix the complex.

Due to the componential or structural heterogeneity, Cryo electron tomography is an alternative for relatively low-resolution but reasonable 3D structure. In sub-tomo average, multiple frames will be recorded for the caspase-4/LipidA complex particles that are vitrified in ice. Because the frames are collected with different tilting angles, structural information from different views will be recorded and processed to a structural model. The cryo-ET structure of caspase-4 LipidA structure will give a conceptual framework of how caspase-4 forms a complex with LPS to activate itself.

Another future study will focus on how the complex is formed. One method is to use the density marker in the EM analyses to identify the arrangement of caspase4/LPS complex. Mtn tag can be added to the caspase-4 protein and we can follow a newly developed method to generate gold particles on protein [159]. The other method is to generate antibody Fab of caspase-4. The Fab bound to caspase-4 in the complex will indicate the location of caspase-4 CARD domain in the EM analyses and provide information on the complex arrangement.

Materials and Methods in detail

Purification of Caspase-4 FL C258A

pET-28(a) Avi-6xHis-SUMO Caspase-4 FL was transformed into Escherichia coli strain Bl21(DE3) on 50 ug/mL Kanamycin LB agar plates overnight. Several colonies from transformation plates were inoculated into 50 mL LB medium with 50 ug/mL Kanamycin, and shaken at 37 °C for 4 hours. Subsequently, all of the 50 mL LB was distributed evenly into 6-liter LB medium with 50 ug/mL kanamycin, and shaken at 37 °C until the OD_{600} reached about 1.0. LB medium was cooled down by ice to 15 °C. 5 mM Biotin and 0.4 mM isopropyl β-D-thiogalactoside (IPTG) was added to the medium. Cells were harvested at 16 hours post-induction and stored at -80°C refrigerator.

SUMO Caspase-4 FL was purified through Ni NTA and Superdex 200 gel filtration columns. Briefly, cell pellets were resuspended by 200 mL lysis buffer (50 mM Tris, 300mM NaCl, 1mM PMSF, pH 7.8) and applied to sonication until the lysate become non-viscous. The non-viscous lysate was centrifuged at 16000g, 4 °C for 30 min to remove the cell debris from the supernatant. The supernatant was loaded to 20 mL packed Ni NTA resin and washed with 200 mL washing buffer (50 mM Tris, 500mM NaCl, 20 mM imidazole, pH 8.0). The target protein was eluted with elution buffer (50 mM Tris, 300mM NaCl, 250 mM imidazole, pH 7.4) and concentrated to 2 mL to Superdex 200 gel filtration column. The peak fractionation was collected and analyzed by SDS-PAGE. Pure fractionation was concentrated, freshly frozen by liquid nitrogen, and stored in a -80°C refrigerator.

Purification of Caspase-4 FL CARD

To generate the CARD domain of caspase-4, a thrombin site was inserted after CARD domain in the caspase-4 FL plasmid. After this thrombin site inserted FL caspase-4 is purified with the same method as the original one. Thrombin enzyme was added at a ratio of 1:100 with 2mM CaCl2. After overnight cleavage, the caspase-4 CARD domain is purified by SEC Superdex 200 gel filtration column.

REFERENCES

1. Medzhitov, R., *Toll-like receptors and innate immunity.* Nat Rev Immunol, 2001. **1**(2): p. 135-45.
2. Akira, S., *Toll-like receptors and innate immunity.* Adv Immunol, 2001. **78**: p. 1-56.
3. Anderson, R.P., *Innate and adaptive immunity in celiac disease.* Curr Opin Gastroenterol, 2020. **36**(6): p. 470-478.
4. Hensen, L., K. Kedzierska, and M. Koutsakos, *Innate and adaptive immunity toward influenza B viruses.* Future Microbiol, 2020. **15**: p. 1045-1058.
5. Turvey, S.E. and D.H. Broide, *Innate immunity.* J Allergy Clin Immunol, 2010. **125**(2 Suppl 2): p. S24-32.
6. Medzhitov, R. and C.A. Janeway, Jr., *Innate immunity: the virtues of a nonclonal system of recognition.* Cell, 1997. **91**(3): p. 295-8.
7. Beutler, B., *Innate immunity: an overview.* Mol Immunol, 2004. **40**(12): p. 845-59.
8. Bianchi, M.E., *DAMPs, PAMPs and alarmins: all we need to know about danger.* J Leukoc Biol, 2007. **81**(1): p. 1-5.
9. Joncker, N.T. and D.H. Raulet, *Regulation of NK cell responsiveness to achieve self-tolerance and maximal responses to diseased target cells.* Immunol Rev, 2008. **224**: p. 85-97.
10. Takeda, K., T. Kaisho, and S. Akira, *Toll-like receptors.* Annu Rev Immunol, 2003. **21**: p. 335-76.
11. Takeuchi, O. and S. Akira, *Pattern recognition receptors and inflammation.* Cell, 2010. **140**(6): p. 805-20.
12. Nie, L., et al., *Toll-Like Receptors, Associated Biological Roles, and Signaling Networks in Non-Mammals.* Front Immunol, 2018. **9**: p. 1523.
13. Das, S., et al., *Unmethylated CpG motifs in the L. donovani DNA regulate TLR9-dependent delay of programmed cell death in macrophages.* J Leukoc Biol, 2015. **97**(2): p. 363-78.
14. Luo, Z., et al., *Plasmid DNA containing multiple CpG motifs triggers a strong immune response to hepatitis B surface antigen when combined with incomplete Freund's adjuvant but not aluminum hydroxide.* Mol Med Rep, 2012. **6**(6): p. 1309-14.
15. Ramirez-Ortiz, Z.G., et al., *Toll-like receptor 9-dependent immune activation by unmethylated CpG motifs in Aspergillus fumigatus DNA.* Infect Immun, 2008. **76**(5): p. 2123-9.
16. Krieg, A.M., *CpG motifs: the active ingredient in bacterial extracts?* Nat Med, 2003. **9**(7): p. 831-5.
17. Krieg, A.M., *The role of CpG motifs in innate immunity.* Curr Opin Immunol, 2000. **12**(1): p. 35-43.
18. Lu, A., et al., *Plasticity in PYD assembly revealed by cryo-EM structure of the PYD filament of AIM2.* Cell Discov, 2015. **1**.

19. Franchi, L., et al., *The inflammasome: a caspase-1-activation platform that regulates immune responses and disease pathogenesis.* Nature Immunology, 2009. **10**(3): p. 241-247.
20. Girardin, S.E., et al., *Nod1 detects a unique muropeptide from Gram-negative bacterial peptidoglycan.* Science, 2003. **300**(5625): p. 1584-1587.
21. Keestra-Gounder, A.M. and R.M. Tsolis, *NOD1 and NOD2: Beyond Peptidoglycan Sensing.* Trends in Immunology, 2017. **38**(10): p. 758-767.
22. Viala, J., et al., *Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island.* Nature Immunology, 2004. **5**(11): p. 1166-1174.
23. Morre, S.A., et al., *The true ligand of the NOD2 receptor is peptidoglycan instead of lipopolysaccharide: A schematic representation of ligand-receptor interactions and NF-kappa B activation.* Gastroenterology, 2004. **126**(1): p. 371-372.
24. Girardin, S.E., et al., *Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection.* Journal of Biological Chemistry, 2003. **278**(11): p. 8869-8872.
25. Lamkanfi, M. and V.M. Dixit, *Mechanisms and functions of inflammasomes.* Cell, 2014. **157**(5): p. 1013-22.
26. Broz, P. and V.M. Dixit, *Inflammasomes: mechanism of assembly, regulation and signalling.* Nat Rev Immunol, 2016. **16**(7): p. 407-20.
27. Guo, H., J.B. Callaway, and J.P. Ting, *Inflammasomes: mechanism of action, role in disease, and therapeutics.* Nat Med, 2015. **21**(7): p. 677-87.
28. Rathinam, V.A., S.K. Vanaja, and K.A. Fitzgerald, *Regulation of inflammasome signaling.* Nat Immunol, 2012. **13**(4): p. 333-42.
29. Munoz-Planillo, R., et al., *K(+) efflux is the common trigger of NLRP3 inflammasome activation by bacterial toxins and particulate matter.* Immunity, 2013. **38**(6): p. 1142-53.
30. Chen, J. and Z.J. Chen, *PtdIns4P on dispersed trans-Golgi network mediates NLRP3 inflammasome activation.* Nature, 2018. **564**(7734): p. 71-76.
31. Sharif, H., et al., *Structural mechanism for NEK7-licensed activation of NLRP3 inflammasome.* Nature, 2019. **570**(7761): p. 338-343.
32. Burckstummer, T., et al., *An orthogonal proteomic-genomic screen identifies AIM2 as a cytoplasmic DNA sensor for the inflammasome.* Nature Immunology, 2009. **10**(3): p. 266-272.
33. Hornung, V., et al., *AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC.* Nature, 2009. **458**(7237): p. 514-8.
34. Fernandes-Alnemri, T., et al., *AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA.* Nature, 2009. **458**(7237): p. 509-U5.
35. Panchanathan, R., et al., *Aim2 Deficiency Stimulates the Expression of IFN-Inducible Ifi202, a Lupus Susceptibility Murine Gene within the Nba2 Autoimmune Susceptibility Locus.* Journal of Immunology, 2010. **185**(12): p. 7385-7393.
36. Vanaja, S.K., et al., *AIM2, NLRP3, AND NLRP12 FUNCTION IN A NON-REDUNDANT MANNER TO DRIVE IL-1 beta IN RESPONSE TO ENTEROHEMORRHAGIC ESCHERICHIA COLI.* Inflammation Research, 2010. **59**: p. S302-S302.

37. Rathinam, V.A.K., et al., *The AIM2 inflammasome is essential for host defense against cytosolic bacteria and DNA viruses.* Nature Immunology, 2010. **11**(5): p. 395-403.
38. Zhang, W.J., et al., *AIM2 Facilitates the Apoptotic DNA-induced Systemic Lupus Erythematosus via Arbitrating Macrophage Functional Maturation.* Journal of Clinical Immunology, 2013. **33(**5): p. 925-937.
39. Ablasser, A., et al., *RIG-I-dependent sensing of poly(dA:dT) through the induction of an RNA polymerase III-transcribed RNA intermediate.* Nat Immunol, 2009. **10**(10): p. 1065-72.
40. Kato, H., et al., *Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5.* J Exp Med, 2008. **205**(7): p. 1601-10.
41. Gebhardt, A., B.T. Laudenbach, and A. Pichlmair, *Discrimination of Self and Non-Self Ribonucleic Acids.* J Interferon Cytokine Res, 2017. **37**(5): p. 184-197.
42. Yoneyama, M., et al., *Viral RNA detection by RIG-I-like receptors.* Curr Opin Immunol, 2015. **32**: p. 48-53.
43. Hornung, V., et al., *OAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids.* Nat Rev Immunol, 2014. **14**(8): p. 521-8.
44. Platanias, L.C., *Mechanisms of type-I- and type-II-interferon-mediated signalling.* Nat Rev Immunol, 2005. **5**(5): p. 375-86.
45. Lazear, H.M., J.W. Schoggins, and M.S. Diamond, *Shared and Distinct Functions of Type I and Type III Interferons.* Immunity, 2019. **50**(4): p. 907-923.
46. Morita, K., et al., *Expression of interferon receptor genes (IFNAR1 and IFNAR2 mRNA) in the liver may predict outcome after interferon therapy in patients with chronic genotype 2a or 2b hepatitis C virus infection.* J Clin Gastroenterol, 1998. **26**(2): p. 135-40.
47. Pay, S., et al., *IFNAR1 and IFNAR2 polymorphisms in patients with Behcet's disease.* Clin Exp Rheumatol, 2009. **27**(2 Suppl 53): p. S110-1.
48. Kaur, S. and L.C. Platanias, *IFN-beta-specific signaling via a unique IFNAR1 interaction.* Nat Immunol, 2013. **14**(9): p. 884-5.
49. Ball, E.A., et al., *IFNAR1 controls progression to cerebral malaria in children and CD8+ T cell brain pathology in Plasmodium berghei-infected mice.* J Immunol, 2013. **190**(10): p. 5118-27.
50. Marijanovic, Z., et al., *TYK2 activity promotes ligand-induced IFNAR1 proteolysis.* Biochem J, 2006. **397**(1): p. 31-8.
51. Pfeffer, L.M., et al., *STAT3 as an adapter to couple phosphatidylinositol 3-kinase to the IFNAR1 chain of the type I interferon receptor.* Science, 1997. **276**(5317): p. 1418-20.
52. Reis, L.F., T. Ho Lee, and J. Vilcek, *Tumor necrosis factor acts synergistically with autocrine interferon-beta and increases interferon-beta mRNA levels in human fibroblasts.* J Biol Chem, 1989. **264**(28): p. 16351-4.
53. Li, S.F., et al., *Type I Interferons: Distinct Biological Activities and Current Applications for Viral Infection.* Cell Physiol Biochem, 2018. **51**(5): p. 2377-2396.

54. Barber, G.N., *Innate immune DNA sensing pathways: STING, AIMII and the regulation of interferon production and inflammatory responses.* Curr Opin Immunol, 2011. **23**(1): p. 10-20.
55. Kato, H., K. Takahasi, and T. Fujita, *RIG-I-like receptors: cytoplasmic sensors for non-self RNA.* Immunol Rev, 2011. **243**(1): p. 91-8.
56. Keating, S.E., M. Baran, and A.G. Bowie, *Cytosolic DNA sensors regulating type I interferon induction.* Trends Immunol, 2011. **32**(12): p. 574-81.
57. Schattgen, S.A. and K.A. Fitzgerald, *The PYHIN protein family as mediators of host defenses.* Immunol Rev, 2011. **243**(1): p. 109-18.
58. Stetson, D.B. and R. Medzhitov, *Recognition of cytosolic DNA activates an IRF3-dependent innate immune response.* Immunity, 2006. **24**(1): p. 93-103.
59. Ishii, K.J., et al., *A Toll-like receptor-independent antiviral response induced by double-stranded B-form DNA.* Nat Immunol, 2006. **7**(1): p. 40-8.
60. Tao, J., X. Zhou, and Z. Jiang, *cGAS-cGAMP-STING: The three musketeers of cytosolic DNA sensing and signaling.* IUBMB Life, 2016. **68**(11): p. 858-870.
61. Chen, Q., L. Sun, and Z.J. Chen, *Regulation and function of the cGAS-STING pathway of cytosolic DNA sensing.* Nat Immunol, 2016. **17**(10): p. 1142-9.
62. Roers, A., B. Hiller, and V. Hornung, *Recognition of Endogenous Nucleic Acids by the Innate Immune System.* Immunity, 2016. **44**(4): p. 739-54.
63. Paludan, S.R. and A.G. Bowie, *Immune sensing of DNA.* Immunity, 2013. **38**(5): p. 870-80.
64. Bhat, N. and K.A. Fitzgerald, *Recognition of cytosolic DNA by cGAS and other STING-dependent sensors.* Eur J Immunol, 2014. **44**(3): p. 634-40.
65. Burdette, D.L. and R.E. Vance, *STING and the innate immune response to nucleic acids in the cytosol.* Nat Immunol, 2013. **14**(1): p. 19-26.
66. Barbalat, R., et al., *Nucleic acid recognition by the innate immune system.* Annu Rev Immunol, 2011. **29**: p. 185-214.
67. Sun, L., et al., *Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway.* Science, 2013. **339**(6121): p. 786-91.
68. Xiao, T.S. and K.A. Fitzgerald, *The cGAS-STING pathway for DNA sensing.* Mol Cell, 2013. **51**(2): p. 135-9.
69. Wu, J. and Z.J. Chen, *Innate immune sensing and signaling of cytosolic nucleic acids.* Annu Rev Immunol, 2014. **32**: p. 461-88.
70. Gao, D., et al., *Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses.* Science, 2013. **341**(6148): p. 903-6.
71. Wu, J., et al., *Cyclic GMP-AMP is an endogenous second messenger in innate immune signaling by cytosolic DNA.* Science, 2013. **339**(6121): p. 826-30.
72. Zhang, X., et al., *Cyclic GMP-AMP containing mixed phosphodiester linkages is an endogenous high-affinity ligand for STING.* Mol Cell, 2013. **51**(2): p. 226-35.
73. Ablasser, A., et al., *cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING.* Nature, 2013. **498**(7454): p. 380-4.
74. Gao, P., et al., *Cyclic [G(2',5')pA(3',5')p] is the metazoan second messenger produced by DNA-activated cyclic GMP-AMP synthase.* Cell, 2013. **153**(5): p. 1094-107.

75. Diner, E.J., et al., *The innate immune DNA sensor cGAS produces a noncanonical cyclic dinucleotide that activates human STING.* Cell Rep, 2013. **3**(5): p. 1355-61.
76. Ishikawa, H. and G.N. Barber, *STING is an endoplasmic reticulum adaptor that facilitates innate immune signalling.* Nature, 2008. **455**(7213): p. 674-8.
77. Ishikawa, H., Z. Ma, and G.N. Barber, *STING regulates intracellular DNA-mediated, type I interferon-dependent innate immunity.* Nature, 2009. **461**(7265): p. 788-92.
78. Barber, G.N., *STING: infection, inflammation and cancer.* Nat Rev Immunol, 2015. **15**(12): p. 760-70.
79. Tanaka, Y. and Z.J. Chen, *STING specifies IRF3 phosphorylation by TBK1 in the cytosolic DNA signaling pathway.* Sci Signal, 2012. **5**(214): p. ra20.
80. Sharma, S., et al., *Triggering the interferon antiviral response through an IKK-related pathway.* Science, 2003. **300**(5622): p. 1148-51.
81. Hemmi, H., et al., *The roles of two IkappaB kinase-related kinases in lipopolysaccharide and double stranded RNA signaling and viral infection.* J Exp Med, 2004. **199**(12): p. 1641-50.
82. Fitzgerald, K.A., et al., *IKKepsilon and TBK1 are essential components of the IRF3 signaling pathway.* Nat Immunol, 2003. **4**(5): p. 491-6.
83. Liu, S., et al., *Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induces IRF3 activation.* Science, 2015. **347**(6227): p. aaa2630.
84. Yoneyama, M., et al., *Direct triggering of the type I interferon system by virus infection: activation of a transcription factor complex containing IRF-3 and CBP/p300.* EMBO J, 1998. **17**(4): p. 1087-95.
85. Lin, R., et al., *Virus-dependent phosphorylation of the IRF-3 transcription factor regulates nuclear translocation, transactivation potential, and proteasome-mediated degradation.* Mol Cell Biol, 1998. **18**(5): p. 2986-96.
86. Gonzalez-Navajas, J.M., et al., *Immunomodulatory functions of type I interferons.* Nat Rev Immunol, 2012. **12**(2): p. 125-35.
87. Stetson, D.B. and R. Medzhitov, *Type I interferons in host defense.* Immunity, 2006. **25**(3): p. 373-81.
88. Trinchieri, G., *Type I interferon: friend or foe?* J Exp Med, 2010. **207**(10): p. 2053-63.
89. Schneider, W.M., M.D. Chevillotte, and C.M. Rice, *Interferon-stimulated genes: a complex web of host defenses.* Annu Rev Immunol, 2014. **32**: p. 513-45.
90. Barber, G.N., *STING-dependent cytosolic DNA sensing pathways.* Trends Immunol, 2014. **35**(2): p. 88-93.
91. Barber, G.N., *STING-dependent signaling.* Nat Immunol, 2011. **12**(10): p. 929-30.
92. Stetson, D.B., *Endogenous retroelements and autoimmune disease.* Curr Opin Immunol, 2012. **24**(6): p. 692-7.
93. Stetson, D.B., *Connections between antiviral defense and autoimmunity.* Curr Opin Immunol, 2009. **21**(3): p. 244-50.
94. Gall, A., et al., *Autoimmunity initiates in nonhematopoietic cells and progresses via lymphocytes in an interferon-dependent autoimmune disease.* Immunity, 2012. **36**(1): p. 120-31.

95. Ahn, J., et al., *STING manifests self DNA-dependent inflammatory disease.* Proc Natl Acad Sci U S A, 2012. **109**(47): p. 19386-91.
96. Gray, E.E., et al., *Cutting Edge: cGAS Is Required for Lethal Autoimmune Disease in the Trex1-Deficient Mouse Model of Aicardi-Goutieres Syndrome.* J Immunol, 2015. **195**(5): p. 1939-43.
97. Gao, D., et al., *Activation of cyclic GMP-AMP synthase by self-DNA causes autoimmune diseases.* Proc Natl Acad Sci U S A, 2015. **112**(42): p. E5699-705.
98. Jeremiah, N., et al., *Inherited STING-activating mutation underlies a familial inflammatory syndrome with lupus-like manifestations.* J Clin Invest, 2014. **124**(12): p. 5516-20.
99. Warner, J.D., et al., *STING-associated vasculopathy develops independently of IRF3 in mice.* J Exp Med, 2017. **214**(11): p. 3279-3292.
100. Corrales, L., et al., *Direct Activation of STING in the Tumor Microenvironment Leads to Potent and Systemic Tumor Regression and Immunity.* Cell Rep, 2015. **11**(7): p. 1018-30.
101. Li, T., et al., *Antitumor Activity of cGAMP via Stimulation of cGAS-cGAMP-STING-IRF3 Mediated Innate Immune Response.* Sci Rep, 2016. **6**: p. 19049.
102. Fu, J., et al., *STING agonist formulated cancer vaccines can cure established tumors resistant to PD-1 blockade.* Sci Transl Med, 2015. **7**(283): p. 283ra52.
103. Civril, F., et al., *Structural mechanism of cytosolic DNA sensing by cGAS.* Nature, 2013. **498**(7454): p. 332-7.
104. Li, X.D., et al., *Pivotal roles of cGAS-cGAMP signaling in antiviral defense and immune adjuvant effects.* Science, 2013. **341**(6152): p. 1390-4.
105. Yamamoto, M., et al., *Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway.* Science, 2003. **301**(5633): p. 640-3.
106. Zhao, B., et al., *Structural basis for concerted recruitment and activation of IRF-3 by innate immune adaptor proteins.* Proc Natl Acad Sci U S A, 2016. **113**(24): p. E3403-12.
107. Volkman, H.E., et al., *Tight nuclear tethering of cGAS is essential for preventing autoreactivity.* Elife, 2019. **8**.
108. Dou, Z.X., et al., *Cytoplasmic chromatin triggers inflammation in senescence and cancer.* Nature, 2017. **550**(7676): p. 402-406.
109. Gluck, S., et al., *Innate immune sensing of cytosolic chromatin fragments through cGAS promotes senescence.* Nature Cell Biology, 2017. **19**(9): p. 1061-+.
110. Harding, S.M., et al., *Mitotic progression following DNA damage enables pattern recognition within micronuclei.* Nature, 2017. **548**(7668): p. 466-+.
111. Mackenzie, K.J., et al., *cGAS surveillance of micronuclei links genome instability to innate immunity.* Nature, 2017. **548**(7668): p. 461-+.
112. Yang, H., et al., *cGAS is essential for cellular senescence.* Proceedings of the National Academy of Sciences of the United States of America, 2017. **114**(23): p. E4612-E4620.
113. Lahaye, X., et al., *NONO Detects the Nuclear HIV Capsid to Promote cGAS-Mediated Innate Immune Activation.* Cell, 2018. **175**(2): p. 488-501 e22.

114. Zierhut, C., et al., *The Cytoplasmic DNA Sensor cGAS Promotes Mitotic Cell Death.* Cell, 2019. **178**(2): p. 302-+.
115. Zhao, B., et al., *The molecular basis of tight nuclear tethering and inactivation of cGAS.* Nature, 2020. **587**(7835): p. 673-677.
116. Michalski, S., et al., *Structural basis for sequestration and autoinhibition of cGAS by chromatin.* Nature, 2020. **587**(7835): p. 678-682.
117. Pathare, G.R., et al., *Structural mechanism of cGAS inhibition by the nucleosome.* Nature, 2020. **587**(7835): p. 668-672.
118. Boyer, J.A., et al., *Structural basis of nucleosome-dependent cGAS inhibition.* Science, 2020. **370**(6515): p. 450-454.
119. Kujirai, T., et al., *Structural basis for the inhibition of cGAS by nucleosomes.* Science, 2020. **370**(6515): p. 455-458.
120. Li, X., et al., *Cyclic GMP-AMP Synthase Is Activated by Double-Stranded DNA-Induced Oligomerization.* Immunity, 2013. **39**(6): p. 1019-31.
121. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 4): p. 486-501.
122. Liebschner, D., et al., *Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix.* Acta Crystallogr D Struct Biol, 2019. **75**(Pt 10): p. 861-877.
123. Drenth, J., *Principles of protein X-ray crystallography*. Springer advanced texts in chemistry. xiii, 305 pages.
124. Bai, X.C., G. McMullan, and S.H. Scheres, *How cryo-EM is revolutionizing structural biology.* Trends Biochem Sci, 2015. **40**(1): p. 49-57.
125. Kimple, M.E., A.L. Brill, and R.L. Pasker, *Overview of affinity tags for protein purification.* Curr Protoc Protein Sci, 2013. **73**: p. Unit 9 9.
126. Bornhorst, J.A. and J.J. Falke, *Purification of proteins using polyhistidine affinity tags.* Methods Enzymol, 2000. **326**: p. 245-54.
127. Einhauer, A. and A. Jungbauer, *Affinity of the monoclonal antibody M1 directed against the FLAG peptide.* J Chromatogr A, 2001. **921**(1): p. 25-30.
128. Einhauer, A. and A. Jungbauer, *The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins.* J Biochem Biophys Methods, 2001. **49**(1-3): p. 455-65.
129. Muyldermans, S., et al., *Camelid immunoglobulins and nanobody technology.* Vet Immunol Immunopathol, 2009. **128**(1-3): p. 178-83.
130. Frenzel, A., M. Hust, and T. Schirrmann, *Expression of recombinant antibodies.* Front Immunol, 2013. **4**: p. 217.
131. Chen, L.H., et al., *Expression, purification, and in vitro refolding of a humanized single-chain Fv antibody against human CTLA4 (CD152).* Protein Expr Purif, 2006. **46**(2): p. 495-502.
132. Rothbauer, U., et al., *A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins.* Mol Cell Proteomics, 2008. **7**(2): p. 282-9.
133. Kubala, M.H., et al., *Structural and thermodynamic analysis of the GFP:GFP-nanobody complex.* Protein Sci, 2010. **19**(12): p. 2389-401.

134. Beckett, D., E. Kovaleva, and P.J. Schatz, *A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation.* Protein Sci, 1999. **8**(4): p. 921-9.
135. Zhao, B., et al., *A conserved PLPLRT/SD motif of STING mediates the recruitment and activation of TBK1.* Nature, 2019. **569**(7758): p. 718-722.
136. Weber, P.C., et al., *Structural origins of high-affinity biotin binding to streptavidin.* Science, 1989. **243**(4887): p. 85-8.
137. Holmberg, A., et al., *The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures.* Electrophoresis, 2005. **26**(3): p. 501-10.
138. Mossessova, E. and C.D. Lima, *Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast.* Mol Cell, 2000. **5**(5): p. 865-76.
139. Lee, C.D., et al., *An improved SUMO fusion protein system for effective production of native proteins.* Protein Sci, 2008. **17**(7): p. 1241-8.
140. Shu, C., et al., *Structural Insights into the Functions of TBK1 in Innate Antimicrobial Immunity.* Structure, 2013. **21**(7): p. 1137-1148.
141. Li, X., et al., *Cyclic GMP-AMP Synthase Is Activated by Double-Stranded DNA-Induced Oligomerization.* Immunity, 2013. **39**(6): p. 1019-1031.
142. Raetz, C.R. and C. Whitfield, *Lipopolysaccharide endotoxins.* Annu Rev Biochem, 2002. **71**: p. 635-700.
143. Schumann, R.R., et al., *Structure and function of lipopolysaccharide binding protein.* Science, 1990. **249**(4975): p. 1429-31.
144. Pugin, J., et al., *Lipopolysaccharide activation of human endothelial and epithelial cells is mediated by lipopolysaccharide-binding protein and soluble CD14.* Proc Natl Acad Sci U S A, 1993. **90**(7): p. 2744-8.
145. Chow, J.C., et al., *Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction.* J Biol Chem, 1999. **274**(16): p. 10689-92.
146. Triantafilou, M. and K. Triantafilou, *Lipopolysaccharide recognition: CD14, TLRs and the LPS-activation cluster.* Trends Immunol, 2002. **23**(6): p. 301-4.
147. Dauphinee, S.M. and A. Karsan, *Lipopolysaccharide signaling in endothelial cells.* Lab Invest, 2006. **86**(1): p. 9-22.
148. Gurung, P., et al., *Toll or interleukin-1 receptor (TIR) domain-containing adaptor inducing interferon-beta (TRIF)-mediated caspase-11 protease production integrates Toll-like receptor 4 (TLR4) protein- and Nlrp3 inflammasome-mediated host defense against enteropathogens.* J Biol Chem, 2012. **287**(41): p. 34474-83.
149. Rathinam, V.A., et al., *TRIF licenses caspase-11-dependent NLRP3 inflammasome activation by gram-negative bacteria.* Cell, 2012. **150**(3): p. 606-19.
150. Man, S.M., et al., *IRGB10 Liberates Bacterial Ligands for Sensing by the AIM2 and Caspase-11-NLRP3 Inflammasomes.* Cell, 2016. **167**(2): p. 382-396 e17.
151. Yamamoto, M., et al., *A cluster of interferon-gamma-inducible p65 GTPases plays a critical role in host defense against Toxoplasma gondii.* Immunity, 2012. **37**(2): p. 302-13.
152. Shi, J., et al., *Inflammatory caspases are innate immune receptors for intracellular LPS.* Nature, 2014. **514**(7521): p. 187-92.

153. Shi, J., et al., *Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death.* Nature, 2015. **526**(7575): p. 660-5.
154. Ding, J., et al., *Pore-forming activity and structural autoinhibition of the gasdermin family.* Nature, 2016. **535**(7610): p. 111-6.
155. Liu, Z., et al., *Crystal Structures of the Full-Length Murine and Human Gasdermin D Reveal Mechanisms of Autoinhibition, Lipid Binding, and Oligomerization.* Immunity, 2019. **51**(1): p. 43-49 e4.
156. Ruan, J., et al., *Cryo-EM structure of the gasdermin A3 membrane pore.* Nature, 2018. **557**(7703): p. 62-67.
157. Liu, X., et al., *Inflammasome-activated gasdermin D causes pyroptosis by forming membrane pores.* Nature, 2016. **535**(7610): p. 153-+.
158. Aglietti, R.A., et al., *GsdmD p30 elicited by caspase-11 during pyroptosis forms pores in membranes.* Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(28): p. 7858-7863.
159. Jiang, Z., et al., *Genetically encoded tags for direct synthesis of EM-visible gold nanoparticles in cells.* Nat Methods, 2020. **17**(9): p. 937-946.