THE STRUCTURAL BASIS OF IRF-3 ACTIVATION UPON PHOSPHORYLATION

A Dissertation

by

TAO JING

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,	Margaret Glasner
Committee Members,	Pingwei Li
	Xiuren Zhang
	Robert O. Watson
Head of Department,	A.Joshua Wand

May 2021

Major Subject: Biochemistry

ABSTRACT

Humans and animals are constantly facing a variety of potential pathogens including bacteria and viruses daily. In order to survive and fight against their “enemies”, they have developed both innate immune system and adaptive immune system to eliminate infection. The latter one remembers the previous specific pathogens and gets rid of them when they invade again whereas the innate immune system serves as the first line to defend against a new pathogen in an efficient and timely manner.

The innate immune system is able to detect the pathogen-associated molecular patterns (PAMPs) such as nucleic acids and lipopolysaccharide (LPS) through pattern recognition receptors (PRRs), which leads to the induction of type-I interferons (IFN-I) by activating a transcription factor known as IRF-3. Although extensive studies on the functions of IRF-3 have been reported over years, the mechanism of IRF-3 activation still remains not fully understood.

IRF-3 belongs to the interferon regulatory factor (IRF) family and contains a DNA binding domain (DBD), IRF-3 association domain (IAD), and auto-inhibitory elements flanking on both sides of IAD. Within the C-terminal auto-inhibitory segment, there is a serine rich repeat containing seven serine/threonine residues that has been shown to be the potential sites of phosphorylation. In this study, we solved the crystal structures of both phosphorylated human and mouse IRF-3 bound to CBP (cAMP response element binding protein (CREB)-binding protein), which demonstrate that phosphorylated IRF-3 forms a dimer through phosphorylated Ser386 (pSer379 in mouse IRF-3) and a downstream pLxIS motif. Besides, size-exclusion chromatography and cell-based studies show that mutations of key residues

interacting with pSer386 severely impair IRF-3 activation and IFN-β induction. By contrast, phosphorylation of Ser396 within the pLxIS motif of human IRF-3 only plays a moderate role in IRF-3 activation.

The mouse IRF-3/CBP complex structure reveals that Arg373 and Arg205 of mIRF-3, corresponding to Arg211 and Arg380 of hIRF-3, are critical for the dimerization of mIRF-3. However, the replacement of Glu388 and Asn389 of human IRF-3 by mouse Lys381 makes this region restructured and distinct from human IRF-3. Biochemical studies also confirmed this discrepancy. Therefore, the mechanism of mouse IRF-3 activation is similar but distinct from human IRF-3.

Taken together, these structural and functional studies reveal the detailed mechanism of IRF-3 activation upon phosphorylation.

ACKNOWLEDGEMENTS

I would like to sincerely 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 my committee chair, Dr. Margaret Glasner for her valuable suggestions on my project and constructive comments on my dissertation. I also really appreciated my committee members, Dr. Robert O. Watson and Dr.Xiuren Zhang. I have collaborated with Dr. Watson's lab on a few projects and two of them have been published on good journals. I also asked Dr.Zhang for help when I had trouble in my research and he is always willing to help me out.

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. Pengbiao Xu, a graduate student, is always talking with me about my project and helped me with some experiments. Xinsheng Gao, a former post-doc, trained me in fluorescence experiment.

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 Berkeley Center for Structural Biology for collecting the X-Ray diffraction data, the Welch Foundation and National Institute of Health for funding my research.

Finally, I want to give many thanks to my Mom for her consistent encouragement and unconditional love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Pingwei Li [Advisor], Professor Margaret Glasner [Committee Chair] and Professor Xiuren Zhang of the Department of Biochemistry and Biophysics and Professor Robert O. Watson of the Department of Microbial Pathogenesis and Immunology, Texas A&M University Health Science Center.

The phosphorylated human IRF-3 structure was solved by the post-doc Baoyu Zhao in my lab. The fluorescence experiment in Chapter 4 were conducted in part by Xinsheng Gao in my lab. Quantification of nuclear translocation of hIRF-3 in Figure 46 was performed by Huajun Han of the Department of Biochemistry and Biophysics. Mutagenesis studies were carried out in part by Pengbiao Xu in my lab. Discussion of IRF-3 project was revised by the post-doc Baoyu Zhao and Professor Pingwei Li.

All other work conducted for the dissertation was completed by me independently.

Funding sources

This work was supported in part by the Welch Foundation (Grant A-1931 to P.L.) and the National Institutes of Health, National Institute of Allergy and Infectious Diseases (Grant R01 AI-145287 to P.L.).

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
cGAMP	Cyclic guanosine monophosphate–adenosine monophosphate
cGAS	Cyclic GMP-AMP synthase
CTD	Carboxyl-terminal domain
CTT	Carboxyl-terminal tail
D	Aspartic acid
DAPI	4′-6-diamidino-2-phenylindole
DBD	DNA binding domain
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
Gln	Glutamine
Glu	Glutamic acid
H	Histidine
HEK	Human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
I	Isoleucine
IAD	IRF association domain
IBiD	IRF-binding domain
IDT	Integrated DNA technologies
IFN	Interferon
IFN-I	Type-I interferons
IKK	Inhibitory-κB kinase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRF	Interferon regulatory factor
ISRE	Interferon-stimulated response element
K	Lysine

KD	Kinase domain
L	Leucine
LB	Luria Bertani broth
LPS	Lipopolysaccharide
Lys	Lysine
MAD	Multi-wavelength anomalous dispersion
MALDI-TOF	Matrix-assisted laser desorption ionization time-of-flight
MAVS	Mitochondrial antiviral signaling
MR	Molecular replacement
MS	Mass spectrometry
NDV	Newcastle disease virus
NF-κB	Nuclear factor κB
NP	Nucleoprotein
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
phIRF-3	Phosphorylated human IRF-3

pmIRF-3	Phosphorylated mouse IRF-3
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
SAD	Single-wavelength anomalous dispersion
SDD	Scaffold/dimerization domain
SDS	Sodium dodecyl sulphate
SEC	Size-exclusion chromatography
Ser	Serine
Sf9	Spodoptera frugiperda 9
SPR	Surface plasmon resonance
SRR	Serine-rich repeat
STING	Stimulator of interferon genes
SUMO	Small ubiquitin-like modifiers
TBD	Tank-binding domain
TBK1	TANK-binding kinase 1
TF	Transcription factor

Thr	Threonine
TLR	Toll-like receptor
TRIF	TIR domain-containing adaptor inducing IFN-β
TRIM 14	Tripartite motif-containing protein 14
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
ULD	Ubiquitin-like domain
V	Valine
VPS	Vacuolar protein sorting
W	Tryptophan
WT	Wild type
X	Any amino acid residue
Y	Tyrosine

TABLE OF CONTENTS

Page

LIST OF FIGURES

CHAPTER I. INTRODUCTION

I.1. Innate immune system

Vertebrates are constantly exposed to a variety of potential pathogens including bacteria and viruses daily. Their ability to eliminate infection depends on two ways: innate immune system and adaptive immune system(1-4). The latter one remembers the previous specific pathogens and gets rid of them when they invade again whereas the innate immune system serves as the first line to defend against a new pathogen in an efficient and timely manner, thus protecting us from infection(2, 3).

Specifically, the innate immune system is able to detect the pathogen-associated molecular patterns (PAMPs) such as DNA, RNA or lipopolysaccharide (LPS) through various pattern recognition receptors (PRRs), which then triggers the induction of a variety of cytokines including type-I interferons (IFN-I) to initiate host defense against pathogens(5-10). IFN-Is are secreted cytokines that activate a signal transduction cascade which results in the induction of hundreds of interferon-stimulated genes (ISGs). Specifically, Type I IFNs signal through the JAK/STAT pathway to activate a transcription factor complex ISGF3 containing phosphorylated STAT1/STAT2 and IRF9. Activated ISGF3 translocates to the nucleus and binds IFN-stimulated response elements (ISRE) in the upstream promoter regions of ISGs to initiate the transcription of ISGs. ISG-encoded proteins play very critical roles in antiviral infection by fighting against a particular step in the infection/replication cycle of a virus. For example, some ISG effectors including IFITM and NCOA7 proteins can impair entry of several enveloped viruses. In addition, another ISG effector TRIM5α can target viral capsid after retroviral entry, which then mediates proteasome-dependent destruction of the viral core. Moreover, the ISG system has evolved a variety of distinct mechanisms to target viral gene expression, protein production, and genome amplification. For example, APOBECs and IFI16 interfere with viral mRNA synthesis; PKR and ZAP target viral protein synthesis; Viperin and IFI6 inhibit viral replication. The viral particle assembly and egress can also be targeted by ISG effectors although not very common. The most well characterized ISG involved in this process is tetherin which is a transmembrane protein that is able to anchor budding virions to the cell surface and therefore prevent them from being released.

I.2. cGAS-STING, RIG-I-MAVS, TLR3/4-TRIF pathways

There are several signaling pathways involved in innate immunity that have been studied for years, as described below (9, 11-18).

Viral or bacterial DNA in cytosol are recognized by cyclic GMP-AMP synthase (cGAS), which catalyzes the synthesis of a cyclic dinucleotide cGAMP. cGAMP binds to the adaptor STING (Stimulator of Interferon Genes) located on the endoplasmic reticulum (ER) membrane and mediates the recruitment and activation of TBK1 (TANK-Binding Kinase 1) and IRF-3. Activated IRF-3 dimerizes, translocates to the nucleus and initiates the transcription of IFN-β gene with other transcription factors such as NF-κB(9, 11, 12, 19, 20).

Double-stranded RNA and LPS can be recognized by retinoic acid-inducible protein 1 (RIG-I)-like receptors (RLRs) and Toll-like receptors (TLRs), respectively, to induce the expression of IFN-Is through the adaptors MAVS (mitochondrial antiviral signaling) and TRIF (TIR domain-containing adaptor inducing IFN-β). Interestingly, it has been found that all these three signaling pathways converge at the recruitment of IRF-3 via a conserved pLxIS (p, hydrophilic residue, x, any residue, S, phosphorylation site) motif located within these adaptor proteins(10, 17) (**Figure 1**).

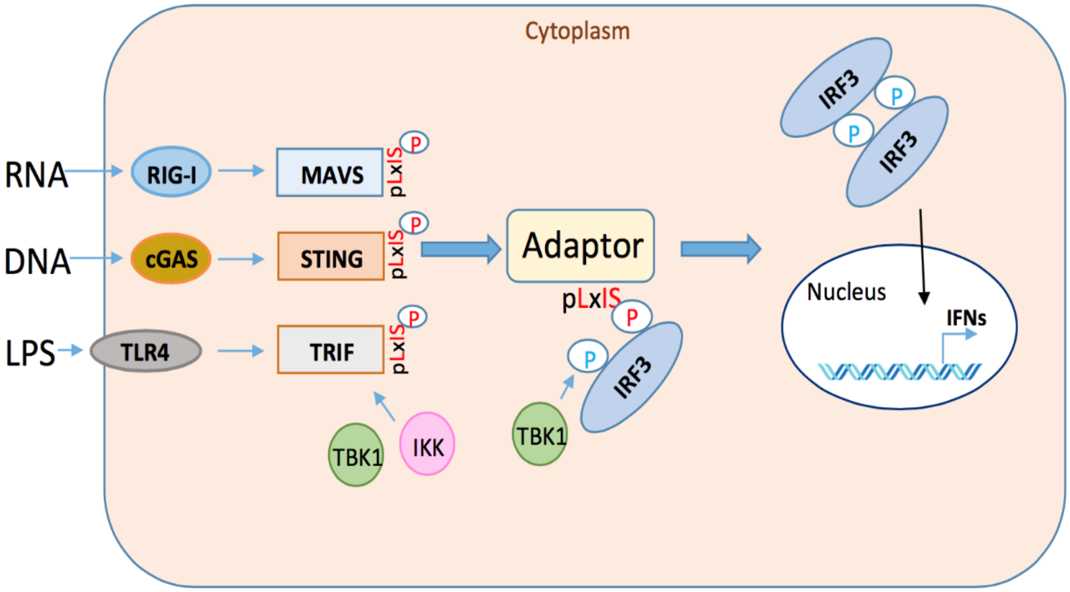


Figure 1 Schematic diagram showing the signaling pathways involved in innate immunity. PAMPs: RNA, DNA, LPS PRRs: RIG-I, cGAS, TLR4 Adaptors: MAVS, STING, TRIF

I.3. cGAS-STING pathway

In recent years, our lab has been focusing on the cGAS-STING pathway and trying to elucidate the mechanism of this pathway involving many critical proteins from a structural perspective.

Up to now, we have determined the structures of cGAS alone or in complex with an 18 bp DNA, revealing that cGAS binds to dsDNA via two binding sites resulting in a 2:2 complex(18). Also, we have solved the structures of STING ligand binding domain alone and in complex with cGAMP, elucidating the mechanism of cyclic dinucleotide recognition by STING(15). Moreover, mTBK1 structures (N657) that are bound to two inhibitors SU6668 and BX795 have been solved, indicating the mechanism of TBK1 activation and the role it plays in IRF-3 activation(21). Later on, we also determined a set of phosphorylated adaptor peptides bound to IRF-3 C-terminal domain (pSTING-IRF-3, pMAVS-IRF-3, pTRIF-IRF-3), demonstrating a concerted mechanism for IRF-3 recruitment through a *p*L*x*IS motif(17). Last year, we determined the structures of TBK1 in complex with STING C-terminal domain or C-terminal tail, suggesting PLPLRT/SD motif located in the STING CTT mediates TBK1 recruitment and activation(22). These complicated structural studies facilitate our understanding of the mechanism of the cGAS-STING pathway (**Figure 2**) and may provide us with some ways to eliminate undesired innate immune responses caused by some viral infection or autoimmune disorders.

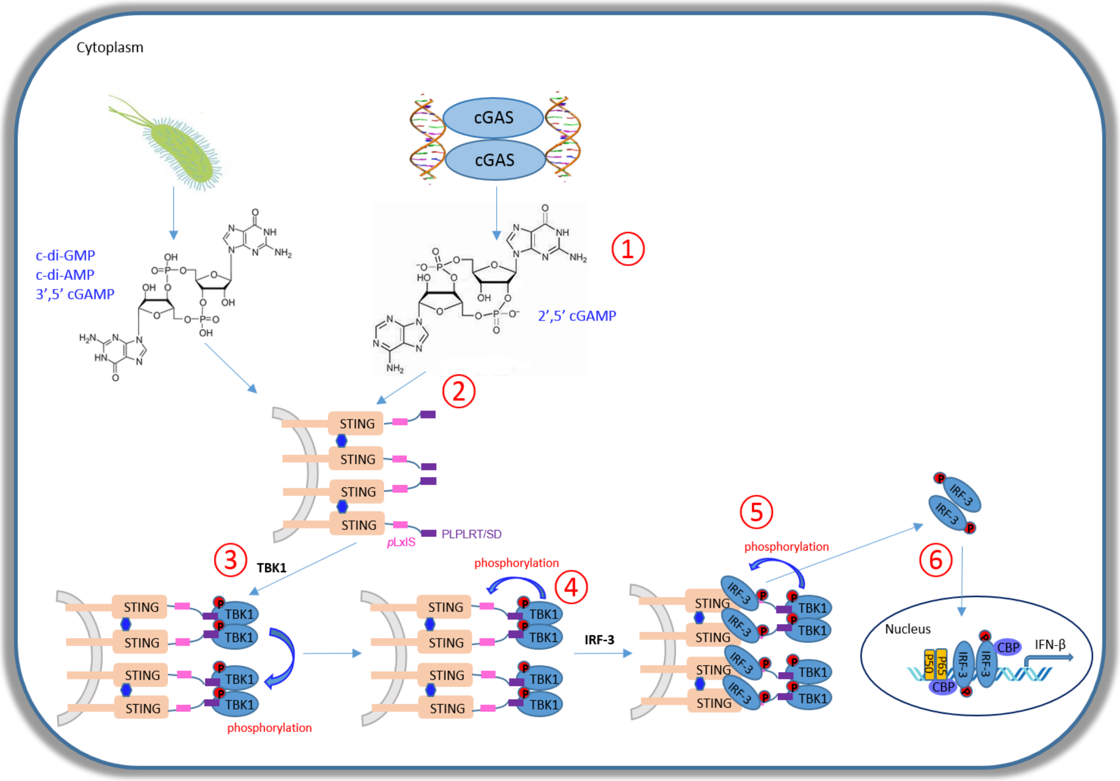


Figure 2 cGAS-STING pathway.

(1). cGAS is activated by dsDNA and catalyses cGAMP synthesis. (2). The newly synthesized cGAMP binds to STING and leads to its oligomerization at the ER or Golgi membranes. (3). TBK1 is recruited to the STING oligomers through PLPLRT/SD motif and is further activated due to induced proximity. (4). The phosphorylation of STING via TBK1 facilitates TBK1 recruitment and activation. The activated TBK1 phosphorylates STING at its *p*L*x*IS motif, facilitating the recruitment of IRF-3 to the signaling complex. (5). The proximity between TBK1 and IRF-3 bound to STING molecules leads to IRF-3 phosphorylation. (6). Phosphorylated IRF-3 then dissociates from STING, dimerizes, translocates to the nucleus, and transcribes *IFN-β* gene.

I.4. IRF family and IRF-3

The interferon regulatory factor (IRF) family transcription factors contain nine members (IRF-1 through IRF-9). These transcription factors contain a highly conserved N-terminal DNA-binding domain known for the recognition of IFN-stimulated response element (ISRE) and a relatively divergent C-terminal regulatory domain mediating the interaction of a IRF with another IRF family member or other transcription factors, suggesting that most members function non-redundantly (**Figure 3**). The IRF family members play important roles in multiple areas ranging from innate immunity, immune cell development, the regulation of the cell cycle and apoptosis and oncogenesis(23-27). Among all the members, IRF-3 is a key transcription factor that regulates the expression of type I interferon genes. Under resting conditions, IRF-3 adopts an auto-inhibited conformation and is ubiquitously accumulated in the cytoplasm. Bacterial or viral infection triggers the activation of IRF-3 through various innate immune sensing pathways. Phosphorylated IRF-3 binds to p300/CBP, translocates to the nucleus, and initiates the transcription of IFN-I genes(28-30). Interestingly, IRF-7, which is closely related to IRF-3 in terms of the conserved regulatory domain, can also be phosphorylated, translocate to the nucleus and regulate the expression of IFN-Is(31-33). Compared to IRF-3, which is mainly responsible for the induction of IFN-β in the early stage of an infection, IRF-7 induced by IFN-β is primarily responsible for the induction of IFN-α in the later stage of a response(23, 24, 34, 35).

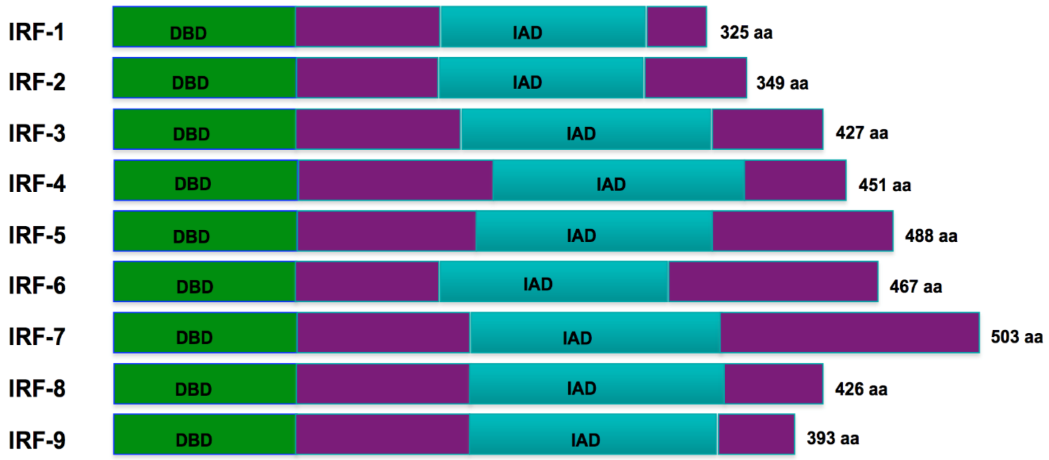


Figure 3 Schematic illustration of human IRF family members.
All nine members possess an N-terminal DNA binding domain (DBD) colored in green, and an IRF association domain (IAD) colored in cyan.

In terms of IRF-3 alone, it is a 427 amino acid long transcription factor, which not only possesses the DBD and IAD domains but also contains auto-inhibitory elements flanking on both sides of IAD. Within the C-terminal auto-inhibitory segment, there is a serine rich repeat containing seven serine/threonine residues that has been shown to be the potential sites of phosphorylation (**Figure 4A and B**). The mechanism of IRF-3 activation has been extensively studied, which showed that IRF-3 is activated through phosphorylation of the C-terminal serine-rich repeat (SRR)(28, 29, 36, 37). Previous studies by the Fujita lab showed that the phosphorylation site 1 (residues Ser385, Ser386 of human IRF-3) plays a key role in IRF-3 activation. The phosphorylation of Ser386 induced by viral infection has been detected by a specific antibody. They also observed that the dimerization of IRF-3 was abolished by the mutation of Ser386(38-40). In addition, the Hiscott lab observed that the phosphorylation site 2, which includes residues Ser396, Ser398, Ser402, Thr404, Ser405 of human IRF-3, plays a critical role in IRF-3 activation. The phosphomimetic mutation of these residues (IRF-3 5D) results in a constitutively active phenotype. Moreover, they observed that the S396D mutation alone induces IFN-I expression(41, 42), suggesting that Ser396 also plays a critical role in IRF-3 activation. Another study by the Harrison lab proposed a two-step phosphorylation and activation model, which suggest that phosphorylation at site 2 leads to the alleviation of IRF-3 auto-inhibition that facilitates the phosphorylation at site 1 and eventually leads to the activation of IRF-3(43). Later on, studies by the Lin group showed that IRF-3 mutant S386D/S396D bound to CBP forms a stable oligomer(44), suggesting that the phosphorylation of both Ser386 and Ser396 is essential for human IRF-3 activation. From a structural perspective, the crystal structures of human IRF-3 (173-427aa) (**Figure 4C**)and human IRF-3(173-394aa) in complex with IBiD of CBP have been solved by Lin group, elucidating the mechanism of auto-inhibition(28, 29). Based on all these cellular and structural studies, our lab mutated Ser386 and Ser396 to glutamic acid in a truncated form of IRF-3 (residues 189 to 398) to mimic the phosphorylation and determined the structure of the phosphomimetic IRF-3 in complex with CBP, which reveals that the phosphomimetic IRF-3 mutant forms a dimer(17). However, the higher affinity between phosphorylated IRF-3/CBP complex than that of the S386/396E mutant indicates that the phosphomimetic mutation does not fully recapitulate the interactions between phosphorylated IRF-3.

Overall, the extensive research on the IRF-3 activation done by all the groups so far has indicated that Ser386 and Ser396 of human IRF-3 play important roles in IRF-3 activation.

However, how do these two sites contribute to the activation of IRF-3 remains unclear. From a structural perspective, although human IRF-3(173-394aa) bound to IBiD of CBP and phosphomimetic double mutant S386/396E bound to CBP have been determined, there is no real phosphorylated IRF-3 structures reported yet, thus making the exact mechanism of IRF-3 activation not fully clear. Therefore, in order to elucidate the exact mechanism of IRF-3 activation upon phosphorylation, our goal is to get phosphorylated IRF-3 structures containing both phosphorylated Ser386 and Ser396 and clearly uncover the roles of these two sites in IRF-3 activation using biochemical studies combined with cell-based assays.

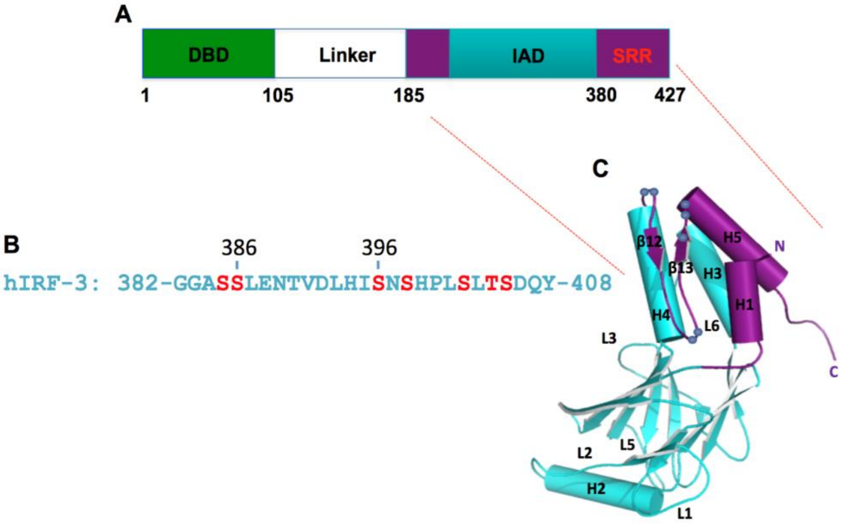


Figure 4 Schematic diagram of IRF-3 and its SRR within the C-terminal tail and crystal structure of human IRF-3 (173-427aa) adapted from Bin Y Qin.
(A) Human IRF-3 contains an N-terminal DNA binding domain (DBD) colored in green, and an IRF association domain (IAD) colored in cyan and auto-inhibitory segments flanking on both sides of IAD colored in purple. In between DBD and IAD is a long linker. (B) The C-terminal auto-inhibitory segment containing serine-rich repeat (SRR). The amino acids in red color are the potential phosphorylation sites (C) Human IRF-3 (173-427aa) structure that includes IAD domain and the flanking auto-inhibitory segments that stabilizes the hydrophobic surface of H3 and H4 helices of the IAD. The circles between H4 and H5 helices represent the putative phosphorylation sites.

This dissertation contains six chapters. The first five chapters are tightly related to the structural basis of IRF-3 activation upon phosphorylation. The last chapter is about three other projects I have participated in. Specifically, the first chapter is the introduction concerning innate immune system, three signaling pathways involved in innate immunity, background information on IRF family and IRF-3. The second chapter talks about the preparation work I did in order to get phosphorylated IRF-3 crystals and IRF-3 crystals I obtained after careful optimization of crystallization conditions, which is the basis for my project. The third chapter mainly depicts the phosphorylated mouse IRF-3 structure and compare and contrast this structure to the phosphorylated human structure. In addition, comparisons of phosphorylated human structure and phosphomimetic double mutant S386/396E; phosphorylated human structure and autoinhibited human structure have been discussed. Besides, the roles of Ser386 and Ser396 have been studied by size-exclusion chromatography, native gel electrophoresis and luciferase reporter assay. Together, this chapter provides critical insights into the mechanism of IRF-3 activation upon phosphorylation. The fourth chapter continues to elucidate the mechanism of IRF-3 activation by mutating residues that are interacting with phosphorylated Ser386 and Ser396. The cell-based functional studies together with biochemical studies indicated that mutations of residues interacting with phosphorylated Ser386 severely impaired IRF-3 dimerization, inhibited IRF-3 translocation to the nucleus and diminished IRF-3 mediated signaling whereas mutations of residues interacting with phosphorylated Ser396 moderately affect IRF-3 dimerization, its translocation and mediated signaling, indicating that Ser386 plays a more critical role in IRF-3 activation. The fifth chapter talks about some future studies on IRF-3 project, which includes 1). how could the structural differences between mouse and human IRF-3 lead to the differential induction of IFN-I in cells? 2). What is the molecular mechanism of IFN-β transcription by the phosphorylated IRF3 in the presence of full-length CBP? The sixth chapter involves the other three projects I participated in: 1). The TRIM 14 project which elucidated that TRIM 14 is a key regulator of the type I IFN response during mycobacterium tuberculosis infection 2). The cGAS-nucleosome project which demonstrated the molecular basis of cGAS inactivation by nuclear tethering 3). The SseC project which delineated quantitative genetic interaction profiles called E-MAP, through which an uncharacterized SseC effector protein has been discovered and the role of SseC has been studied.

CHAPTER II. PREPARATION WORK TO GET IRF-3 CRYSTALS

II.1. Introduction

As we discussed in the introduction chapter, IRF-3 contains a highly conserved N-terminal DNA-binding domain for the recognition of ISRE and a C-terminal IRF association domain for the interaction of a IRF with another IRF family member or other transcription factors. There have been a few structural studies on the DNA-binding domain of IRF-3 which elucidated that IRF-3 molecules bind to variably spaced consensus and nonconsensus sequences of the PRDIII-I element on the IFN-β enhancer. However, the activation of IRF-3 is dependent on its transactivation domain including IRF association domain and the flanking autoinhibitory elements. Lin group previously reported a crystal structure of IRF-3 transactivation domain (173-427aa), which indicated that H3 and H4 helices of IRF-3 association domain interacted with the H1 and H5 helices of the autoinhibited region to form a condensed hydrophobic core to make IRF-3 inactive. They further suggested that phosphorylation that would happen on β12 and β13 sheets could reorganize the autoinhibitory elements and therefore unmask the hydrophobic core for IRF-3 activation. However, our goal is to obtain the phosphorylated IRF-3 structure and further explore the mechanism of IRF-3 activation. In this chapter, we wanted to get single crystals of IRF-3 that could diffract well. In order to achieve this goal, we need to make IRF-3 constructs suitable for getting crystals; then we need to express and purify SUMO protease that would help cleave the SUMO tag off the IRF-3 protein; we also need to express and purify TBK1 kinase that would be used to phosphorylate IRF-3 in vitro. Finally, we need to set up for crystallization screen/optimization.

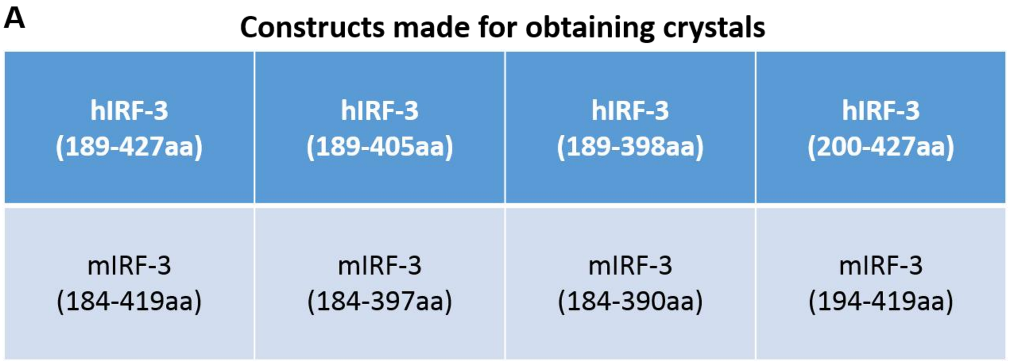
II.2. IRF-3 constructs selection and preparation

By closely looking at the solved structure by Lin group, we found that this structure started with Glu189 and ended with Ser427. The sequence shown in the structure is different from the construct (173-427aa) they used due to the fact that the loop region between amino acid 173 and 189 is very flexible. Besides, the phosphomimetic double mutant S386/396E IRF-3 (189-398aa) structure solved by the post-doc Baoyu in our lab indicated that it forms a stable dimer. Based on these two structures, we decided to make three human constructs with Glu189 as the starting amino acid. The shortest construct is hIRF-3 (189-398aa) that contains both Ser386 and Ser396 phosphorylation sites. The longest construct is hIRF-3 (189-427aa) that includes the last amino acid of human IRF-3. As an alternative, we also designed a human construct starting from Glu200 which lacks a small helix as compared to the other three human constructs based on the phosphomimetic double mutant structure. we then aligned the C-terminal sequences of IRF-3 across different species and observed that Ser386 and the pLxIS motif containing Ser396 of hIRF-3 are highly conserved in other species (**Figure 5**). In order to increase the chance to get crystals, we also designed four mouse IRF-3 constructs for the following experiment. The shortest mouse construct is mIRF-3 (184-390aa) that contains phosphorylation sites corresponding to both Ser386 and Ser396 in human IRF-3. The longest construct is mIRF-3 (184-419aa) that includes the last amino acid Ile419 (**Figure 6A**).

```
                               386       396
                               *|         |  *   * **
IRF-3 HUMAN KVVPTCLRALVEMARVGGASSLENTVDLHISNSHPLSLTSDQYKAYLQDLVEGMDFQGPGES
IRF-3 MOUSE KVVPTCLKELLEMAREGGASSLK-TVDLHISNSQPISLTSDQYKAYLQDLVEDMDFQATGNI
IRF-3 PIG   KVVPMCLRALVDMARDGGASSLENTVDLHISNSHPLSLTSDQYKACLRDLVEDMDF------
IRF-3 HORSE KVVPTCLKALLEMARIQGASSMENTVDLHISNSHPLSLTSDQYKAYLEDLVEDMDI------
IRF-3 RAT   KVVPTCLKELLEMAREGGASSLK-TVDLHISNSQPISLTSDQYKACLQDLVEDMDFQATGET
IRF-3 SHEEP KVVPMCLRVLVDIARQGGASSLENTVDLHISNSQPLSLTSDQYMAYLQDLAEDMDF------
IRF-3 BOVINEKVVPMCLRVLVDIARQGGASSLENTVDLHISNSDPLSLTPDQYMACLQDLAEDMDF------
IRF-3 CAT   KVVPTCLRALLDMARSGGASSLETTVDLHISNSYPLSLTSDQYKAYLQDLVEDMDFVT-GEI
```

Figure 5 Sequence alignment of C-terminal tail of IRF-3 across different species.
The conserved phosphorylation sites and the pLxIS motif are shown in red. The asterisks indicate other potential phosphorylation sites.

The cDNA encoding human and mouse IRF-3 association domains were cloned into a modified pET-28a (+) vector containing an N-terminal His$_6$-SUMO tag with appropriate primers obtained from Integrated DNA Technologies (IDT) using *Eco*RI and *Xho*I restriction sites. All the clones with inserted IRF-3 fragments were digested by *Eco*RI and *Xho*I enzymes and run on the 1% agarose gel to show the IRF-3 fragment bands before sending them for DNA sequencing to avoid empty vectors with no inserted IRF-3 (**Figure 6B**). Sequences of all the constructs were confirmed by DNA Sequencing.



A

Constructs made for obtaining crystals

hIRF-3 (189-427aa)	hIRF-3 (189-405aa)	hIRF-3 (189-398aa)	hIRF-3 (200-427aa)
mIRF-3 (184-419aa)	mIRF-3 (184-397aa)	mIRF-3 (184-390aa)	mIRF-3 (194-419aa)

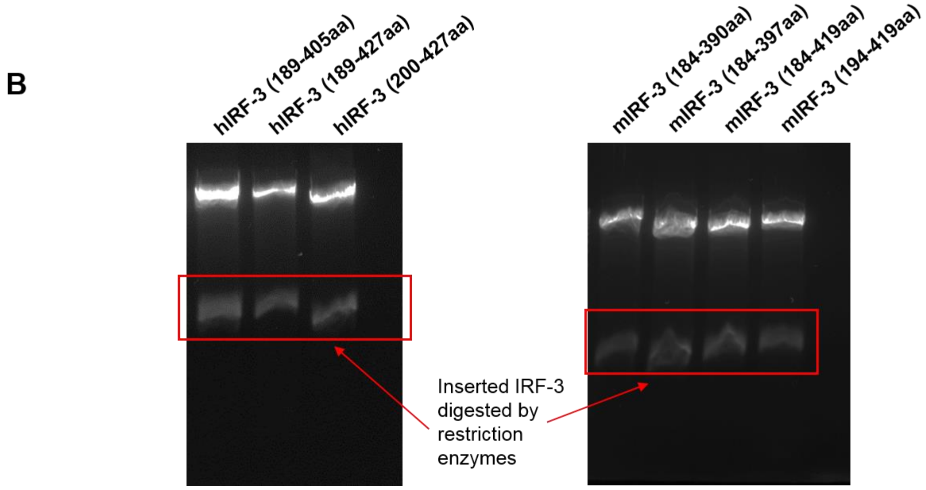


Figure 6 Constructs selected and made for getting IRF-3 crystals.
(A) Different truncated IRF-3 constructs are selected. (B) IRF-3 constructs are digested by EcoRI and XhoI restriction enzymes to show the IRF-3 bands before DNA sequencing.

II.3. Expression and purification of SUMO protease

SUMO protease known as Ulp is an active cysteinyl protease that recognizes the tertiary structure of SUMO and cleaves it in a highly specific manner(45-47). Since all of our IRF-3 constructs were cloned into pET-28a (+) vector containing an N-terminal His_6-SUMO tag, it means that IRF-3 proteins were expressed as an N-terminal SUMO fusion. In order to cleave the SUMO tag from IRF-3 proteins, SUMO protease is required. We expressed SUMO protease in *Escherichia coli* BL21 (DE3) cells and purified it using Ni^{2+}-NTA column (Qiagen) followed by gel-filtration chromatography (**Figure 7**).

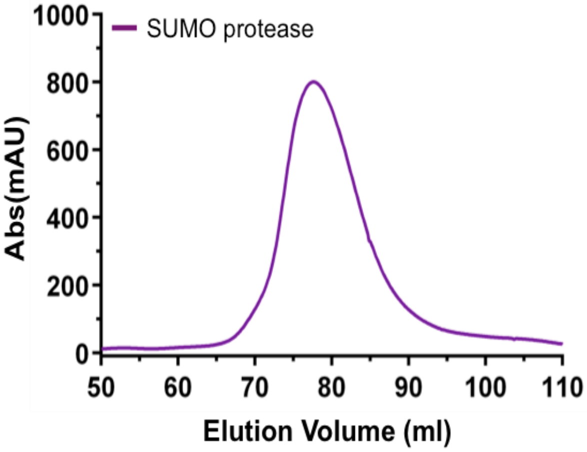


Figure 7 Gel filtration chromatography showing the purified SUMO protease.

II.4. Expression and purification of IRF-3 proteins bound to CBP

Since the truncated IRF-3 is prone to oligomerization and CBP is required for IRF-3 function in the nucleus(39, 42, 48, 49), we co-expressed truncated forms of human and mouse IRF-3 C-terminal domains with a CBP fragment in *Escherichia coli* BL21 (DE3) cells. All the proteins were purified using Ni^{2+}-NTA column (Qiagen) followed by gel-filtration chromatography after the His_6-SUMO-tag was cleaved by sumo protease (**Figure 8 and Figure 9**). Compared to all the purified human proteins, mouse proteins shown on gel-filtration chromatography indicated that they are not as pure as the human ones and therefore we only collected several fractions around the peaks (mIRF-3 (184-390aa): fractions collected between A11 and B9; the other three mIRF-3: fractions collected between B12 and B8). The collected fractions are pure enough for the following experiments.

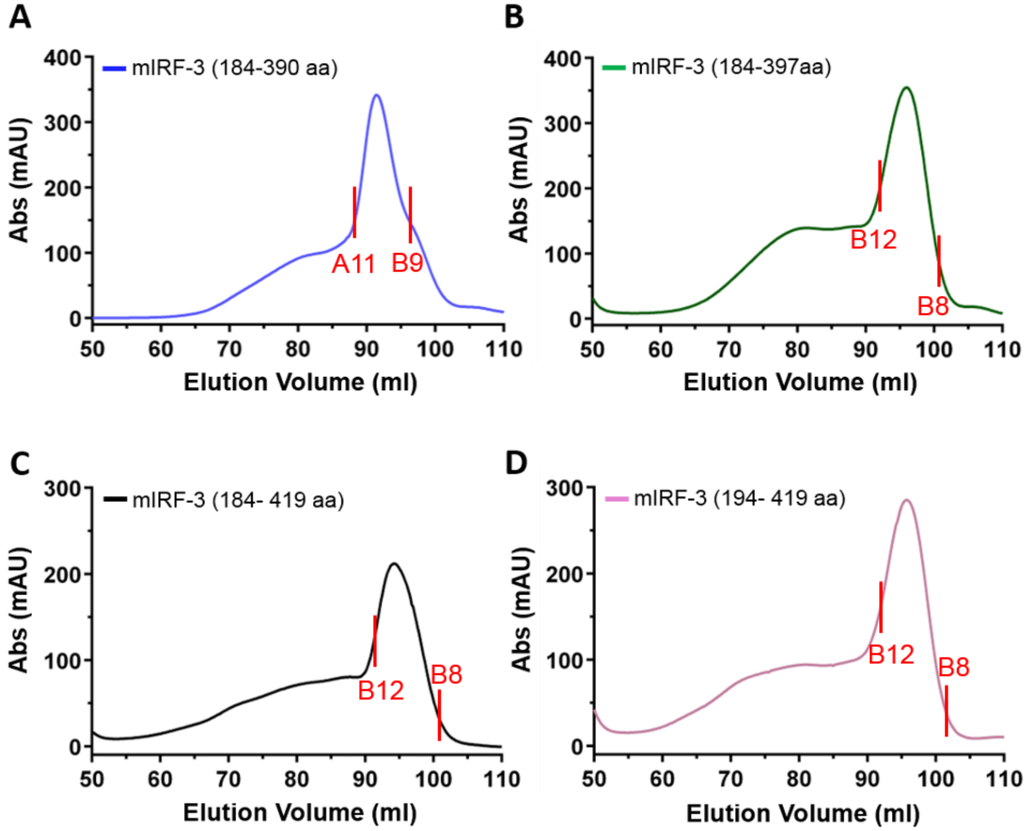


Figure 8 Gel filtration chromatography showing the purified mouse IRF-3 proteins.
(A) mIRF-3 (184-390 aa)/CBP (B) mIRF-3 (184-397 aa)/CBP (C) mIRF-3 (184-419 aa)/CBP (D) mIRF-3 (194-419 aa)/CBP

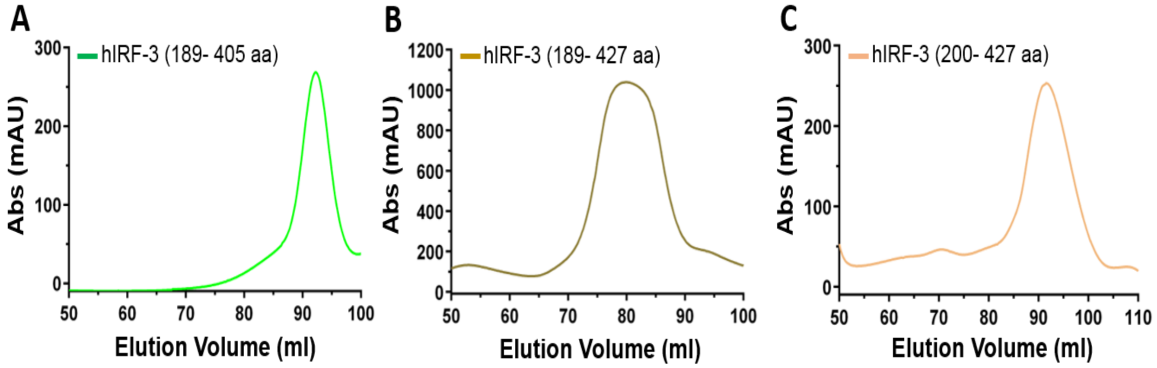


Figure 9 Gel filtration chromatography showing the purified human IRF-3 proteins. (A) hIRF-3 (189-405 aa)/CBP (B) hIRF-3 (189-427 aa)/CBP (C) hIRF-3 (200-427 aa)/CBP

II.5. Expression and purification of mIRF3(184aa-397aa)/CBP in M9 media

After mIRF3(184aa-397aa) together with CBP was co-expressed and purified using the exact same procedure as other IRF3 truncations, it was noticed that this protein after cleaved with SUMO protease overnight, appeared to be heavily degraded which was clearly observed on SDS-PAGE gel. The upper left band was the sumo-tagged protein with no SUMO protease cleavage and the two bands in the red box were mIRF3 that has been degraded (**Figure 10 left gel**). In order to avoid this protein to be degraded, I used M9 minimal medium to culture bacterial cells instead of using regular LB medium. Also, all the purification procedure was carried out on ice. After purification through gel-filtration, small samples were again run on the SDS-PAGE gel, which indeed showed that much less protein was degraded (**Figure 10 right gel**).

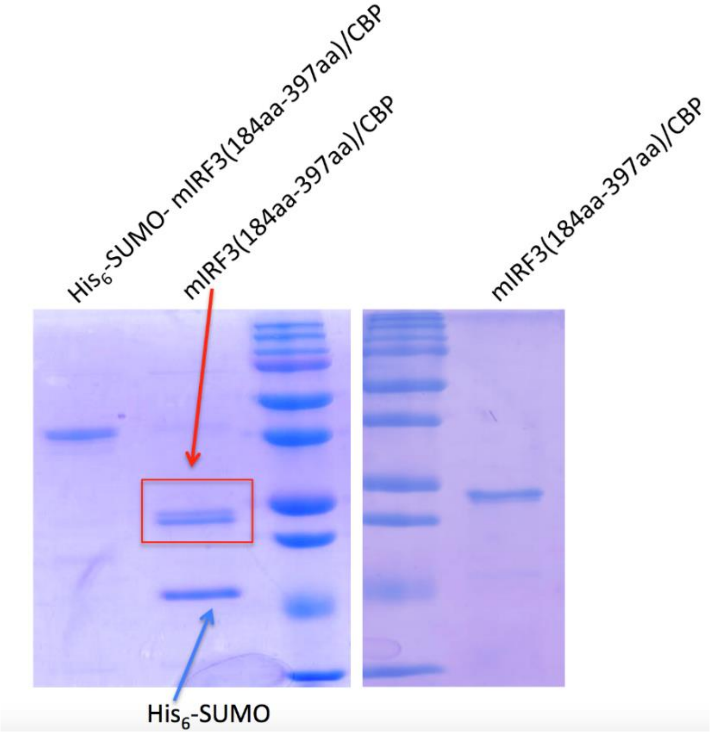


Figure 10 mIRF3(184aa-397aa)/CBP protein is shown on the SDS-PAGE gel.
The left gel shows the protein expressed in normal LB medium. The bands in the red box are degraded mIRF3(184aa-397aa). The gel on the right shows the same protein expressed in M9 minimal medium.

II.6. Expression and purification of mouse TBK1 (N657)

Tank-binding kinase 1 (TBK1) is a 729 amino acid long protein with four domains including kinase domain, ubiquitin-like domain, scaffold/dimerization domain and tank-binding domain (**Figure 11**). It is a serine/threonine protein-kinase that is playing a critical role in a variety of signaling pathways involved in innate immunity(50-52). Once the PAMPs are recognized by PRRs, TBK1 can be autophosphorylated at Ser172, which leads to the phosphorylation of IRF-3(21, 22, 53).

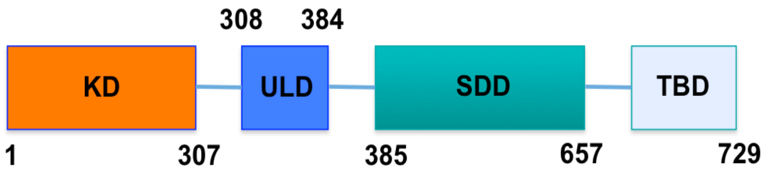


Figure 11 Schematic diagram of TBK1 with four domains.
KD: kinase domain, ULD: ubiquitin-like domain, SDD: scaffold/dimerization domain, TBD: tank-binding domain.

In order to get the purified IRF-3 phosphorylated by TBK1, I expressed and purified mouse TBK1 (1-657aa) with no C-terminal TBD from *sf9* insect cells (**Figure 12A and B**) because this truncation does not affect its kinase activity and the protein yield is relatively large compared to the full-length TBK1.

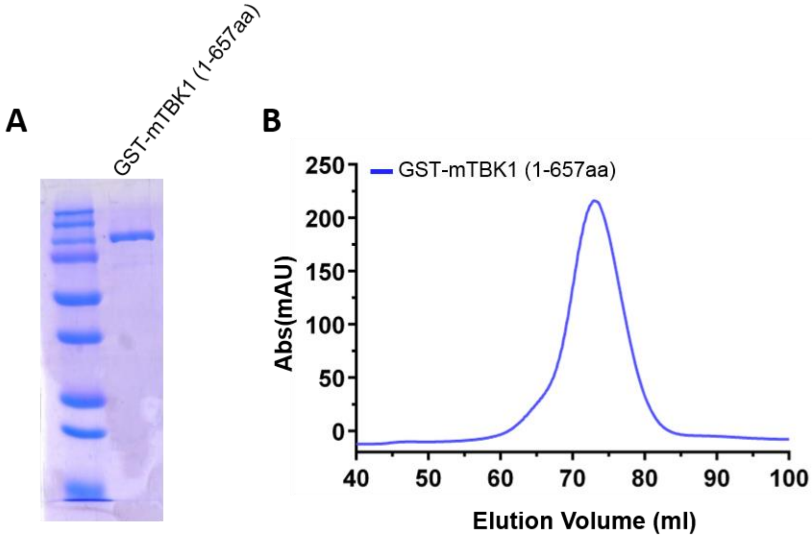


Figure 12 Characterization of GST-mTBK1 (1-657aa).
(A) The purified GST-mTBK1 (1-657aa) from *sf9* cells shown on SDS-PAGE. (B) Gel filtration showing the purified GST-mTBK1 (1-657aa).

II.7. IRF-3 phosphorylation by TBK1

With both purified IRF-3 and TBK1 in hand, I performed the TBK1 phosphorylation assay in vitro. Basically, I mixed IRF-3 with TBK1 in a TBK1 kinase buffer and incubated the reaction at 27 °C for ~24 hours. The next day, the phosphorylated IRF-3 was purified by gel filtration chromatography (**Figure 13 and Figure 14**).

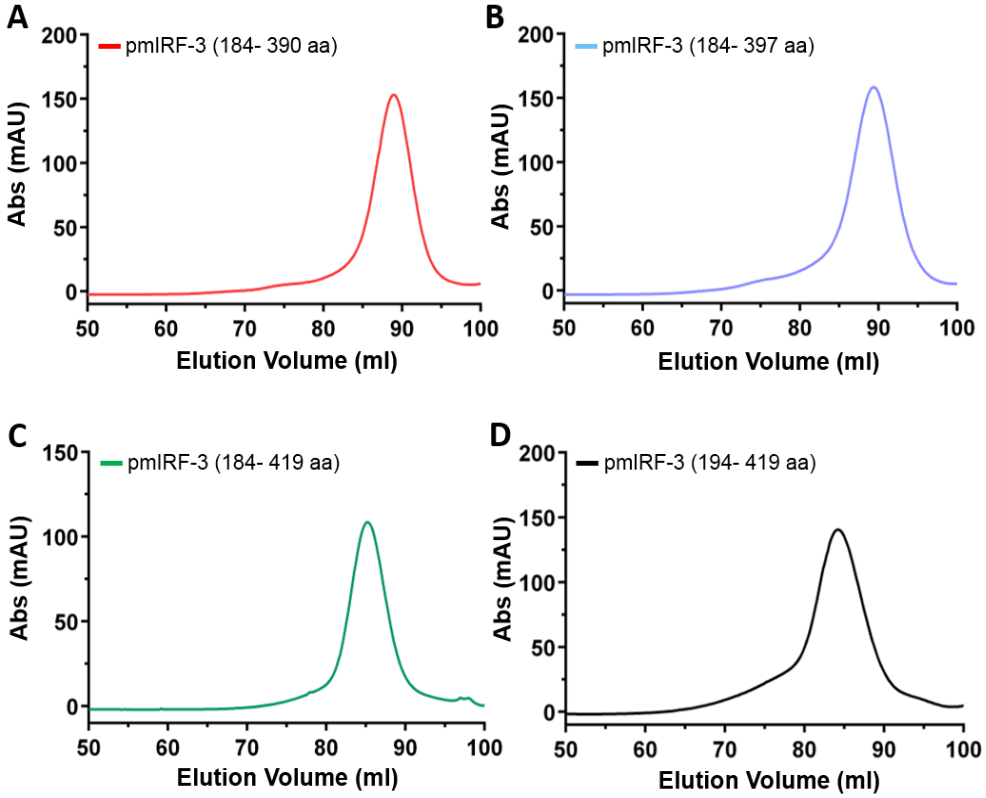


Figure 13 Gel filtration chromatography showing the phosphorylated mouse IRF-3 proteins after 24 h incubation with TBK1.
(A) pmIRF-3 (184-390 aa)/CBP (B) pmIRF-3 (184-397 aa)/CBP (C) pmIRF-3 (184-419 aa)/CBP (D) pmIRF-3 (194-419 aa)/CBP

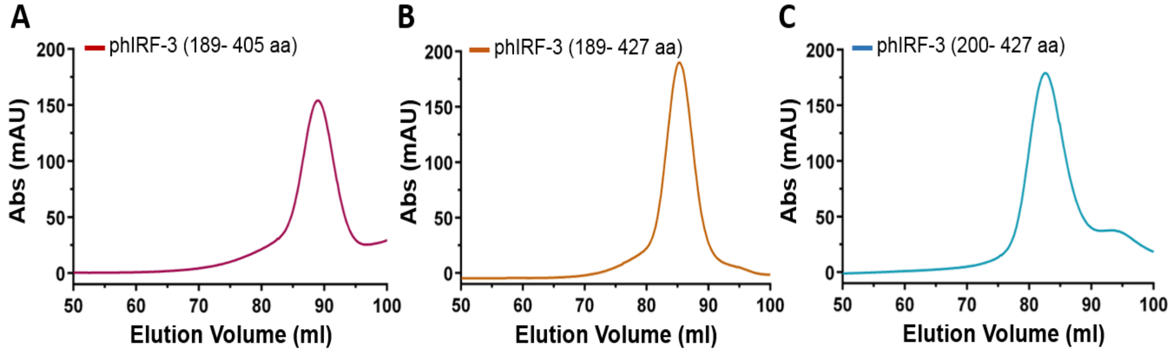
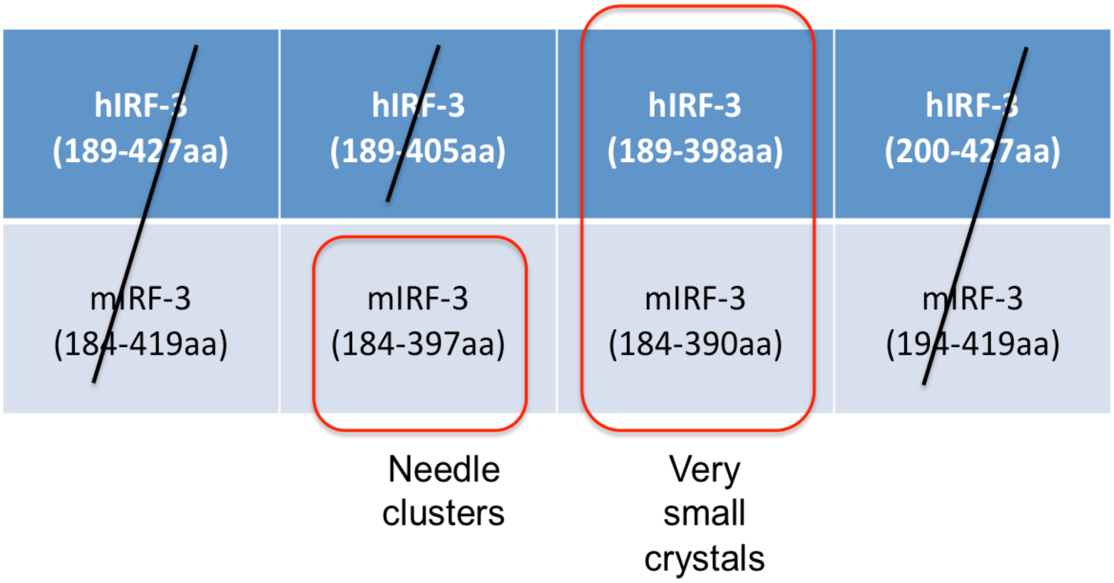


Figure 14 Gel filtration chromatography showing the phosphorylated human IRF-3 proteins after 24 h incubation with TBK1.
(A) phIRF-3 (189-405 aa)/CBP (B) phIRF-3 (189-427 aa)/CBP (C) phIRF-3 (200-405 aa)/CBP

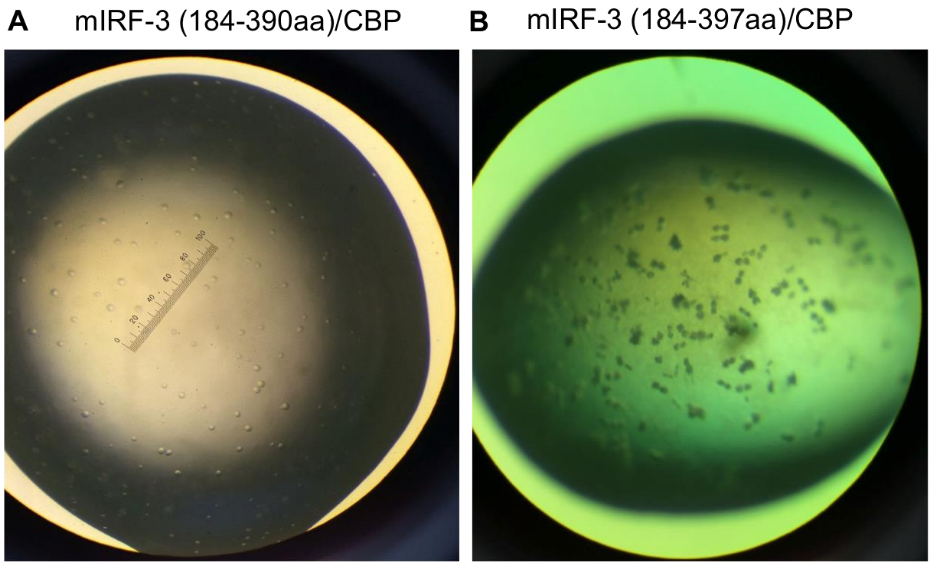
II.8. Crystallization screen and optimization of protein crystallization

With all purified phosphorylated IRF-3, it was ready for setting up the crystallization screen. I mixed IRF-3 and crystallization conditions in equal amounts and set up the screen by using the vapor diffusion method in a cold room. After one week, I checked out all the plates in the cold room and found that only mIRF-3 (184-390aa)/CBP and mIRF-3 (184-397aa)/CBP proteins turned into crystals (**Figure 15**). Specifically, mIRF-3 (184-390aa)/CBP was crystallized under conditions of index 86, index 88, index 89, index 90 respectively whereas mIRF-3 (184-397aa)/CBP was crystallized under conditions of index 26, index 35, index 39, index 63. (Note: the post-doc Baoyu set up the crystallization screen of the human IRF-3 (189-398aa)/CBP, which also turned into crystals). The crystals of mIRF-3 (184-390aa)/CBP under index 86 and index 88 were single crystals but they were very small (**Figure 16A**). In contrast, the crystallized mIRF-3 (184-397aa)/CBP were all needle clusters (**Figure 16B**).



~~hIRF-3 (189-427aa)~~	~~hIRF-3 (189-405aa)~~	hIRF-3 (189-398aa)	~~hIRF-3 (200-427aa)~~
~~mIRF-3 (184-419aa)~~	mIRF-3 (184-397aa)	mIRF-3 (184-390aa)	~~mIRF-3 (194-419aa)~~

Figure 15 Summary of the IRF-3 constructs crystallized.
The constructs in the rounded rectangular box were crystallized under several conditions whereas the constructs crossed by black lines indicate that no crystals were observed.



Index 88: 0.2 M ammonium citrate tribasic pH=7.0, 20% PEG 3350

Index 63: 5% v/v Tacsimate pH 7.0, 0.1 M HEPES pH 7.0, 10% w/v Polyethylene glycol monomethyl ether 5,000

Figure 16 Mouse IRF-3/CBP crystals observed under light microscope.
(A) mIRF-3 (184-390aa)/CBP were crystallized into very small single crystals under the condition of index 88. (B) mIRF-3 (184-397aa)/CBP were crystallized into needle clusters under the condition of index 63.

Since either very small single crystals or needle cluster crystals were not ideal for X-ray diffraction, I decided to optimize the conditions where these crystals were grown and tried to get large single crystals.

I optimized the protein concentrations, set up various buffers at pH gradient values, changed the precipitant concentrations, and used various salts(54). Eventually, mIRF-3 (184-390aa)/CBP crystals turned out to be big enough for X-ray diffraction (**Figure 17**).

Unfortunately, mIRF-3 (184-397aa)/CBP could not become single crystals after several rounds of optimization.

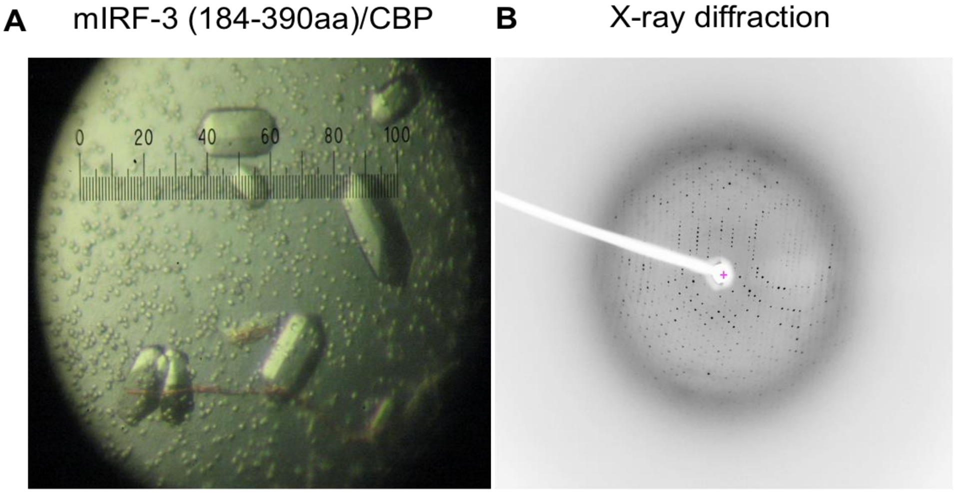


Figure 17 Crystals of mIRF3 (184-390aa)/CBP and X-ray diffraction of one crystal. (A) mIRF3 (184-390aa)/CBP were grown in the condition of 0.2 M ammonium citrate tribasic at pH 7.0, ~12% PEG 3350. (B) X-ray diffraction of one single crystal of mIRF3 (184-390aa)/CBP.

II.9. Discussion of results

All of the human and mouse IRF-3 constructs we designed were cloned into a modified pET-28a (+) vector that contains an N-terminal His$_6$-SUMO tag. The SUMO tag can promote the solubility and stability of the fusion protein due to the fact that SUMO is rapidly folded and relatively stable even when expressed at high levels in *E.coli* cells. In addition, SUMO tag can be easily and efficiently cleaved by SUMO protease which recognizes the tertiary structure of SUMO and cleaves at the C-terminal end of a conserved –Gly-Gly sequence. However, as compared to the purified human IRF-3 proteins, the mouse IRF-3 we purified by using this vector was heterogeneous. Although this problem was solved by only collecting a couple of fractions eluted from the gel-filtration chromatography, other fusion tags might be good options. For example, we can use an expression vector containing His_6-MBP-TEV for mouse IRF-3 expression and purification. MBP has been reported to promote target protein solubility by showing its chaperone intrinsic activity. In addition, MBP can be used as an affinity tag for protein purification besides His_6 tag which means that mouse IRF-3 can be purified by both nickel and amylose column, therefore leading to pure proteins. The His_6-MBP tag can be eventually removed by TEV protease that recognizes a linear peptide sequence ENLYFQS at a specific manner. As an alternative, another expression vector containing NusA-His_6-TEV can be used. Similar to SUMO, NusA has been reported to confer stability and high solubility to its target protein. Therefore, the fusion protein can be purified by nickel column followed by NusA-His_6 tag removal by TEV protease.

All of the IRF-3 proteins were phosphorylated through in-vitro TBK1 phosphorylation assay and the phosphorylated proteins were used for crystallization screen. However, it has been reported that IRF-3 can be activated through phosphorylate by either TBK1 or IKKε. Several studies indicated that both TBK1 and IKKε specifically phosphorylated Ser386 but selectively phosphorylated other residues. Therefore, it would be interesting to phosphorylate IRF-3 using IKKε in-vitro and try to obtain phosphorylated crystals by IKKε.

II.10. Materials and methods in detail

II.10.1. IRF-3 constructs preparation, protein expression and purification

The cDNA encoding human IRF-3, mouse IRF-3 dimerization domains were cloned into a modified pET-28a (+) vector containing an N-terminal His6-SUMO tag by using EcoRI and xhoI restriction sites with appropriate primers obtained from Integrated DNA Technologies (IDT). SUMO fusion of human CBP (residues 2065 to 2111) was cloned into the pET-22b (+) vector using appropriate primers from IDT. Sequences of all the constructs were confirmed by DNA Sequencing. The plasmid containing human IRF-3 or mouse IRF-3 dimerization domain was co-transformed with CBP plasmid into Escherichia coli BL21 (DE3) cells. The cells were grown on LB agar plates containing both kanamycin and ampicillin. Next day, the cell colonies from the plates were transferred to 6 liters of LB liquid medium in flasks with kanamycin and ampicillin in an incubator shaker at 37 °C under 225 rpm. When OD600 reaches ~1.2, BL21 cells were induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) overnight at 16 °C. The cells were harvested by centrifugation at 4000 rpm for 10 minutes and then suspended in a 200 mL lysis buffer containing 300 mM NaCl and 50 mM Tris·HCl (pH 8.0). The cells were lysed by sonication for 10 minutes with 0.5 sec pulse and 0.5 sec rest and the cell lysate was centrifuged at 16,000 rpm for 30 minutes. The supernatant was loaded onto a Ni2+-NTA column (Qiagen). Then a 200 mL of washing buffer containing 500 mM NaCl, 20 mM Tris·HCl, 25 mM Imidazole at pH 7.5 was used to wash non-specific binding proteins off the Ni2+-NTA column. The target proteins were then eluted with 75 mL of elution buffer containing 150 mM NaCl, 20 mM Tris·HCl, and 250 mM Imidazole at pH 7.5. The His6-

SUMO-tag was cleaved with SUMO protease at a concentration of 10 μg/mL at 4 °C overnight and removed using a Ni2+-NTA column. The target proteins in the flow-through were centrifuged to ~2 mL and further purified by gel-filtration chromatography using a HiLoad 16/60 Superdex 75 column (GE Healthcare) equilibrated with a running buffer containing 150 mM NaCl and 20 mM Tris·HCl at pH 7.5.

II.10.2. SUMO protease expression and purification

The cDNA encoding SUMO protease was cloned into a pET-28a (+) vector containing an N-terminal His6 tag. The expression and purification steps are the same as IRF-3 dimerization domains except that SUMO protease was further purified by a HiLoad 16/60 Superdex 75 column (GE Healthcare) after elution without any cleavage. 200 μl of the purified protein was aliquoted into each 500 μl tubes in the concentration of ~2 mg/mL. The tubes were flash-frozen in liquid nitrogen and stored in -80 °C.

II.10.3. IRF-3 expression and purification in M9 minimal media

On day 1, the plasmid containing mouse IRF-3 (184-397aa) was co-transformed with CBP plasmid into Escherichia coli BL21 (DE3) cells. The cells were grown on LB agar plates containing kanamycin and ampicillin. On day 2, the cell colonies from the plates were transferred to 50 mL of LB liquid medium in 125 mL flasks with kanamycin and ampicillin in an incubator shaker at 37 °C under 200 rpm overnight. On the same day, two liters of 5X pre-M9 media containing Na2HPO4, NaCl, KH2PO4 were prepared first. Then one liter of 1X M9 media was made by adding 200 mL 5X pre-M9, 768 mL Milli-Q ddH2O and 1 g NH4Cl. This mixture was autoclaved using liquid cycle 20. On day 3, filter-sterilized 20 mL of 20% D-glucose, 2 mL of 1 M MgSO4, 0.1 mL of 1 M CaCl2, 10 mL 100X Kao and Michayluk vitamin solution, 1 mL kanamycin and 1 mL ampicillin were added to the autoclaved 1X M9 media. Next, 50 mL of LB liquid medium with bacteria was centrifuged at 4000 rpm for 10 minutes. The cell pellet was resuspended in 20 mL of fresh-made M9 media. This step was repeated twice to remove any residual LB media. After two washes by M9 media, the cells were transferred to M9 media for the large scale expression. When OD600 reaches ~0.9, the shaker temperature was changed to 15 °C and BL21 cells were induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) overnight. The purification procedure was very similar to the LB media method except that all the steps were performed on ice and PMSF protease inhibitor was added to the lysis buffer to avoid protein degradation.

II.10.4. TBK1 expression and purification in sf9 insect cells

Mouse TBK1 (mTBK1) was cloned into the pAcGHLTc vector with an N-terminal GST tag and a His6 tag. The 2 μg plasmid was transfected together with 2.5 μL Baculo-Gold bright linearized baculovirus DNA (BD Biosciences) into sf9 insect cells to generate recombinant baculovirus. The recombinant viruses were amplified for at least two rounds (4-6 days/round) before the large-scale protein expression. The insect cells at a density of 2.5 × 106 cells/mL were infected with the TBK1 recombinant baculovirus and cultured at 27°C and harvested 72 hours post infection by centrifugation at 3000 rpm for 10 minutes. The cells were lysed in a buffer containing 150 mM NaCl, 0.2 M Tris·HCl, 1% NP-40, 1 mM PMSF at pH 8.0 in a shaker at 4 °C for 2 hours. The cell lysate was centrifuged at 16,000 rpm for 30 minutes. The GST-TBK1 protein in the supernatant was mixed with 6 mL Ni2+-NTA beads and incubated in a shaker at 4 °C for 2 hours. The beads were then spun down and washed three times using a buffer containing 500 mM NaCl, 20 mM Tris·HCl, 25 mM Imidazole at pH 7.5. The target protein was eluted with a buffer containing 150 mM NaCl, 20 mM Tris·HCl, and 250 mM Imidazole at pH 7.5. The eluted protein was further purified by gel-filtration chromatography using a HiLoad 16/60 Superdex 200 column.

II.10.5. IRF-3 phosphorylation by in-vitro TBK1 assay

Purified human and mouse IRF-3 proteins were mixed with GST-mTBK1 in a ratio of 10:1 (w/w) in a 1 mL reaction buffer with 20 mM HEPES pH 7.5, 10 mM MgCl2, 100 mM NaCl, 5 mM ATP, 0.1 mM Na3VO4, 5 mM NaF, 5 mM DTT at 27 °C for ~24 hours. The final concentration of the proteins was about 1 mg/mL. After ~24-hour incubation, the phosphorylated IRF-3 proteins were further purified using a HiLoad 16/60 Superdex 200 column eluted with a buffer containing 20 mM Tris·HCl and 150 mM NaCl at pH 7.5.

II.10.6. Crystallization screen and optimization of crystallization conditions

The crystallization screen was performed by hanging drop vapor diffusion technique at 4 °C using Index, Crystal Screen and Crystal Screen 2 reagent kits from Hampton Research. Purified phosphorylated IRF-3 proteins were concentrated to ~5 mg/mL. 500 μL of crystallization reagent/condition was pipetted into each reservoir of a VDX plate. Then 2µL of proteins were pipetted into the center of a siliconized 22mm circle cover slide. 2µL of reagent from each reservoir was pipetted into the drop on the cover slide containing the proteins. The cover slide with the mixture drop was then inverted and positioned onto the bead of grease on each reservoir. The plates were kept on shelf in the cold room and examined after one week. The mouse crystals grown at several conditions were then optimized by changing protein concentrations (2.5 mg/mL, 7.5 mg/mL, 10 mg/mL), adjusting precipitant concentrations (18% PEG 3350, 15% PEG 3350, 12% PEG 3350, 10% PEG 3350), using different buffer concentrations at various pH gradient (0.25 M ammonium citrate tribasic at pH 7.0, 0.25 M ammonium citrate tribasic at pH 7.5, 0.25 M ammonium citrate tribasic at pH 6.5, 0.15 M ammonium citrate tribasic at pH 7.0, 0.15 M ammonium citrate tribasic at pH 7.5, 0.15 M ammonium citrate tribasic at pH 6.5). Crystals of mouse IRF-3 in complex with CBP were grown in 0.2 M ammonium citrate tribasic at pH 7.0 with ~12% PEG 3350 after several rounds of careful optimization.

CHAPTER III. THE MOLECULAR BASIS OF IRF-3 ACTIVATION UPON PHOSPHORYLATION

III.1. Introduction

Type I IFNs play very important roles in antiviral immunity and several signaling pathways have been reported to induce the expression of Type I IFNs through the activation of IRF-3 by TBK1 or IKK phosphorylation. Both our lab and Zhijian Chen lab found that phosphorylation of the pLxIS motif located within three adaptor proteins STING, MAVS and TRIF are required for the recruitment of IRF-3. In addition, IRF-3 itself has a pLxIS motif that mediates its activation. Specifically, phosphorylation of Ser366 in the pLxIS motif of STING is critical for IRF-3 recruitment and mutation of this residue to alanine abolished IRF-3 binding and IFN-β reporter activation. The structure of phosphorylated STING C-terminal tail bound to IRF-3 C-terminal domain indicated that phosphorylated Ser366 interacts with a cluster of positively charged residues including Arg285, His288, His290, and Lys313 through electrostatic interactions. Besides, Pro361, Leu363, and I365 upstream of pSer366 reach into a hydrophobic groove of IRF-3 and interact with it through hydrophobic interactions. The structure of phosphomimetic double mutant S386/396E IRF-3/CBP dimer has demonstrated that phosphomimetic pLxIS motif interacts with a neighboring IRF-3 in a manner similar to that of the phosphorylated pLxIS motif of STING. The structure of phosphorylated MAVS and TRIF bound to IRF-3 also indicated that The pLxIS motifs of pMAVS and pTRIF interact with the same ligand-binding surface on IRF-3 as pSTING. In this chapter, we want to investigate the mechanism of IRF-3 activation upon phosphorylation by taking a close look at the real phosphorylated mouse and human structures; by comparing and contrasting the phosphorylated IRF-3 structures with the previously solved autoinhibited structure and phosphomimetic double mutant; by using biochemical studies together with functional studies.

III.2. The structure of phosphorylated mouse IRF-3/CBP

As we mentioned in the introduction, in order to investigate how phosphorylation activates IRF-3, we expressed, purified mouse IRF-3 (residues 184 to 390) /CBP complex and phosphorylated it using TBK1. Figure 18A, shows that phosphorylated mouse IRF-3 shifted to the left after 24h incubation with TBK1 meaning that it becomes larger due to phosphorylation indicating that it may not be a monomer (**Figure 18A**). The two samples in Figure 18A were collected and then resolved on the gradient gel, which indicated that the phosphorylated protein was pure enough for crystallization (**Figure 18B**). After several rounds of optimization of crystals, we got big, single mouse IRF-3/CBP complex (pmIRF-3/CBP) crystals. It was sent to the advanced light source (ALS) located at Berkeley, CA to collect the X-ray diffraction data. Then we solved phosphorylated IRF-3/CBP complex structure at 2.4 Å resolution (**Figure 18C and Table1**). The IRF-3/CBP structure showed that only Ser379 was phosphorylated (Figure 18C). The samples with and without phosphorylation in Figure 18A were also sent out for MALDI-TOF Mass spectrometry (MS) analysis and the results showed that only one residue of IRF-3 is phosphorylated, which is consistent with our structure (**Figure 18D**).

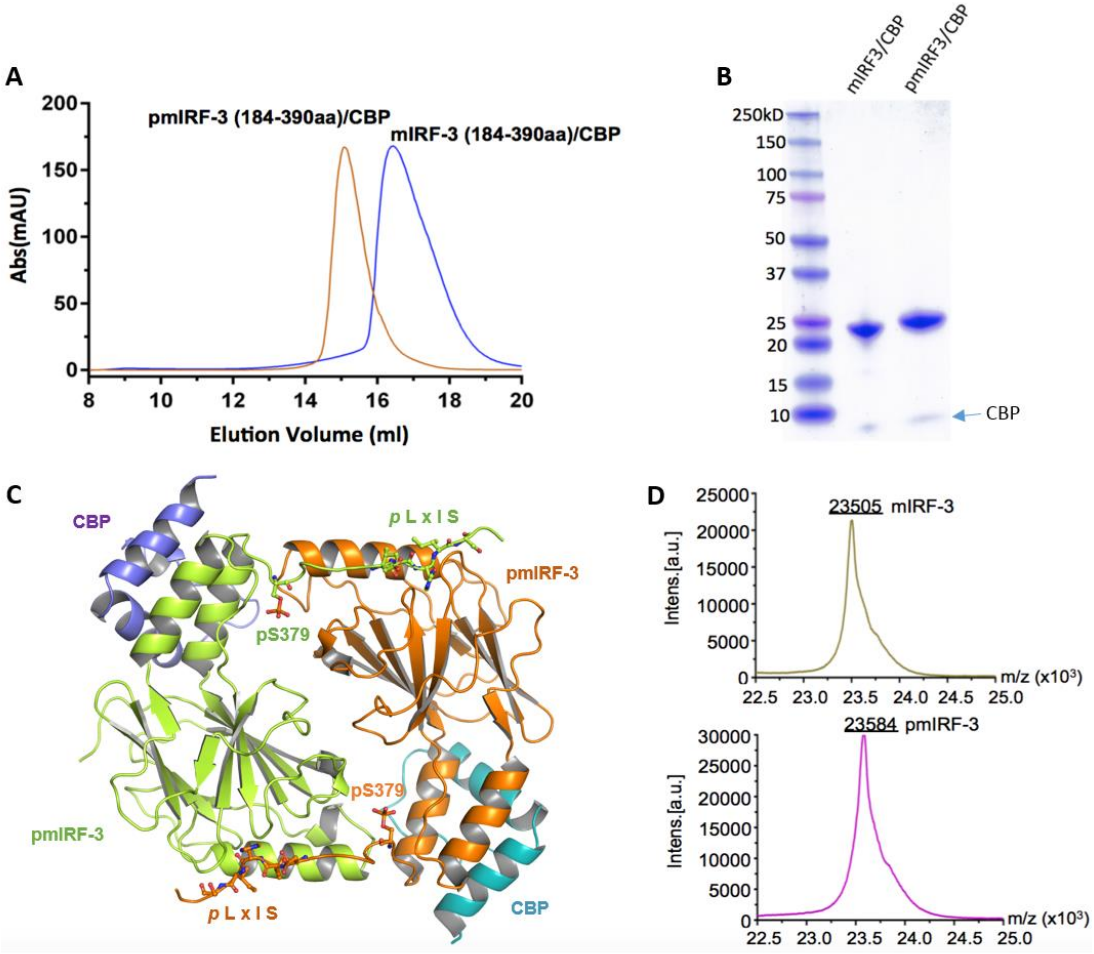


Figure 18 Characterization and structure of phosphorylated mouse IRF-3 bound to CBP.
(A) Size-exclusion chromatography showing the mouse IRF-3 (residues 184 to 390) /CBP complex with and without TBK1 phosphorylation. The phosphorylated complex shifts to the left. (B) Gradient gel showing purified mIRF-3/CBP and pmIRF-3/CBP. (C) The structure of phosphorylated mouse IRF-3/CBP dimer. Phosphorylated IRF-3 are colored lime and orange. CBP are in purple and teal. (D) MS analyses of mouse IRF-3 (residues 184-390) before and after TBK1 phosphorylation.

Overall, phosphorylated IRF-3/CBP (pmIRF-3/CBP) complex forms a dimer mainly through phosphorylated Ser379 and the downstream DLHIS sequence (known as pLxIS motif) (**Figure 19**).

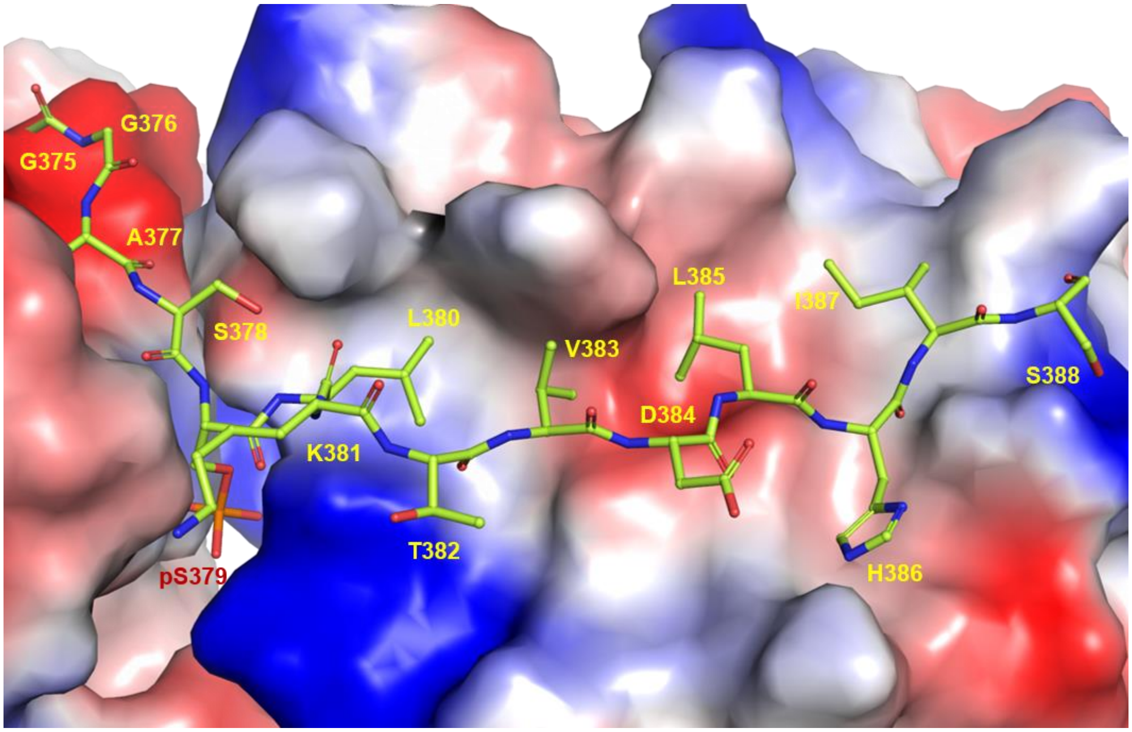


Figure 19 Structure of mouse IRF-3 C-terminal region containing phosphorylated Ser379 and the pLxIS motif (DLHIS sequence).
One IRF-3 molecule is shown by limon ball-and-stick model. The other IRF-3 in the IRF-3 dimer is shown by the surface representation with positively charged and negatively charged surfaces in blue and red, respectively.

Specifically, pSer379 reaches into a positively charged pocket containing residues Arg373 from the same IRF-3 molecule and Arg205 from the other IRF-3 and interacts with them via electrostatic interactions (**Figure 20**). In addition, Asp247 of the other IRF-3 molecule and Ser378 upstream pSer379 also interacts with pSer379 via hydrogen bonds. Moreover, Ser332 forms a hydrogen bond with pSer379 through its side chain hydroxyl group. Based on these structural analyses, it is obvious that Arg373, Arg205, Ser332 play key roles in promoting mouse IRF-3 dimerization by interacting with pSer379 through electrostatic interactions and hydrogen bonds (**Figure 20**).

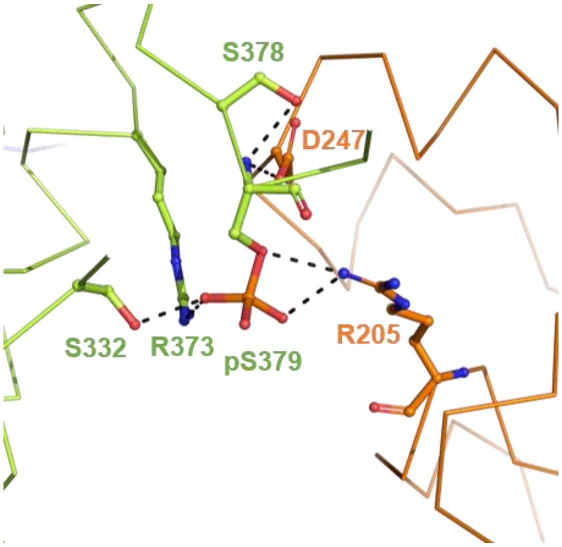


Figure 20 The mouse structure showing the interactions between phosphorylated Ser379 and the surrounding residues.

The dimer interface also entails the tail-mediated interactions and a central core region (**Figure 21**). Specifically, the tail containing residues L380, V383, L385, I387 reaches into a hydrophobic groove of another IRF-3 surface, interacting with V250, Y253, L355, V257, L261, C282 through hydrophobic interactions. Moreover, the central core region of the dimer is stabilized by F203, L292, L293, P350, W351, V352 residues via hydrophobic interactions and hydrogen bonds. (**Figure 21A to C**).

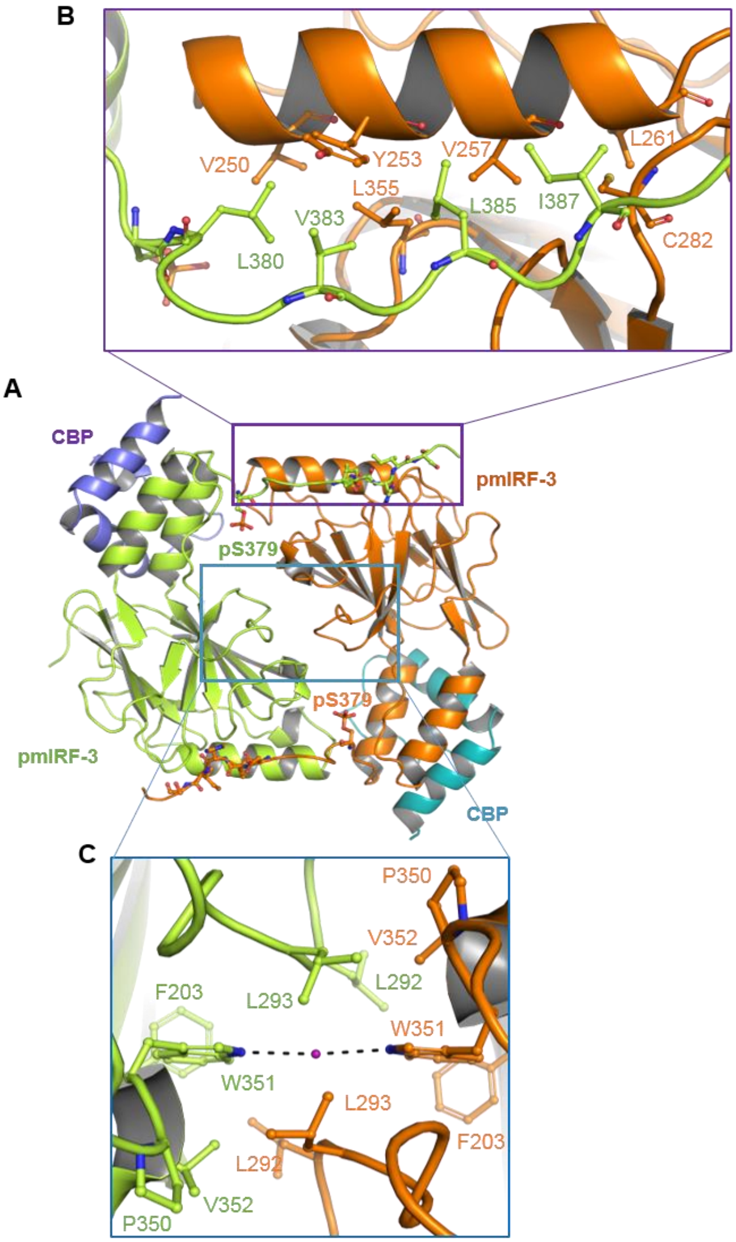


Figure 21 Intermolecular interactions between the phosphorylated mouse IRF-3 dimer. (A) Overall structure of mouse IRF-3/CBP dimer. (B) The tail of IRF-3 mediating the interactions with another IRF-3 molecule through hydrophobic interactions. (C) The core region of IRF-3 stabilized by hydrophobic interactions and hydrogen bonds.

III.3. Comparison of mouse phosphorylated IRF-3/CBP and human phosphorylated IRF-3/CBP complexes

Based on the structure of phosphorylated mouse IRF-3/CBP complex, the C-terminal tail of IRF-3 including Ser379 mainly mediates the IRF-3 dimerization upon phosphorylation. Since the post-doc Baoyu got the human IRF-3 (residues 189-398)/CBP crystal and structure, I wanted to compare both mouse and human structures. For the human IRF-3/CBP structure, I noticed that there is only one phosphorylated residue which is pSer386. This phosphorylated residue corresponded to phosphorylated Ser379 in mouse. I also sent both phosphorylated and non-phosphorylated samples out for Mass Spectrometry and the result supported the structure indicating that there is only one site that has been phosphorylated (**Figure 22**).

The overall structures of the pmIRF-3/CBP and phIRF-3/CBP complexes are similar (r.m.s.d. 1.3 Å). We observed that Ser386, which corresponds to Ser379 of mouse IRF-3, is phosphorylated in the structure. Besides, the C-terminal tail containing pLxIS motif reaches into another molecule of human IRF-3 (**Figure 23**). The two phosphorylated sites pSer379 and pSer386 were then closed examined. Generally, there are a lot more interactions between pSer386 and its surrounding residues as compared to pSer379 in the mouse structure (**Figure 24A and B**). Specifically, pSer386 reaches into a highly positively charged pocket surrounded by residues Arg211, Arg380, Arg341 and Lys360 and interacts with these residues via electrostatic interactions and an extensive network of hydrogen bonds (**Figure 24B**). pSer386 interacts with Arg211 from another IRF-3 molecule through a network of three hydrogen bonds.

In addition, Arg211 also interacts with Lys360 and Glu388 through hydrogen bonds, thus making critical contribution to the formation of IRF-3 dimer. Arg380 forms a hydrogen bond with pSer386 within the same IRF-3 molecule via its side chain guanidinium group. A water molecule forms a network of three hydrogen bonds with Arg380, pSer386 and Asp254, making additional contributions to the interactions between pSer386 and Arg380. In addition, Asp254 of the other IRF-3 molecule interacts with Ser385 via two hydrogen bonds. Arg341 is within 4.0 Å from pSer386 and interacts with pSer386 through electrostatic interactions. Moreover, Arg341 also interacts with the phosphate group of pSer386 via a solvent mediated hydrogen bond. Similarly, the sidechain of Lys360 from the other IRF-3 in the dimer is within 4.0 Å from the phosphate group and interacts with pSer386 via electrostatic interaction and a solvent mediated hydrogen bond. In addition, Ser339 forms a hydrogen bond with pSer386 through its side chain hydroxyl group. Ser339 also interacts with the phosphate group of pSer386 via a solvent mediated hydrogen bond through its main-chain amine group. By contrast, pSer379 only interacts with residues Arg373 and Arg205 via electrostatic interactions in the mouse structure (**Figure 24A**). Asp247 of the other IRF-3 molecule and Ser378 upstream pSer379 also interacts with pSer379 via hydrogen bonds. Moreover, Ser332 forms a hydrogen bond with pSer379 through its side chain hydroxyl group. The superposition of the two structures clearly revealed that Arg373 is structurally conserved and interacts with pSer379 in a similar fashion as Arg380 in phIRF-3. The side chain of Arg205 adopts a slightly different conformation compared to Arg211 of phIRF-3 and interacts with pSer379 through electrostatic interaction and hydrogen bonds. The interactions between Asp247 and pSer379, Ser378 are also well preserved in both mouse and human IRF-3. By contrast, the sidechains of Arg334 and Lys353 move away from pSer379 and do not interact with pSer379 directly (**Figure 25**).

Based on these structural analyses, it is obvious that Arg373, Arg205 and Ser332 in mouse IRF-3 (Arg380, Arg211, Ser339 in human IRF-3) play key roles in promoting IRF-3 dimerization by interacting with pSer379 (pSer386 in human) through electrostatic interactions and hydrogen bonds (**Figure 24 and Figure 25**).

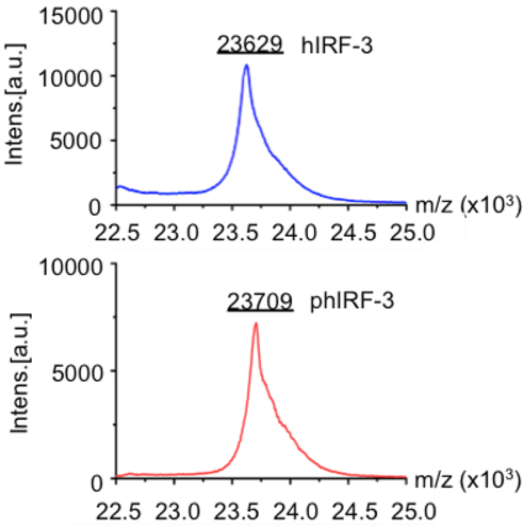


Figure 22 Mass spectrometry analyses of human IRF-3 (residues 189-398) with and without TBK1 phosphorylation.

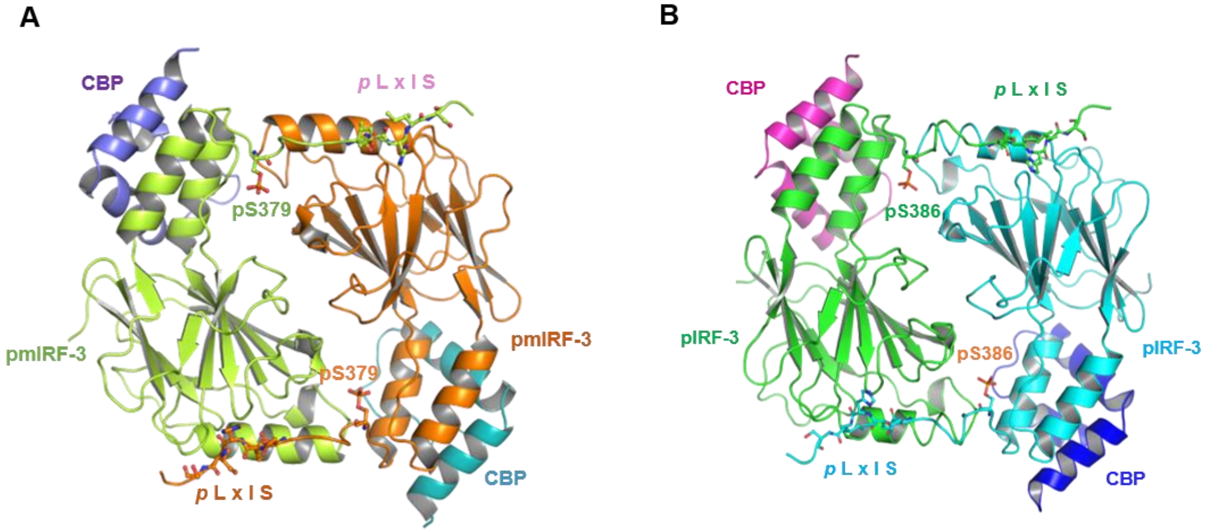


Figure 23 Comparison of structures of phosphorylated mouse and human IRF-3 bound to CBP.

(A) The structure of phosphorylated mouse IRF-3/CBP dimer. Phosphorylated IRF-3 are colored lime and orange. CBP are in purple and teal. Phosphorylated Ser379 and residues of the pLxIS motif are indicated by ball-and-stick models (B) The structure of phosphorylated human IRF-3/CBP dimer. Phosphorylated Ser386 and residues of the pLxIS motif are indicated by ball-and-stick models. IRF-3 are shown as green and cyan ribbons. CBP are shown by magenta and blue ribbons.

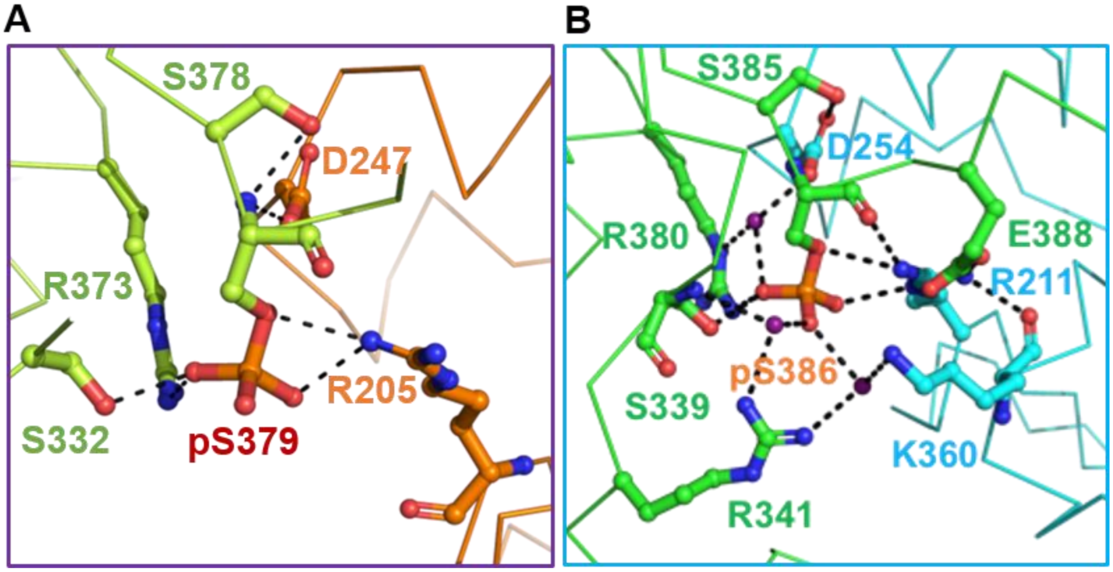


Figure 24 Comparison of pSer379 and pSer386 in mouse and human structures.
(A) Interactions between pSer379 and surrounding residues in phosphorylated mouse IRF-3/CBP structure. (B) Interactions between pSer386 and surrounding residues in phosphorylated human IRF-3/CBP structure.

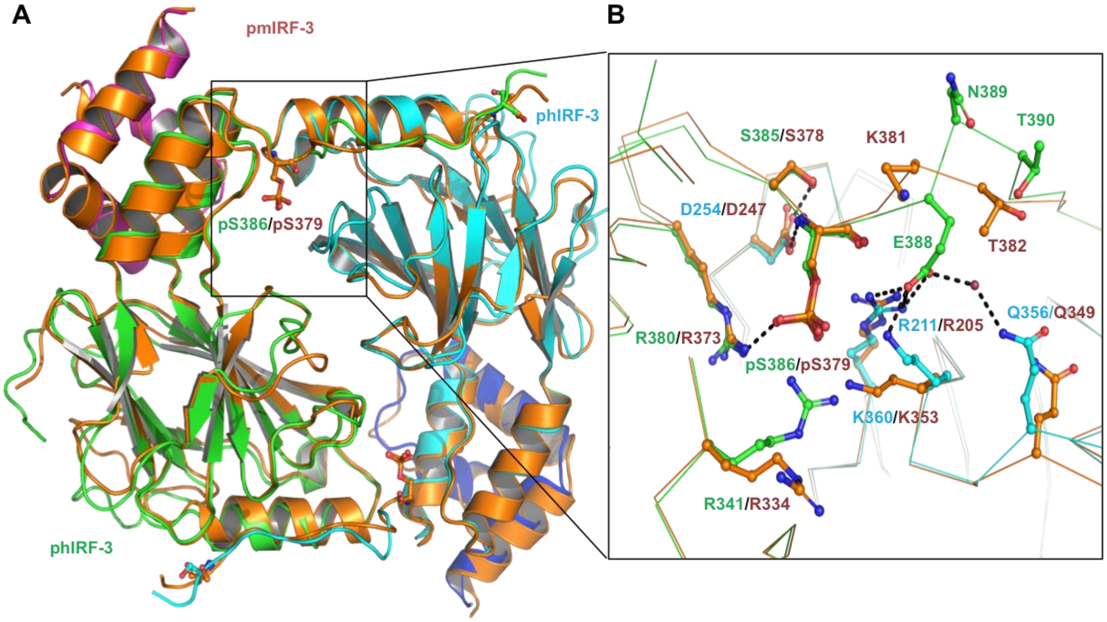


Figure 25 Superposition of phosphorylated mouse and human IRF-3/CBP complexes.
(A) The superposition of structures of phosphorylated mouse and human IRF-3/CBP complexes. Phosphorylated mouse IRF-3/CBP is shown by the orange ribbon. Human IRF-3 dimer is colored in green and cyan with CBP bound to hIRF-3 in magenta and blue. Phosphorylated Ser379 of mIRF-3 and Ser386 of hIRF-3 are shown by the ball-and-stick models. (B) Distinct interactions between phosphorylated mouse and human IRF-3. Key residues mediating human and mouse IRF-3 dimerization are colored in green and orange, respectively. Residues interacting with Glu388 of hIRF-3 are in cyan.

To our surprise, due to the replacement of Glu388 and Asn389 of human IRF-3 by Lys381, this region of phosphorylated mouse IRF-3 is restructured. Unlike Glu388 in human IRF-3, which contributes to hIRF-3 activation by interacting with Arg211, Lys360 and Gln356 through electrostatic interactions and the solvent-mediated hydrogen bond, Lys381 of mIRF-3 flips into the solvent and does not interact with any residues nearby (**Figure 25**). This discrepancy between mouse and human structures may indicate that Glu388 may play additional roles in human IRF-3 activation. Indeed, the luciferase reporter assay showed that mutating Glu388 to alanine or serine in human IRF-3 reduced the IFN-β reporter signal by about 45% or 40% respectively (**Figure 26A and B**). Furthermore, phosphorylated E388A and E388S mutants both showed two bands on native gel compared to phosphorylated wild-type IRF-3, indicating a mixture of monomer and dimer (**Figure 27A and B**). In agreement with these results, SEC shows that both E388A and E388S mutants of phIRF-3 eluted as two peaks, demonstrating that the dimerization of hIRF-3 was compromised by those two mutations (**Figure 27C and D**). Taken together, these structural and functional analyses reveal that mouse IRF-3 is activated in a similar but distinct manner compared to human IRF-3.

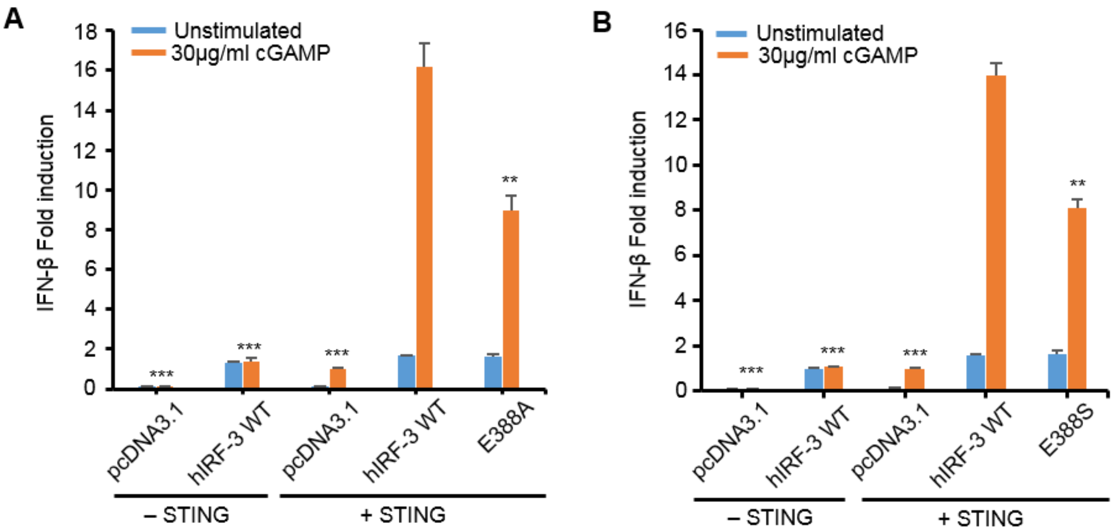


Figure 26 IFN-β luciferase reporter assays showing the effect of E388A and E388S mutations on the signaling mediated by hIRF-3.

(A) Luciferase assay showing the effect of E388A on the IRF-3 mediated signaling. (B) Luciferase assay showing the effect of E388S on the IRF-3 mediated signaling. The data are mean ± s.e.m. and representative of three independent assays. **P < 0.01, ***P < 0.001 values were calculated by comparisons of signals in cells transfected with E388A, E388S mutants and those transfected with wild-type IRF-3.

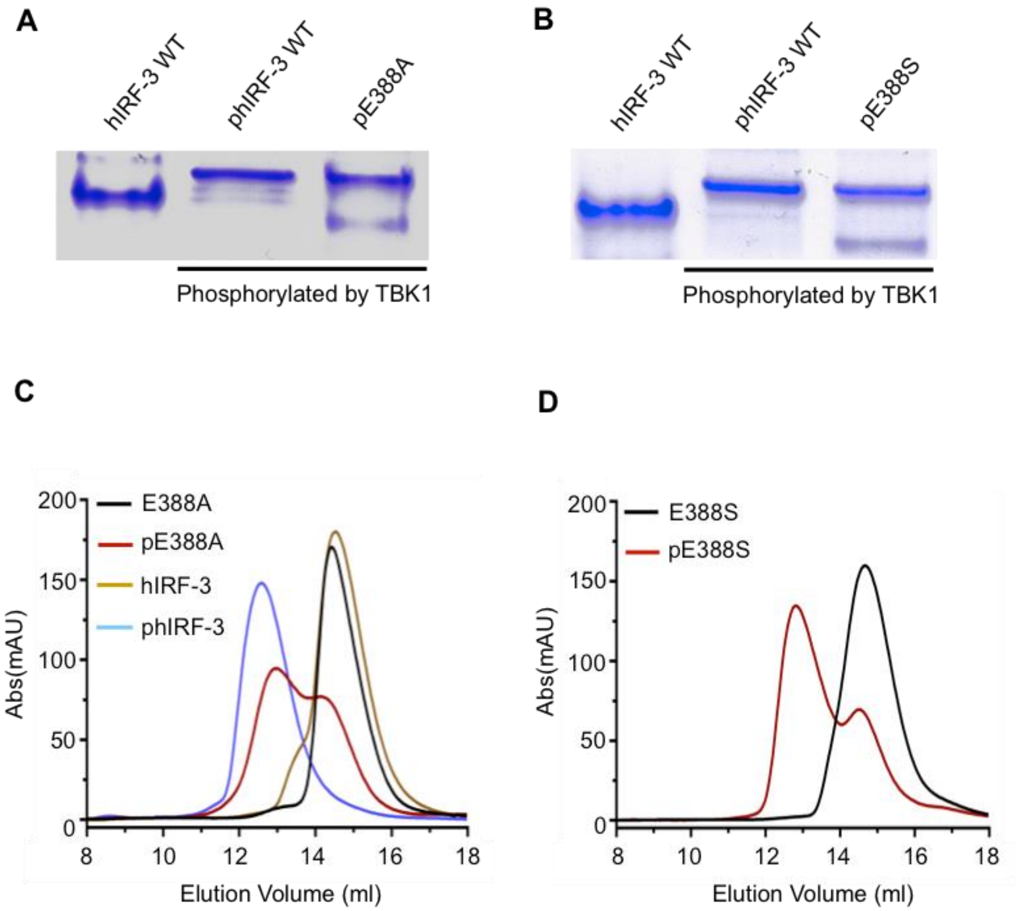


Figure 27 Native gel electrophoreses and Size-exclusion chromatography analyses of E388A and E388S mutations upon phosphorylation.

(A and B) Native gel electrophoresis showing the dimerization of wild-type human IRF-3, E388A and E388S mutants upon phosphorylation. (C and D) Size-exclusion chromatography showing the effect of E388A and E388S mutations on the dimerization of hIRF-3 upon phosphorylation as compared to wild-type IRF-3.

III.4. Comparison of phosphorylated human IRF-3/CBP complex and auto-inhibited IRF-3 (PDB: 1QWT) structure

Previously, Qin group determined the structure of the IRF-3 transactivation domain which suggested an autoinhibition and virus induced phosphoactivation mechanism. In order to clearly see the reorganization of IRF-3 structure upon phosphorylation, I superimposed the human phosphorylated IRF-3/CBP complex and auto-inhibited IRF-3 (PDB: 1QWT) structure determined by Qin group. The superposition structure reveals a dramatic conformational change of the C-terminal tail upon phosphorylation. In auto-inhibited IRF-3, the C-terminal tail is folded and blocks the binding of CBP whereas in the phosphorylated IRF-3/CBP complex, the C-terminal tail of IRF-3 containing the pLxIS motif unfolds and extends into another IRF-3 molecule mediating the formation of IRF-3 dimer through pSer386 and the pLxIS motif. The two red arrows indicated that buried Ser386 and pLxIS motif in the auto-inhibited IRF-3 are released and extends to the other IRF-3 molecule upon phosphorylation (**Figure 28A and B**).

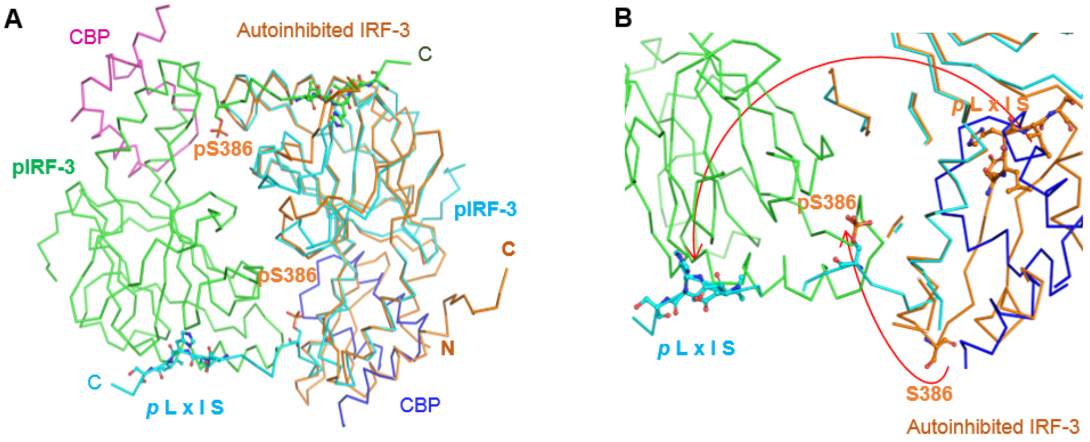


Figure 28 Conformational changes of IRF-3 upon phosphorylation.
(A and B) Superposition of the phosphorylated human IRF-3/CBP and auto-inhibited IRF-3 (PDB: 1QWT) showing the conformational changes of IRF-3 upon phosphorylation. The IRF-3/CBP dimer is colored the same as in Figure 23B. Auto-inhibited IRF-3 is colored orange.

III.5. Comparison of phosphomimetic S386/396E IRF-3/CBP dimer and phosphorylated IRF-3/CBP dimer

Previously, our lab also determined the phosphomimetic double mutant S386/396E IRF-3/CBP dimer. Although the overall structure of the phosphomimetic S386/396E IRF-3/CBP dimer is similar to the phIRF-3/CBP dimer (**Figure 29A**), Glu386, which mimics pSer386, contributes much less significantly to IRF-3 dimerization (**Figure 29B**). Similar to the phIRF-3/CBP dimer, Arg380 is less than 4 Å from Glu386 and likely interacts with each other via electrostatic interactions. By contrast, the closest distance between the sidechains of Arg211 and Glu386 is over 5 Å and Arg211 forms no hydrogen bonds directly with Glu386. Instead, Arg211 stabilizes the phosphomimetic IRF-3 dimer mainly through its interaction with the sidechain of Glu388. In addition, Arg341 and Lys360 are further away from Glu386 and do not interact with Glu386 directly (**Figure 29B**). Moreover, Lys360, Gln356 move away from Glu388 and don't interact with Glu388 in the phosphomimetic dimer whereas these two residues are involved in the interaction with Glu388 through solvent-mediated hydrogen bond and electrostatic interaction (**Figure 30A and B**). Based on these structural comparisons, the phosphomimetic dimer does not fully recapitulate the extensive intermolecular interactions observed in the phIRF-3/CBP dimer, explaining why phosphorylated IRF-3 dimer is more stable compared to the phosphomimetic dimer which was previously demonstrated by binding affinities measured using sedimentation equilibrium ultracentrifugation analysis (Kd : 167 nM for phosphorylated IRF-3; Kd : 1.26 μM for phosphomimetic dimer).

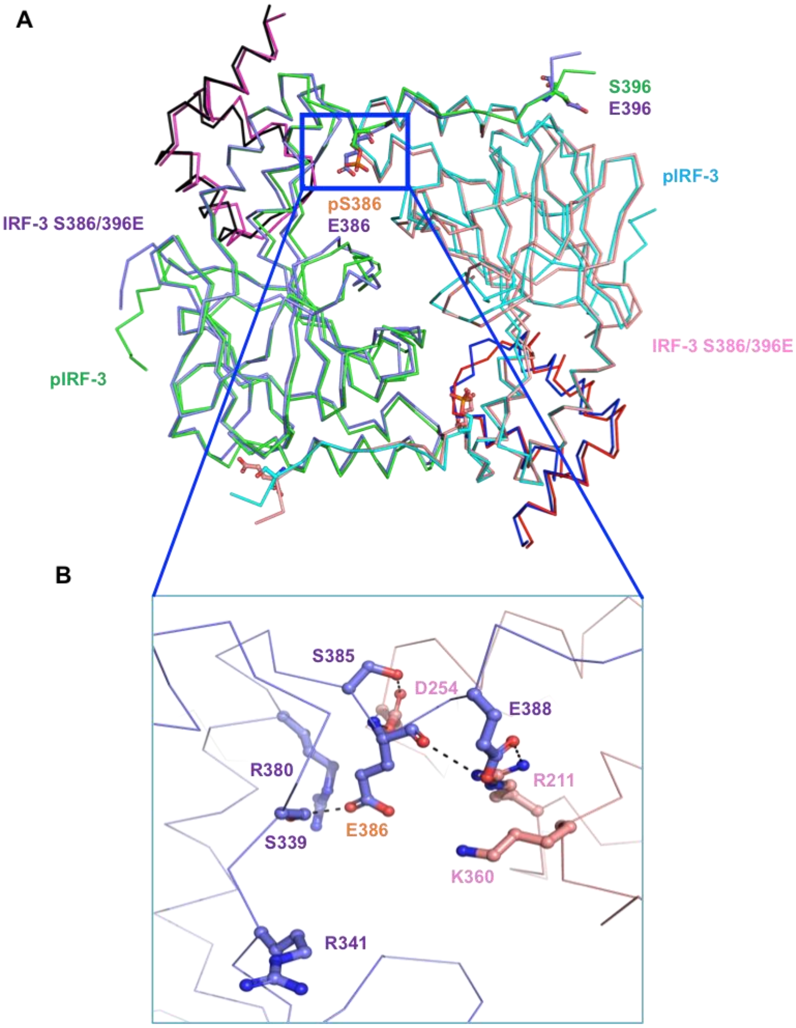


Figure 29 Comparison of structures of phosphorylated human IRF-3 and its S386/396E mutant bound to CBP.
(A) Superposition of structures of phosphorylated human IRF-3/CBP complex and phosphomimetic S386/396E mutant bound to CBP (PDB: 5JEM). Phosphorylated IRF-3 are in green and cyan. The phosphomimetic mutant are in purple and pink. (B) Interactions between Glu386 and surrounding residues in the phosphomimetic S386/396E IRF-3 dimer (PDB: 5JEM).

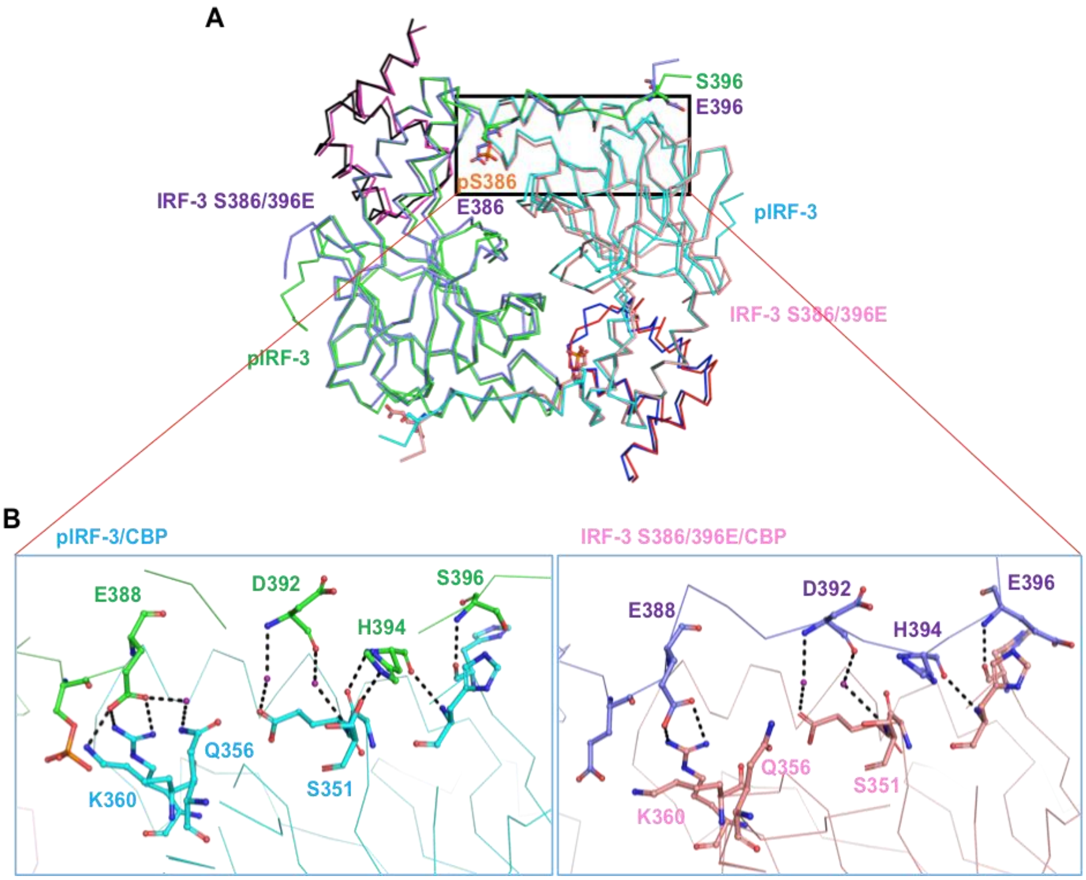


Figure 30 Comparison of the interactions of the C-terminal tail between phosphorylated human IRF-3 and its S386/396E double mutant bound to CBP.
(A) Superposition of phIRF-3/CBP and S386/396E mutant bound to CBP (PDB: 5JEM). (B) Comparison of intermolecular interactions downstream of E386 and pS386 within phosphorylated and phosphomimetic IRF-3 dimers.

III.6. Mimicry of phosphorylated Ser396 by phosphorylated Ser366 of pSTING/IRF-3

It has been reported that in addition to Ser386, Ser396 within the pLxIS motif in human IRF-3 is also involved in IRF-3 activation and can be phosphorylated by TBK1(20, 41, 43, 44, 55, 56). Our western blot indeed showed that both Ser386 and Ser396 have been phosphorylated upon cGAMP stimulation in HEK293T cells transfected with wild-type IRF-3 whereas no phosphorylation on either Ser386 or Ser396 site has been detected with no cGAMP treatment in those cells transfected with wild-type IRF-3. As controls, the cells transfected with the pcDNA3.1 vector did not show the phosphorylated bands on both Ser386 and Ser396 sites whether they were treated with cGAMP or not (**Figure 31**). In the phIRF-3/CBP complex structure, the electron density for Ser396 and surrounding residues was well defined. However, we did not observe the phosphorylation of Ser396 in the structure, most likely due to the truncation at residue Ser398 that prevents the phosphorylation of Ser396 by TBK1.

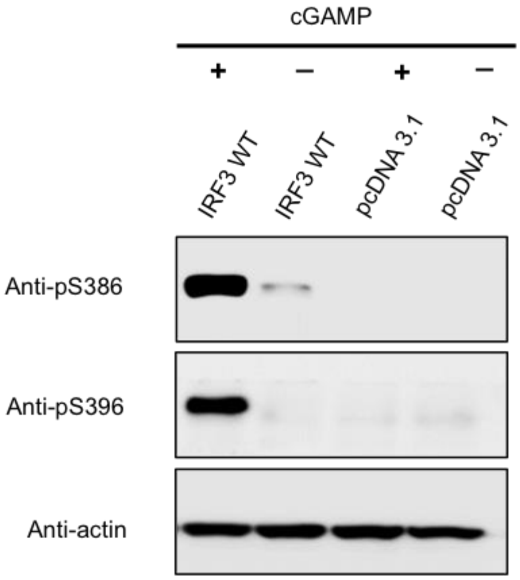


Figure 31 Western blot showing the phosphorylation of Ser386 and Ser396 in HEK293T cells transfected with wild-type IRF-3.

To investigate how phosphorylation of Ser396 contributes to IRF-3 activation, we superimposed the structure of pSTING/IRF-3 (PDB: 5JEJ) complex over the pIRF-3/CBP complex structure (**Figure 32**). The pLxIS motif of the phosphorylated STING is well aligned with the pIRF-3 pLxIS motif. Thus, it is likely that phosphorylated Ser396 could reach into the positively charged pocket surrounded by Arg285, His288, His290 and Lys313 and interact with them through electrostatic interactions in a similar fashion as pSer366 of pSTING.

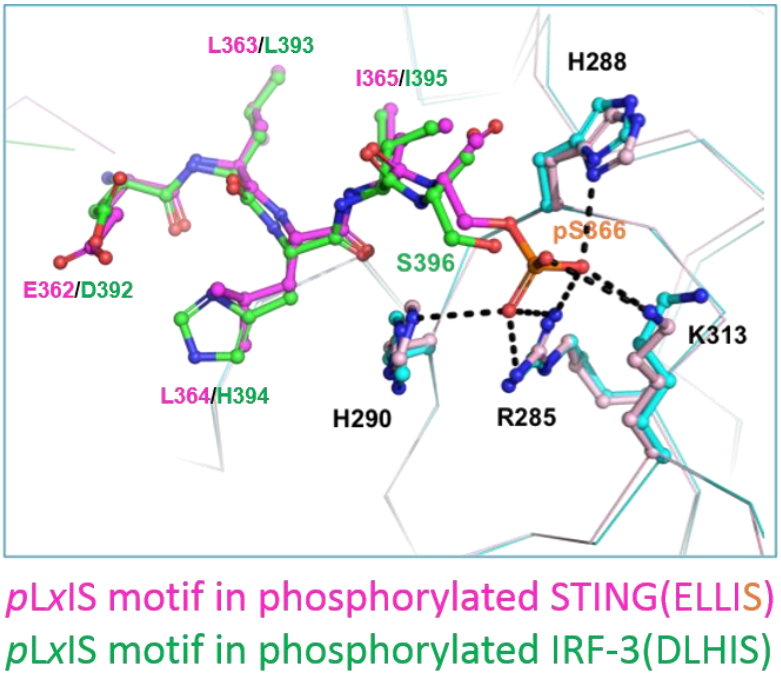


Figure 32 Superposition of the phIRF-3/CBP dimer and pSTING/IRF-3 complex structures (PDB: 5JEJ).
The green and cyan colored ribbons represent IRF-3 in the IRF-3/CBP dimer. The magenta ribbon represents phosphorylated STING and the pink colored ribbon indicates IRF-3 in pSTING/IRF-3 complex.

III.7. Distinct roles of Ser386 and Ser396 in IRF-3 activation

Since both Ser386 and Ser396 can be phosphorylated in cells, in order to distinguish the roles of Ser386 and Ser396 in IRF-3 activation, we expressed and purified both S386A and S396A mutants of human IRF-3 (**Figure 33A**), phosphorylated them by TBK1, and analyzed them by native PAGE. Interestingly, the phosphorylated S386A mutant showed a single lower band, which is indicative of a monomer whereas the phosphorylated S396A mutant exhibited two bands indicating a mixture of both monomer and dimer (**Figure 33B**). Consistent with the native gel result, size-exclusion chromatography (SEC) showed that phosphorylated S386A mutant was eluted at the same position as un-phosphorylated S386A and wild type IRF-3, while the phosphorylated S396A mutant showed two peaks, which correspond to a mixture of IRF-3 monomer and dimer (**Figure 33C and D**). These results demonstrate that both Ser386 and Ser396 are involved in IRF-3 dimerization but Ser386 plays a more important role in IRF-3 activation compared to Ser396.

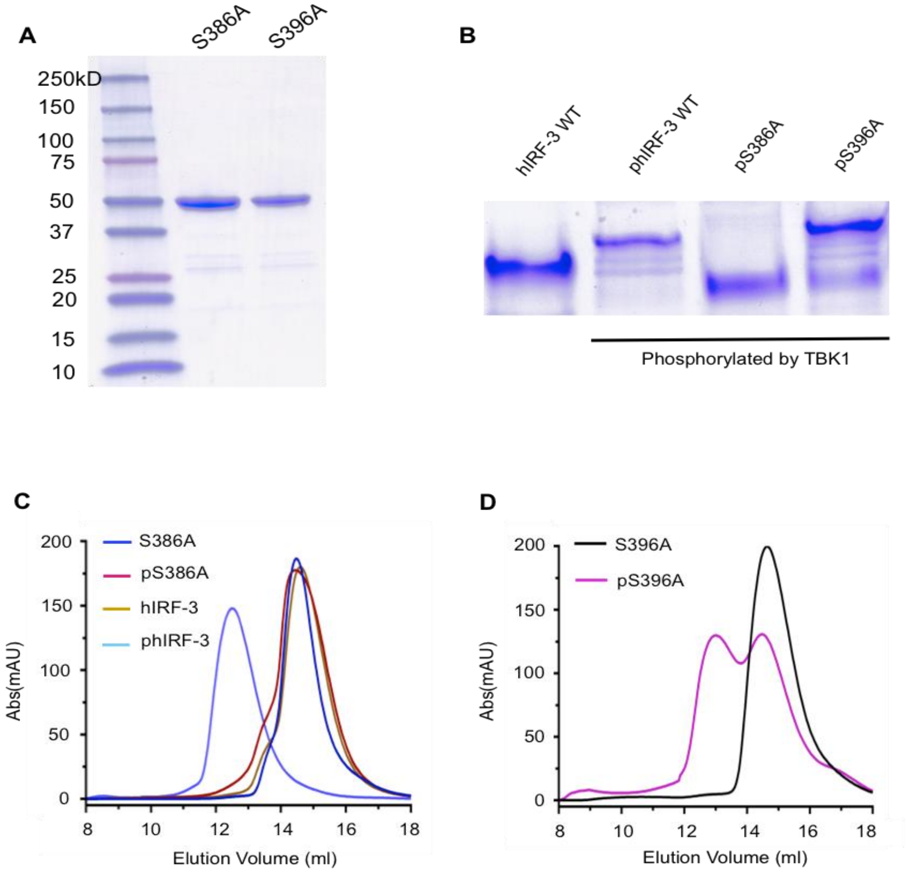


Figure 33 Distinct roles of Ser386 and Ser396 in IRF-3 activation.
(A) Gradient gel electrophoreses of purified S386A and S396A mutants (B) Native gel electrophoresis showing the dimerization state of wild-type IRF-3 and its S386A and S396A mutants upon phosphorylation by TBK1. (C) Size-exclusion chromatography showing how mutation S386A affects the dimerization of phosphorylated IRF-3 as compared to wild-type IRF-3. (D) Size-exclusion chromatography showing how mutation S396A affects the dimerization of phosphorylated IRF-3.

To further explore how these two residues affect IRF-3 mediated signaling, we conducted IFN-β luciferase reporter assays in cells transfected with STING and IRF-3. We observed that the S386A mutation blocked the IFN-β reporter activation and the S396A mutation reduced the reporter signal by about 50%, demonstrating that both Ser386 and Ser396 are involved in IRF-3 mediated signaling but Ser386 is more crucial (**Figure 34**).

Altogether, these extensive structural and functional studies provide critical insights into the detailed mechanism of IRF-3 activation upon phosphorylation.

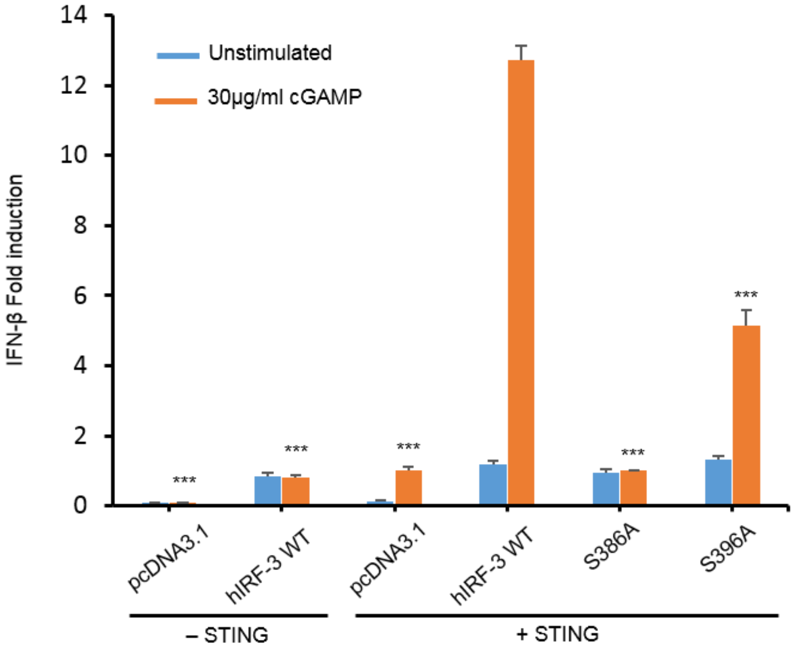
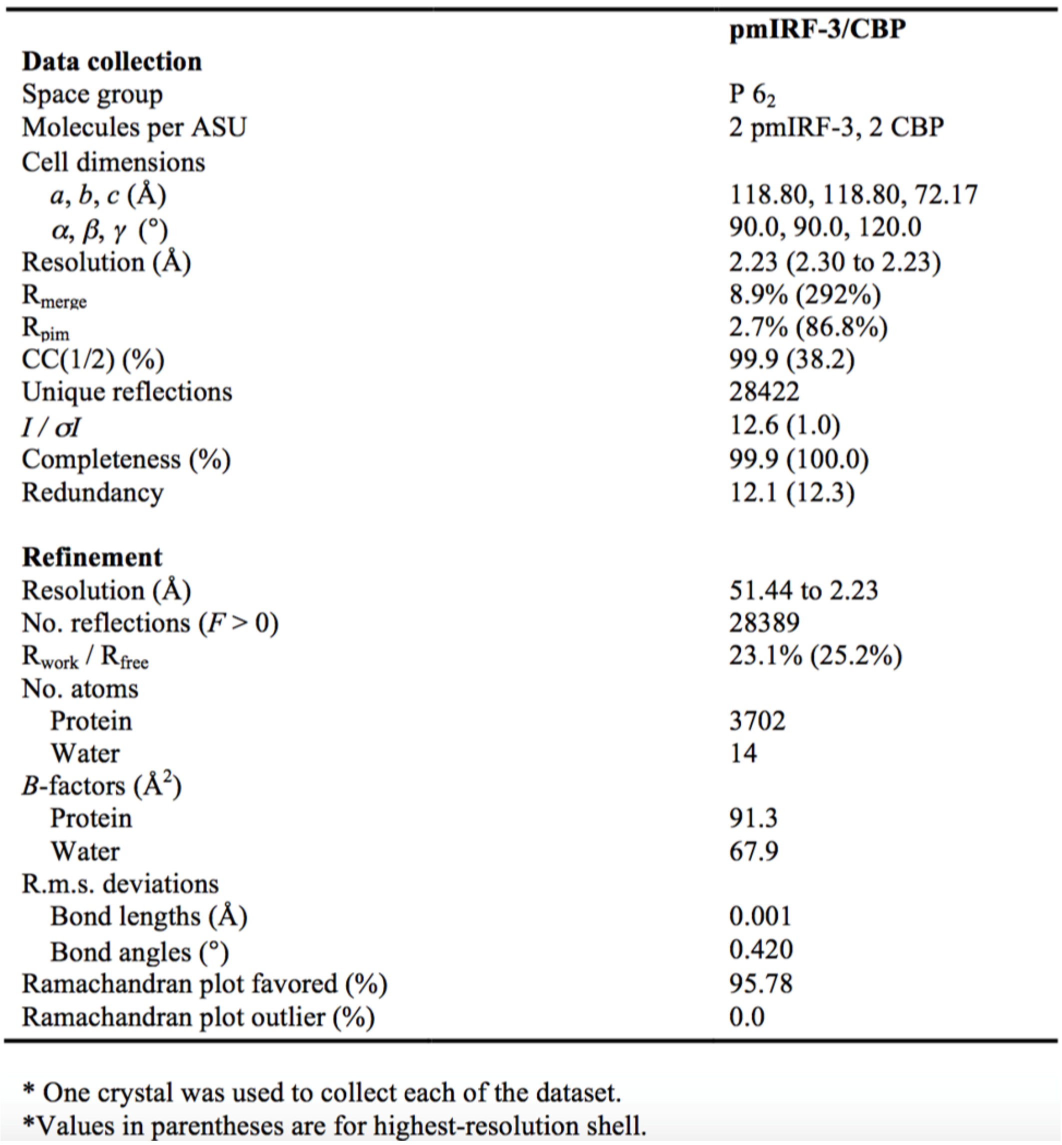


Figure 34 IFN-β luciferase reporter assays showing the effects of S386A and S396A mutations on IRF-3 mediated signaling.
The data are mean ± s.e.m. and representative of three independent assays. **P < 0.01, ***P < 0.001 values were calculated by comparisons of signals in cells transfected with S386A, S396A mutants and those transfected with wild-type IRF-3.

Table1. Data collection and refinement statistics for pmIRF-3/CBP complexes.



	pmIRF-3/CBP
Data collection	
Space group	P 6_2
Molecules per ASU	2 pmIRF-3, 2 CBP
Cell dimensions	
a, b, c (Å)	118.80, 118.80, 72.17
α, β, γ (°)	90.0, 90.0, 120.0
Resolution (Å)	2.23 (2.30 to 2.23)
R_{merge}	8.9% (292%)
R_{pim}	2.7% (86.8%)
CC(1/2) (%)	99.9 (38.2)
Unique reflections	28422
$I / \sigma I$	12.6 (1.0)
Completeness (%)	99.9 (100.0)
Redundancy	12.1 (12.3)
Refinement	
Resolution (Å)	51.44 to 2.23
No. reflections ($F > 0$)	28389
R_{work} / R_{free}	23.1% (25.2%)
No. atoms	
Protein	3702
Water	14
B-factors (Å^2)	
Protein	91.3
Water	67.9
R.m.s. deviations	
Bond lengths (Å)	0.001
Bond angles (°)	0.420
Ramachandran plot favored (%)	95.78
Ramachandran plot outlier (%)	0.0

* One crystal was used to collect each of the dataset.

*Values in parentheses are for highest-resolution shell.

III.8. Discussion of results

The purified mouse IRF-3 (residues 184 to 390) /CBP complex with and without TBK1 phosphorylation were run using the gel-filtration chromatography, and the phosphorylated sample clearly shifted to the left indicating it becomes larger in size. Ideally both of the samples with and without phosphorylation should be examined by ultracentrifugation analysis to investigate their oligomerization state. However, in a PNAS paper our lab published, the authors ran three samples including the phosphomimetic double mutant S386/396E/CBP (189-398aa) dimer, phosphorylated human IRF-3/CBP complex (189-398aa), and the non-phosphorylated IRF-3/CBP (189-398aa) using gel-filtration chromatography which showed that both the phosphomimetic double mutant and the phosphorylated IRF-3 shifted to the left compared to the non-phosphorylated IRF-3. These three samples were then analyzed by sedimentation velocity ultracentrifugation and the results indicated that non-phosphorylated IRF-3/CBP was mostly monomeric whereas a majority of the phosphomimetic double mutant formed a dimer and phosphorylated human IRF-3/CBP complex were almost all dimeric. Based on these previous data, we believe that the phosphorylated mouse IRF-3/CBP shifted to the left on the gel-filtration column should be a dimer which is supported by the mouse structure.

We also compared the phosphorylated human IRF-3/CBP structure to the phosphomimetic double mutant S386/396E/CBP dimer and found that Glu386, mimicking pSer386, contributes much less significantly to IRF-3 dimerization than pSer386. In addition, Glu388 downstream of Glu386 does not interact with Lys360, Gln356 in the phosphomimetic dimer whereas these two residues are involved in the interaction with Glu388 through solvent-mediated hydrogen bond and electrostatic interaction in the phosphorylated IRF-3/CBP structure. These observations were supported by the sedimentation equilibrium AUC analyses, which demonstrated that the S386/396E double mutant had a binding affinity of 1.26 μM between IRF-3 molecules whereas the phosphorylated IRF-3/CBP led to a stable dimer with a Kd of 167 nM. Therefore, we conclude that the phosphomimetic dimer does not fully recapitulate the extensive intermolecular interactions observed in the phIRF-3/CBP dimer.

The comparison between mouse and human phosphorylated IRF-3/CBP structures elucidated that pSer379 only interacts with residues Arg373, Arg205 (corresponding to Arg380, Arg211 in human IRF-3) via electrostatic interactions and with Ser332 (corresponding to Ser339 in human) via hydrogen bonds in the mouse structure whereas Arg334 and Lys353 move away from pSer379 and are not involved in any interactions in mouse structure. By contrast, the corresponding residues Arg341 and Lys360 in human IRF-3 interact with pSer386 via solvent mediated hydrogen bonds and electrostatic interactions. The interactions present in human but missing in mouse IRF-3 are the same amino acid residues. We checked the crystal packing of both human and mouse IRF-3 by using coot software and observed that no surrounding IRF-3 molecules pulled Arg334 and Lys353 residues away from pSer379. Therefore, the interactions missing in mouse structure are not due to the crystal packing. However, the missing interactions might be due to Lys381 of mIRF-3 and Glu388 of hIRF-3. Specifically, Glu388 present in human IRF-3 pulls Lys360 up through electrostatic interaction which created enough space for Arg341 to be in the proper position to interact with pSer386. By contrast, Lys381 flips into the solvent and does not interact with any residues nearby leading to the fact that Lys353 is not pulled up. Since both Arg334 and Lys353 are positively charged, Lys353 may push Arg334 away from pSer379. Therefore, the restructured region around

Lys381 in mouse IRF-3 might be the reason why some of the interactions present in human IRF-3 are missing.

III.9. Materials and methods in detail

III.9.1. Crystallization, data collection, and structure determination

Purified human IRF-3 (residues 189-398) and mouse IRF3 (residues 184-390) bound to the CBP fragment (residues 2065-2111) were phosphorylated by GST-mTBK1. After 24 hour incubation, the proteins were purified using a HiLoad 16/60 Superdex 75 column eluted with 20 mM Tris·HCl and 150 mM NaCl at pH 7.5. The purified phosphorylated proteins were concentrated to a final concentration of ~5 mg/mL. The crystallization screen was performed by hanging drop vapor diffusion technique at 4 °C using Index, Crystal Screen and Crystal Screen 2 reagent kits from Hampton Research. Crystals of human IRF-3 bound to CBP were grown in 0.1 M sodium acetate pH 5.0, 0.2 M MgCl2, ~5% PEG 3350. Crystals of mouse IRF-3 in complex with CBP were grown in 0.2 M ammonium citrate tribasic at pH 7.0 with ~12% PEG 3350. The crystals were flash-frozen in liquid nitrogen in the reservoir solution containing 25% (vol/vol) glycerol. Diffraction data were collected at the Advanced Light Source beamlines 5.0.1 using a Quantum 315R CCD detector. The diffraction data were indexed and integrated with iMosflm and merged with Aimless in the CCP4 package (57). The structures of the pIRF-3/CBP complex were determined by molecular replacement (MR) using the structure of our phosphomimetic IRF-3/CBP complex (PDB ID code 5JEM) as the search model using Phaser in the Phenix package (58). The structures were manually rebuilt using Coot and refined with

Phenix. Details of data quality and structure refinement are summarized in Table 1. The structural figures were generated with PyMOL (https://www.pymol.org).

III.9.2. Mass spectrometry

Molecular mass of human IRF-3 (189-398aa) and mouse IRF-3 (184-390aa) with and without TBK1 phosphorylation were determined by MALDI-TOF mass spectrometry using a Bruker Ultraflextreme TOF-TOF mass spectrometer (Protein Chemistry Laboratory, Texas A&M University). The samples were solid phase extracted using Protea LithTip C4 and analyzed using alpha-cyanohydroxycinnamic acid as matrix using the dried drop method. The mass spectrometer was operated in reflector mode and calibrated with angiotensin II, fibrinopeptide, renin substrate and ACTH (18-39 fragment).

III.9.3. Analysis of phosphorylated IRF-3 by size-exclusion chromatography (SEC)

Purified full-length human IRF-3 proteins (wild-type and mutants) were mixed with GST-mTBK1 in a ratio of 10:1 (w/w) in a 1 mL reaction buffer with 20 mM HEPES pH 7.5, 10 mM MgCl2, 100 mM NaCl, 5 mM ATP, 0.1 mM Na3VO4, 5 mM NaF, 5 mM DTT at 27 °C for ~24 hours. The final concentration of the proteins was about 1 mg/mL. After ~24-hour incubation, the phosphorylated IRF-3 proteins were analyzed using a Superdex 200 (10/300 GL) column eluted with a buffer containing 20 mM Tris·HCl and 150 mM NaCl at pH 7.5.

III.9.4. Native gel electrophoresis of phosphorylated IRF-3

The purified full-length IRF-3 proteins were phosphorylated with GST-mTBK1 using the method described above and each of the phosphorylated proteins with non-phosphorylated wild-type was resolved on 10% native gels running in a buffer containing 25 mM Tris and 192 mM glycine pH 8.4 at 4°C at 100V for 30 minutes. The gels were stained with Coomassie blue for one hour and destained with a solution containing H2O, methanol, and acetic acid in a ratio of 50/40/10 (v/v/v) until the bands were clearly seen. The gel image was taken using Bio-Rad imager.

III.9.5. IFN-β luciferase reporter assays

The cDNA encoding wild-type human IRF-3 was cloned into a pcDNA3.1(-) vector using appropriate primers. Mutants of hIRF-3 were generated using the QuikChange site-directed mutagenesis kit (Agilent). Sequences of the mutants were confirmed by DNA Sequencing. HEK293T cells were plated in CoStar White 96-well plates at 4 × 104 cells per well and each well contains 100 μl DMEM (1×) + GlutaMAX medium (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco). After ~24h incubation at 37°C, the cells were transfected with the IRF-3 plasmids (10 ng per transfection) using Lipofectamine 2000 reagent (Invitrogen) and Opti-MEM medium (Gibco) together with constant amount of IFN-β firefly luciferase reporter plasmids (20 ng per transfection), phRL-TK–Renilla luciferase plasmids (2 ng per transfection) (Promega), and human STING plasmids (0.2 ng per transfection). Transfections with empty pcDNA3.1(-) and WT hIRF-3 with no STING plasmid were used as controls. The cells were incubated for another 24 h to allow the expression of the genes. The half of the cells in the plates were treated with 30 µg/mL cGAMP dissolved in DMEM (1×) + GlutaMAX medium and the other half were treated with the medium only. After ~16 h incubation, the cells were analyzed using the Dual-Glo luciferase reporter assay kit (Promega). Luminescence was quantified with the BioTek Synergy HTX Multi-Mode microplate reader. The relative firefly luciferase activity was normalized by the Renilla luciferase activity. The relative IFN-β reporter fold of induction represents the ratio normalized to control plasmid values with the same treatment.

III.9.6. Cell culture

HEK293T cells (ATCC, CRL-3216) were cultured in DMEM (1×) + GlutaMAX medium (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco), streptomycin (100 µg/mL) and penicillin (100 U/mL) at 37 °C in a humidified atmosphere containing 5% CO2.

III.9.7. Statistical analysis

Statistical analyses for the luciferase reporter assays were carried out by Microsoft Excel. All of the data are presented as mean ± SEM. A two-tailed Student's t test assuming equal variances was used to compare two groups. The statistical significance between the indicated samples and the control is designated as $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, or NS ($P > 0.05$).

III.9.8. Data deposition

The atomic coordinates and structural factors of the phosphorylated human and mouse IRF-3/CBP complexes have been deposited in the Worldwide Protein Data Bank, www.wwpdb.org (PDB ID: 7JFL and 7JFM respectively).

CHAPTER IV. RESIDUES INTERACTING WITH PSER386 AND PSER396 AFFECT IRF-3 ACTIVATION, IRF-3 MEDIATED SIGNALING AND IRF-3 TRANSLOCATION TO THE NUCLEUS

IV.1. Introduction

Up to now, there are several structures concerning IRF-3 activation domain: 1). The autoinhibited IRF-3 transactivation domain (173-427aa) solved by Lin group, which suggested an autoinhibitory mechanism through which phosphorylation reorganizes the autoinhibitory elements leading to the uncovering of the folded IRF-3. 2). The IRF-3 (173-394aa) bound to CBP solved by Lin group which identified the hydrophobic binding surface for CBP that were previously covered by autoinhibited elements of IRF-3. 3). The phosphomimetic double mutant S386/396E bound to CPB which demonstrated that IRF-3 forms a dimer through S386E and its pLxIS motif. In addition, some functional studies have reported that R285Q mutation in human IRF-3 impairs innate immune responses to herpes simplex virus infection and R285D mutation disrupts IFN-β reporter activation by Newcastle disease virus in IRF-3−/− fibroblasts. Since we solved the phosphorylated human IRF-3 structure and observed that many residues are involved in the interactions with pSer386 and pSer396, we wanted to investigate how these residues affect IRF-3 activation, IRF-3 mediated signaling and IRF-3 translocation to the nucleus.

IV.2. Mutations of residues interacting with phosphorylated Ser386 and Ser396 impair IRF-3 activation

Our data has supported that both Ser386 and Ser396 are involved in IRF-3 dimerization but Ser386 plays a more crucial role compared to Ser396. Based on our structure, we also identified a few residues that are interacting with pSer386 and pSer396. In order to investigate how these residues interacting with pSer386 and pSer396 contribute to IRF-3 activation, we generated twelve mutants of full-length hIRF-3, which include mutations R211A, R380A, S339A, R211A/R380A, R211A/R380A/S339A of residues interacting with pSer386, and mutations R285A, H288A, H290A, K313A, R285A/K313A, H290A/K313A, and H288A/H290A/K313A of residues that are likely to interact with pSer396. Each of these mutants was expressed and purified for in vitro phosphorylation (**Figure 35A and B**).

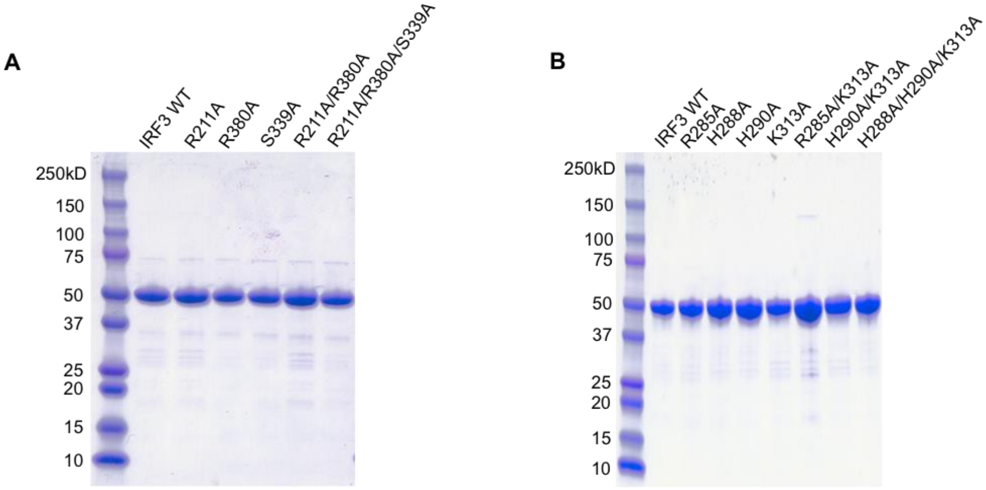


Figure 35 Gradient gel electrophoreses showing purified full-length human IRF-3 mutants. (A) Gradient gel electrophoreses showing purified full-length human IRF-3 mutants around pSer386 (B) Gradient gel electrophoreses showing purified full-length human IRF-3 mutants around pSer396.

To make sure that mouse TBK1 (N657) and TBK1 phosphorylation assay are working in an efficient and effective manner, freshly purified wild type full-length IRF-3 with and without TBK1 phosphorylation were sent out for Mass spectrometry. The MS analyses showed that WT IRF-3 could be efficiently phosphorylated by TBK1 (**Figure 36**).

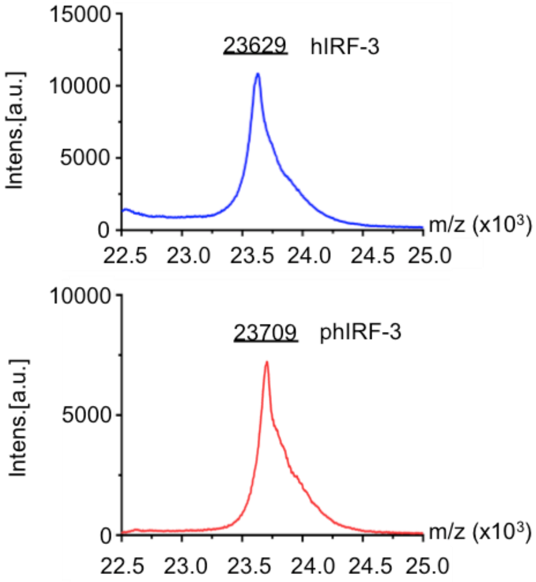


Figure 36 MS analyses of full-length human IRF-3 before and after TBK1 phosphorylation.

Then each of the IRF-3 mutants was phosphorylated and analyzed by SEC. Strikingly, the R211A, R211A/R380A, and R211A/R380A/S339A mutants failed to form dimers upon phosphorylation compared to the wild-type control since they only showed one single peak eluted at the same position as wild-type IRF-3. In addition, mutation R380A severely impaired the dimerization of IRF-3 upon phosphorylation and mutation S339A moderately inhibited IRF-3 dimerization (**Figure 37A to F**).

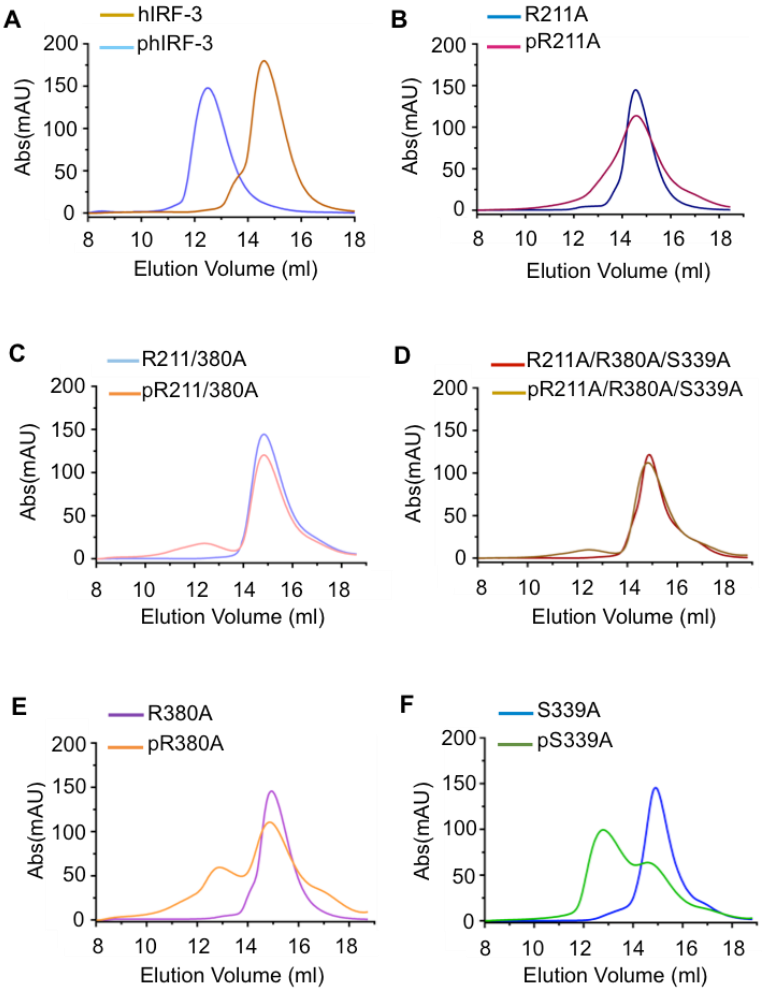


Figure 37 Mutations of key residues interacting with pSer386 of hIRF-3 affect IRF-3 dimerization upon phosphorylation.

(A to F) Size-exclusion chromatography analyses showing how mutations of key residues interacting with pSer386 affect IRF-3 dimerization upon phosphorylation.

By contrast, mutations R285A, R285A/K313A, H290A/K313A, and H288A/H290A/K313A have moderate effects on IRF-3 dimer formation since all of them showed two peaks indicating a mixture of monomer and dimer (**Figure 38A to D**). Moreover, mutations H288A, H290A, and K313A alone have little effects on IRF-3 dimerization because only the dimer peak was observed (**Figure 39A, B and C**).

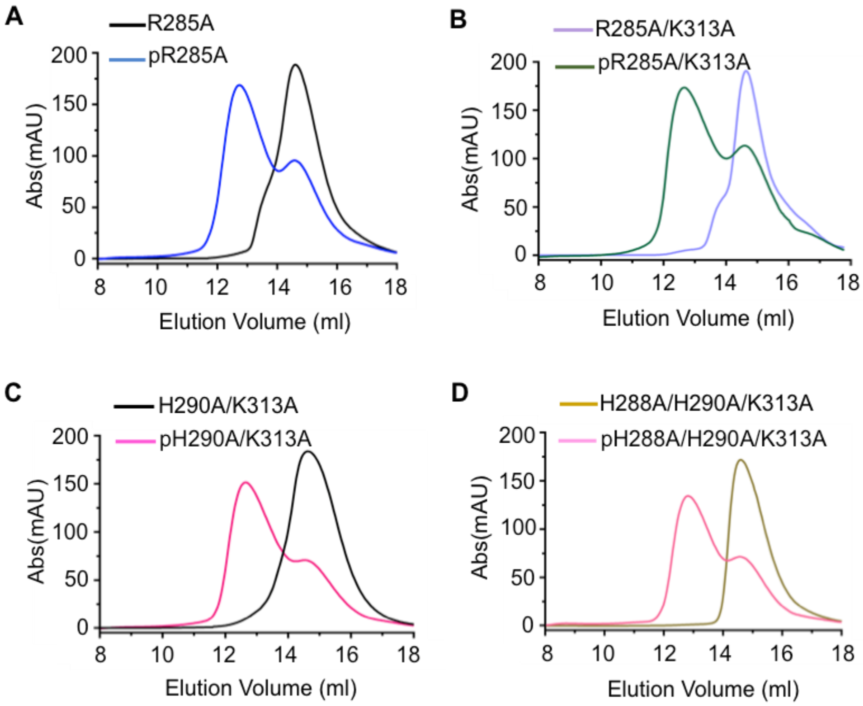


Figure 38 Mutations of key residues interacting with pSer396 of hIRF-3 affect IRF-3 dimerization upon phosphorylation.
(A to D) Size-exclusion chromatography analyses showing how mutations of key residues interacting with pSer396 affect IRF-3 dimerization upon phosphorylation.

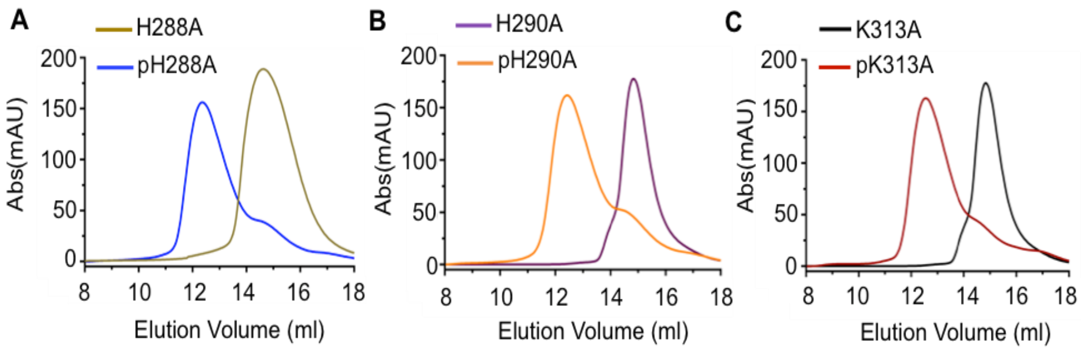


Figure 39 Mutations of key residues interacting with pSer396 of hIRF-3 affect IRF-3 dimerization upon phosphorylation.
(A to C) Size-exclusion chromatography analyses showing how mutations of key residues interacting with pSer396 affect IRF-3 dimerization upon phosphorylation.

To further investigate the effect of these mutations on IRF-3 activation, we also analyzed these IRF-3 mutants by native gel electrophoresis (**Figure 40A and B**). We observed that phosphorylated R211A, R211A/R380A, and R211A/R380A/S339A mutants only showed a single band similar to un-phosphorylated IRF-3, suggesting the activation of IRF-3 was disrupted by these mutations. Phosphorylated R380A mutant appeared as two bands and only a small fraction of this mutant formed dimers, indicating that the activation of IRF-3 was dramatically impaired by this mutation. By contrast, mutations S339A, R285A, R285A/K313A, H290A/K313A, and H288A/H290A/K313A moderately affect the dimerization of IRF-3 upon phosphorylation. Three other mutations H288A, H290A, and K313A individually do not affect the dimerization of IRF-3. These results are consistent with the SEC analyses of phosphorylated IRF-3 mutants. Taken together, these studies show that mutations of key residues interacting with pSer386 significantly impair the activation of IRF-3. By contrast, mutations of residues that interact with pSer396 have moderate effects on IRF-3 activation, demonstrating that phosphorylation of Ser386 is more critical for the activation of IRF-3.

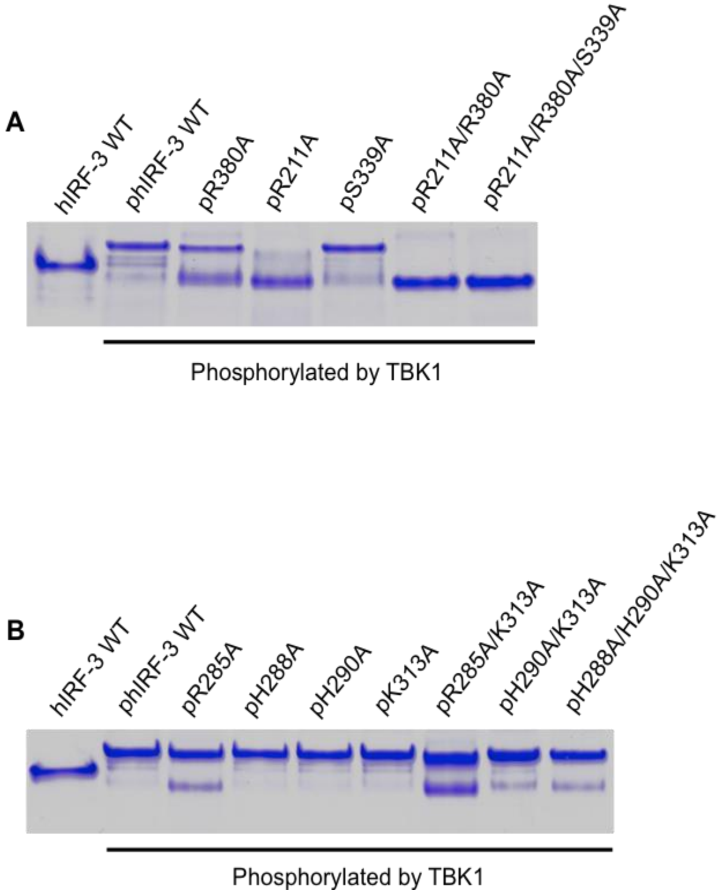


Figure 40 Mutations of key residues interacting with pSer386 and pSer396 of hIRF-3 affect IRF-3 dimerization upon phosphorylation.
(A and B) Native gel electrophoreses showing how IRF-3 mutations affect its dimerization upon phosphorylation.

IV.3. Residues mediating IRF-3 dimerization play critical roles in IRF-3 mediated signaling

It has been reported that IRF-3 dimerizes in the cytosol upon phosphorylation and then translocates to the nucleus to initiate the transcription of the IFN-β gene(26, 34, 59-61). In order to investigate how IRF-3 mediated signaling is affected by mutations of residues involved in IRF-3 dimerization, we conducted IFN-β luciferase reporter assay in cells transfected with full-length IRF-3 mutants. As is shown in Fig. 41A, mutations R211A, R380A, R211A/R380A and R211A/K380A/S339A blocked the induction of the IFN-β reporter, whereas mutation S339A reduced the IFN-β reporter signal by ~50%. As controls, the cells transfected with pcDNA3.1(-) or WT IRF-3 plasmid in the absence of STING plasmid showed almost no signals. The cells co-transfected with pcDNA3.1(-) and STING plasmids also showed very little signals (**Figure 41A**). In order to make sure that all the mutants were expressed normally in cells as the wild-type IRF-3, HEK293T cells were transfected with empty pcDNA3.1(-), WT IRF-3 or IRF-3 mutants together with human STING plasmid, then the cells were stimulated by 30 μg/mL cGAMP added to the culture media 24h post transfection. After ~16 h incubation, protein expression levels were analyzed by western blot analysis (**Figure 41B**). Western blot showed that the expression level of the IRF-3 mutants was similar to the wild type IRF-3 whereas only a very faint band was observed in the vector control, which corresponds to endogenous IRF-3.

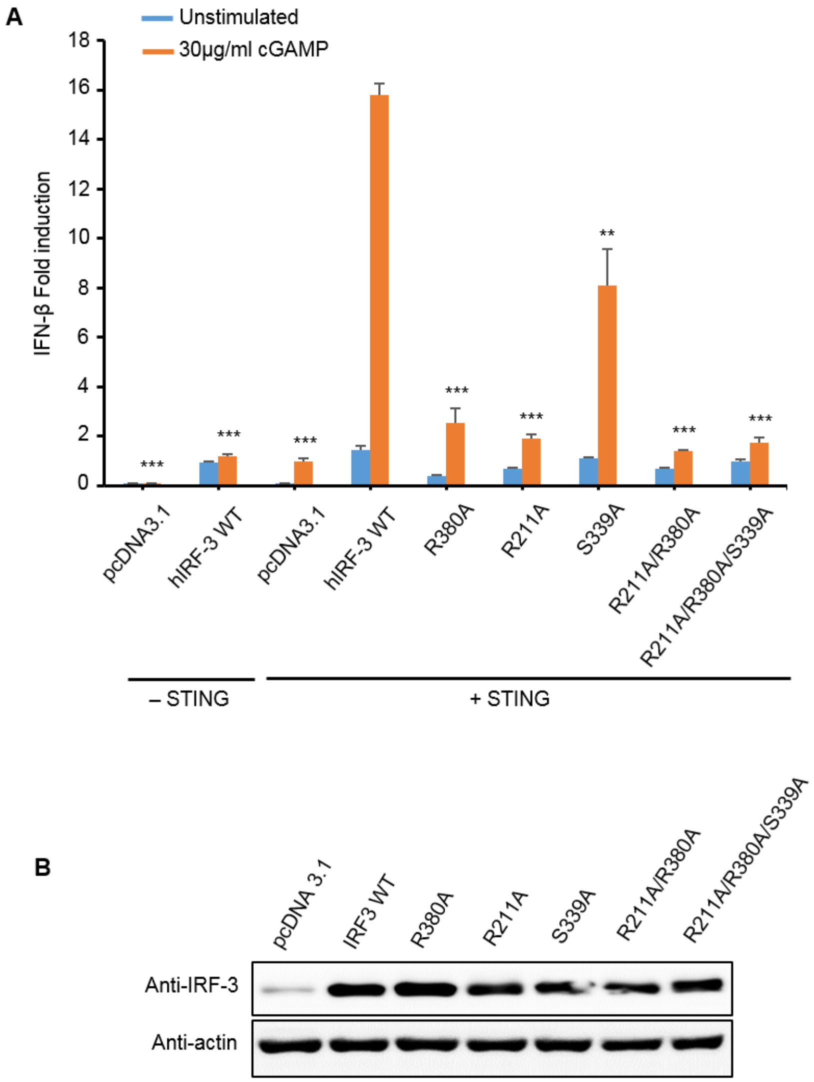


Figure 41 Mutations of residues interacting with pSer386 of hIRF-3 affect IRF-3 mediated signaling.

(A) IFN-β luciferase reporter assays showing mutations of residues interacting with pSer386 affect IRF-3 mediated signaling. The data are mean ± s.e.m. and representative of three independent assays. **P < 0.01, ***P < 0.001, NS (P > 0.05) values were calculated by comparisons of signals in cells transfected with IRF-3 mutants and those transfected with wild-type IRF-3. (B) Western blot showing the expression of IRF-3 mutants and wild type in HEK293T cells.

As to the mutations of residues surrounding pSer396, mutations R285A, R285A/K313A, H290A/K313A, and H288A/H290A/K313A reduced the reporter signal by 60-70% compared to the wild type. By contrast, mutations H288A, H290A, and K313A individually barely impacted the activation of the reporter (**Figure 42A**). Similarly, western blot of cells transfected with WT IRF-3 and mutants indicated that these IRF-3 mutants are expressed at similar levels (**Figure 42B**). These results demonstrate that the mutations of key residues interacting with pSer386 abrogated IFN-β reporter activation whereas the mutations of residues interacting with pSer396 only partially inhibited IFN-β reporter activation, suggesting that residues mediating IRF-3 dimerization are critical in IRF-3 mediated signaling.

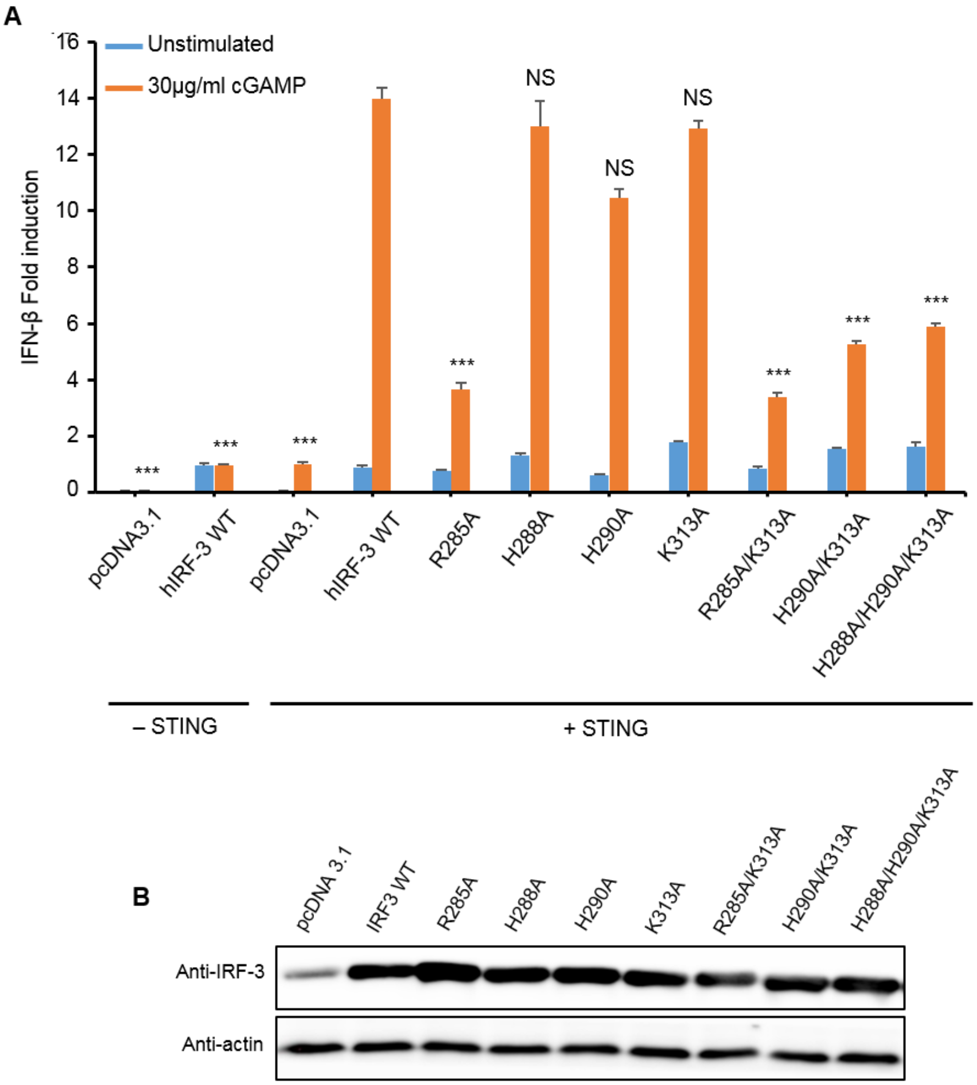


Figure 42 Mutations of residues interacting with pSer396 of hIRF-3 affect IRF-3 mediated signaling.

(A) IFN-β luciferase reporter assays showing mutations of residues interacting with pSer386 affect IRF-3 mediated signaling. The data are mean ± s.e.m. and representative of three independent assays. **P < 0.01, ***P < 0.001, NS (P > 0.05) values were calculated by comparisons of signals in cells transfected with IRF-3 mutants and those transfected with wild-type IRF-3. (B) Western blot showing the expression of IRF-3 mutants and wild type in HEK293T cells.

IV.4. Residues mediating IRF-3 dimerization affect IRF-3 translocation to the nucleus

The subcellular localization is important for the function of a variety of proteins(62-64). For example, many transcription factors that reside in the cytoplasm can be transmitted to the nucleus by responding to some external signals and they induce the transcription of some genes rapidly and transiently. However, persistent presence in the nucleus is not desirable because the continuous induction of target genes might be detrimental to the cells(65-67). In terms of IRF-3, it is cytoplasmic in normal cells. Upon infection by pathogens, IRF-3 dimerizes through phosphorylation, translocates to the nucleus and binds to CBP/p300 to initiate the transcription of target genes. In order to investigate how the residues mediating IRF-3 dimerization affect the subcellular localization of IRF-3, we conducted confocal microscopy analyses of cells transfected with the IRF-3 mutants (**Figure 43**). We observed that WT IRF-3 efficiently translocated to the nuclei after cGAMP treatment, whereas the R211A/R380A, R211A/R380A/S339A, R211A, and R380A mutants are mostly localized in the cytosol. The S339A, R285A, R285A/K313A, H290A/K313A and H288A/H290A/K313A mutants partially entered the nuclei. By contrast, the H288A, H290A, and K313A mutants individually behaved similarly as WT IRF-3.

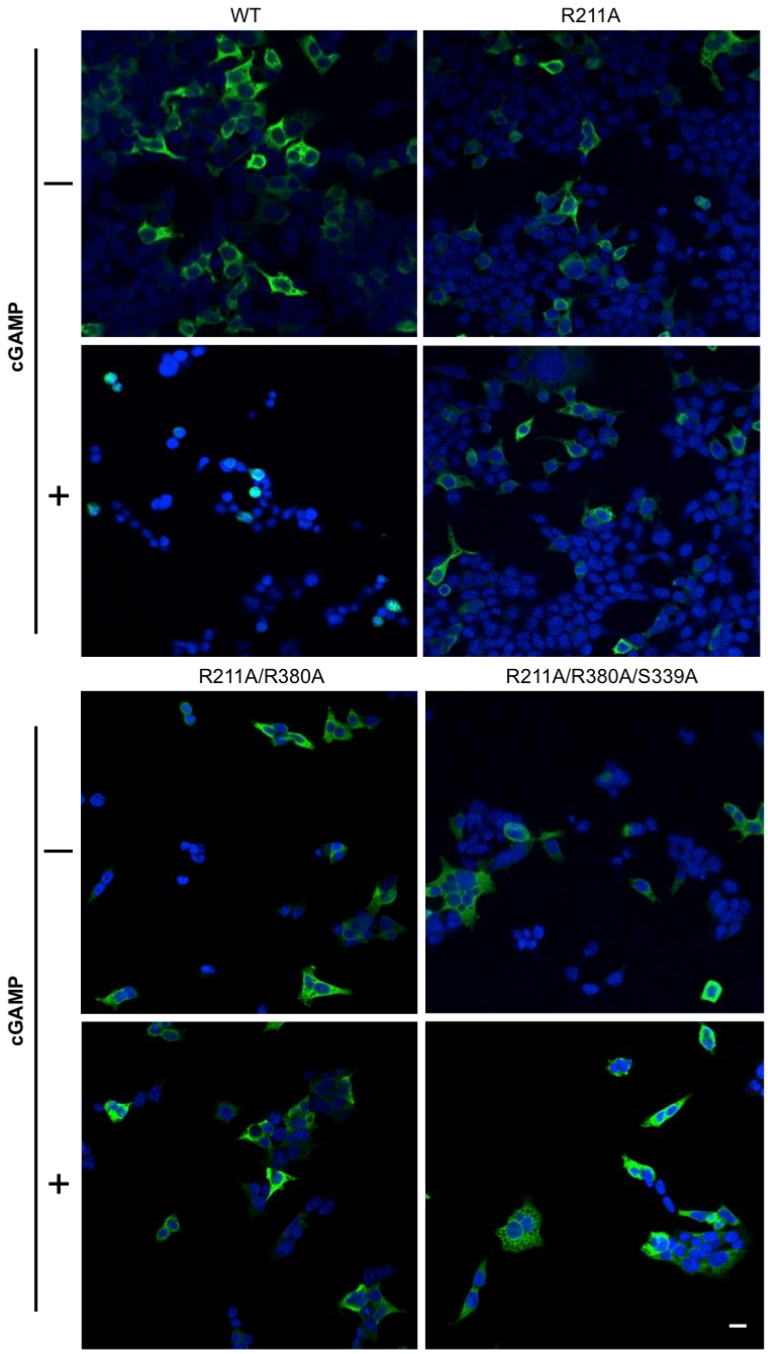


Figure 43 Mutations of residues mediating IRF-3 dimerization affect IRF-3 nuclear localization.

Confocal microscopy of HEK293T cells transfected with IRF-3 mutants and STING upon cGAMP stimulation. The green color represents IRF-3 and the blue color represents DAPI indicating where the nucleus is in the cells. Scale bars, 20 μm.

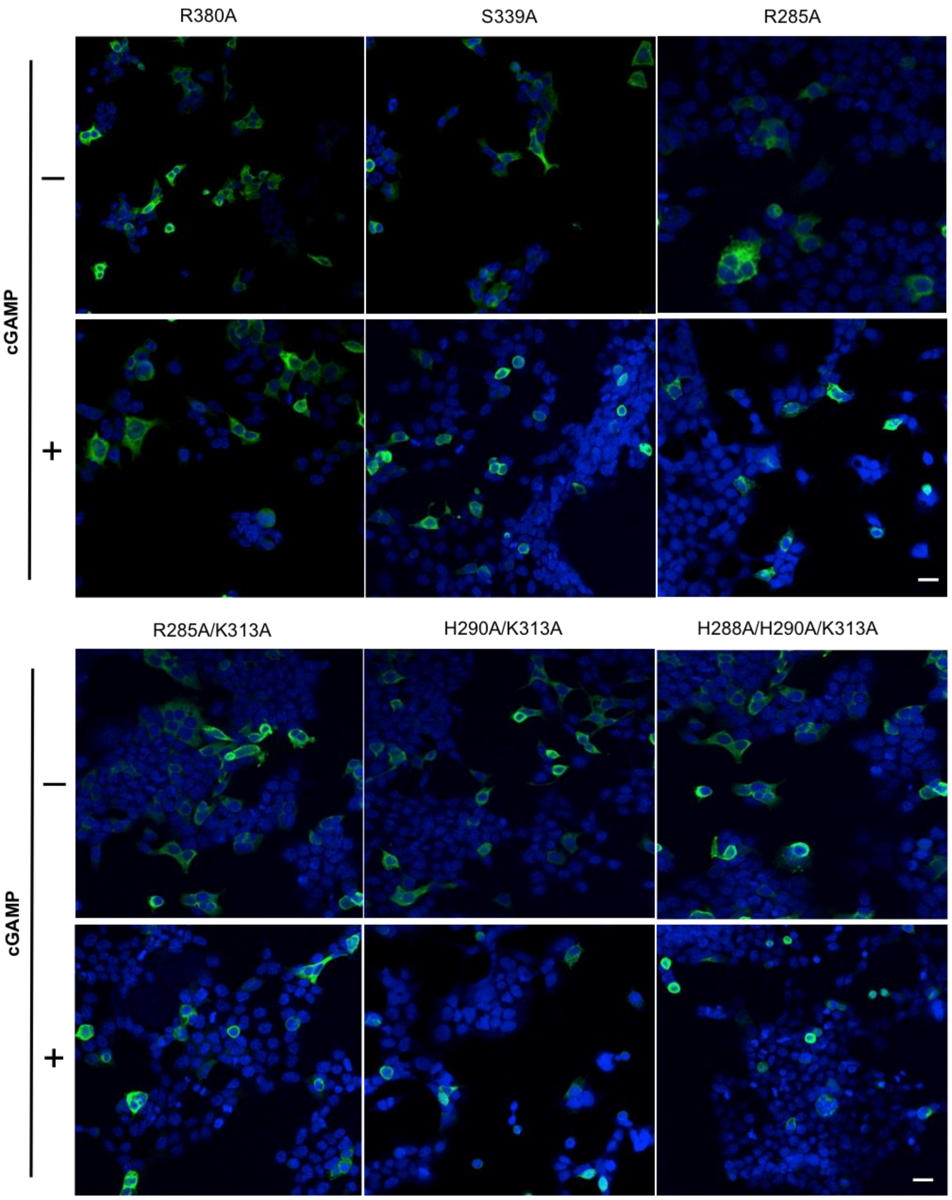


Figure 43 Continued.

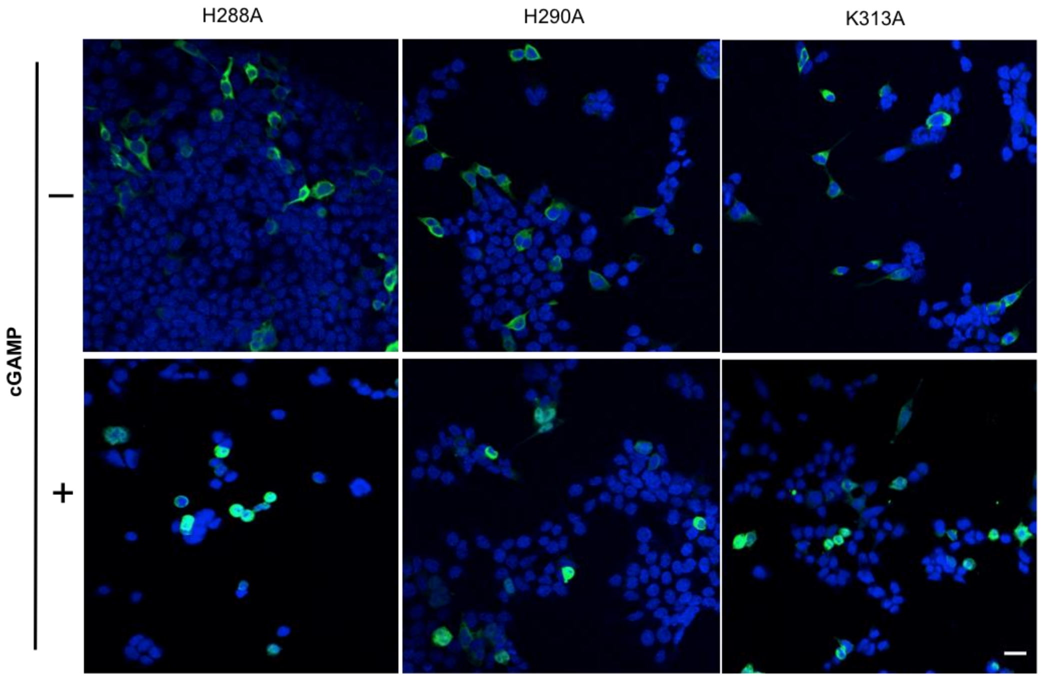


Figure 43 Continued.

We also used ImageJ to quantify the amount of nuclear translocation of IRF-3 in the cells and observed that mutations R211A, R380A, R211A/R380A, R211A/R380A/S339A greatly impaired the amount of IRF-3 translocated to the nucleus after cGAMP treatment whereas mutations H288A, H290A, K313A individually did not affect IRF-3 nuclear translocation (**Figure 44**). These results demonstrate that the mutations that affect IRF-3 dimerization impair nuclear translocation of IRF-3.

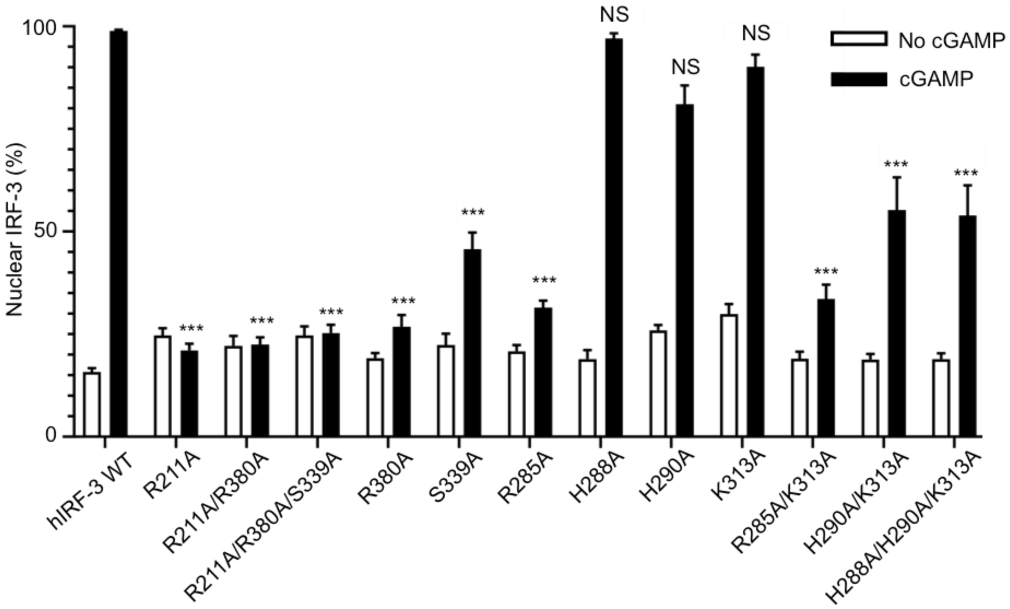


Figure 44 Quantification of nuclear translocation of hIRF-3 in HEK293T cells.
Nuclear hIRF-3 in wild-type or IRF-3 mutants with or without cGAMP stimulation is quantified by ImageJ.

IV.5. Discussion of results

The residues Arg211 and Arg380 interacting with pS386 are critical for IRF-3 dimerization, IRF-3 mediated signaling and its translocation to the nucleus since the mutations of these two residues severely impaired IRF-3 dimerization, greatly reduced IFN-β reporter signal and largely inhibited IRF-3 translocation to the nucleus. It has been reported that IRF-3 subcellular localization is mediated by a nuclear export signal (NES) that is located between its DNA-binding domain and IRF association domain. In uninfected cells, IRF-3 is inactive and stays in the cytosol due to the continuous nuclear export mediated by NES. Upon viral infection, IRF-3 is activated through phosphorylation, dimerizes and translocates to the nucleus where it binds to CBP and eventually leads to the induction of IFN-Is. How our mutations R211A and R380A inhibited IRF-3 translocation to nucleus upon phosphorylation remains unclear so far. Further studies that link dimerization to the function of NES has to be done.

The structures of the phosphorylated IRF-3 dimer and the adaptors bound to IRF-3 reveal that Arg285 (Arg278 in mouse) is a key residue interacting with the phosphorylated serine residue within the pLxIS motif. Mutation of Arg285 to alanine moderately impairs IRF-3 activation. These results explain why the R285Q mutation impaired IFN responses to HSV-1 infection and the R285E mutation in either HEK293 cells or fibroblasts of IRF-3-deficient mice showed much less activity upon NDV infection compared to WT IRF-3.

IV.6. Materials and methods in detail

IV.6.1. Gradient gel electrophoresis of purified IRF-3

Full-length WT IRF-3 and all the mutant proteins were purified using Ni2+-NTA column followed by gel-filtration chromatography as described above. 10 μg of each sample was mixed with 5× loading dye of 250 mM Tris·HCl, pH 6.8, 10% SDS, 30% (v/v) Glycerol, 10 mM DTT, 0.05% (w/v) bromophenol blue and then resolved on 4-20% gradient gels in a buffer containing 25 mM Tris, 192 mM glycine and 0.1% SDS at pH8.4 at 100V for one hour. The gel was stained with Coomassie blue for one hour and destained with a solution containing

H2O, methanol, and acetic acid in a ratio of 50/40/10 (v/v/v) until the bands were clearly seen. The gel image was taken using Bio-Rad imager.

IV.6.2. Western blot

HEK293T cells were transfected with empty pcDNA3.1(-), WT IRF-3 or IRF-3 mutants together with human STING plasmid. The cells were stimulated by 30 μg/mL cGAMP added to the culture media 24h post transfection. After ~16 h incubation, the cells were washed and suspended in PBS and then lysed in 150 mM NaCl, 200 mM Tris·HCl (pH 7.5), 1 mM EDTA, and 1% Nonidet P-40 supplemented with one complete EDTA-free protease inhibitor mixture tablet (Roche) and one PhosSTOP phosphatase inhibitor mixture tablet (Roche) for each 10 mL of lysis buffer. The proteins were resolved on 10% SDS-PAGE at 100V for 1.5h and transferred to PVDF membrane in a transfer buffer containing 1x Tris-Glycine + 20% methanol for another 1.5 hours. Then 5% milk in 1xPBST solution was used to block the membrane for 1 hour followed by three rinses with PBST. Next the membrane was incubated with primary antibodies dissolved in 1xPBST with 5% BSA overnight at 4 °C. The membrane was washed three times using PBST next day and further incubated with the corresponding HRP-conjugated secondary antibodies dissolved in PBST. The following antibodies were used in the Western blot experiment: anti–IRF-3 (1:1,000; sc-9082; Santa Cruz) and anti-Actin (1: 4,000; HHF35; Pierce). The proteins were visualized using the Western Lightening Plus ECL (PerkinElmer) according to the manufacturer’s protocol. For the detection of phosphorylated IRF-3, half of the cells transfected with empty pcDNA3.1(-) and WT IRF-3 were stimulated by 30 μg/mL cGAMP and the other half were treated with the medium only. Anti-IRF-3 phospho-Ser386 (1:2,500; ab76493; Abcam) and anti-IRF-3 phospho-Ser396 (1:1,000; 4947S; Cell Signaling) were used.

IV.6.3. Immunocytochemistry

HEK293T cells were grown on poly-L-lysine coated coverslips placed in 12-well plates for 24 h and then co-transfected with 20 ng WT IRF-3-HA and 1ng STING plasmids or 20 ng mutant IRF-3-HA and 1ng STING respectively using lipofectamine 2000 reagent mixed with Opti-MEM medium (Gibco). 24 hours post transfection; the medium was replaced with fresh DMEM (1×) + GlutaMAX medium with or without 30 μg/mL cGAMP. After 12 h incubation, cells were washed using PBS, then fixed by 4% paraformaldehyde in PBS for 15 min at room temperature and permeabilized with PBST containing 0.5% Triton X-100 in PBS. Cells were washed and blocked with 5% fetal bovine serum (FBS) in PBST and then incubated with anti-HA tag primary antibody (Cell signaling, 3724; 1:100) overnight. The cells were washed three times with PBS and incubated with Alexa Fluor 488 Goat Anti-Rabbit IgG (Thermo Fisher Scientific, A11034, 1:1000) at room temperature for 1 hour. The coverslips were washed by PBS, mounted on slides with ProLong Gold antifade reagent with 4′-6-diamidino-2-phenylindole (DAPI) and then imaged under Olympus FV1000 fluorescence microscope. The scale bars in the images correspond to 20 μm in length. To quantify the amount of nuclear translocation, 12 IRF-3-HA highly expressed cells or cell clusters were randomly selected in each field. The ROI of nuclear or total IRF-3-HA fluorescence in the same cell or cell cluster was manually drawn. The area integrated fluorescence intensity was calculated by ImageJ (Version 1.51n).

IV.6.4. Mass spectrometry

Molecular mass of full-length WT IRF-3 and phosphorylated IRF-3 were determined by MALDI-TOF mass spectrometry using a Bruker Ultraflextreme TOF-TOF mass spectrometer (Protein Chemistry Laboratory, Texas A&M University). The samples were solid phase extracted using Protea LithTip C4 and analyzed using alpha-cyanohydroxycinnamic acid as matrix using the dried drop method. The mass spectrometer was operated in reflector mode and calibrated with angiotensin II, fibrinopeptide, renin substrate and ACTH (18-39 fragment).

IV.6.5. Statistical analysis

Statistical analyses for quantification of fluorescence were carried out by Prism. All of the data are presented as mean ± SEM. A two-tailed Student's t test assuming equal variances was used to compare two groups. The statistical significance between the indicated samples and the control is designated as *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, or NS ($P > 0.05$).

IV.7. Discussion of IRF-3 project

IRF-3 is the key transcription factor regulating the expression of type I IFNs in response to various pathogens(68-70). Here, we determined the crystal structures of phosphorylated human and mouse IRF-3/CBP complexes, which reveal that IRF-3 forms a dimer upon phosphorylation. Compared to auto-inhibited IRF-3, the C-terminal tail of IRF-3 undergoes a dramatic conformational change upon phosphorylation, extending to the binding surface on another IRF-3 molecule and mediates the dimerization of IRF-3. Phosphorylated Ser386 interacts with several residues in a positively charged pocket through extensive electrostatic interaction and hydrogen bonds. Cell-based studies combined with in vitro phosphorylation assays demonstrate that mutations of Ser386 and the residues surrounding pSer386 abrogate IRF-3 dimerization, block its translocation to the nuclei and abolish IRF-3 mediated signaling. By contrast, phosphorylation of Ser396 within the pLxIS motif likely plays a moderate role in IRF-3 activation. Mutations of Ser396 or residues that may interact with pSer396 only partially impair IRF-3 activation and signaling. Moreover, the structural analyses reveal that Glu388 plays additional roles in the activation of human IRF-3. These structural and functional studies established the molecular basis of IRF-3 activation upon phosphorylation.

In previous studies, Ser386 and the adjacent Ser385 (Ser378 and Ser379 in mouse) were considered as two important phosphorylation sites and were identified as critical residues for IRF-3 mediated signaling(38, 39, 71). However, the structures of human and mouse IRF-3 dimer clearly show that only Ser386 is phosphorylated. The phIRF-3/CBP structure reveals that Ser385 interacts with Asp254 via hydrogen bonds to stabilize the IRF-3 dimer. Phosphorylated Ser386 reaches into a large positively charged pocket formed between two molecules of IRF-3 and contributes significantly to IRF-3 dimerization. Mutations of Ser386 and residues interacting with pSer386 dramatically impair IRF-3 dimerization, nuclear translocation and signaling. These results suggest that phosphorylation of Ser386, but not Ser385, is essential for IRF-3 mediated signaling. Consistent with these results, two previous studies showed that mutation of Ser385 to aspartic acid impairs IRF-3 dimerization and no phosphorylated Ser385 was detected with specific antibodies(40, 44).

Although Ser396 (Ser388 in mouse IRF-3) within the pLxIS motif is not phosphorylated in the human and mouse IRF-3 dimer structures presented here, these residues can be phosphorylated in vivo. Mutations of Ser396 and key residues that may interact with pSer396 impair IRF-3 dimerization and its functions, suggesting that phosphorylation of Ser396 also plays an important role in IRF-3 activation. Our previous studies demonstrated that the adaptors STING, MAVS, and TRIF employ the conserved pLxIS motif to recruit IRF-3 upon phosphorylation(10, 17, 72). Interestingly, the pLxIS motif from IRF-3 itself also mediates its own dimerization and this motif binds to the same surface on the other IRF-3 molecule. Based on the current human phosphorylated IRF-3 structure, the phosphorylated STING bound to IRF-3 structure and the structure of TBK1 in complex with STING C-terminal tail, we proposed the following steps for cGAS-STING-TBK1-IRF-3 pathway: 1). cGAS is activated by dsDNA and catalyses cGAMP synthesis; 2). cGAMP binds to STING and leads to its oligomerization at the ER or Golgi membranes; 3). TBK1 is recruited to the STING oligomers through PLPLRT/SD motif and is further activated due to induced proximity; 4). STING is activated at its pLxIS motif via activated TBK1; 5). IRF-3 is recruited to the signaling complex through pLxIS motif of STING; 6). The proximity between TBK1 and IRF-3 bound to STING molecules leads to IRF-3 phosphorylation at both Ser386 and its pLxIS motif containing Ser396; 7). Phosphorylated IRF-3 unfolds and its phosphorylated tail containing the pLxIS motif displaces the original phosphorylated pLxIS motif of STING and therefore extends into the pLxIS motif-binding surface of the other IRF-3 molecule to form a stable dimer; 8). IRF-3 dimer dissociates from the STING, translocates to the nucleus to initiate the induction of IFN-Is. What is interesting is that the binding affinity of pSTING/IRF-3 is similar to that of the pIRF-3382-409aa/IRF-3, both of which showed a micromolar affinity. However, the phosphorylated IRF-3 dimer has been reported to have an affinity of 167nM, which is much higher than the affinity of pSTING/IRF-3, which may explain why phosphorylated IRF-3 forms a stable dimer and then dissociates from STING. However, further studies including the ternary structure of STING-TBK1-IRF-3 need to be determined to explain how the tail of IRF-3 displaces the tail of STING upon TBK1 phosphorylation.

Besides Ser386 and the pLxIS motif, Glu388 of human IRF-3 also plays a role in human IRF-3 activation. Size-exclusion chromatography and cell-based studies confirm the contribution of Glu388 in IRF-3 dimerization and IRF-3 mediated signaling. Interestingly, Glu388 is highly conserved in various species except for mouse and rat (**Figure 45**). The replacement of two residues Glu388 and Asn389 in human by Lys381 in mouse IRF-3 restructures this region and likely reduces the intermolecular interactions of mouse IRF-3 upon phosphorylation that may affect the kinetics of IFN-I induction by mice. In summary, these extensive structural and functional studies provide critical insights into the molecular basis of IRF-3 activation upon phosphorylation.

```
                   386 388          396
                    |   |            |  *     *  * *
IRF-3 Human     LVEMARVGGASSLENTVDLHISNSHPLSLTSDQYKAYLQDLVEGMDFQGPGES
IRF-3 Mouse     LLEMAREGGASSLK-TVDLHISNSQPISLTSDQYKAYLQDLVEDMDFQATGNI
IRF-3 Pig       LVDMARDGGASSLENTVDLHISNSHPLSLTSDQYKACLRDLVEDMDF------
IRF-3 Horse     LLEMARIQGASSMENTVDLHISNSHPLSLTSDQYKAYLEDLVEDMDI------
IRF-3 Rat       LLEMAREGGASSLK-TVDLHISNSQPISLTSDQYKACLQDLVEDMDFQATGET
IRF-3 Sheep     LVDIARQGGASSLENTVDLHISNSQPLSLTSDQYMAYLQDLAEDMDF------
IRF-3 Bovine    LVDIARQGGASSLENTVDLHISNSDPLSLTPDQYMACLQDLAEDMDF------
```

Figure 45 Sequence alignment of C-terminal tail of IRF-3 across different species with EN highlighted in blue.

The conserved phosphorylation sites and the pLxIS motif are shown in red. Residues corresponding to Glu388 and Asn389 of human IRF-3 are highlighted in blue. K381 in mouse and K383 in rat replacing E and N residues in other species are also colored blue. Other potential phosphorylation sites are indicated by the asterisks.

CHAPTER V. FUTURE STUDIES ON IRF-3 PROJECT

V.1. The induction of IFN-I in response to human and mouse IRF-3

The comparison of mouse and human phosphorylated IRF-3 structures indicated that Glu388 in human is involved in the IRF-3 dimer formation by interacting with Arg211, Lys360 and Gln356 residues through electrostatic interactions and the solvent-mediated hydrogen bond (**Figure 25**). In addition, the phosphorylated Ser386 in human interacts with more surrounding residues as compared to the phosphorylated Ser379 in mouse (**Figure 24**). We wonder if these structural differences between mouse and human IRF-3 could lead to the differential induction of IFN-I in cells.

In order to investigate how mouse and human IRF-3 induce IFN-I, luciferase reporter assay will be carried out. Specifically, dose-dependent (IRF-3 transfection: 0.1ng, 1ng, 10ng) and time-course (2h, 4h, 8h, 12h, 16h, 20h) reporter assays need to be done in HEK 293T cells to figure out how IFN-I is induced in response to different amounts of human and mouse IRF-3 at different time points.

Our data including luciferase assays, native-gel electrophoresis and size-exclusion chromatography indicated that mutations E388S and E388A in human IRF-3 indeed impaired the formation of IRF-3 dimer and affected the IRF-3 mediated signaling (**Figure 26 and Figure 27**). In addition, the structural comparison between human and mouse IRF-3 together with the C-terminal alignment of IRF-3 across different species demonstrated that due to the replacement of Glu388 and Asn389 of human IRF-3 by Lys381, this region of phosphorylated mouse IRF-3 is restructured (**Figure 25 and Figure 45**). In order to further explore how this structural change affects IRF-3 activation and IRF-mediated signaling, we wanted to mutate both Glu388 and Asn389 to a single Lysine residue in human IRF-3 and do the luciferase assays, native-gel electrophoresis and size-exclusion chromatography experiments. Likewise, we wanted to mutant a single Lys381 to Glutamic acid and Asparagine in mouse IRF-3 and do the same set of experiments.

Currently, I made a construct of full-length mouse IRF-3 in pcDNA3.1(-) vector (**Figure 46**). Also, I already mutated Lys381 to Glutamic acid and Asparagine in this vector. I also mutated Glu388 and Asn389 to a single Lysine in pcDNA3.1(-) containing full-length human IRF-3.

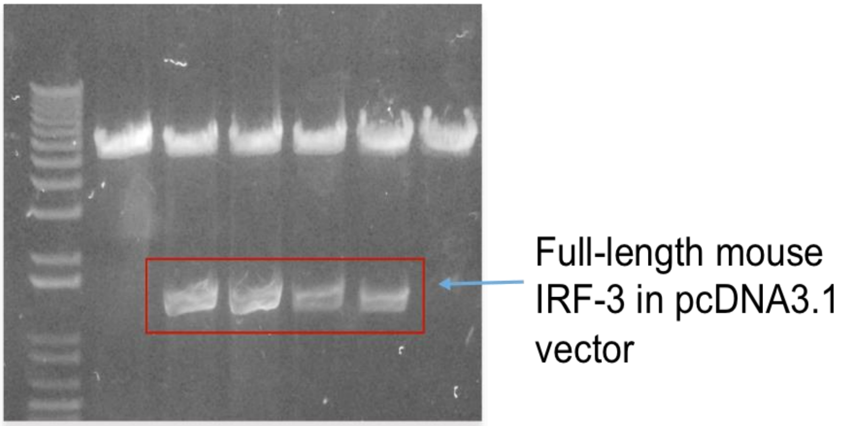


Figure 46 Full length mouse IRF-3 constructs in pcDNA3.1(-) digested by XhoI and EcoRI. The full-length mouse IRF-3 sequence has been confirmed by DNA sequencing.

V.2. Uncover the molecular mechanism of IFN-β transcription by the phosphorylated IRF3 in the presence of full-length CBP

CREB-binding protein (CBP)/p300 is a transcriptional co-activator with multi-domains that has 2442 residues in length. It has been demonstrated that CBP engages in a variety of biological functions including innate immune response, differentiation, cell cycle, apoptosis and so on. A distinct property of CBP/p300 is that it can interact with various transcription factors (TFs) via its multiple domains, which allows diverse interactions to happen between CBP/p300 and many activators(48, 75-78). There have been extensive studies on the multiple interactions between the CBP/P300 and a variety of transcription factors (TFs) assembled on the IFN-β enhanceosome(79, 80). For example, the IRF-binding domain (IBiD) in the C terminal region of CBP binds to IRF-3(81, 82); the KIX domain at the N-terminal region interacts with c-Jun and RelA(83, 84); also the CH2 domain in between IBiD and KIX domain is bound to ATF-2(85). However, in order to illuminate the mechanism of how IFN-β transcription is initiated by the full length pIRF-3 in the nucleus, two structures will have to be determined 1). The structure of pIRF-3 (full-length) in complex with full-length CBP. 2). The structure of full-length pIRF-3-DNA in complex with full-length CBP

In order to solve the structure of full-length pIRF-3 in complex with full-length CBP, human CBP (1 to 2442aa) will be expressed and purified. The purified sample will then be mixed with excess full-length pIRF-3 and incubated for several hours. The resulting complex will be purified by size-exclusion column where the excess pIRF-3 will be separated from the pIRF-3-CBP complex. The negative staining of the complex will be carried out for initial structural analysis. If the 2D average images are satisfactory, single-particle Cryo-EM will be used to generate a 3D model. Otherwise, buffer conditions will have to be adjusted to get a homogenous complex. Or the chemical cross-linking can be applied to the complex with great care. Alternatively, initial crystallization screen will be set up for the complex. If crystals with good quality can be obtained, the complex structure can be determined by x-ray crystallography using molecular replacement since several domain structures of CBP with different transcription factors are available. As an alternative, Se-Met derivative crystals will be generated for experimental phasing by SAD or MAD.

Our ultimate goal is to solve the structure of full-length pIRF-3-DNA-CBP to uncover the molecular mechanism of IFN-β transcription by the phosphorylated IRF3 in the presence of full-length CBP. In order to achieve this goal, the sample of full-length pIRF-3 in complex with full-length CBP is mixed with dsDNA in different molar ratios (6:1, 8:1, 10:1) to obtain the best ternary complex. Next the pIRF-3-DNA-CBP can be applied to the negative staining for initial structure analysis. When the complex appears to be homogenous in 2D model, single-particle Cryo-EM will be used to generate a 3D model. Otherwise, several methods can be done to improve the complex quality: 1). We can generate pIRF-3-DNA subcomplex first, and then mix it with CBP in a ratio of 2:1 or 3:1. After FPLC, the excess subcomplex will be separated from the ternary complex. 2) Chemical cross-linking can be also used as a backup. Alternatively, the complex can be used for crystallization screen under 200 conditions. If single crystals can be obtained after optimization, the complex structure can be determined by x-ray crystallography using MR since several domain structures of CBP with different transcription factors and the structure of IRF-3-DNA are available. As an alternative, Se-Met derivative crystals will be generated for experimental phasing by SAD or MAD. When the structure of the complex is available, the whole IFN-β enhanceosome model can be generated based on the previous model of CBP domains with different transcription factors as well as the model of the transcription factors with DNA.

When the structure of the pIRF-3-DNA-CBP complex is available, cell-based functional studies can be performed to validate the structure. We will design IRF-3 mutants at the interfaces of pIRF-3-DNA and pIRF-3-CBP and then test how they will be affecting the downstream IFN-β activation using luciferase reporter assay.

CHAPTER VI. PARTICIPATION IN OTHER PROJECTS*

VI.1. TRIM14 is a key regulator of the type I IFN response during mycobacterium tuberculosis infection

VI.1.1. Summary and introduction to TRIM14

Our lab had collaboration with Dr.Watson's lab on this project. This project has been published in The Journal of Immunology(86).

Tripartite motif-containing protein 14 (TRIM 14) contains a coiled-coil, a B-box, and a C-terminal PRY/ SPRY domain and has been reported to play an important role in antiviral innate immunity(87-91). For example, TRIM 14 can interact with the viral nucleoprotein (NP) and therefore inhibit replication of influenza A virus(92). Also, it can bind to hepatitis B virus X (HBx) protein to impair HBV replication(93). Moreover, TRIM 14 can inhibit the replication of HCV through the interaction with HCV encoded non-structural protein (NS5A)(94). However, there is little study on the functions of TRIM 14 during bacterial infections. Dr.Watson's lab aimed to study the roles of TRIM 14 in regulating macrophage responses to Mycobacterium tuberculosis. Their immunofluorescence microscopy data together with co-IP results indicated that TRIM14 associated with both cGAS and TBK1.

∗ Reprinted with permission from "The molecular basis of tight nuclear tethering and inactivation of cGAS" by Baoyu Zhao *et al.*, 2020. Nature, 587, 673-677, Copyright [2021] by Nature.

* Reprinted with permission from "Quantitative Yeast Genetic Interaction Profiling of Bacterial Effector Proteins Uncovers a Role for the Human Retromer in Salmonella Infection" by Kristin L. Patrick *et al.*, 2018. Cell Systems, 7, 323-338, Copyright [2021] by Cell Systems.

VI.1.2. Expression and purification of TRIM14

In order to determine if these associations in cells are physical interactions, I firstly made three constructs of TRIM 14 (full-length TRIM 14, TRIM 14 (1-246aa), TRIM 14 (247-440aa)) using a modified pET-28a (+) vector containing an N-terminal avi-His6-SUMO tag. Then I tried to express and purify these proteins in bacterial cells. Unfortunately, both full-length TRIM 14 and TRIM 14 (1-246aa) were not expressed in bacteria. Only TRIM 14 (247-440aa) was expressed and purified out of bacteria under high salt condition. In order to get full-length TRIM 14 protein, I also cloned it into pEGFP-N1 vector and tried to express and purify it from 293F cells. Unfortunately, I failed. As is shown in Figure 46, the full-length TRIM 14 was expressed well in 293F cells two days post transfection, but no matter how I purified it, it always tended to be in the precipitate instead of being in the supernatant (**Figure 47**).

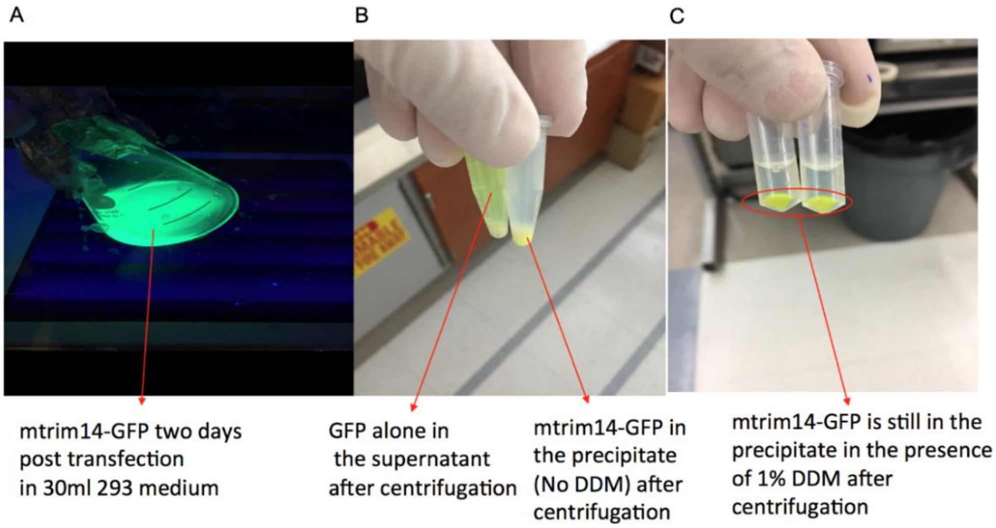


Figure 47 Mouse TRIM 14-GFP expression and purification from 293F cells.
(A) Mouse TRIM 14-GFP expression two days post transfection. (B) Purification of Mouse TRIM 14-GFP in the absence of DDM. Purification of GFP alone was a control. (C) Purification of Mouse TRIM 14-GFP in the presence of 1% DDM.

VI.1.3. Expression and purification of other proteins used for SPR

In order to do the binding studies, besides biotin-labeled sumo-mTRIM14 (247-440 aa), I also expressed and purified mouse TBK1 (1-657 aa), human TBK1 (1-657 aa), full-length human TBK1, human cGAS (157-522 aa) and mouse IRF-3 (184-419 aa) (**Figure 48 and Figure 49**).

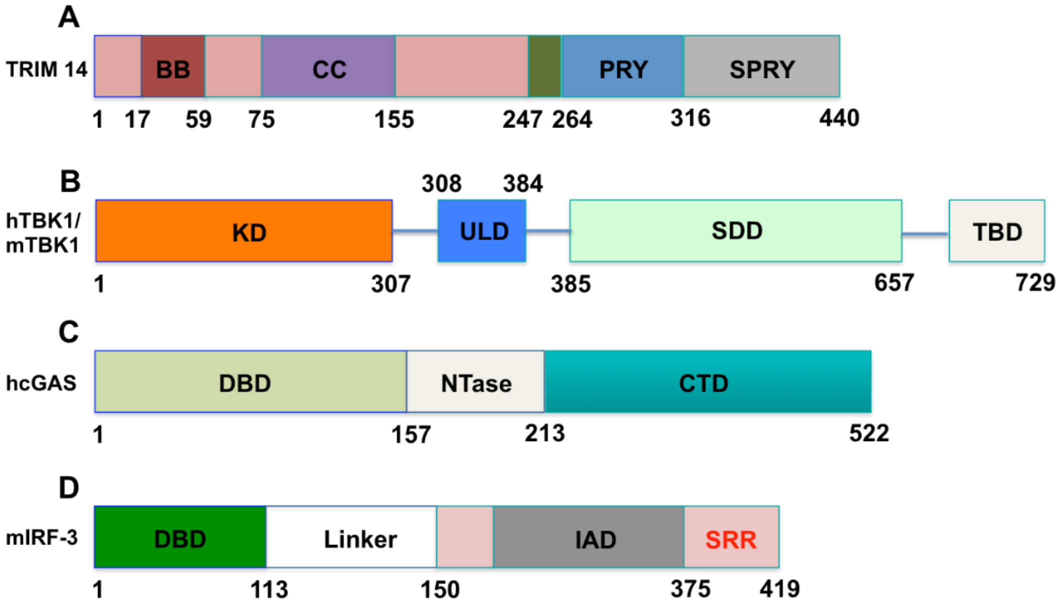


Figure 48 Schematic diagram of TRIM 14, human and mouse TBK1, human cGAS and mouse IRF-3 with their own domains.

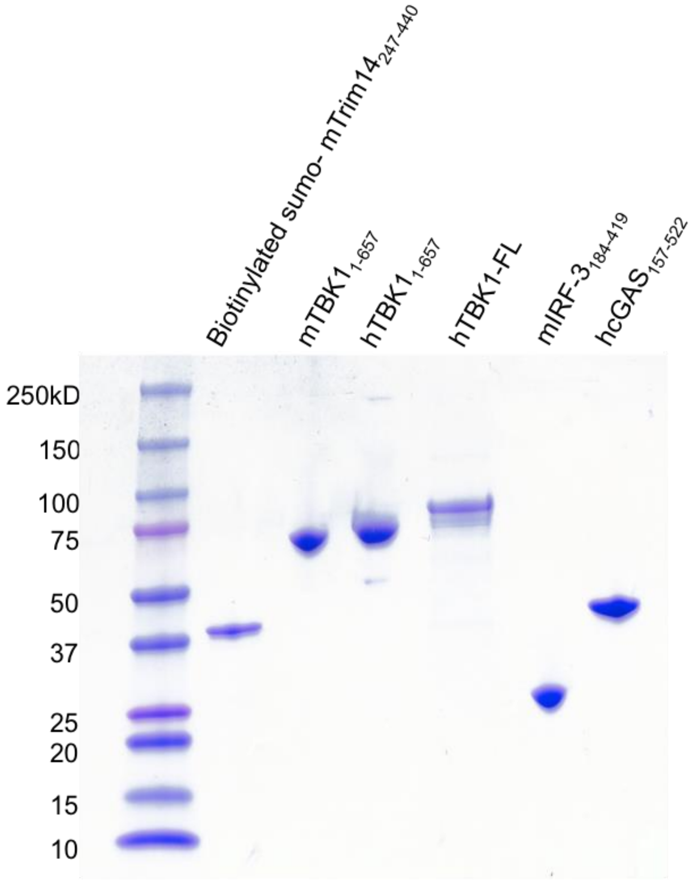


Figure 49 SDS-PAGE showing all the purified proteins for SPR studies.

VI.1.4. Surface plasmon resonance (SPR) binding studies

With all the purified proteins, we did the SPR studies. We found that TRIM 14 indeed physically interacted with both TBK1 and cGAS. Specifically, TRIM14 interacted with both mouse and human TBK1 domains at affinities of 24 μM and 42 μM respectively (**Figure 50A to D**). The equilibrium binding studies also indicated that TRIM14 directly interacted with human full-length TBK1 and human cGAS domain at affinities of 11 μM and 26 μM respectively whereas there is no binding observed between TRIM14 and mouse IRF-3 (**Figure 51A to D**). All of those data showed the evidence that there are direct associations between TRIM14 and TBK1.

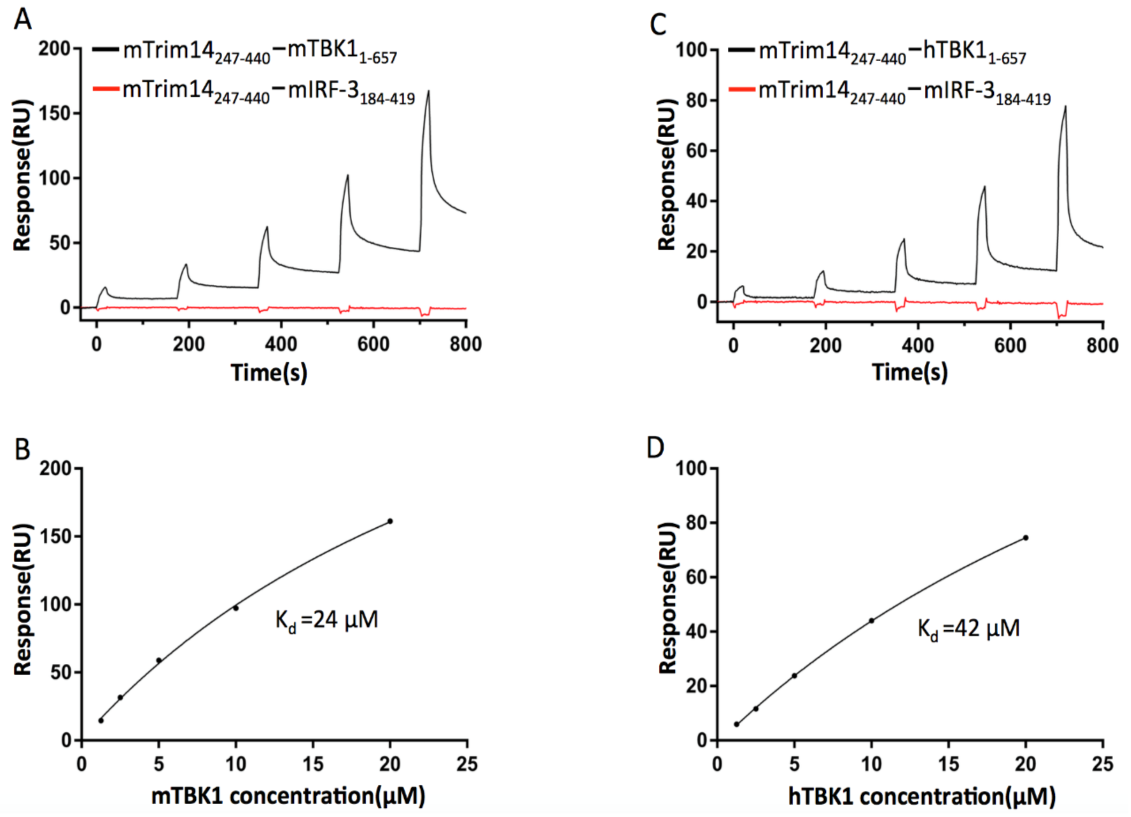


Figure 50 SPR binding studies between mouse TRIM14 (247-440aa) and TBK1 domains. (A and B) The interaction between TRIM14 (247-440aa) and mouse TBK1 domain at an affinity of 24 µM (C and D) The interaction between TRIM14 (247-440aa) and human TBK1 domain at an affinity of 42 µM. No binding is observed between TRIM 14 and mouse IRF-3.

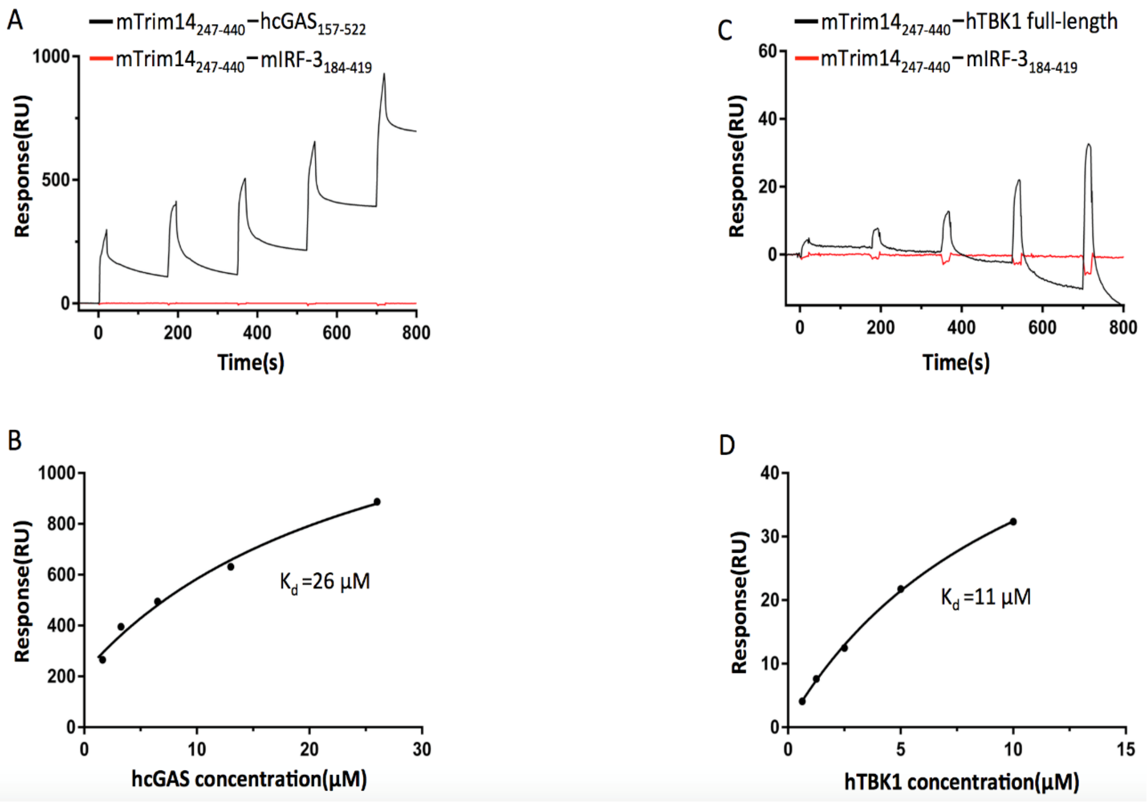


Figure 51 SPR binding studies between mouse TRIM14 (247-440aa) and full-length human TBK1 or human cGAS domain.
(A and B) The interaction between TRIM14 (247-440aa) and human cGAS domain at an affinity of 26 μM (C and D) The interaction between TRIM14 (247-440aa) and full-length human TBK1 at an affinity of 11 μM. No binding is observed between TRIM 14 and mouse IRF-3.

VI.1.5. Materials and methods in detail

The cDNA encoding mouse Trim14 (residues 247 to 440), mouse IRF-3 (residues 184-419) and human cGAS (residues 157-522) were cloned into a modified pET28(a) vector containing an N-terminal Avi-His6-SUMO tag. Sequences of the plasmids were confirmed by DNA sequencing. The BL21 (DE3) cells were co-transformed with the trim14 plasmid 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 Biotin-labelled-Avi-His6-SUMO proteins were purified using a nickel-NTA column followed by gel-filtration chromatography using a HiLoad 16/60 Superdex 75 column (GE Healthcare). For mouse IRF-3 and human cGAS, they are expressed in Escherichia coli BL21 (DE3) cells induced with 0.4 mM IPTG overnight at 16 °C and purified using a Ni2+-NTA column (Qiagen). The purified proteins were cleaved with SUMO protease at 4 °C overnight. The His6-SUMO tag was removed using a Ni2+-NTA column, and the flow-through proteins were further purified by gel-filtration chromatography using a HiLoad 16/60 Superdex 75 column (GE Healthcare). Mouse TBK1 (residues 1 to 657), human TBK1 (residues 1 to 657) and full-length human TBK1 (residues 1 to 729) were cloned into the pAcGHLTc baculovirus transfer vector. The plasmids were transfected together with Baculo-Gold bright linearized baculovirus DNA (BD Biosciences) into sf9 insect cells to generate recombinant baculovirus. The original recombinant viruses were amplified for at least two rounds before the large-scale protein expression. The insect cells at a density of 2.5 × 106 cells/ml were infected by TBK1 recombinant baculovirus and cultured at 27°C and harvested 72 hours post infection. The cells were lysed in the buffer containing 150 mM NaCl, 0.2 M Tris-HCl, 1% NP-40, 1 mM PMSF at pH 8.0. The target proteins in the supernatant were purified using nickel affinity chromatography followed by size-exclusion chromatography.

The binding studies were performed using a Biacore X100 SPR instrument (GE Healthcare). Biotin-labeled SUMO-fusion trim14 was coupled on the sensor chip SA (GE Healthcare). Dilution series of TBK1, IRF-3 (1.25, 2.5, 5, 10, 20 μM), full-length TBK1 (0.625, 1.25, 2.5, 5, 10 μM) and cGAS (1.63, 3.25, 6.5, 13, 26 μM) in 1× HBS-EP+ buffer (GE

Healthcare) were individually injected over the sensor chip at a flow rate of 30 μL/min. The single-cycle kinetic/affinity protocol was used in all binding studies. All measurements were duplicated under the same conditions. The equilibrium Kd was determined by fitting the data to a steady-state 1:1 binding model using Biacore X100 Evaluation software version 2.0 (GE Healthcare).

VI.2. The molecular basis of tight nuclear tethering and inactivation of cGAS

VI.2.1.Introduction to cGAS

This is another project I am involved in, which turned out to be published on NATURE. cGAS has been reported to be a critical cytosolic DNA sensor that mediates the induction of type I interferons that are crucial for innate immune responses to pathogen infection. Specifically, cGAS can be activated by foreign dsDNA from bacteria and viruses and then catalyzes the synthesis of a cyclic dinucleotide 2', 5' cGAMP, which eventually leads to the production of type I interferons via cGAS-STING signaling pathway(7, 14, 22, 95, 96). However, how cGAS successfully eludes self-DNA and avoids autoimmune disorders remains unknown. Currently, most researchers proposed that cGAS is localized to the cytosol and therefore cannot have access to the DNA in the nucleus(16, 30, 97). By contrast, some other researchers provide some evidence showing that cGAS is localized to the nucleus and binds to the chromatin, by which way cGAS activity is suppressed(97-99).

VI.2.2. Expression and purification of biotin-labeled human full-length cGAS

Our lab has successfully purified mouse cGAS-nucleosome complex using either purified nucleosomes from 293 cells or reconstituted nucleosomes and determined the complex structures at 4 Å and 3 Å respectively. In order to determine the binding affinity between cGAS and nucleosome, I firstly made a construct of pET28a-avi-His6-sumo-full length human cGAS and the sequence has been confirmed by DNA sequencing (**Figure 52**).

Figure 52 pET28a-avi-His6-sumo-full length human cGAS constructs digested by EcoRI and XhoI.
The full-length human cGAS sequence has been confirmed by DNA sequencing.

Then I expressed biotin-labeled full-length human cGAS by co-transforming pET28a-avi-His6-sumo-FL hcGAS with the pBirAcm plasmid coding for BirA(100, 101) and purified it under 500 mM NaCl condition. The gel filtration chromatography shows the purified cGAS

(**Figure 53A**). Since it has multiple peaks, I run each of the fractions on the SDS-PAGE gel to collect the pure cGAS (**Figure 53B**).

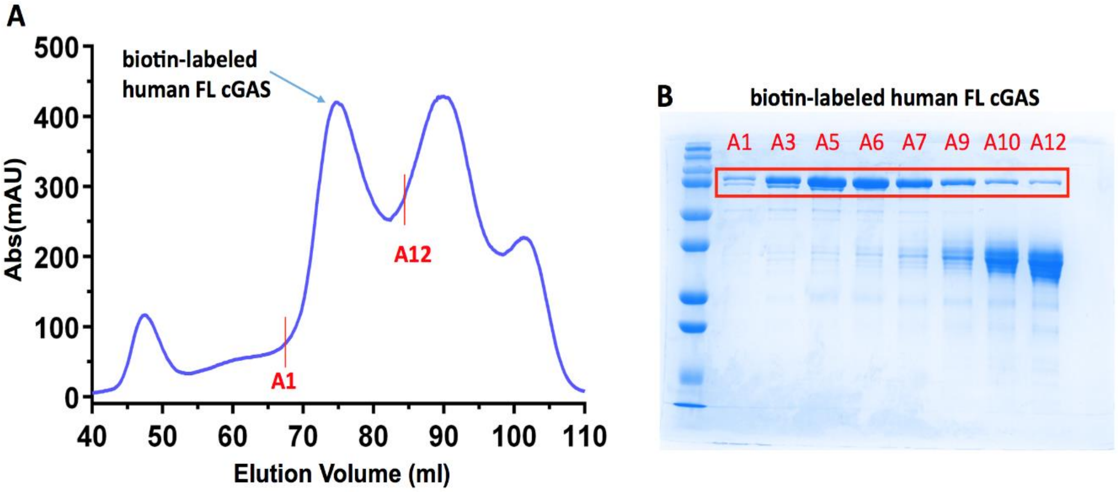


Figure 53 Expression and purification of biotin-labeled full-length human cGAS.
(A) Gel filtration showing the purified cGAS (B) SDS-PAGE showing each fraction of cGAS on the gel filtration column.

VI.2.3. Binding studies between cGAS and nucleosomes by biolayer interferometry (BLI)

To determine the binding affinity between full-length human cGAS and the nucleosomes, biolayer interferometry (BLI)-based assay using the Octet RED96 instrument(102) was carried out. The binding study indicated that full-length human cGAS binds to the nucleosome with an affinity of 3.2 nM (**Figure 54**).

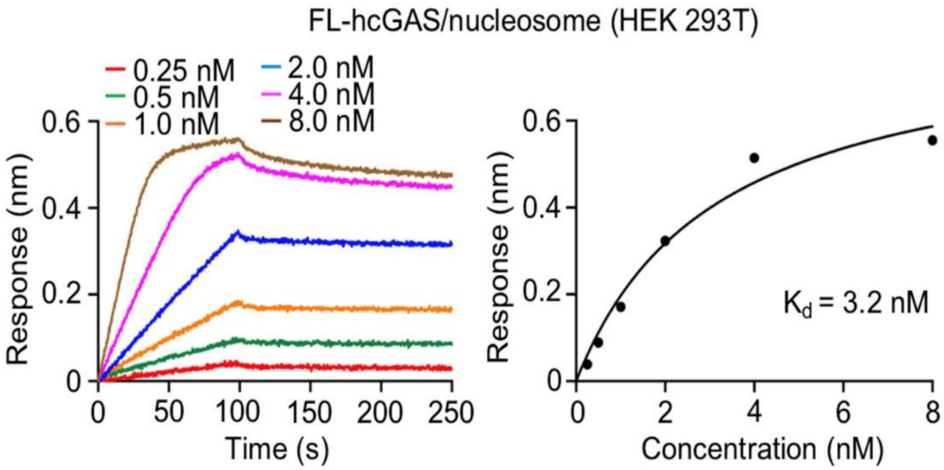


Figure 54 Biolayer interferometry (BLI)-based assay showing the interaction between full-length human cGAS and the nucleosome with a nanomolar affinity.

Besides the binding study between full-length human cGAS and the nucleosome, we also determined the binding affinities between full-length mouse cGAS-nucleosome, mouse cGAS domain-nucleosome, and human cGAS domain-nucleosome. The equilibrium binding studies demonstrated that cGAS whether it is the catalytic domain or full-length bound to the nucleosomes with similar nanomolar high affinities (**Figure 55A, B and C**).

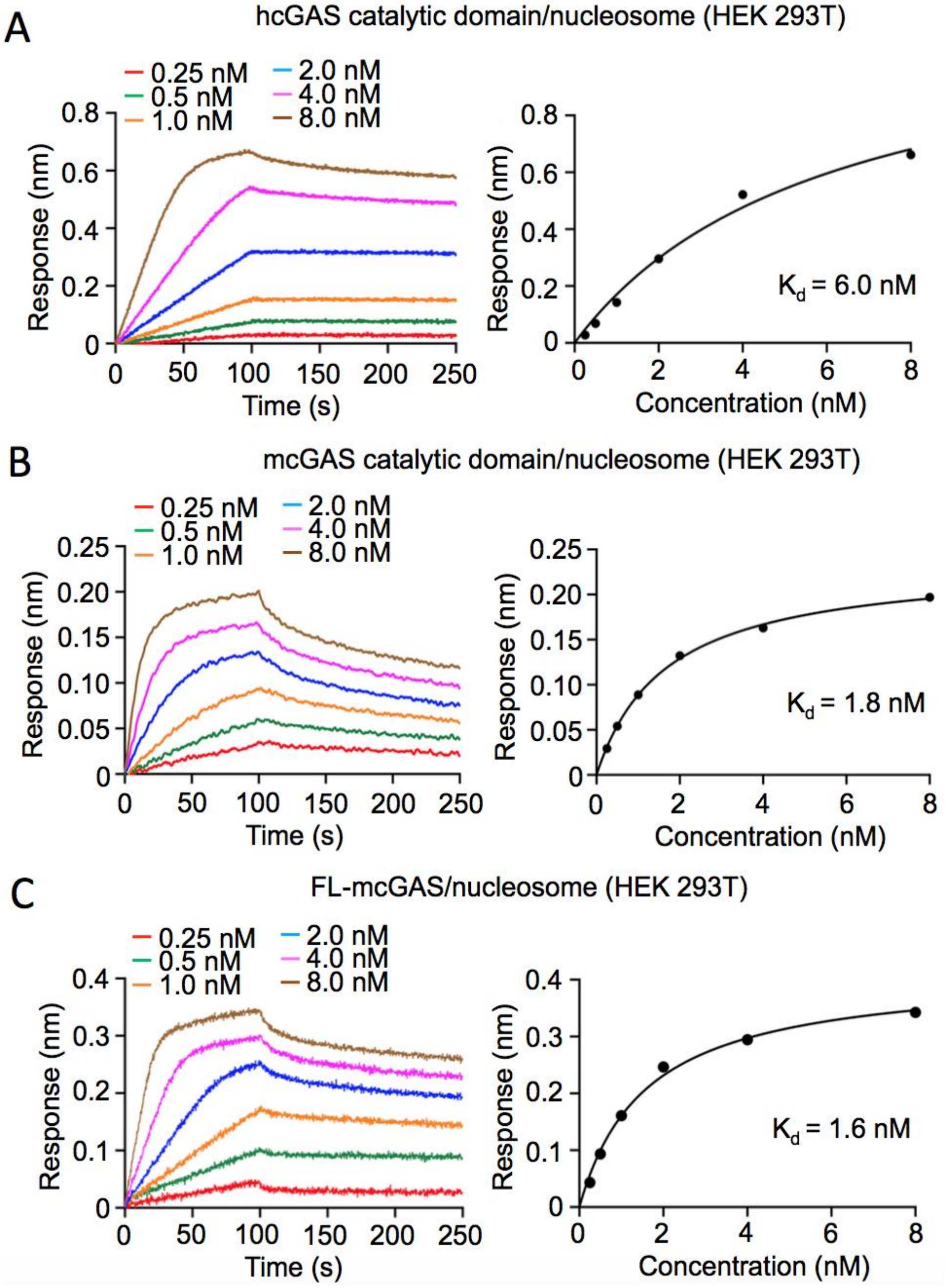


Figure 55 cGAS-nucleosome binding studies by Biolayer interferometry (BLI).
(A-C) Biolayer interferometry binding studies of human cGAS domain, mouse cGAS domain and mouse full-length cGAS with nucleosomes purified from HEK 293T cells.

VI.2.4. Generation of mutations of full-length human cGAS and expression and purification of the mutants

Based on the cGAS-nucleosome complex structures we determined, we found that a patch of positively charged residues of cGAS binds to a patch of negatively charged acidic residues formed by histones. In order to study how these positively charged residues of cGAS affect nucleosome binding, I have generated thirteen mutants of human full-length cGAS by the QuikChange site-directed mutagenesis method (R236E, R246E, K252E, K254E, R255E, K258E, K327E, K347E, R349E, K350E, R353E, K355E and K394E). Then I expressed and purified the biotin-labeled human cGAS mutants for SPR binding studies (**Figure 56A and B**). (I did not do the SPR binding studies). Based on the SPR binding studies done by Pengbiao, we found that R236E, R353E, K254E and R255E mutations abolished nucleosome binding. Besides, K350E dramatically impaired nucleosome binding whereas R349E and K347E reduced the binding affinity by only a few folds. By contrast, the binding to the nucleosome was not affected by mutations K327E, K355E, K394E, R246E, K252E and K258E.

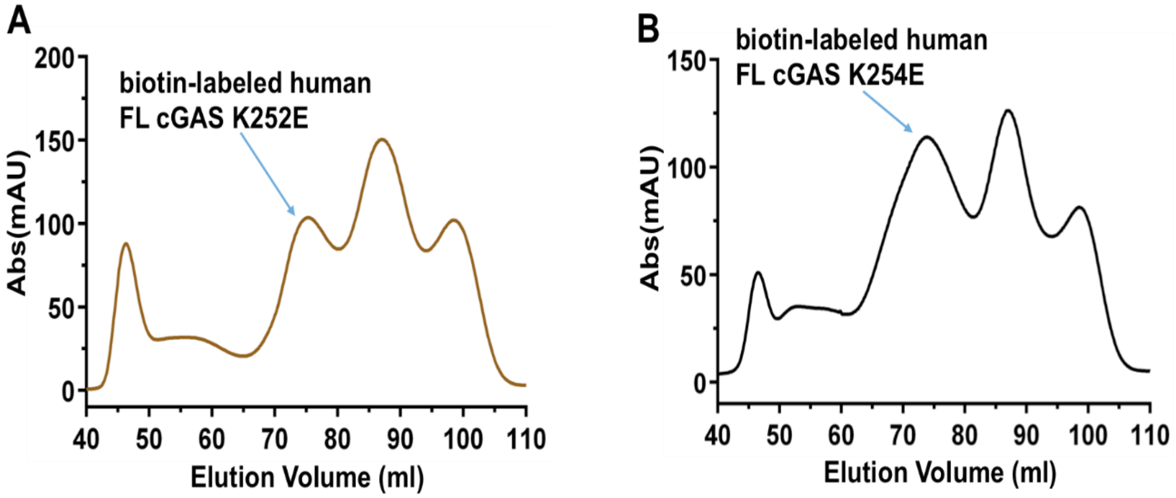


Figure 56 Size exclusion chromatography showing two mutants of biotin-labeled full-length human cGAS.
(A) biotin-labeled full-length human Cgas K252E (B) biotin-labeled full-length human cGAS K254E.

VI.2.5. Materials and methods in detail

The cDNA encoding full-length human cGAS (residues 1 to 522) was cloned into a modified pET28(a) vector containing an N-terminal Avi-His6-SUMO tag. Mutants of human cGAS were generated using the QuikChange site-directed mutagenesis kit (Agilent). Sequences of the plasmids were confirmed by DNA sequencing. The BL21 (DE3) cells were co-transformed with the cGAS plasmid 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 Biotin-labelled-Avi-His6-SUMO proteins were purified using a nickel-NTA column followed by gel-filtration chromatography using a HiLoad 16/60 Superdex 200 column (GE Healthcare).

Biolayer interferometry (BLI)-based assays using the OctetRED96 instrument (ForteBio, Inc.)(102) were performed to detect the interaction between cGAS and nucleosome. Specifically, streptavidin biosensors were loaded with 5 μg/mL biotin-labeled cGAS proteins. The cGAS-immobilized tips were dipped into 2-fold serial dilutions of nucleosome (0.25, 0.5, 1, 2, 4, and 8 nM) in 1× HBS-EP buffer supplemented with 5 mM MgCl2. The association and dissociation phases were measured for 100s and 150s respectively. The buffer control was subtracted from raw data and curves were aligned to the baseline. All data were analyzed using Octet Data Analysis 11.1 software (Forte Bio, Inc.) and the binding affinities (Kd) were determined by fitting the data to a steady-state 1:1 binding model.

VI.3. Quantitative yeast genetic interaction profiling of bacterial effector proteins uncovers a role for the human retromer in Salmonella infection

VI.3.1. Introduction and summary of this project

Our lab had collaboration with Dr.Watson's lab on this project. This project has been published on Cell Systems(103). Over millions of year's evolutions, intracellular bacterial pathogens have been able to secrete a variety of effector proteins into host cells to help them survive(104-109). However, the function of most effector proteins remains largely unknown. In this study, Dr.Watson's group developed quantitative genetic interaction profiles called E-MAP, through which they not only recapitulated some well-known effectors but also predicted an uncharacterized SseC, a component of the Salmonella SPI-2 translocon, that targeted human retromer complex. In addition, a role of SseC in regulating maintenance of the Salmonella vacuole has been validated by experiments in human cells.

The retromer complex plays an important role in trafficking proteins from endosomes to the plasma membrane or the trans-Golgi network and it contains two parts: a core of the complex: vps26-vps35-vps29 and sorting nexin proteins(110-115). Since E-MAP identified the human retromer as a target of SseC and both co-IP and mass spectrometry indicated that SseC interacted with the retromer in vivo, we would like to further examine the precise nature of this interaction.

VI.3.2. Expression and purification of SseC/Ssca and each component of retromer

Specifically, we individually expressed and purified SseC bound to Ssca, (a chaperone that helps protein folding and stability)(116) and each component of the retromer: vps35, vps26, vps29. Then we made different combinations of the retromer components using 200b analytical column (**Figure 57A to F**). Then in-vitro pull-down assay was carried out. (Data not shown because it was pengbiao who did the pull-down experiment). Surprisingly, we found that SseC binding to the retromer requires both vps35 and vps26. SseC cannot bind to any single retromer components or other combinations of retromer.

VI.3.3. Binding studies between SseC and the retromer by SPR

To further confirm this result and to determine the binding affinity between SseC and the retromer, surface plasmon resonance (SPR) was performed (**Figure 58A, B and C**). We observed that SseC only binds to the retromer when both vps35 and vps26 were present. As a control, Ssca doesn't bind to the retromer. In addition, the equilibrium binding studies indicated that SseC binds to the vps35-vps26 sub-complex with a strong affinity of 380 nM.

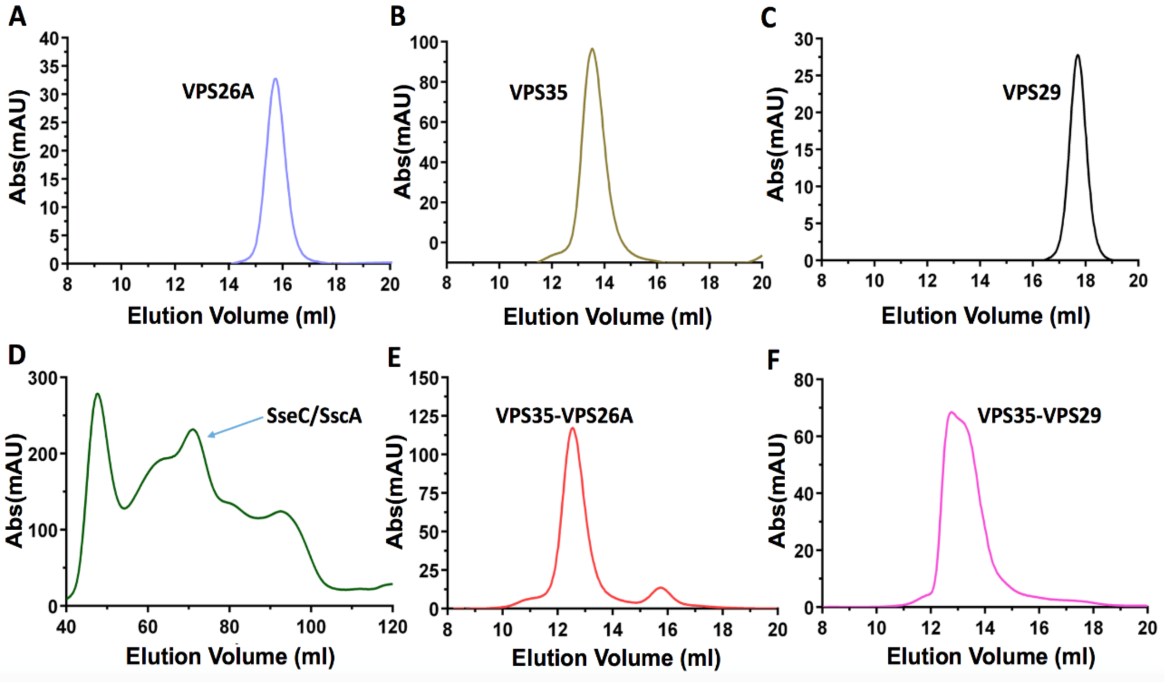


Figure 57 Gel filtration chromatography showing the purified components of the retromer and the effector protein SseC bound to SscA.

(A-C) individual components of the retromer (D) effector protein SseC bound to SscA (E and F) Sub-complexes of the retromer.

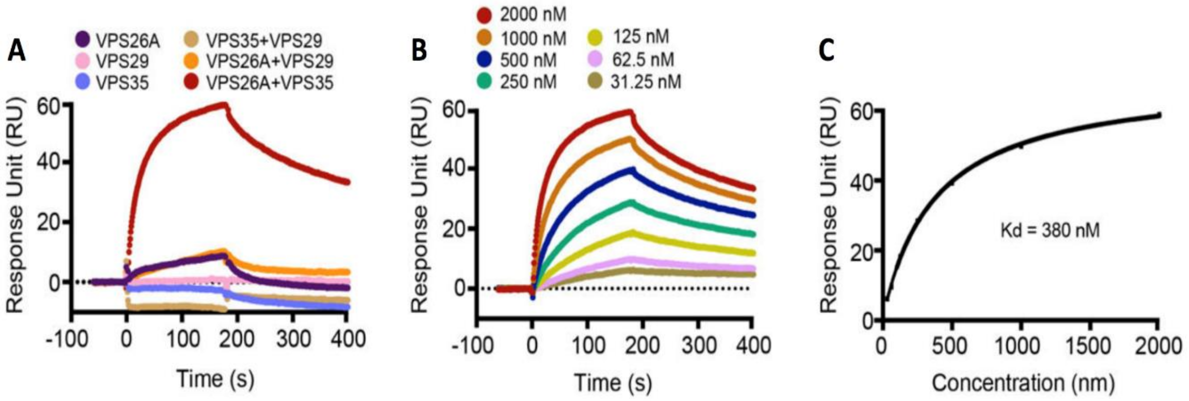


Figure 58 SPR binding studies of retromer components with the effector protein SseC. (A) The binding studies of SseC with components of retromer (B) The binding studies of SseC with VPS35-VPS26A sub-complex (C) The equilibrium binding study showing that SseC binds to VPS35-VPS26A at an affinity of 380 nM.

VI.3.4. Materials and methods in detail

Constructs of individual VPS35, VPS26A, VPS29 and ST SseC-SscA complex for protein expression were cloned into pET28a-SUMO or pET22b vectors. All the proteins were expressed in Escherichia coli strain BL21 (DE3) overnight at 15 °C. The recombinant proteins were purified by nickel affinity chromatography followed by gel-filtration chromatography. All the proteins were eluted with a buffer containing 20 mM Tris-HCl, 150 mM NaCl at pH 7.50 (buffer A). Purified proteins were flash-frozen in liquid nitrogen after concentration and stored at –80 °C.

The binding affinities between ST SseC-SscA and the retromer complexes were determined by SPR using a Biacore X100 SPR instrument (GE Healthcare). Biotin-labeled

SUMO-fusion sseC-sscA was coupled on the sensor chip using the Biotin CAPture Kit (GE Healthcare). Dilution series of retromer complex VPS35 and VPS26 (31.25, 62.5, 125, 250, 500, 100031.25, 62.5, 125, 250, 500, 2000 nM) in 1× HBS-EP+ buffer (GE Healthcare) were injected over the sensor chip at a flow rate of 30 μL/min. The multi-cycle kinetic/affinity protocol was used in all binding studies. The sensor chip was regenerated with a buffer containing 6 M guanidine hydrochloride and 0.25 M NaOH. The concentration of all controls was set to 2μM. All measurements were duplicated under the same conditions. The equilibrium Kd was determined by fitting the data to a steady-state 1:1 binding model using Biacore X100 Evaluation software version 2.0 (GE Healthcare).

CHAPTER VII. SUMMARY AND CONCLUSIONS

The innate immune system plays a key role in defending against bacterial and viral infections. Several innate immune sensing pathways, such as cGAS-STING, RIG-I like receptors, and TLR pathways, lead to the expression of type-I interferons through activation of IRF-3 via the kinase TBK1. Although the mechanism of IRF-3 activation has been extensively studied, the structural basis of IRF-3 activation upon phosphorylation remains not fully understood. In the IRF-3 project, we have determined the crystal structures of phosphorylated human and mouse IRF-3/CBP complexes for the first time. These structures reveal that phosphorylated IRF-3 forms a domain-swapped dimer via pSer386 (pSer379 in mouse) and its internal pLxIS motif. Compared to the auto-inhibited IRF-3, the C-terminal tail of phosphorylated IRF-3 unfolds and interacts with another IRF-3 molecule through extensive hydrogen bonds, hydrophobic interaction, and electrostatic interaction. The phosphorylated Ser386 reaches into a large positively charged pocket formed between two molecules of IRF-3 and contributes significantly to IRF-3 dimerization. The phIRF-3/CBP structure also shows that Ser385 is not phosphorylated but stabilizes the IRF-3 dimer, which explains why mutation of Ser385 impairs IRF-3 dimerization. In addition, Arg285 is a key residue interacting with pSer396 within the pLxIS motif. Mutation of R285 to alanine moderately impairs IRF-3 activation. These results explain why the R285Q mutation impaired IFN responses to HSV-1 infection and the R285E mutation in either HEK293 cells or fibroblasts of IRF-3-deficient mice showed much less activity upon NDV infection. Furthermore, the comparison of phosphorylated human and mouse IRF-3/CBP structures demonstrate that the mechanism of mouse IRF-3 activation is similar but distinct from human IRF-3 due to the replacement of Glu388 and Asn389 in human IRF-3 by mouse Lys381.

Based on the structure of human phosphorylated IRF-3 dimer, the biochemical and functional studies reveal that pSer386 and the residues surrounding pSer386 are critical for IRF-3 activation and downstream signaling. By contrast, phosphorylation of Ser396 within the pLxIS motif plays a moderate role in IRF-3 activation. Mutations of Ser396 and residues interacting with pSer396 only partially impair IRF-3 activation and signaling. In contrast to mouse IRF-3, the cell-based studies show that Glu388 of human IRF-3 plays additional roles in IRF-3 activation. In summary, these comprehensive structural and functional studies about IRF-3 activation upon phosphorylation advance our understanding of the molecular mechanisms of key signaling pathways in innate immunity.

In the TRIM14 project I participated in, Dr.Watson's group demonstrates that TRIM14 is a critical negative regulator of the type I IFN response in Mus musculus macrophages. They show that TRIM14 interacts with both cGAS and TBK1 and that macrophages with no TRIM14 greatly hyper-induce IFN stimulated gene expression after M. tuberculosis infection. In addition, macrophages with Trim14 knocked out show more phospho-Ser754 STAT3 relative to phospho-Ser727 and cannot upregulate the STAT3 target Socs3. These data indicate that TRIM14 plays a scaffold role between STAT3 and TBK1 to promote STAT3 phosphorylation at Ser727 and resolve ISG expression.

In the cGAS-nucleosome project I was involved in, our lab firstly demonstrates that cGAS tightly binds to nucleosomes, which potently inhibits the catalytic activity of cGAS. Later on, we solved the structure of mouse cGAS in complex with human nucleosome by cryo-

EM showing that cGAS binds to a negatively charged acidic patch that is formed by histone H2A and H2B. We found that nucleosome bound to cGAS blocks dsDNA binding and therefore maintains cGAS in an inactive conformation. In addition, mutagenesis studies together with cell-based studies indicate that mutations of cGAS that disrupt the interactions between nucleosome and cGAS greatly affect cGAS-mediated signaling.

The SseC project I collaborated on with Dr.Watson's lab generated an E-MAP, through which an uncharacterized SseC, a component of the Salmonella SPI-2 translocon, is identified. In addition, a role of SseC in regulating maintenance of the Salmonella vacuole has been validated by experiments in human cells.

REFERENCES

1. G. W. Volcheck, Overview of the Human Immune Response. *Curr Clin Pract* 10.1007/978-1-59745-315-8_1, 1-39 (2009).
2. M. R. Romo, D. Perez-Martinez, C. C. Ferrer, Innate immunity in vertebrates: an overview. *Immunology* **148**, 125-139 (2016).
3. J. Parkin, B. Cohen, An overview of the immune system. *Lancet* **357**, 1777-1789 (2001).
4. K. L. Medina, Overview of the immune system. *Hand Clinic* **133**, 61-76 (2016).
5. S. R. Paludan, A. G. Bowie, Immune sensing of DNA. *Immunity* **38**, 870-880 (2013).
6. N. Bhat, K. A. Fitzgerald, Recognition of Cytosolic DNA by cGAS and other STING-dependent sensors. *Eur J Immunol* 10.1002/eji.201344127 (2013).
7. D. L. Burdette, R. E. Vance, STING and the innate immune response to nucleic acids in the cytosol. *Nat Immunol* **14**, 19-26 (2013).
8. R. Barbalat, S. E. Ewald, M. L. Mouchess, G. M. Barton, Nucleic acid recognition by the innate immune system. *Annu Rev Immunol* **29**, 185-214 (2011).
9. C. Shu, X. Li, P. Li, The mechanism of double-stranded DNA sensing through the cGAS-STING pathway. *Cytokine Growth Factor Rev* **25**, 641-648 (2014).
10. S. Liu *et al.*, Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induces IRF3 activation. *Science* **347**, aaa2630 (2015).
11. T. S. Xiao, K. A. Fitzgerald, The cGAS-STING pathway for DNA sensing. *Mol Cell* **51**, 135-139 (2013).

12. Q. Chen, L. Sun, Z. J. Chen, Regulation and function of the cGAS-STING pathway of cytosolic DNA sensing. *Nat Immunol* **17**, 1142-1149 (2016).

13. E. J. Diner *et al.*, The Innate Immune DNA Sensor cGAS Produces a Noncanonical Cyclic Dinucleotide that Activates Human STING. *Cell Rep* 10.1016/j.celrep.2013.05.009 (2013).

14. A. Ablasser *et al.*, cGAS produces a 2'-5'-linked cyclic dinucleotide second messenger that activates STING. *Nature* 10.1038/nature12306 (2013).

15. C. Shu, G. Yi, T. Watts, C. C. Kao, P. Li, Structure of STING bound to cyclic di-GMP reveals the mechanism of cyclic dinucleotide recognition by the immune system. *Nat Struct Mol Biol* **19**, 722-724 (2012).

16. L. Sun, J. Wu, F. Du, X. Chen, Z. J. Chen, Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science* **339**, 786-791 (2013).

17. B. Zhao *et al.*, Structural basis for concerted recruitment and activation of IRF-3 by innate immune adaptor proteins. *Proc Natl Acad Sci U S A* **113**, E3403-3412 (2016).

18. X. Li *et al.*, Cyclic GMP-AMP synthase is activated by double-stranded DNA-induced oligomerization. *Immunity* **39**, 1019-1031 (2013).

19. F. Civril *et al.*, Structural mechanism of cytosolic DNA sensing by cGAS. *Nature* **498**, 6 (2013).

20. Y. Tanaka, Z. J. Chen, STING specifies IRF3 phosphorylation by TBK1 in the cytosolic DNA signaling pathway. *Sci Signal* **5**, ra20 (2012).

21. C. Shu *et al.*, Structural insights into the functions of TBK1 in innate antimicrobial immunity. *Structure* **21**, 1137-1148 (2013).

22. B. Y. Zhao *et al.*, A conserved PLPLRT/SD motif of STING mediates the recruitment and activation of TBK1. *Nature* **569**, 718-+ (2019).

23. H. Yanai, H. Negishi, T. Taniguchi, The IRF family of transcription factors Inception, impact and implications in oncogenesis. *Oncoimmunology* **1**, 1376-1386 (2012).

24. Anonymous, Interferons and IRF family of transcription factors in the regulation of immunity. *Endocr J* **52**, 16-16 (2005).

25. H. Negishi, T. Taniguchi, H. Yanai, The Interferon (IFN) Class of Cytokines and the IFN Regulatory Factor (IRF) Transcription Factor Family. *Csh Perspect Biol* **10** (2018).

26. T. Taniguchi, A. Takaoka, Type I interferon system and IRF family of transcription factors in host defense regulation. *P Jpn Acad B-Phys* **81**, 1-13 (2005).

27. T. Taniguchi, Interferon system and IRF family transcription factors in oncogenesis and immunity. *Cancer Res* **66** (2006).

28. B. Y. Qin *et al.*, Crystal structure of IRF-3 in complex with CBP. *Structure* **13**, 1269-1277 (2005).

29. B. Y. Qin *et al.*, Crystal structure of IRF-3 reveals mechanism of autoinhibition and virus-induced phosphoactivation. *Nat Struct Biol* **10**, 913-921 (2003).

30. D. B. Stetson, R. Medzhitov, Recognition of cytosolic DNA activates an IRF3-dependent innate immune response. *Immunity* **24**, 93-103 (2006).

31. S. Ning, J. S. Pagano, G. N. Barber, IRF7: activation, regulation, modification and function. *Genes Immun* **12**, 399-414 (2011).

32. W. C. Au, P. A. Moore, D. W. LaFleur, B. Tombal, P. M. Pitha, Characterization of the interferon regulatory factor-7 and its potential role in the transcription activation of interferon A genes. *J Biol Chem* **273**, 29210-29217 (1998).

33. I. Marie, J. E. Durbin, D. E. Levy, Differential viral induction of distinct interferon-alpha genes by positive feedback through interferon regulatory factor-7. *Embo J* **17**, 6660-6669 (1998).

34. T. Taniguchi, K. Ogasawara, A. Takaoka, N. Tanaka, IRF family of transcription factors as regulators of host defense. *Annu Rev Immunol* **19**, 623-655 (2001).

35. K. Honda, A. Takaoka, T. Taniguchi, Type I interferon [corrected] gene induction by the interferon regulatory factor family of transcription factors. *Immunity* **25**, 349-360 (2006).

36. R. T. Lin, Y. Mamane, J. Hiscott, Structure-function analysis of IRF-3: Localization of the C-terminal transactivation and autoinhibitory domains. *Eur Cytokine Netw* **9**, 332-332 (1998).

37. K. Takahasi *et al.*, X-ray crystal structure of IRF-3 and its functional implications. *Nat Struct Biol* **10**, 922-927 (2003).

38. M. Yoneyama, W. Suhara, T. Fujita, Control of IRF-3 activation by phosphorylation. *J Interf Cytok Res* **22**, 73-76 (2002).

39. M. Yoneyama *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* **17**, 1087-1095 (1998).

40. M. Mori *et al.*, Identification of Ser-386 of interferon regulatory factor 3 as critical target for inducible phosphorylation that determines activation. *J Biol Chem* **279**, 9698-9702 (2004).

41. M. J. Servant *et al.*, Identification of the minimal phosphoacceptor site required for in vivo activation of interferon regulatory factor 3 in response to virus and double-stranded RNA. *J Biol Chem* **278**, 9441-9447 (2003).

42. R. T. Lin, Y. Mamane, J. Hiscott, Structural and functional analysis of interferon regulatory factor 3: Localization of the transactivation and autoinhibitory domains. *Mol Cell Biol* **19**, 2465-2474 (1999).

43. D. Panne, S. M. McWhirter, T. Maniatis, S. C. Harrison, Interferon regulatory factor 3 is regulated by a dual phosphorylation-dependent switch. *J Biol Chem* **282**, 22816-22822 (2007).

44. W. Chen *et al.*, Contribution of Ser386 and Ser396 to activation of interferon regulatory factor 3. *J Mol Biol* **379**, 251-260 (2008).

45. C. M. Hickey, N. R. Wilson, M. Hochstrasser, Function and regulation of SUMO proteases. *Nat Rev Mol Cell Bio* **13**, 755-766 (2012).

46. D. Mukhopadhyay, M. Dasso, Modification in reverse: the SUMO proteases. *Trends Biochem Sci* **32**, 286-295 (2007).

47. D. Reverter, C. D. Lima, A basis for SUMO protease specificity provided by analysis of human Senp2 and a Senp2-SUMO complex. *Structure* **12**, 1519-1531 (2004).

48. M. G. Wathelet *et al.*, Virus infection induces the assembly of coordinately activated transcription factors on the IFN-beta enhancer in vivo. *Mol Cell* **1**, 507-518 (1998).

49. B. K. Weaver, K. P. Kumar, N. C. Reich, Interferon regulatory factor 3 and CREB-binding protein/p300 are subunits of double-stranded RNA-activated transcription factor DRAF1. *Mol Cell Biol* **18**, 1359-1368 (1998).

50. X. L. Ma *et al.*, Molecular basis of Tank-binding kinase 1 activation by transautophosphorylation. *P Natl Acad Sci USA* **109**, 9378-9383 (2012).

51. J. F. Clement, S. Meloche, M. J. Servant, The IKK-related kinases: from innate immunity to oncogenesis. *Cell Res* **18**, 889-899 (2008).

52. H. Hacker, M. Karin, Regulation and function of IKK and IKK-related kinases. *Sci STKE* **2006**, re13 (2006).

53. A. Larabi *et al.*, Crystal Structure and Mechanism of Activation of TANK-Binding Kinase 1. *Cell Rep* **3**, 734-746 (2013).

54. J. J. Liu, Y. D. Hu, X. Z. Wang, Optimization and Control of Crystal Shape and Size in Protein Crystallisation Process. *Comput-Aided Chem En* **31**, 1160-1164 (2012).

55. J. Wu, Z. J. Chen, Innate immune sensing and signaling of cytosolic nucleic acids. *Annu Rev Immunol* **32**, 461-488 (2014).

56. R. Lin, C. Heylbroeck, P. M. Pitha, J. Hiscott, Virus-dependent phosphorylation of the IRF-3 transcription factor regulates nuclear translocation, transactivation potential, and proteasome-mediated degradation. *Mol Cell Biol* **18**, 2986-2996 (1998).

57. H. R. Powell, T. G. G. Battye, L. Kontogiannis, O. Johnson, A. G. W. Leslie, Integrating macromolecular X-ray diffraction data with the graphical user interface iMosflm. *Nat Protoc* **12**, 1310-1325 (2017).

58. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).

59. J. Hiscott *et al.*, Triggering the interferon response: The role of IRF-3 transcription factor. *J Interf Cytok Res* **19**, 1-13 (1999).

60. M. J. Servant, N. Grandvaux, J. Hiscott, Multiple signaling pathways leading to the activation of interferon regulatory factor 3. *Biochem Pharmacol* **64**, 985-992 (2002).

61. S. Bakshi, J. Taylor, S. Strickson, T. McCartney, P. Cohen, Identification of TBK1 complexes required for the phosphorylation of IRF3 and the production of interferon beta. *Biochem J* **474**, 1163-1174 (2017).

62. C. Charlot, H. Dubois-Pot, T. Serchov, Y. Tourrette, B. Wasylyk, A Review of Post-translational Modifications and Subcellular Localization of Ets Transcription Factors: Possible Connection with Cancer and Involvement in the Hypoxic Response. *Transcription Factors: Methods and Protocols* **647**, 3-30 (2010).

63. M. C. Hung, W. Link, Protein localization in disease and therapy. *J Cell Sci* **124**, 3381-3392 (2011).

64. J. R. Davis, M. Kakar, C. S. Lim, Controlling protein compartmentalization to overcome disease. *Pharm Res-Dordr* **24**, 17-27 (2007).

65. T. Boulikas, Putative Nuclear-Localization Signals (Nls) in Protein Transcription Factors. *J Cell Biochem* **55**, 32-58 (1994).

66. N. C. Bauer, P. W. Doetsch, A. H. Corbett, Mechanisms Regulating Protein Localization. *Traffic* **16**, 1039-1061 (2015).

67. J. F. Lau, J. P. Parisien, C. M. Horvath, Interferon regulatory factor subcellular localization is determined by a bipartite nuclear localization signal in the DNA-binding domain and interaction with cytoplasmic retention factors. *P Natl Acad Sci USA* **97**, 7278-7283 (2000).

68. C. A. Jefferies, Regulating IRFs in IFN Driven Disease. *Frontiers in Immunology* **10** (2019).

69. S. Sharma *et al.*, Triggering the interferon antiviral response through an IKK-related pathway. *Science* **300**, 1148-1151 (2003).

70. K. A. Fitzgerald *et al.*, IKKepsilon and TBK1 are essential components of the IRF3 signaling pathway. *Nat Immunol* **4**, 491-496 (2003).

71. W. Suhara *et al.*, Analyses of virus-induced homomeric and heteromeric protein associations between IRF-3 and coactivator CBP/p300. *J Biochem-Tokyo* **128**, 301-307 (2000).

72. S. Q. Liu *et al.*, Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induces IRF3 activation. *Science* **347**, 1217-U1217 (2015).

73. L. L. Andersen *et al.*, Functional IRF3 deficiency in a patient with herpes simplex encephalitis. *J Exp Med* **212**, 1371-1379 (2015).

74. W. Chen *et al.*, Insights into interferon regulatory factor activation from the crystal structure of dimeric IRF5. *Nat Struct Mol Biol* **15**, 1213-1220 (2008).

75. M. Merika, A. J. Williams, G. Y. Chen, T. Collins, D. Thanos, Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. *Mol Cell* **1**, 277-287 (1998).

76. J. M. Yie, K. Senger, D. Thanos, Mechanism by which the IFN-beta enhanceosome activates transcription. *P Natl Acad Sci USA* **96**, 13108-13113 (1999).

77. H. H. Zhong, R. E. Voll, S. Ghosh, Phosphorylation of NF-kappa B p65 by PKA stimulates transcriptional activity by promoting a novel bivalent interaction with the coactivator CBP/p300. *Mol Cell* **1**, 661-671 (1998).

78. R. N. De Guzman, J. M. Wojciak, M. A. Martinez-Yamout, H. J. Dyson, P. E. Wright, CBP/p300 TAZ1 domain forms a structured scaffold for ligand binding. *Biochemistry-Us* **44**, 490-497 (2005).

79. S. Contreras-Martos *et al.*, Linking functions: an additional role for an intrinsically disordered linker domain in the transcriptional coactivator CBP. *Sci Rep-Uk* **7** (2017).

80. T. K. Kim, T. H. Kim, T. Maniatis, Efficient recruitment of TFIIB and CBP-RNA polymerase II holoenzyme by an interferon-beta enhanceosome in vitro. *P Natl Acad Sci USA* **95**, 12191-12196 (1998).

81. E. Prinarakis, E. Chantzoura, D. Thanos, G. Spyrou, S-glutathionylation of IRF3 regulates IRF3-CBP interaction and activation of the IFN beta pathway. *Embo J* **27**, 865-875 (2008).

82. S. Y. Wang *et al.*, AGO2 Negatively Regulates Type I Interferon Signaling Pathway by Competition Binding IRF3 with CBP/p300. *Front Cell Infect Mi* **7** (2017).

83. K. M. Campbell, K. J. Lumb, Structurally distinct modes of recognition of the KIX domain of CBP by Jun and CREB. *Biochemistry-Us* **41**, 13956-13964 (2002).

84. N. K. Goto, T. Zor, M. Martinez-Yamout, H. J. Dyson, P. E. Wright, Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP) - The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. *J Biol Chem* **277**, 43168-43174 (2002).

85. S. Park, M. A. Martinez-Yamout, H. J. Dyson, P. E. Wright, The CH2 domain of CBP/p300 is a novel zinc finger. *Febs Lett* **587**, 2506-2511 (2013).

86. C. T. Hoffpauir *et al.*, TRIM14 Is a Key Regulator of the Type I IFN Response during Mycobacterium tuberculosis Infection. *J Immunol* **205**, 153-167 (2020).

87. M. van Gent, K. M. J. Sparrer, M. U. Gack, TRIM Proteins and Their Roles in Antiviral Host Defenses. *Annu Rev Virol* **5**, 385-405 (2018).

88. R. Rajsbaum, A. Garcia-Sastre, G. A. Versteeg, TRIMmunity: The Roles of the TRIM E3-Ubiquitin Ligase Family in Innate Antiviral Immunity. *J Mol Biol* **426**, 1265-1284 (2014).

89. G. A. Versteeg *et al.*, The E3-Ligase TRIM Family of Proteins Regulates Signaling Pathways Triggered by Innate Immune Pattern-Recognition Receptors. *Immunity* **38**, 384-398 (2013).

90. M. G. Koliopoulos, D. Esposito, E. Christodoulou, I. A. Taylor, K. Rittinger, Functional role of TRIM E3 ligase oligomerization and regulation of catalytic activity. *Embo J* **35**, 1204-1218 (2016).

91. J. G. Sanchez *et al.*, The tripartite motif coiled-coil is an elongated antiparallel hairpin dimer. *P Natl Acad Sci USA* **111**, 2494-2499 (2014).

92. X. W. Wu *et al.*, Inhibition of Influenza A Virus Replication by TRIM14 via Its Multifaceted Protein-Protein Interaction With NP. *Front Microbiol* **10** (2019).

93. G. Y. Tan *et al.*, Identification of TRIM14 as a Type I IFN-Stimulated Gene Controlling Hepatitis B Virus Replication by Targeting HBx. *Frontiers in Immunology* **9** (2018).

94. S. S. Wang *et al.*, TRIM14 inhibits hepatitis C virus infection by SPRY domain-dependent targeted degradation of the viral NS5A protein. *Sci Rep-Uk* **6** (2016).

95. G. N. Barber, Innate immune DNA sensing pathways: STING, AIMII and the regulation of interferon production and inflammatory responses. *Curr Opin Immunol* **23**, 10-20 (2011).

96. P. Gao *et al.*, Cyclic [G(2',5')pA(3',5')p] Is the Metazoan Second Messenger Produced by DNA-Activated Cyclic GMP-AMP Synthase. *Cell* **153**, 14 (2013).

97. K. P. Hopfner, V. Hornung, Molecular mechanisms and cellular functions of cGAS-STING signalling. *Nat Rev Mol Cell Bio* **21**, 501-521 (2020).

98. H. E. Volkman, S. Cambier, E. E. Gray, D. B. Stetson, Tight nuclear tethering of cGAS is essential for preventing autoreactivity. *Elife* **8** (2019).

99. C. Zierhut *et al.*, The Cytoplasmic DNA Sensor cGAS Promotes Mitotic Cell Death. *Cell* **178**, 302-+ (2019).

100. K. Kobza, G. Sarath, J. Zempleni, Prokaryotic BirA ligase biotinylates K4, K9, K18 and K23 in histone H3. *Bmb Rep* **41**, 310-315 (2008).

101. Y. F. Li, R. Sousa, Expression and purification of E. coli BirA biotin ligase for in vitro biotinylation. *Protein Expres Purif* **82**, 162-167 (2012).

102. D. O. Apiyo, Biolayer Interferometry (Octet) for Label-free Biomolecular Interaction Sensing. *Handbook of Surface Plasmon Resonance, 2nd Edition*, 356-397 (2017).

103. K. L. Patrick *et al.*, Quantitative Yeast Genetic Interaction Profiling of Bacterial Effector Proteins Uncovers a Role for the Human Retromer in Salmonella Infection. *Cell Syst* **7**, 323-+ (2018).

104. J. D. Thiriot, Y. B. Martinez-Martinez, J. J. Endsley, A. G. Torres, Hacking the host: exploitation of macrophage polarization by intracellular bacterial pathogens. *Pathog Dis* **78** (2020).

105. J. Du *et al.*, The type III secretion system apparatus determines the intracellular niche of bacterial pathogens. *P Natl Acad Sci USA* **113**, 4794-4799 (2016).

106. C. G. Winchell, S. Steele, T. Kawula, D. E. Voth, Dining in: intracellular bacterial pathogen interplay with autophagy. *Curr Opin Microbiol* **29**, 9-14 (2016).

107. M. D. Siqueira, R. D. Ribeiro, L. H. Travassos, Autophagy and its interaction with intracellular Bacterial Pathogens. *Frontiers in Immunology* **9** (2018).

108. V. S. Harley, B. S. Drasar, B. Forrest, B. Krahn, G. Tovey, Invasion Strategies and Intracellular Growth of Bacterial Pathogens. *Biochem Soc T* **17**, 1118-1118 (1989).

109. J. A. Theriot, The cell biology of infection by intracellular bacterial pathogens. *Annu Rev Cell Dev Bi* **11**, 213-239 (1995).

110. C. Burd, P. J. Cullen, Retromer: A Master Conductor of Endosome Sorting. *Csh Perspect Biol* **6** (2014).

111. N. Attar, P. J. Cullen, The retromer complex. *Adv Enzyme Regul* **50**, 216-236 (2010).

112. L. Martinez-Nunez, M. Munson, Retro Is Cool: Structure of the Versatile Retromer Complex. *Structure* **28**, 387-389 (2020).

113. S. J. Norwood *et al.*, Assembly and Solution Structure of the Core Retromer Protein Complex. *Traffic* **12**, 56-71 (2011).

114. C. Trousdale, K. Kim, Retromer: Structure, function, and roles in mammalian disease. *Eur J Cell Biol* **94**, 513-521 (2015).

115. B. M. Collins, The Structure and Function of the Retromer Protein Complex. *Traffic* **9**, 1811-1822 (2008).

116. C. A. Cooper, D. T. Mulder, S. E. Allison, A. V. C. Pilar, B. K. Coombes, The SseC translocon component in Salmonella enterica serovar Typhimurium is chaperoned by SscA. *Bmc Microbiol* **13** (2013).