

Exploring the antiviral defense mechanisms of IFITs in human innate immunity

Zixian Li

Department of Biochemistry

McGill University

Montreal, Quebec, Canada

December 2023

A thesis submitted to McGill University in partial fulfilment of the requirements
for the degree of Doctor of Philosophy.

Table of Contents

Abstract	vi
Résumé.....	vii
Acknowledgement	ix
Preface.....	xiii
Original contribution of knowledge.....	xiv
Contribution of Authors	xvi
List of Figures	xviii
List of Tables	xx
List of Abbreviations	xxi
Chapter 1 Introduction	1
1.1 Pathogens	1
1.1.1 Worms.....	1
1.1.2 Protozoa	1
1.1.3 Fungi	2
1.1.4 Bacteria	2
1.1.5 Viruses	3
1.1.5.1 Major viral outbreaks	3
1.1.5.2 Virus classifications	4
1.1.5.3 Influenza.....	7
1.1.5.4 Coronaviruses.....	8
1.1.5.5 Hepatitis viruses	9
1.1.5.6 The genome replication and transcription mechanism of viruses	11
1.2 Innate immune responses	16
1.2.1 Host defense forces restricting viral replication	16
1.2.2 IFITs.....	17
1.2.3 IFIT antiviral modes of action	18
1.2.4 ZAP	21
1.2.5 Core set of ISGs in action	22
1.3 Functional structure RNA elements regulating mRNA translation	23
1.4 The RNA capping and 2'O-methylation pathway.....	26
1.5 Viral evasion strategies	28

1.5.1	Mimic host mRNA caps.....	29
1.5.2	Secondary structure elements	29
1.5.3	Cap-independent and IRES-dependent translation pathway.....	32
1.5.4	Viral nsp1 mediated inhibition on host mRNA	33
1.5.5	microRNA induced protection.....	36
1.5.6	Target IFITs	36
1.6	Remaining questions to be answered	37
1.7	Outline of the thesis.....	37
1.8	Bridge to Chapter 2	40
Chapter 2 A method to isolate short RNA at single nucleotide resolution using quaternary anion exchange chromatography		41
2.1	Introduction	41
2.2	Results	43
2.2.1	Separation of <i>in vitro</i> transcribed short RNA by Q anion exchange chromatography 43	
2.2.2	Separation of capped RNA from uncapped RNA.....	46
2.2.3	Crystallization of purified capped RNA with a protein complex	47
2.3	Discussion	49
2.4	Methods.....	52
2.4.1	Purification of bacteriophage T7 RNA polymerase (RNAP)	52
2.4.2	<i>In vitro</i> transcription	53
2.4.3	Q column chromatography to isolate <i>in vitro</i> transcribed short RNA	54
2.4.4	Mass spectrometry	55
2.4.5	Vaccinia capping enzyme Purification	55
2.4.6	Capping reaction	57
2.4.7	Q column chromatography to isolate short capped RNA	58
2.4.8	Crystallization	58
2.5	Acknowledgement.....	58
2.6	Supporting information	59
2.7	Bridge to Chapter 3	67
Chapter 3 Human IFIT2 shows distinct RNA binding from IFIT1 and IFIT5		68
3.1	Introduction	68
3.2	Results	70

3.2.1	IFIT2 does not bind to dsRNA.....	70
3.2.2	IFIT2 specifically binds to ssRNA	73
3.2.3	AU-rich sequences enhance IFIT2 binding to ssRNA.....	74
3.2.4	The IFIT2-RNA crystal structure.....	75
3.2.5	IFIT2 shows distinguishable RNA binding patterns from IFIT1 and IFIT5	78
3.2.6	Biological function roles of IFIT2 dimerization	82
3.2.7	Structure guided mutation analysis and translation reporter assay	86
3.3	Discussion	88
3.4	Materials and Methods	91
3.4.1	IFIT2 cloning, expression and purification.....	91
3.4.2	IFIT2 mutagenesis	93
3.4.3	IFIT1, IFIT3, IFIT5, and IFIT2/3 cloning, expression and purification.....	93
3.4.4	In vitro transcription, RNA capping and 2'O-methylation, and RNA mass spectrometry.....	94
3.4.5	Electrophoretic mobility shift assay.....	95
3.4.6	Size exclusion chromatography binding assay	95
3.4.7	Crystallization of IFIT2 with RNA.....	96
3.4.8	Isothermal titration calorimetry	96
3.4.9	In vitro transcription and translation for firefly luciferase reporter assay	97
3.4.10	Cell culture.....	97
3.4.11	Cryo-EM sample preparation, data collection and processing	98
3.4.12	Cryo-EM Modeling and refinement.....	99
3.5	Acknowledgement.....	99
3.6	Supporting information	101
3.7	Bridge to Chapter 4	115
Chapter 4 The effect of methylation and secondary structure elements on viral RNA binding by IFIT1 protein and IFIT complexes.....		116
4.1	Introduction	116
4.2	Results	118
4.2.1	The methylation effect on RNA binding affinity of IFIT1	118
4.2.2	Structure basis of 2'O-methylation on IFIT1-Cap1-RNA interactions	119
4.2.3	The secondary structure effect on IFIT1-RNA binding.....	121
4.2.4	IFIT complex synergistically binds to RNA.....	124

4.2.5	RNA bound IFIT1/2/3 complex revealed by Cryo-EM.....	125
4.3	Discussion	127
4.4	Material and Methods.....	130
4.4.1	IFIT1 cloning, expression and purification.....	130
4.4.2	IFIT1/2/3 complex purification.....	131
4.4.3	TbMTr2 purification	132
4.4.4	Cap2 methylation and electrophoretic mobility shift assay	132
4.4.5	Isothermal titration calorimetry	133
4.4.6	Crystallization of IFIT1 with capped RNA	133
4.4.7	Size Exclusion Chromatography-coupled Multi-Angle Light Scattering (SEC-MALS) 134	
4.4.8	High-Performance Liquid Chromatography (HPLC) and Mass Spectrometry	135
4.4.9	Cryo-EM sample preparation, data collection and processing	135
4.4.10	Cryo-EM Modeling and refinement.....	136
4.5	Acknowledgement.....	137
4.6	Supporting information	137
Chapter 5	General discussion and outlook.....	144
5.1	Conclusions	144
5.2	Future work	148
5.2.1	Function of IFIT5 needs to be redefined.....	148
5.2.2	Molecular basis for the IFIT2 involved translation enhancement	148
5.2.3	The destiny of IFIT-bound transcript.....	149
References		151

Abstract

The life-threatening nature of viral infections was brought to the forefront by the Covid-19 pandemic. As the first line of host defense, the innate immune system detects pathogen-associated molecular patterns (PAMPs) upon viral invasion. PAMPs, such as single-stranded and double-stranded viral nucleic acids, are sensed by specific pattern recognition receptors, which trigger signaling cascades resulting in the expression of virus-responsive type I interferons (IFNs). IFNs further induce the expression of diverse IFN-stimulated genes (ISGs) that inhibit viral replication. Amongst the most potently expressed ISGs are the interferon-induced proteins with tetratricopeptide repeats (IFITs). There are four relatively well-characterized human IFITs (each ~55 kDa): IFIT1, IFIT2, IFIT3, and IFIT5. Generally, IFITs work by binding to viral RNAs to prevent their translation.

This thesis aims to further the structural characterization of IFIT proteins initiated in our lab in hopes of providing a deeper understanding of their role in antiviral defense. The work presented in this thesis complements previous studies on the antiviral defense mechanism of IFITs. The functions of IFITs may not be well appreciated for immune healthy people to defeat pathogens in daily life. However, a deeper understanding of IFITs' role can guide treatment interventions for immunocompromised people when infection becomes more severe and hospitalization is required. For instance, the double-edged function of IFIT2 could make IFIT2 a potential therapeutic target under the circumstances when viruses take over IFIT2 to promote viral protein synthesis. IFIT complexes show stronger interactions with RNA than IFIT1 alone, which needs to be taken into consideration when designing mRNA-based vaccines to minimize innate immune responses so that vaccine mRNAs can be largely translated to trigger the production of antibodies against a certain virus.

Résumé

En tant que première ligne de défense de l'hôte, le système immunitaire inné détecte les modèles moléculaires associés aux agents pathogènes (PAMP) lors d'une invasion virale. Les PAMP, tels que les acides nucléiques viraux simples brin et double brin, sont détectés par des récepteurs de reconnaissance de formes spécifiques, qui déclenchent des cascades de signalisation entraînant l'expression d'interférons de type I (IFNs) sensibles aux virus. Les interférons induisent en outre l'expression de divers gènes stimulés par les interférons qui inhibent la réplication virale. Parmi les ISG les plus puissamment exprimés figurent les protéines induites par l'interféron avec des répétitions tétratricopeptides (IFITs). Il existe quatre interférons avec des répétitions tétratricopeptides humains relativement bien caractérisés (chacun ~55 kDa): IFIT1, IFIT2, IFIT3 et IFIT5. Généralement, les IFITs fonctionnent en se liant aux ARN viraux pour empêcher leurs traduction.

Cette thèse vise à approfondir la caractérisation structurale des protéines IFIT initiées dans notre laboratoire dans l'espoir de mieux comprendre leurs rôles dans la défense antivirale. Le travail présenté dans cette thèse complètent des études antérieures sur le mécanisme de défense antivirale des IFIT. Les fonctions des IFIT peuvent ne pas être aussi bien appréciées par les personnes immunitairement en bonne santé pour vaincre les agents pathogènes dans la vie quotidienne. Cependant, une compréhension plus approfondie du rôle des IFIT peut orienter les interventions thérapeutiques destinées aux personnes immunodéprimées lorsque l'infection s'aggrave et qu'une hospitalisation est nécessaire. Par exemple, la fonction à double tranchant d'IFIT2 pourrait faire d'IFIT2 une cible thérapeutique potentielle dans les circonstances où les virus prennent le relais d'IFIT2 pour favoriser la synthèse des protéines virales. Les complexes IFIT présentent des interactions plus fortes avec l'ARN que IFIT1 seul, ce qui doit être pris en compte lors de la conception de vaccins à base d'ARNm afin de minimiser les réponses immunitaires innées afin

que les ARNm des vaccins puissent être largement traduits pour déclencher la production d'anticorps contre un certain virus.

Acknowledgement

I would like to thank my supervisor, Dr. Bhushan Nagar, for offering me the opportunity to learn and work in the lab, especially considering my very limited wet lab experience when I joined his group. I am so indebted to his continual support and robust patience. The only stress I need to persevere is from the work itself; thankfully, I don't need to endure unnecessary pressure as his student. I appreciate his simple personality and straightforward character/style over the years. I would also like to thank previous members of the lab, especially Heidi and Yazan, to give me training at the beginning. I am also grateful to Katalin, Alexei, Danilo, and Garvit for their kindness and support. Particularly, I had many discussions with Katalin about experiments, family, and life on many early mornings when we two early birds were in the lab. I received invaluable advice and encouragement from her; I hope my deep listening to her struggles dealing with some ridiculous invoices and orders could also comfort her to some extent. Furthermore, I want to extend my gratitude to the following:

To our collaborator Dr. Khanh Huy Bui, for guidance on cryo-EM data processing and model building.

To the Schmeing lab, for letting me join many synchrotron trips with them.

To our facility manager, Kim, for teaching me ITC and much help on ITC data processing.

To our collaborators Dr. Jerry Pelletier (deceased) and Francis Robert, for several discussions in Jerry's office and Francis's big effort to carry on the experiment after Jerry left.

To my RAC members, Dr. Kalle Gehring, Dr. Jason Young (deceased), and Dr. Selena Sagan, for their advice.

To Dr. Kelly Sears and Dr. Kaustuv Basu, for their training on operating electron microscopes.

To Venkat Mahadevan, for setting up cryoSPARC on the Compute Canada clusters and troubleshooting many program errors and cluster bugs.

To John and Lloydetta, for their encouragement and support throughout the journey, especially, John carefully proofread the initial thesis.

To Brother Marc-André Cantin-Proulx, for his help on the French translation of the thesis abstract.

Dedicated to my wife, parents, sister, and church family...

LORD, make me to know my end, and what is the measure of my days, that I may know how frail I am.

Psalm 39: 4

To everything there is a season, a time for every purpose under heaven.

Ecclesiastes 3: 1

Through the purification of biomacromolecules, my heart is purified; through the refinement of protein-RNA structures, my character is refined...

Preface

This is a manuscript-based thesis comprising one manuscript submitted for publication, one manuscript to be submitted upon adding several small-scale experiments, and one manuscript currently in preparation.

Chapter 2

A method to isolate short RNAs at single nucleotide precision using quaternary anion exchange chromatography, Li, Zixian; Bilic, Mia; Nagar, Bhushan (submitted)

Chapter 3

Human IFIT2 shows distinct RNA binding from IFIT1 and IFIT5, Li, Zixian; Robert, Francis; Munro, Kim; Pelletier, Jerry; Bui, Khanh Huy; Illes, Katalin; Nagar, Bhushan (to be submitted)

Chapter 4

The effect of methylation and secondary structure elements on viral RNA binding by IFIT1 protein and IFIT complexes, Li, Zixian; Abbas, Yazan; Munro, Kim; Yang, Sky; Bui, Khanh Huy; Nagar, Bhushan (in preparation)

Original contribution of knowledge

Chapter 2:

I re-engineered the tag of the original capping enzyme plasmid, solved the problem of removing the co-purified contaminants, established an expression and purification protocol for generating homemade high quality capping enzyme. I developed a chromatography purification method to obtain highly homogeneous short RNA for structure and biochemical studies. The quality of the short RNA, purified by the proposed method, was verified by solving a crystal structure of the RNA bound protein.

Chapter 3:

I tested IFIT2 binding with various single-stranded (ss) and double-stranded (ds) RNAs using binding assays and clarified a misleading belief about RNA binding by IFIT2. IFIT2 does not show specific interaction with dsRNA but specifically binds to ssRNA. I did a series of ITC binding affinity measurement, which showed the effect of RNA with different 5' ends (Cap0, PPP-, P-, and HO-) on interacting with the three RNA binding IFITs for the first time. I also quantified the binding affinity for IFIT2 with AU-rich RNA and non AU-rich RNA for the first time, demonstrating the AU-rich sequence preference by IFIT2. Further, I determined the first RNA-bound IFIT2 molecular structure. Using viral RNA sequence, binding assay, and cryo-EM, I revealed a different binding mode from that observed in the crystal structure, which explained the biological function of IFIT2 domain-swapped dimerization.

Chapter 4:

I acquired a set of binding affinity comparison of IFIT1 with short Cap0-, Cap1-, and Cap2-RNA, respectively, using ITC, which showed how much Cap2 methylation significantly weakens IFIT1-RNA binding for the first time. I further determined the first Cap1-RNA bound IFIT1 structure, demonstrating the molecular basis for the weakened IFIT1-RNA interaction caused by Cap1-methylation. I performed binding studies for IFIT1 and IFIT1/3 as well as IFIT1/2/3 complex with Zika 5'UTR RNA with the same folding and methylation status as the real Zika mRNA, which showed the effect of Cap1 methylation in the presence of secondary structure elements on RNA binding by IFIT1 and IFIT complexes for the first time. The synergy of IFIT1/2/3 complex observed in binding assays was further revealed by cryo-EM for the first time. The cryo-EM model clearly showed the assembly of RNA bound IFIT1/2/3 complex and different roles contributing to the synergy of this complex: the primary RNA binding role of IFIT1, the modulatory role of IFIT3 to enhance RNA binding by the central player IFIT1, and the auxiliary (secondary) RNA binding role of IFIT2 to stabilize the bound viral transcript.

Contribution of Authors

Chapter 2

I performed all the experiments, except that RNA LC-MS was run by Wahba at the mass spectrometry facility of McGill University and Bilic involved the small-scale expression trials for the re-engineered vaccinia capping enzyme. I designed the experiments and wrote the manuscript with input from Nagar.

Chapter 3

We received pCDNA3 plasmids of NP and juggled NP as a gift from Mehle. I subcloned WT IFIT2 and mutants into pCDNA3 plasmids and passed them to our collaborator Robert from the previous Pelletier group. Robert optimized the reporter assay, quantified the translation level, ran western blot, processed the data, and made the figures for the reporter assay. Munro and I designed ITC experiments. Munro trained me to use the ITC machines and processed ITC data. Wahba did RNA LC-MS. Nagar oversees the project and guides the experiments. I carried out all other biochemical and structural experiments and wrote the manuscript with advice from Nagar.

Chapter 4

Abbas guided me to do the subcloning, co-expression, and purification of the IFIT2/3 heterodimer. Yang from the Bui group helped check cryo-EM grids I prepared before I became an independent

user of the TF20 microscope. Nagar and I conceived the project. Nagar oversees the project and advises the direction of experiments. Munro processed ITC data. Wahba did RNA LC-MS. I carried out all the biochemical and structural experiments. I wrote the current version of the manuscript.

List of Figures

Figure 1.1 Virus classifications and examples.....	7
Figure 1.2 An example of virus induced DMV.	12
Figure 1.3 The channels connecting DMV and cytosol.....	14
Figure 1.4 TPR motifs in IFITs.....	18
Figure 1.5 The structure and function of IFITs.....	20
Figure 1.6 Cis-acting functional structure RNA elements regulating mRNA translation.	26
Figure 1.7 The mRNA capping and methylation pathway.	28
Figure 1.8 The secondary structure models of viral RNAs determined through enzymatic probing.	32
Figure 1.9 HCV IRES binds to human 40S ribosome.	33
Figure 1.10 SARS-CoV-2 nsp1 mediated translation inhibition.	35
Figure 2.1 Purification of 5'PPP-GUAUA by Q column chromatography.	44
Figure 2.2 Purification of 10-nt 5'PPP-GGUAGAAUAU by Q column chromatography.	45
Figure 2.3 Purification of 5'PPP-GUAUA capping reaction by Q column chromatography.	47
Figure 2.4 Crystallization test using 7mG-GUAUA isolated by Q column chromatography.	48
Figure 3.1 Gel shift assay of in vitro transcribed 5'PPP-RNA with IFIT2 and IFIT5.	72
Figure 3.2 The qualitative and quantitative determination of the IFIT2-RNA interactions.	75
Figure 3.3 The crystal structure of RNA-bound IFIT2.....	77
Figure 3.4 The effect of RNA 5'end on IFITs-RNA binding.	79
Figure 3.5 IFIT2 sequesters 5' end of short RNA.	80
Figure 3.6 The interaction of IFIT2 with WNV 5' UTR.....	83
Figure 3.7 The cryo-EM model of an RNA bound by both monomers in IFIT2.	85

Figure 3.8 A firefly luciferase reporter assay showing the effect of WT IFIT2 and mutants on translation stimulation.....	88
Figure 4.1 The 2'O methylation effect on IFIT1-RNA binding.	119
Figure 4.2 The molecular basis of 2'O methyl group weakening IFIT1-RNA interactions.	121
Figure 4.3 The 5'UTR of Zika mRNA and its interaction with IFITs.....	124
Figure 4.4 The 5'UTR of IAV NP mRNA and its interaction with IFITs.....	125
Figure 4.5 The cryo-EM model of RNA-bound IFIT1/2/3 complex at ~3.9 Å.	126

List of Tables

Table 1.1.1 The common viruses infecting humans.	7
Table 1.2 Other main ISGs and their mechanisms of actions on viral inhibition.	23

List of Abbreviations

7mG: 7-methylguanosine

AIDS: acquired immunodeficiency syndrome

APAF1: apoptotic peptidase activating factor 1

βME: β-mercaptoethanol

CCR4-Not: Carbon Catabolite Repression—Negative On TATA-less

circ-ZNF609: circular-zinc-finger protein 609

c-Jun : Jun proto-oncogene

CLAM-Cap-seq: CircLigase-assisted mapping of caps by sequencing

CLIP-seq: Crosslinking immunoprecipitation sequencing

CRISPR: clustered regularly interspaced short palindromic repeats

CrPV: cricket paralysis virus

cryo-EM: cryo-electron microscopy

cryo-ET: cryo-electron tomography

CTF: contrast transfer function

Dcp1a-Dcp2: decapping enzyme complex

DdRp: DNA-dependent RNA polymerase

DEAE: diethyl-aminoethyl

DM: double mutant

DMV: double membrane vesicles

DNA: deoxyribonucleic acid

ds: double-stranded

DTT: dithiothreitol

eEF2: eukaryotic translation elongation factor 2

eIF: eukaryotic translation initiation factor

ER: endoplasmic reticulum

FMRP: fragile X mental retardation protein 1

GMP: guanosine monophosphate

GTase: guanylyltransferase

GTP: guanosine triphosphate

GTP: Guanosine triphosphate

HA: hemagglutinin

HCV: Hepatitis C virus

HEV: Hepatitis E virus

HIV: human immunodeficiency virus

HPLC: high-performance liquid chromatography

HSP70: heat shock protein 70

IAV: influenza A virus

IFITs: interferon-induced proteins with tetratricopeptide repeats

IFNG: interferon gamma

IFNs: interferons

IPTG: isopropyl β -D-1-thiogalactopyranoside

IRESs: Internal ribosome entry sites

IRF: interferon regulatory factors

ISGs: interferon-stimulated genes

ITC: isothermal titration calorimetry

IVT: in vitro transcription

JAK: Janus kinases

K⁺: potassium ion

kb: kilobase

LncRNAs: long non-coding RNAs

m6A: N6-methyl adenosine

MAVS: mitochondrial antiviral signaling

MDA5: melanoma differentiation-associated protein 5

miRNA: microRNA

mRNA: messenger RNA

N: nucleocapsid

N7-MTase: N7 methyltransferase

NA: neuraminidase

NP: nucleoprotein

NP: nucleoprotein

NRAS, N-ras proto-oncogene

nsp: non-structure protein

nt: nucleotides

ODC: ornithine decarboxylase

P: phosphorylation

PA: polyprotein A

PAGE: polyacrylamide gel electrophoresis

PAMPs: pathogen-associated molecular patterns

Pan2-Pan3: poly(A)-nuclease (PAN) deadenylation complex

PARN: poly(A) specific ribonuclease

PB1: polyprotein B1

PB1-F2: polymerase basic protein 1-F2

PB2: polyprotein B2

PDB: Protein Data Bank

PKR: protein kinase R

PKR: protein kinase RNA-activated

PKs: pseudoknots

poly(A): poly-adenylate

PP2AC (also known as PPP2CA): protein phosphatase 2 catalytic subunit alpha

PPi: pyrophosphate

PPP: triphosphate

PRRs: pattern recognition receptors

Q: quaternary

RBPs: RNA-binding proteins

RG4: G-quadruplexes

RIG-I: retinoic acid-inducible gene I

RNA: Ribonucleic acid

RNAP: RNA polymerase

rNTPs: ribonucleoside triphosphates

RT: reverse transcriptase

RTPase: RNA triphosphatase

SAM: S-Adenosyl methionine

SARS: severe acute respiratory syndrome

SARS-Cov-2: severe acute respiratory syndrome coronavirus 2

SEC-MALS: Size exclusion chromatography coupled with multi-angle light scattering

SL: stem-loop

ss: single-stranded

STAT: signal transducer and activator of transcription proteins

SUMO: small ubiquitin-like modifier protein

TBK1: TNFR associated factor family member-associated NF- κ B activator-binding kinase 1

TCEP: Tris (2-carboxyethyl) phosphine

TNFR: tumor necrosis factor receptor

TPR: tetratricopeptide repeats

tRNA: transfer RNA

Uchl1: ubiquitin carboxyl-terminal hydrolase L1

Uchl1AS (commonly known as Uchl1os): ubiquitin carboxy-terminal hydrolase L1, opposite strand

Ulp1: Ubiquitin-like-specific protease 1

uORFs: upstream open reading frames

UTR: untranslated region

WNV: West Nile Virus

Xrn1: 5'-3' exoribonuclease 1

ZAP: Zinc-finger antiviral protein

ZC3HAV1: Zinc finger CCCH-type antiviral protein 1

Chapter 1 Introduction

1.1 Pathogens

Pathogens are organisms that cause infectious diseases. In general, pathogens are classified into five categories: worms, protozoa, fungi, bacteria, and viruses. Worms and protozoa are usually grouped as parasites in the discipline of parasitology, whereas fungi, bacteria, and viruses are the subject of microbiology¹. In the following sections, I will give a brief description of each type of pathogen and one typical example associated with it.

1.1.1 Worms

Worms include intestinal worms, tissue worms, blood and liver worms. Intestinal worms are one of the main parasites in humans. Intestinal parasites are caused by poor sanitation of food and water and poor hygiene. Children and elderly, as well as immunocompromised groups, such as human immunodeficiency virus (HIV) or acquired immunodeficiency syndrome (AIDS) population, are most vulnerable to intestinal worms. Medicines such as albendazole and mebendazole are effective in treating intestinal worms. These medications are widely used to treat millions of people across the globe are infected with intestinal worms without risk to life².

1.1.2 Protozoa

Protozoa are another main parasite in humans. The well-known protozoan parasite is called Plasmodium which causes malaria, a mosquito-borne infectious disease. An estimated 230 million malaria cases occur globally per year. More than 600,000 people die of malaria annually. Most of them are children under five years old, accounting for 70-80% of total deaths. Several medications have been developed to treat malaria effectively, including mefloquine, doxycycline,

atovaquone/proguanil, and artemisinin³. As of 2023, only two vaccines are licensed for malaria, namely RTS, S/AS01 and R21/Matrix-M. They are both recombinant protein-based vaccines, despite modest efficacy with less than 80% protection rate. Vaccination reduces the rate of severe malaria by around 20%⁴. Future work on the prevention of malaria disease focuses on the known drug resistance challenges and more effective vaccine development.

1.1.3 Fungi

Many fungi are parasites on plants. A few of them can cause serious diseases in humans, especially for immune comprised people. The human fungi pathogens mainly include *Cryptococcus neoformans*, *Candida auris*, *Aspergillus fumigatus* and *Candida albicans*, which are listed as four critical pathogen threats to public health⁵. *Cryptococcus neoformans* is a pathogenic yeast that lives in both plants and animals. After inhaling *Cryptococcus neoformans* fungi cells through the respiratory route, the infection causes Cryptococcosis which initially affects the lungs but can spread to the central nervous system and blood. *Cryptococcus neoformans* primarily target immunocompromised groups, such as HIV patients, though a healthy individual can also be infected. Globally, around 150,000 *Cryptococcus neoformans* infected cases occur every year, and ~110,000 deaths among the infected with HIV/AIDS, with a mortality rate range of ~40-60%⁶. Several medications are available for cryptococcosis treatment, including fluconazole, amphotericin B, and flucytosine⁵.

1.1.4 Bacteria

Most bacteria in the gut are beneficial to humans, but a few of them are harmful and cause infectious diseases, such as tuberculosis, cholera, syphilis, anthrax, leprosy, tetanus and bubonic plague. Among all bacteria pathogenic diseases, tuberculosis is the top killer, which is caused by

a bacterium called *Mycobacterium tuberculosis* that mainly affects the lungs. Worldwide, about 1.6 million people die of tuberculosis, where more than 10% of deaths are HIV-associated tuberculosis patients. The estimated total infected population is more than 10 million annually. Tuberculosis is preventable and treatable; several antibiotic medications are available for treating tuberculosis, including isoniazid, rifampin, pyrazinamide, ethambutol, and streptomycin. Due to inappropriate usage of medications, incorrect prescriptions, prematurely stopping medications, and usage of poor-quality drugs, multi-drug resistant tuberculosis poses a long-term challenge for treatment⁷.

1.1.5 Viruses

Viruses infect a wide range of host living cells, including bacteria, archaea, plants, animals, and humans. Several major viral outbreaks have taken place over the past century.

1.1.5.1 Major viral outbreaks

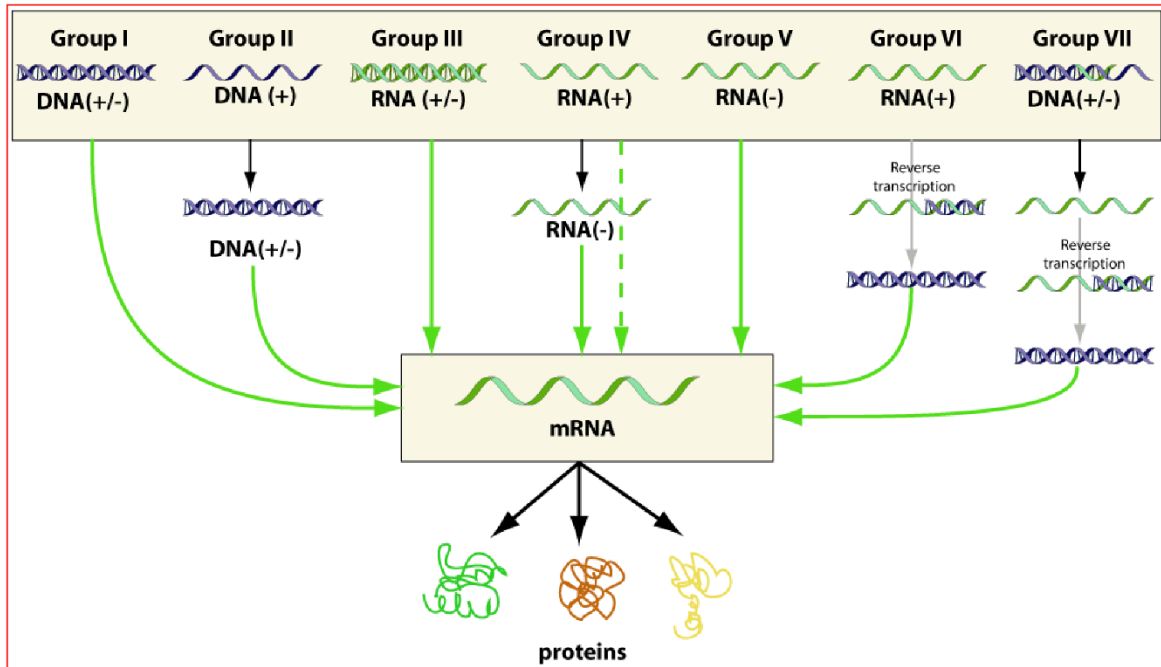
The most devastating viral outbreak is the Spanish flu in 1918-1920 caused by the H1N1 influenza A virus (IAV). Nearly one third of the world's population was infected, and an estimated 50-100 million people died in the global influenza pandemic. In the following years, the Asian flu caused by H2N2 IAV in 1957-58 originated in Guizhou, China, and an estimated 1-2 million people died across the world⁸. About ten years later, the H3N2 IAV descended from H2N2 causing the Hongkong flu in 1968. An estimated 1 million people died globally⁹. The same strain of Spanish flu reappeared and caused the Russian flu in 1977. An estimated 700,000 people died worldwide¹⁰. In 2009-2010, the H1N1 swine flu, originating in Mexico, killed roughly 200,000 people globally¹¹. Apart from Influenza outbreaks, around 28,000 infection cases were confirmed and 11,300 people died, with a fatality rate of ~40% in the 2014-2016 Ebola outbreak in West Africa¹². The 2015-

2016 Zika epidemic took place mainly in Central and South America with around 310,000 people confirmed with Zika infection in Brazil, and ~1950 confirmed cases of microcephaly in babies¹³. At the end of 2019, the Covid-19 pandemic started in Wuhan, China, and spread all over the world. The pandemic was caused by a novel betacoronavirus, namely severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The genome sequence identities between SARS-CoV-2 and two previous coronaviruses, SARS-CoV and MERS-CoV, are 79% and 50%, respectively¹⁴. The closest relative of SARS-Cov-2 is a bat coronavirus found in bats of the genus *Rhinolophus affinis* living in the Yunnan province of China, with a sequence identity of 96.1%¹⁵. At the time of writing this thesis, ~772 million cases infected with Covid-19 have been confirmed and nearly 7 million people's lives have been lost¹⁶.

1.1.5.2 Virus classifications

Depending on the genetic material encapsidated inside a virion, i.e., viral genome, viruses are simply classified into DNA viruses and RNA viruses. Contingent on how genetic information flows during viral replication, viruses have been further categorized into seven classes based on the classical Baltimore classification (**Figure 1.1**). 1) Double-stranded (ds) DNA viruses transcribe messenger RNA (mRNA) using the sense strand +DNA as template either by host- or virus-encoded DNA-dependent RNA polymerase (DdRp). 2) Single-stranded (ss) positive-sense +DNA viruses synthesize the anti-sense strand –DNA and generate dsDNA, then transcribe mRNA as dsDNA viruses do. 3) dsRNA viruses synthesize positive-sense +RNA as mRNA using the dsRNA as template by virus-encoded RNA-dependent RNA polymerase (RdRp). 4) For ss +RNA viruses, the positive-sense genomic RNA can be directly used as mRNA to initialize translation of viral proteins at the early stage of replication. They synthesize the anti-sense strand –RNA first and then synthesizes mRNA based on –RNA by RDRP. The positive-sense RNA viruses account for more

than one-third of viruses¹⁷. 5) ss negative-sense –RNA viruses transcribe mRNA using the complementary –RNA as a template by their own RDRP. 6) For ss +RNA reverse transcriptase (RT) viruses, through reverse transcription by virus-encoded RT, the +RNA is the template to make intermediate dsDNA and then transcribe mRNA from dsDNA. 7) For ds DNA RT viruses, due to a gap in one strand, they use reverse transcription to make the circular genome DNA from the pre-genomic RNA in the capsid. After replication, the replicated genome needs to be sent to the host nucleus for further repair to be a complete dsDNA genome that can be used for transcribing viral mRNAs¹⁸. Some representative examples are described in **Figure 1.1**. Unlike the Covid-19 pandemic or Zika epidemic, seasonal flu caused by the influenza virus occurs every year and it is common that we may not even be aware that up to 650,000 people's lives are lost every year¹⁹. Other common viruses we often hear of, such as hepatitis, HIV, and Papillomaviruses, cause significant deaths each year (**Table 1.1**). In the following paragraphs, I discuss in some details the infectivity, treatment, and vaccination of three examples of common viruses: Influenza, Coronavirus, and Hepatitis.



Baltimore Classification	Group I (dsDNA virus)	Group IV ((+)ssRNA virus)	
Virus	<i>Herpesviridae</i>	<i>Flaviviridae</i>	
Structure	A) Herpes Simplex Virus-1	B) Hepatitis C Virus	C) Dengue virus
	Group V ((-)ssRNA virus)		Group VI ((+)ssRNA-RT virus)
	<i>Orthomyxoviridae</i>		<i>Retroviridae</i>
	D) Influenza A Virus	E) Human Immunodeficiency Virus-1	

Figure 1.1 Virus classifications and examples. (Top) The classification of viruses based on the classical Baltimore six groups and a subsequently found seventh group, figure from ViralZone²⁰ (<https://viralzone.expasy.org/254>). (Bottom) Some examples of viruses with schematic diagrams of virions, figure used with permission from Kaihatsu et al.²¹

Table 1.1.1 The common viruses infecting humans. Infection and death toll are per year.

Names	Family	genome	Virion size (nm)	Estimated Infected cases	Estimated Death
Influenza (Flu) ¹⁹	Orthomyxoviridae	Linear ss(-) RNA	80-120	1 billion	600,000
Hepatitis B ²²	Hepadnaviridae	dsDNA-RT	~42	292 million	820,000
HIV ²³	Retroviridae	ss(+) RNA-RT	80-100	39 million	630,000
Papillomaviruses ²⁴	Papillomaviridae	Circular dsDNA	~60	43 million	342,000
Rotavirus ²⁵	Reoviridae	Linear ds RNA	~80	Nearly all	200,000
Dengue ²⁶	Flaviridae	ss(+) RNA	~50	100-400 million	21,000

1.1.5.3 Influenza

Influenza or flu is not just a common cold; it's caused by influenza viruses that infect the nose, throat, and lung. Influenza belongs to the family of Orthomyxoviridae and comprises four immunological types: A, B, C, and D; Influenza A is the most common. Influenza A is an enveloped virus containing eight segmented linear negative-sense ssRNA(–) genome encapsidated by viral nucleoprotein. Viral RdRP transcribes mRNA that are capped and polyadenylated. The eight segments RNAs encode eleven viral proteins through alternative splicing and ribosomal frame shifting: hemagglutinin (HA), neuraminidase (NA), matrix protein M1, matrix ion channel protein M2, nucleoprotein (NP), viral polymerase complex (polyprotein A (PA), polyprotein B1 and B2 (PB1, PB2)), non-structure protein 1, nuclear export protein (also known as non-structure

protein 2), and polymerase basic protein 1-F2 (PB1-F2)²⁷. There are two important surface glycoproteins found in Influenza A: the rod-like spiking protein called hemagglutinin (HA) and the mushroom-like protein, with less abundance than HA, called neuraminidase (NA). HA proteins are involved in the attachment of sialic acid receptors on host cell surface, NA protein cleaves neuraminic acid from HA so that virus enters into host cells through endocytosis. Influenza is often categorized into subtypes based on the types of HA and NA proteins, such as H1N1 or H2N2.

Influenza A is also widespread in avian species, swine, and horses. The segmentation facilitates genetic reassortment of segment genes in infected cells, which is responsible for periodic antigenic shifts with the influenza A virus and leads to pandemics. The emergence of a new viral strain for which humans lack immunity results from the genetic reassortment of a human strain and an animal strain, yielding a new strain that can replicate in humans but acquires a surface protein from animals. Immunity depends on the antibodies binding to the viral surface protein, particularly hemagglutinin (HA), since HA of the new strain from animals causes no human antibody recognition, leading to immune resistance and, consequently a pandemic. The genetic reassortment between animals and humans, therefore, makes Influenza A most dangerous among all Influenza immunological types. Apart from the genetic reassortment of animal strain, incidental mutations on viral surface proteins also happen over time, producing variants of an existing strain that causes resistance to immunity²⁸. Every year the contagious illness infects millions of people globally, and up to ~650,000 lives are taken away by respiratory-related diseases caused by influenza¹⁹. Most deaths happen in children, elderly, and people with chronic health conditions.

1.1.5.4 Coronaviruses

Coronavirus is an ss(+) RNA virus with a genome size of ~27-32 kilobase (kb), the longest genome among RNA viruses. It belongs to the family of Coronaviridae. Coronavirus virion is enveloped

with a size of ~120 nm. For SARS-Cov-2 causing Covid-19, the viral genome encodes four structural and accessory proteins: spike protein, membrane protein, nucleocapsid protein, and envelope small membrane protein, as well as sixteen non-structural proteins (nsp): nsp1-16 performing a wide range of enzymatic activities, including polymerase, capping, methyltransferase, and nuclease, that are critical for genome replication, transcription, translation, and host innate immune suppression²⁹.

Upon infection, the viral spike glycoprotein binds to the host receptor, such as angiotensin-converting enzyme 2, and attaches to the host cell. The virus enters into host cells through endocytosis. The virus membrane is fused with the endosomal membrane, and ss(+) genomic RNA is released into the cytoplasm. Viral genome replication takes place in the replication organelle, where replication intermediates dsRNAs are synthesized from the genomic RNA, and the newly synthesized dsRNA is a template to transcribe viral mRNAs and genomic RNA. The viral subgenomic mRNAs and genomic RNA are exported from the replication organelle to the host cytoplasm to translate viral proteins and encapsidate new virion packages. New virions are assembled and budded at the membrane of the endoplasmic reticulum, the intermediate compartments, and the Golgi complex. New virions are released by exocytosis to infect the neighboring cells²⁹.

1.1.5.5 Hepatitis viruses

Hepatitis is known to be caused by viral infection, but there are also other possible ways causing hepatitis including heavy alcohol use, toxins, side effects of some medications, and some medical conditions. The most common hepatitis are Hepatitis A, B, and C, though other rare types of hepatitis are also identified such as D and E, which all cause inflammation of the liver.

Hepatitis A is primarily transmitted through fecal and oral route; therefore, consumption of water or food contaminated by feces of an infected person, inadequate sanitation and poor personal hygiene are the main reasons for the infection. The symptoms of Hepatitis A are often mild, and people who fully recover from the infection often develop life-time immunity. However, it can be rarely fatal and cause acute liver failure. Nevertheless, Hepatitis A is usually a short-term infection, it does not become chronic as Hepatitis B and C do. The estimated deaths caused by Hepatitis A are less than 10,000 each year. No specific medications are available for treating Hepatitis A. The safe and effective Hepatitis A vaccination is the best way to prevent Hepatitis A infection³⁰.

Hepatitis B viruses attack the liver and cause acute and chronic disease. The virus is transmitted through contact with blood and other body fluids such as saliva, vaginal fluids and semen, most commonly from mother to child during birth and delivery. The liver cirrhosis and primary liver cancer hepatocellular carcinoma, resulting from chronic hepatitis B infection, cause an estimated 820,000 deaths globally. Hepatitis B can be protected nearly 100% against the virus with a safe and effective vaccine, which is usually given soon after birth followed by two or three doses at least four weeks apart²².

Hepatitis C is a bloodborne virus that causes an inflammation liver disease. It is not spread through breast milk, food, water or casual contact such as hugging and sharing food or drinks with an infected person. The viruses cause both short-term serious and long-term chronic illnesses. Around 30% of infected persons clear out the Hepatitis C virus (HCV) within 6 months of infection without any treatment by the immune system while the remaining ~70% of persons develop chronic illness, such as liver cirrhosis and cancer, leading to approximately 290,000 deaths per year. Currently, no effective vaccine against Hepatitis C, but antiviral medications, including sofosbuvir and daclatasvir, cure over 95% of people with Hepatitis C infection³¹.

1.1.5.6 The genome replication and transcription mechanism of viruses

Vaccination provides the best protection measure against a viral infection. For example, the aforementioned Hepatitis B vaccination gives lifetime protection, and most familiarly, the Covid-19 vaccination, due to emerging variants, offers protection for a period of 3 to 6 months. However, we still haven't found vaccines for many viruses, such as Zika, HIV, and Papillomaviruses. Therefore, efforts for antiviral therapeutic development are important to combat viral infection. Since viral genome replication and transcription are key steps during viral life cycle, it is essential to understand the replication and transcription mechanism of viruses that cause significant health and economic challenges in order to design inhibitor drugs targeting key steps of viral replication for treatment.

1.1.5.6.1 Platforms of viral genome replication: double membrane vesicles

Eukaryotic cells carry out genome replication and transcription in the nucleus, whereas many RNA and DNA viruses carry out genome replication and transcription in the cytoplasm of the infected host cells. Cytoplasm replicating viruses, to ensure efficient genome replication and shield from host intrinsic cytosolic sensing and defense effector proteins, arrange the genome replication and transcription in organelle-like compartments in the cytosol³². Viral RNA synthesis was associated with virus-specific membranes, such as single- and double-membrane vesicles and invaginations³³. Viral replication proteins associate with endoplasmic reticulum (ER) in the case of polio virus replication or mitochondrial membrane in the case of flock house virus³⁴, and they induce ER or mitochondrial membrane invagination for replication vesicles. Using cryo-electron tomography (cryo-ET) to image the coronavirus infected cells, Knoops et al. revealed the interconnected ER-derived double membrane vesicles (DMV) of 200-300 nm in diameter, with the RNA replication intermediates in the form of dsRNA predominantly localized to the interiors of DMVs. A

convoluted single membrane structure was also observed to adjoin and interconnect with DMV. The convoluted membranes seemed to be the major location site of viral replicase subunits and encompassed many compartments with open connections to the cytoplasm (**Figure 1.2**). DMVs might originate from the membrane fusion of interconnected convoluted membranes. As the viral genome replication progresses, DMVs may serve as repositories to store dsRNA and viral mRNA before export³⁵. The operation of RNA synthesis factory, with templates, intermediates, and fully functional mRNA products, is well protected from nuclease and cytoplasmic effectors in the formed replication vesicles.

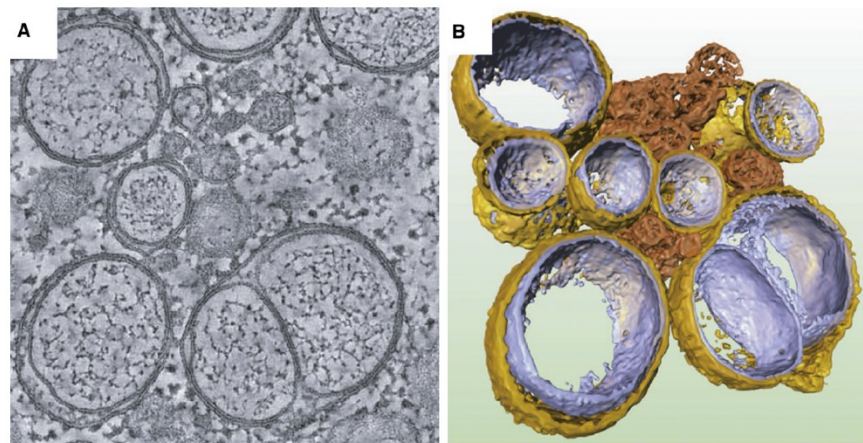


Figure 1.2 An example of virus induced DMV. (A) 2D EM tomographic image and (B) the 3D tomographic reconstruction of severe acute respiratory syndrome (SARS) induced double-membrane vesicles in yellow and blue and convoluted membrane structures in brown. Adapted from den Boon et al.³²

1.1.5.6.2 Gateway from double membrane vesicles to the cytosol

Viruses replicate genomes in a relatively isolated and safe microenvironment inside the host cytosol, to shield from host innate immune sensors and effectors. Nevertheless, mRNAs need to

be exported for translation of viral proteins, and viral genomic RNAs also need to be encapsidated by cytosolic nucleocapsid protein outside DMV.

How are the newly replicated viral RNAs exported from replication vesicles to the cytosol for viral protein translation and new virion packaging? A recent study done by Wolff et al. showed that newly synthesized viral RNA and mRNA are exported to host cytoplasm through the pore complexes on DMV. Inside DMV, dsRNA and subgenomic mRNA are accumulated. Using cryo-ET, the coronavirus-induced replication organelles---DMV at the middle stage of infection were visualized. The DMV lumen was filled with filamentous structures likely corresponding to viral RNA in the form of dsRNA as the replication intermediates, as also seen in **Figure 1.2A**. Several copies of molecular complexes spanned DMV's inner and outer membranes, connecting DMV with cytosol (**Figure 1.3**). These molecular complexes serve as passages to release functional viral mRNA for translation and genomic RNA for virion packaging³⁶.

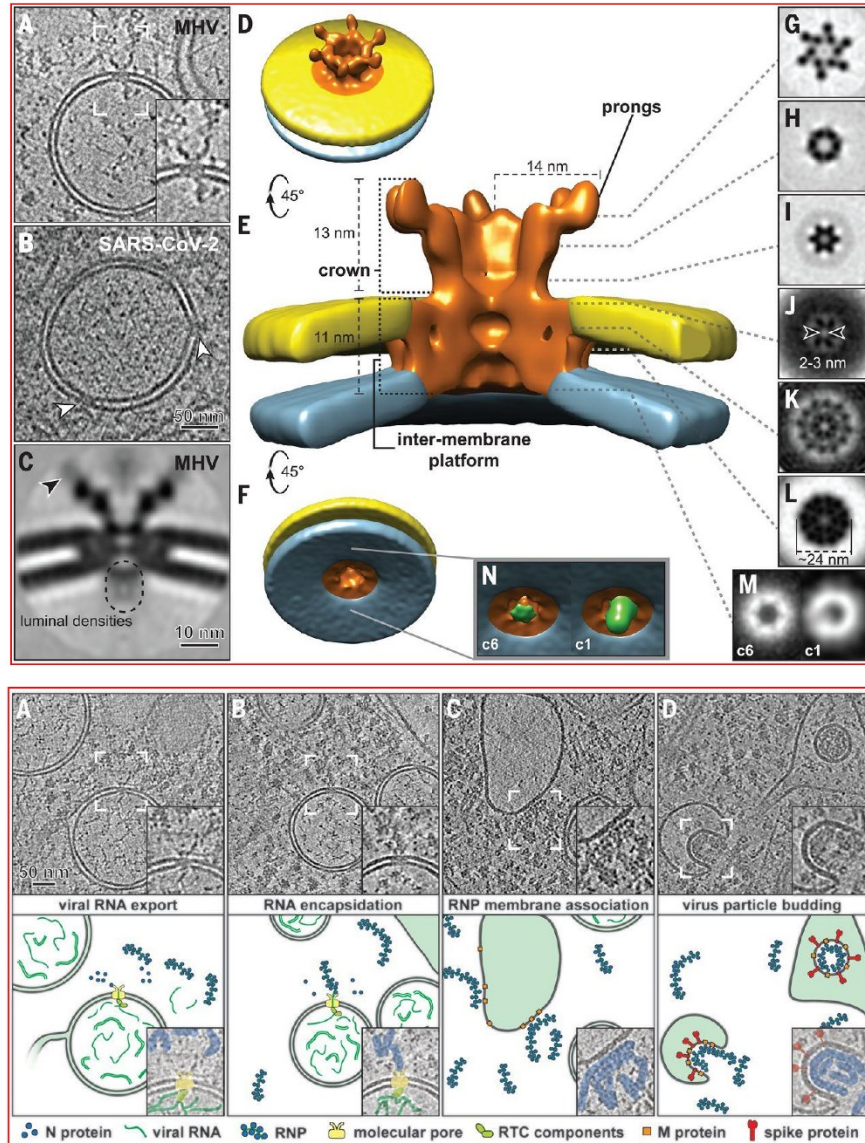


Figure 1.3 The channels connecting DMV and cytosol. (Top) Cryo-ET tomogram and a 3D model of a molecular pore complex embedded in DMV. (Bottom) The tomographic slices of MHV-infected cells capturing important steps during viral replication cycle: (A) Molecular pore complex exports viral RNA into cytosol, (B) RNA encapsidation outside DMV by nucleocapsid (N) protein, (C) Viral ribonucleoprotein complexes reach virus assembly sites for membrane association, and (D) budding of virions. Figure used with permission from Wolff et al.³⁶

Subtomogram averaging of the molecular pore complexes showed an overall sixfold symmetry. A cytosolic crown-like structure extended ~13 nm into the cytosol, and a base platform embedded ~24 nm wide in the DMV membranes. The spacing between the DMV double membranes was ~4.5 nm. The pore complex formed a ~6 nm wide channel on the DMV luminal side, which narrowed down to ~2-3 nm wide toward the cytosol on the DMV outer membrane but still allowed the export of RNA strands. The estimated total molecular mass of the pore complex was about 3 MDa. The complete constituent of the mega pore complex awaits to be elucidated. However, the coronavirus non-structure protein (nsp) 3 labelled with green fluorescence protein was found as the main member of prongs of the crown part³⁶ (**Figure 1.3**).

Specific replication-transcription machinery may associate with the pore complex on the DMV lumen side to guide the newly synthesized viral genomic RNA and subgenomic mRNA entering into the pore channel. In the case of coronavirus, only +RNA would be exported, whereas -RNA templates and dsRNA intermediates would remain for synthesizing more +RNAs. The abundant nucleocapsid (N) proteins outside DMV produced from the translation of N-encoded mRNAs in the cytosol wait for exported RNA, wherein another selection mechanism awaits to be discovered. The short subgenomic mRNAs are passed to host translation machinery for viral protein production and the long genomic RNAs start to associate with N proteins for viral genome packaging. The packed viral ribonucleoprotein complex attaches to single-membrane compartments, derived from the ER to Golgi intermediate compartment, and progresses toward new virion budding to be released into extracellular space to infect other host cells³⁶ (**Figure 1.3**).

1.2 Innate immune responses

When viral RNAs are exported from DMV like replication organelles to host cytoplasm, the opportunity for host cytosolic immune sensor and effector proteins to confront foreign RNA species occurs. One long-standing question is how the host innate immune response is activated by invading viruses, or more specifically, how the interferon (IFN) signaling is turned on upon viral infection. When viral RNAs are inside DMV, they are well protected from host surveillance factors. Once they are released from the gateway to the host cytosol, there is a time window during which the host immune sensors and effectors are accessible to viral genomic RNAs and mRNAs. Another time window is before the formation of DMV during the preparation of viral genome replication, when viral RNAs are exposed in the host cytoplasm. Using the common ss(+)RNA virus as an example, after the endocytosis mediated cell entry, the genomic (+) RNA is used as template for translating essential viral proteins required for replication, such as viral RdRp and nucleocapsid protein. The two time windows before and after massive genome replication are critical for the activation of IFN signaling pathway to launch an antiviral state in the infected cell and alert the surrounding uninfected cells.

1.2.1 Host defense forces restricting viral replication

Pathogen associated molecular patterns (PAMPs), such as viral nucleic acids, are detected by pattern recognition receptors (PRRs) present in cytosol and endosomes, including toll-like receptors, retinoic acid-inducible gene I (RIG-I) like receptors and nucleotide-binding oligomerization domain (NOD)-like receptors. The recognition of PAMPs by PRRs activates interferon regulatory factors (IRF) and leads to the expression of type I interferons (IFN)³⁷. IFNs are cytokines, secreted in paracrine and autocrine manners, and bind to IFN receptors on cell

surface to alert surrounding uninfected cells, by activating a signal transduction cascade through the Janus kinases coupled with signal transducer and activator of transcription proteins (JAK-STAT) signaling pathway. This results in the upregulation of hundreds of IFN stimulated genes (ISGs). ISGs are the innate immune workhorses to fight against the intruding viruses³⁸. The antiviral signaling can also be interferon-independent especially at the very beginning of infection with low numbers of virus particles. IRF3 can directly trigger the expression of a small subset of ISGs in a more efficient way to prime the immune system and block viral replication without activating an elaborate immune response³⁹. Hundreds of ISGs have been identified, and amongst the most prevalent ISGs is a protein family called interferon-induced protein with tetratricopeptide repeats (IFITs)⁴⁰.

1.2.2 IFITs

IFITs were discovered in 1980s-1990s when human cells were treated with interferons: IFIT1 by Michel Revel's lab in 1983⁴¹; IFIT2 by James E. Darnell Jr.'s lab in 1986⁴²; IFIT3 by Sai-Juan Chen's lab in 1997⁴³; IFIT5 by Susan C. Weil's lab in 1997⁴⁴. IFIT1, IFIT2, IFIT3, and IFIT5 are relatively well-studied in humans, each about 55 kDa in size. The sequence identity among IFITs is in the range of 38%-55%. As the name suggests, IFITs are comprised of multiple structure motifs called tetratricopeptide repeats (TPR). The TPR motif is essentially a helix-turn-helix segment, with about 34 amino acids⁴⁵. **(Figure 1.4)**

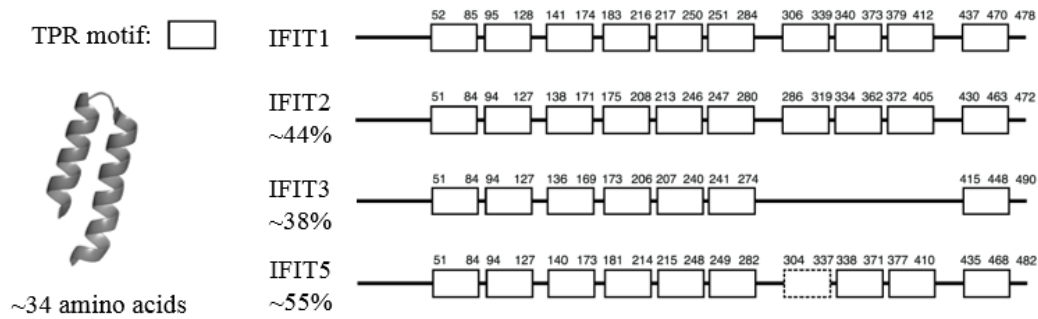


Figure 1.4 TPR motifs in IFITs. Adapted from Abbas et al.⁴⁶

1.2.3 IFIT antiviral modes of action

The tetratricopeptide repeats are well known for mediating protein-protein interactions⁴⁵. Therefore, IFITs were initially thought to perform viral inhibition functions by interactions with translation factors mediated by TPRs. Early work on understanding the antiviral function of IFIT1 and IFIT2 suggested that IFIT1 bound to one subunit of the eukaryotic translation initiation factor eIF3 to block the formation of stable eIF3-ternary complex, and IFIT2 bound two subunits of eIF3 to destabilize the translation ternary complex as well as hinder the formation of 48S pre-initiation complex⁴⁷. Consequently, it was suggested that viral protein synthesis was restrained by IFIT1 and IFIT2, through interactions with eIF3. Several years later, another study discovered IFITs, conventionally thought to bind protein partners, directly engaged viral RNAs and sequestered them from accessing the host translational machinery⁴⁸. The understanding of molecular mechanisms of IFIT-exerted viral inhibition started from the paradigm shift in the discovery of RNA binding ability of IFITs. For example, the 5' ends of viral PPP- and capped-RNA are sequestered by IFIT5 and IFIT1, respectively. The crystal structures of both IFIT5 and IFIT1 in complex with ssRNA were determined in the Nagar lab^{46, 49}. These structures revealed that IFIT1 and IFIT5 use their tetratricopeptide repeats motifs to form a central positively charged pocket that accommodates only ssRNA (**Figure 1.5A-B**). The cavity in IFIT1 is slightly larger to accommodate the additional

cap structure acquired at the 5' end of many viral mRNAs (**Figure 1.5B**). Studies using virus infected cells and an Ifit2 knockout mouse model showed Ifit2 inhibits viral replication for positive-sense^{50, 51} and negative-sense⁵²⁻⁵⁴ ssRNA viruses. Yang et al. solved the apo IFIT2 structure, which revealed a domain-swapped parallel dimer⁵⁵.

IFIT3 cannot directly bind RNA but interacts with IFIT1 to enhance the IFIT1-RNA interactions⁵⁶. Johnson et al. solved the crystal structure of Cap0-RNA bound to IFIT1 in complex with C-terminal IFIT3, showing that IFIT1 and IFIT3 interacted through each C-terminal TPR motif mediated interactions⁵⁷ (**Figure 1.5C**). IFITs also exist in the form of a multi-protein complex consisting of IFIT1/2/3. In vivo sedimentation and Western blot analysis on HeLa cells treated with IFNs revealed that the IFIT1/2/3 complex is ~150-200 kDa in size⁵⁸, suggesting a trimer. The precise role of this complex in inhibiting viral replication is unknown.

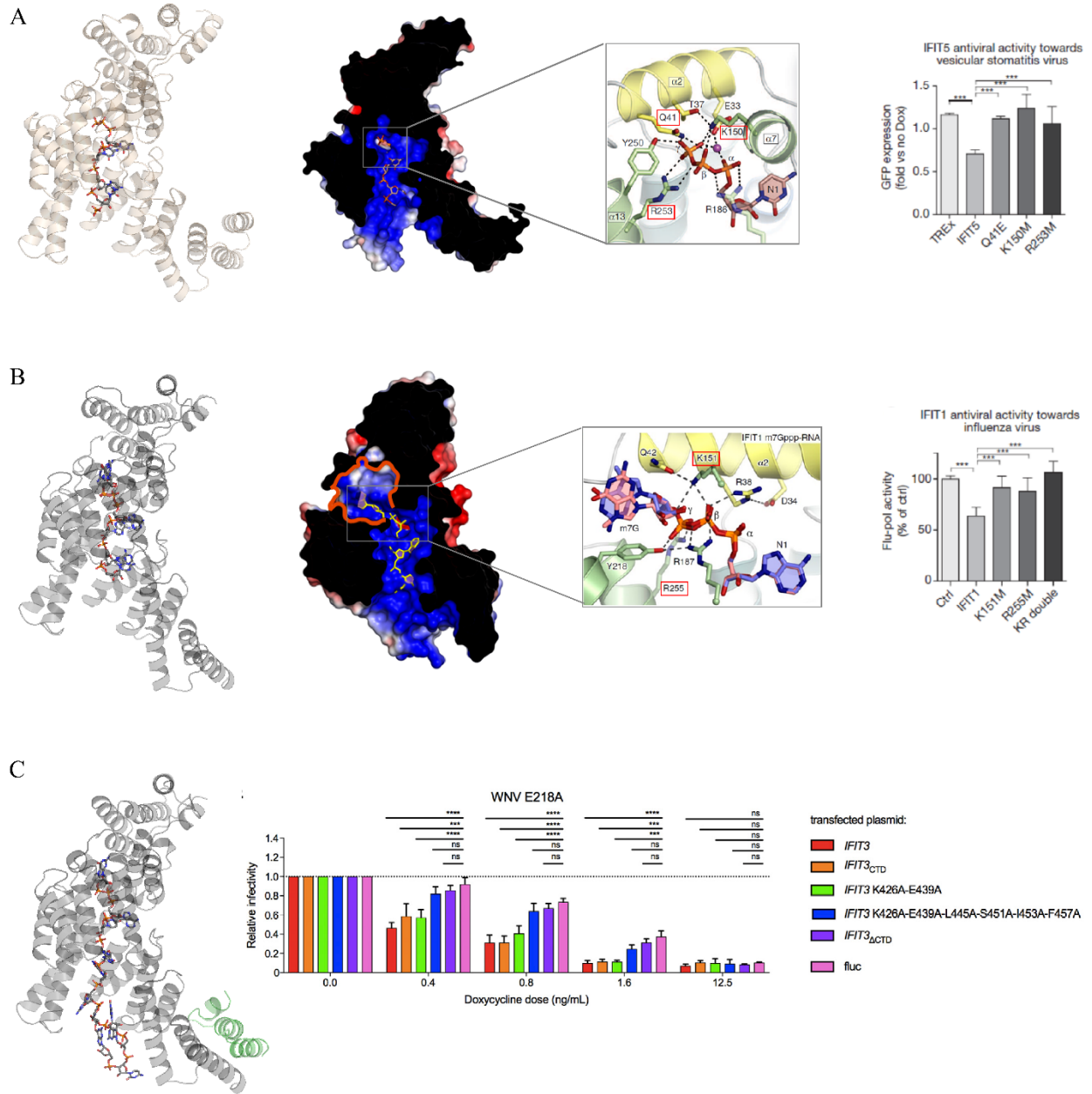


Figure 1.5 The structure and function of IFITs. (A) The molecular basis of IFIT5 sequestering 5'PPP-RNA, RNA binding residues are critical for restraining viruses in viral infectivity assay⁴⁶. (B) The molecular basis of IFIT1 sequestering Cap0-RNA, RNA binding residues are critical for restraining viruses in viral infectivity assay⁴⁹. (C) The molecular basis of IFIT1 sequestering Cap0-RNA in complex with C-terminal IFIT3 (green), residues in IFIT3 interacting with IFIT1 are

critical for restraining viruses in viral infectivity assay⁵⁷. Figures are adapted from Abbas *et al.*⁴⁶ (panel A), Abbas *et al.*⁴⁹ (panel B), and Johnson *et al.*⁵⁷ (panel C).

1.2.4 ZAP

One more example of ISGs is the Zinc-finger antiviral protein (ZAP). Studies showed that ZAP directly binds to viral RNA and induces the degradation of bound viral mRNA. ZAP recruits host mRNA degradation machinery, the deadenylase poly-adenylate poly(A) specific ribonuclease, to remove the poly(A) tail, and then recruits exosome to degrade the bound mRNA from 3' end. ZAP also recruits host decapping enzyme through the cofactor RNA helicase p72 to remove the canonical cap and further recruits Xrn1 in the presence of bound RNA to degrade the target mRNA from 5' end⁵⁹. Crosslinking immunoprecipitation sequencing (CLIP-seq) analysis in cells infected with HIV revealed that ZAP predominantly binds to the GC-rich segment of the HIV genome⁶⁰ and X-ray crystallographic analysis revealed ZAP bound to GC dinucleotides⁶¹.

Mature mammalian mRNAs contain two protective elements from degradation: a 7mG cap at 5' end and a poly(A) tail at 3' end. The mRNA degradation starts with shorting poly(A) tail, by deadenylase, poly(A) specific ribonuclease (PARN), the Carbon Catabolite Repression-Negative On TATA-less (CCR4-NOT) complex, or the poly(A)-nuclease (PAN) deadenylation (Pan2-Pan3) complex⁶². Then the deadenylated RNA body is degraded in a 3'-5' direction by a 3'-5' exoribonuclease complex: the exosome^{63, 64}. The 7mG cap is removed by the decapping enzyme complex Dcp1a-Dcp2⁶⁵, followed by degradation of the decapped RNA body in 5'-3' direction by 5'-3' exoribonuclease 1 (Xrn1)⁶⁶.

1.2.5 Core set of ISGs in action

Back to 1986, only 12 ISGs were identified, Staeheli et al. studied one ISG, namely Mx1 in mice (MxA in humans is the ortholog of Mx1 in mice), showing anti-influenza function. They wrote, “it is not known which or how many of these proteins are required for protection against a particular virus”⁶⁷. Over three decades following this study, the question of “which” ISGs has been explored extensively such that the ISG list is in the hundreds. However, the question of “how many” ISGs are required for protection remains unclear⁶⁸. Upregulating several hundreds of ISGs at the same time is a very energy consuming task, especially when cells face invasion from intruding viruses. Thus, a core set of ISGs might be deployed to fight individually and/or collectively, against the intruders for efficient and economic defense war with viruses.

It was recently found that a small subset of 6 ISGs out of 620, including STAT1, STAT2, IRF9, Zinc finger CCCH-type antiviral protein 1 (ZC3HAV1), IFIT3, and IFIT1, dominantly inhibit alphavirus infection in human U2-OS cells⁶⁹. STAT1, STAT2 and IRF9, identified as ISGs without direct antiviral effect, are involved in the JAK-STAT signaling cascade which activates ISG expression. Only three ISGs directly act on suppressing viral replication: ZAP encoded by the gene ZC3HAV1 binds to CpG dinucleotides present in alphavirus RNA, IFIT1 and IFIT3 associate to strongly bind to 5' end of Cap0-alphavirus mRNAs for inhibiting viral mRNA translation⁶⁹. The dominant set of ISGs against viral infection is virus-specific. For example, different distribution and abundance of CpG dinucleotides can determine whether ZAP would be a major player in restraining a certain virus, and differences in capping and 5' end secondary structure of viral RNA will determine whether IFIT1 and IFIT3 would be essential for suppressing a certain viral replication. Another study using clustered regularly interspaced short palindromic repeats (CRISPR) knockout screens identified as few as five ISGs that play dominant roles in restraining

various HIV strains in primary CD4⁺ T cells⁷⁰. Nevertheless, for each viral infection, the viral inhibition comes from a limited set of major ISGs but not all ISGs, which act primarily based on the characteristics of a certain virus. Apart from the dominant ISGs, other ISGs may only play a minor role on viral inhibition.

Table 1.2 Other main ISGs and their mechanisms of actions on viral inhibition.

Names	Function and mechanism
MxA ⁷¹	Associate with viral nucleoprotein to interfere genome replication
BST2 (tetherin) ⁷²	Tether virion, and restrict virion release after budding
RSAD2 (viperin) ⁷³	Inhibit viral RdRp by generating a replication chain terminator
IFITM3 ⁷⁴	Directly engage virus particles and shuttles them to lysosomes for destruction
ISG20 ⁷⁵	Degradation of viral RNA and deaminated viral DNA, induce IFIT1-mediated viral translation inhibition
IFI6 ⁷⁶	Block the formation of virus induced endoplasmic reticulum membrane invagination in flavivirus replication

1.3 Functional structure RNA elements regulating mRNA translation

The 7-methyl guanosine (7mG) 5' Cap structure and the poly(A) tail (An) at the 3' end stabilize mRNA and stimulate translation. The 5' untranslated region (UTR) contains secondary and tertiary structures and sequence elements. Pseudoknots, stem-loops, hairpins, G-quadruplexes (RG4), upstream open reading frames (uORFs), and upstream start codons (uAUGs) mainly inhibit the cap-dependent translation. Internal ribosome entry sites (IRESs) mediate translation initiation independent of the 5' end cap. RNA modifications such as N6-methyl adenosine (m6A), RNA-

binding proteins (RBPs), long non-coding RNAs (LncRNAs) interacting with mRNA, and the Kozak sequence around the start codon all can regulate translation initiation⁷⁷ (**Figure 1.6**).

The 7mG cap structure is essential to protect mRNA from 5' exoribonuclease degradation, to direct pre-mRNA for splicing, to export mRNA from the nucleus, and to recognize the translation initialization factor eIF4E⁷⁸. The canonical cap-dependent translation initiation requires the recruitment of the 40S ribosomal subunit in association with eukaryotic initiation factors (eIF) to the 5' end 7mG cap structure. It starts with the assembly of the trimeric cap-binding complex at the 7mG cap through the cap-binding protein eIF4E interacting with the scaffolding protein eIF4G and the RNA helicase eIF4A. The mRNA chain is circularized by the interactions between poly(A) binding protein at the 3' poly(A) tail and eIF4G near the 5' cap. Then, the GTP-bound eIF2 and the initiator tRNA form the ternary complex of eIF2-GTP-Met-tRNA. The ternary complex, the 40s subunit, and eIF3 form the 43S pre-initiation complex. Through interactions with eIF3, eIF4G recruits the 43S complex to bind to near cap region and to scan along the 5' UTR. The helicase eIF4A unwinds inhibitory RNA secondary structures, such as stem-loops, so that the initiation machinery reaches the start codon AUG for binding 43S complex and releasing eIFs before proceeding to the elongation phase⁷⁷. Apart from all these functional and structural RNA elements, another functional element is methylation at the 2'O position of the ribose of the first and the second nucleotide adjacent to the 7mG cap. The 2'O-methylation at the first nucleotide is referred to as Cap1 form, while the 2'O-methylation at both the first and second nucleotides is referred to as Cap2 form. If no 2'O-methylation is followed by the 7mG cap, it is referred to as Cap0 form. Viral RNA is commonly in Cap1 form, with a minority in Cap0 and uncapped form. Host mRNA is mainly in Cap1 and Cap2 form. Therefore, 2'O-methylation serves as one signature that distinguishes viral and host mRNA⁷⁹.

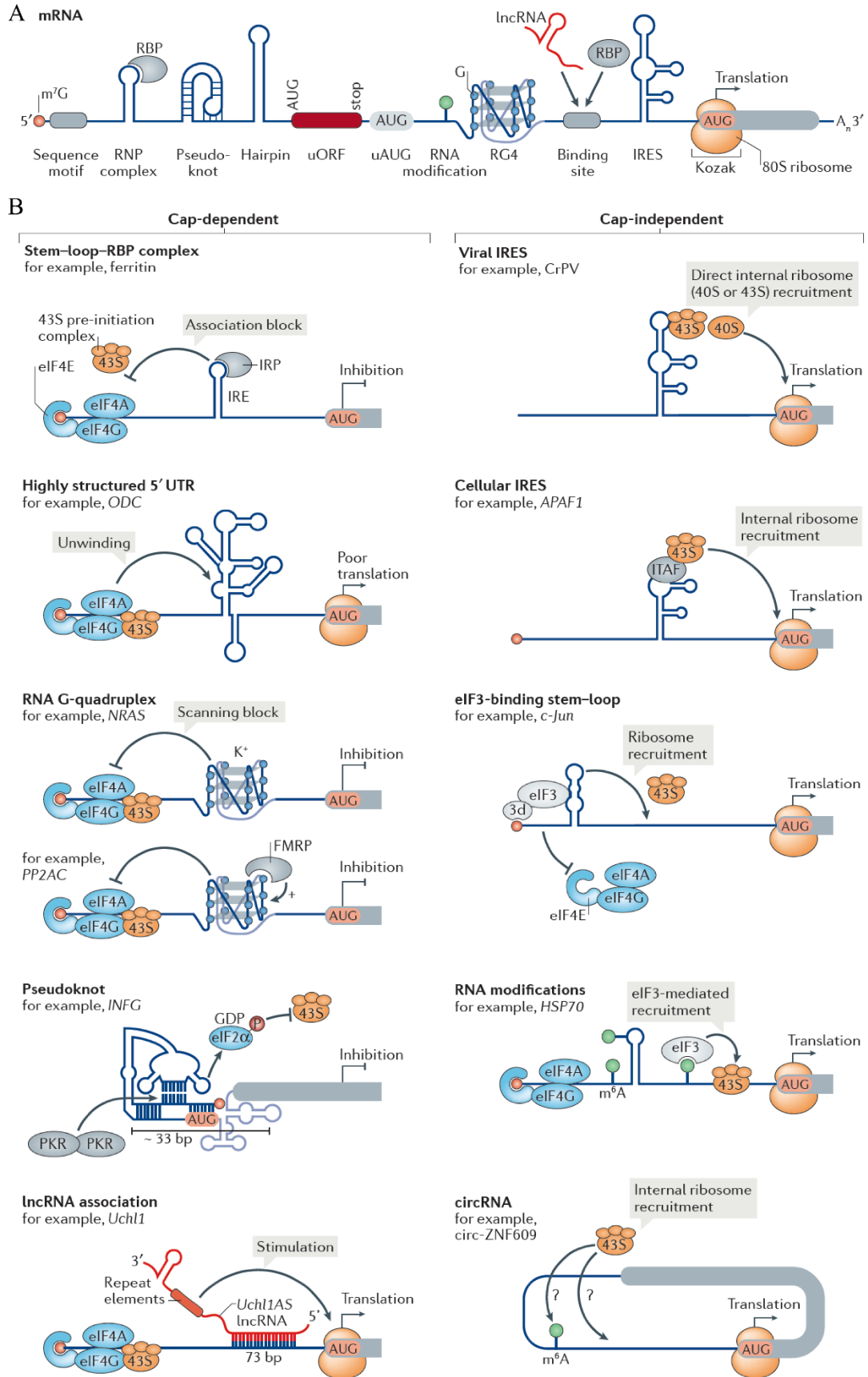


Figure 1.6 Cis-acting functional structure RNA elements regulating mRNA translation. (A)

Cartoon illustration of common mRNA features from 5' end to 3' end. (B) Structure elements that affect cap- or IRES-dependent translation pathway. Abbreviations: CrPV, cricket paralysis virus; ODC, ornithine decarboxylase; APAF1, apoptotic peptidase activating factor 1; NRAS, N-ras proto-oncogene; K⁺, potassium; PP2AC, protein phosphatase 2 catalytic subunit alpha (also known as PPP2CA); FMRP, fragile X mental retardation protein 1; c-Jun, Jun proto-oncogene; IFNG, interferon gamma; P, phosphorylation; PKR, protein kinase RNA-activated; HSP70, heat shock protein 70; Uchl1, ubiquitin carboxyl-terminal hydrolase L1; Uchl1AS: ubiquitin carboxy-terminal hydrolase L1, opposite strand (commonly known as Uchl1os); circ-ZNF609, circular-zinc-finger protein 609. Figure used with permission from Leppek *et al.*⁷⁷

1.4 The RNA capping and 2'O-methylation pathway

The γ -phosphate of the nascent 5'-triphosphorylated RNA (5'ppp-RNA) is removed by RNA triphosphatase (RTPase) to yield a 5'-diphosphorylated RNA (5'pp-RNA). Then the β - and γ -phosphate of guanosine triphosphate (GTP) are removed to yield a guanosine monophosphate (GMP) and pyrophosphate (PPi). GMP is ligated to 5'pp-RNA forming a 5' to 5' triphosphate linker and resulting in a guanosine capped RNA. The two-step cleavage and subsequent ligation reactions are catalyzed by guanylyltransferase (GTase). With methyl donor from S-Adenosyl methionine (SAM), the N7 position of guanosine cap is methylated by guanine N7 methyltransferase (N7-MTase) to yield 7-methylguanosine capped RNA (7mGppp-RNA), namely Cap0-RNA. Furthermore, the ribose 2'-OH position of the first nucleotide adjacent to the cap is methylated to yield Cap1-RNA (7mGppp-Nm-RNA, where N denotes any nucleotide). Many viruses are capable of producing Cap1-RNA. For the host mRNA, the ribose 2'-OH positions of the first nucleotide or the first and second nucleotides adjacent to the cap can be methylated to

yield Cap1-RNA and Cap2-RNA (7mGppp-NmNm-RNA), respectively (**Figure 1.7**). Cap1 methylation has been thought as a measure used by host to prevent action by innate immune sensors such as RIG-I and effector such as IFIT1 to avoid autoimmune attacks. Nevertheless, the function of Cap2 methylation remains unclear except for an additional safety layer to protect host from innate immune response. In addition, the population of Cap1 versus Cap2 host mRNA is unclear. A very recent study eventually revealed the mechanism of Cap2 methylation on host mRNA. Despici and Jaffrey developed a method, namely CircLigase-assisted mapping of caps by sequencing (CLAM-Cap-seq), and performed a transcriptome-wide mapping of Cap2 methylation. Compared with the rapidly co-transcriptional events of N7 and Cap1 methylation in the nucleus, the cytoplasmic Cap2 methylation was slow and gradually enriched over the lifetime of mRNA. Slow Cap2 methylation provides time for innate immune effectors to respond to rapidly replicating viral Cap1 mRNA before viral mRNA acquires a high level of Cap2 methylation. Meanwhile, it suppresses the autoimmune response to host Cap1 mRNAs that are gradually modified to Cap2 RNA as they age⁸⁰.

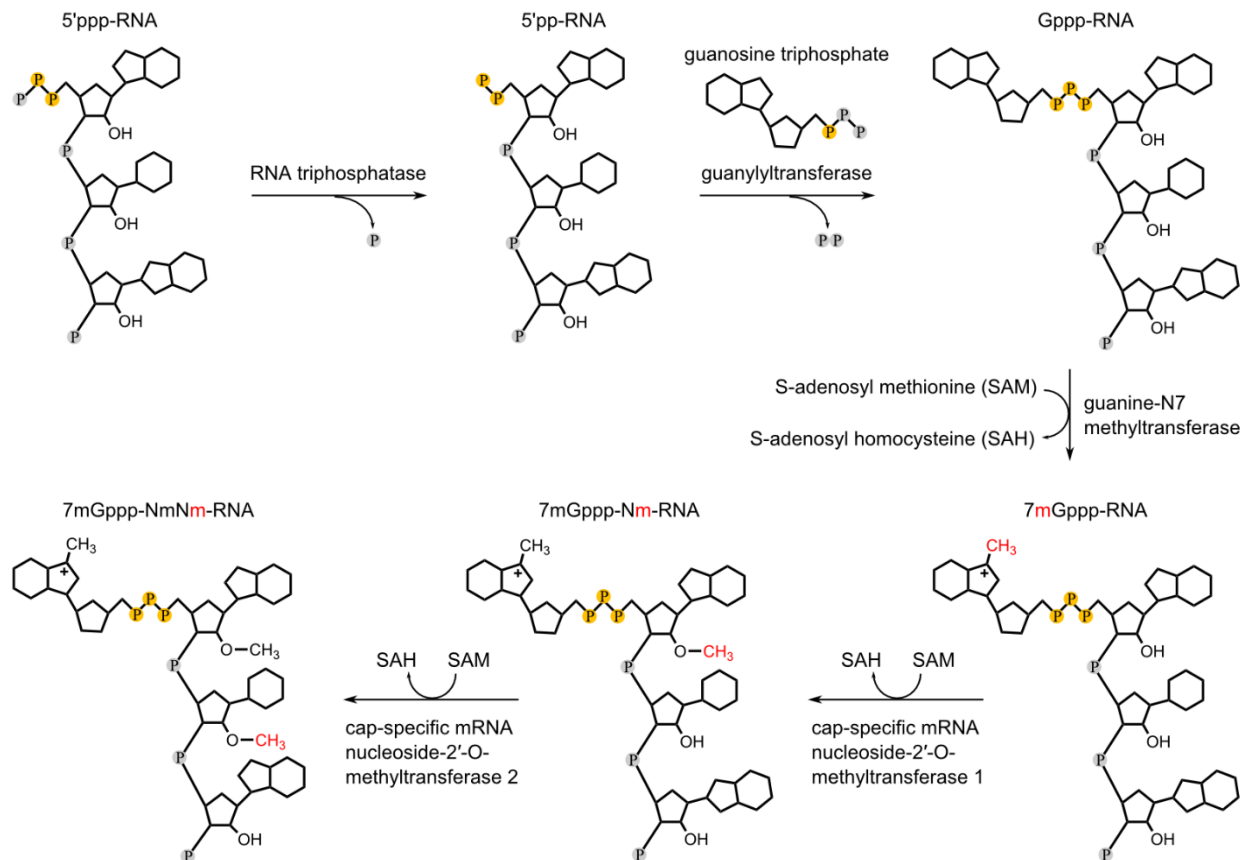


Figure 1.7 The mRNA capping and methylation pathway. A diagram showing the enzymatic modification steps from 5' ppp-RNA to Cap0-, Cap1-, and Cap2-RNA.

1.5 Viral evasion strategies

Facing the potent viral inhibition acted by IFITs and other antiviral effectors, viruses are driven to counteract to survive and thrive. A deeper understanding of novel mechanisms that viruses exploit to evade human immune defense helps us discover how and when viruses hijack host machineries to dysregulate the function of the immune system, and ultimately, help develop therapeutic interventions so that IFITs and other dominant ISGs are reactivated to restrict viral counteracting measures.

1.5.1 Mimic host mRNA caps

Mimicking the cap of host mRNAs is a straightforward way for virus to co-opt host translation machinery. For example, the family of dsDNA Hepadnaviridae and ss(+)RNA Retroviridae use the host capping machinery in the nucleus to generate Cap1 viral mRNA. Many viruses, including the family of dsDNA Poxviridae, dsRNA Reoviridae, ss(+)RNA Flaviviridae and Coronaviridae, encode their own capping enzymes to generate Cap1 viral mRNA in cytoplasm⁷⁸. Some viruses such as Influenza A virus and the family of Arenaviridae and Bunyavirales utilize a mechanism described as “cap snatching” to steal caps from host mRNA. The viral protein with nuclease activity, RdRp or nucleocapsid protein, cleaves the 5’ end of host mRNA including the cap and the adjacent 10-15 nucleotides; the cleaved host short Cap1- or Cap2-RNA is used as a primer for the transcription of viral mRNA by viral RdRp⁷⁸.

1.5.2 Secondary structure elements

Viral RNAs are highly structured, as shown in the example of 5’UTR of Zika and SARS-Cov-2 genomic RNA (**Figure 1.8**). Zika genome is a positive-sense RNA with approximately 10.8 kb. Through nuclease digestion coupled with sequencing, the 5’UTR of Zika and SARS-Cov-2 have been enzymatically mapped to reveal complex secondary structure elements, such as the stem-loop A (SLA) and SLB in 5’UTR of Zika as well as several SLs and pseudoknots (PKs) in 5’UTR of SARS-Cov-2. Pseudoknots (PKs) are one type of higher-order mRNA structure. PKs consist of at least two intercalated stem-loops that form a knot-like 3D intramolecular structure. The *cis*-regulatory PKs in the coding sequences found in many RNA viruses can directly interact with translating ribosomes to induce programmed frameshifting for the synthesis of different proteins⁷⁷,⁸¹. PKs in 5’ UTR fold into a functional tertiary structure that controls translation in a cap-independent manner⁷⁷. These complex secondary structure elements not only act as barriers to

block the recognition of innate sensors and effectors, therefore, downplaying the host defense, but they are also functional elements critical for virus infectivity and pathogenicity. For example, a long-range interaction between the Zika 5'UTR and viral Envelope protein coding region is significant for ribosome binding and scanning on 5' Cap1 mRNA⁸². The complex SL5 contains a four-way junction which is present in all coronavirus, and the start codon of nsp1 located in SL5. The 5' UTR of SARS-Cov-2 contains an upstream open reading frame (uORF). The presence of the 7mG cap as well as the SL1-4 before the translation start site of uORF suggests the cap-dependent translation initiation would require RNA helicase eIF4A to unfold the strand for ribosome scanning. On the other hand, the structure features in the 5' UTR of SARS-Cov-2 and the AUG start codon location downstream of a four-way junction make it reminiscent of the internal ribosome entry site (IRES) found in HCV⁸³.

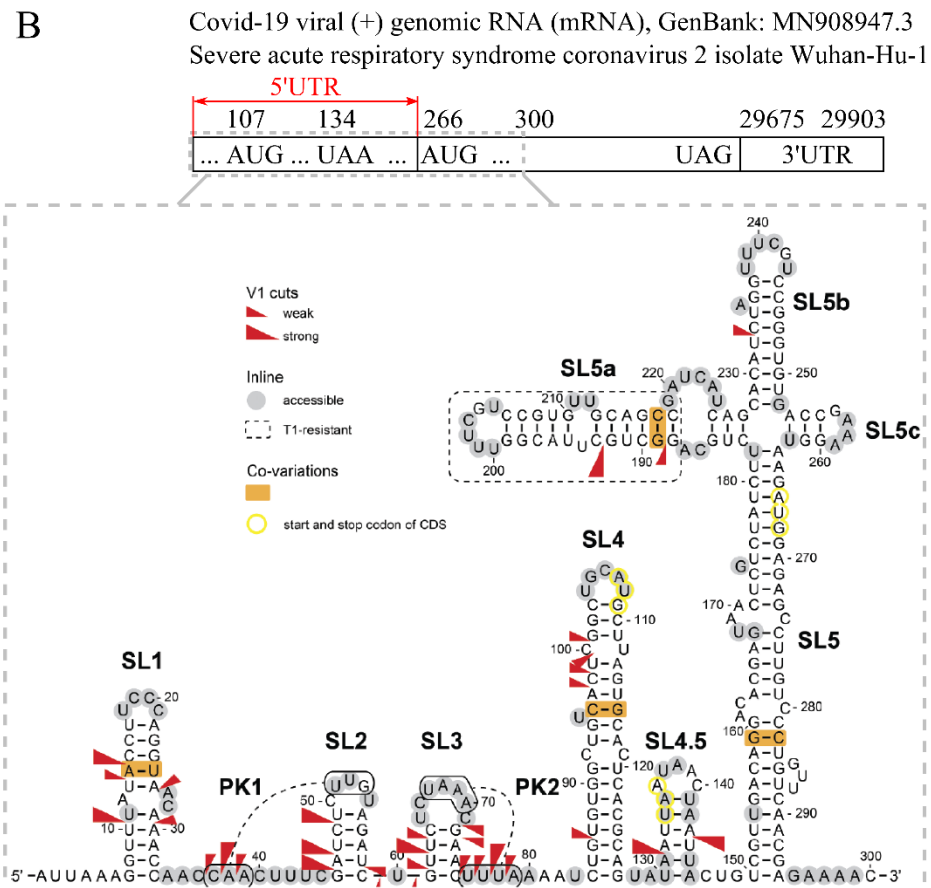
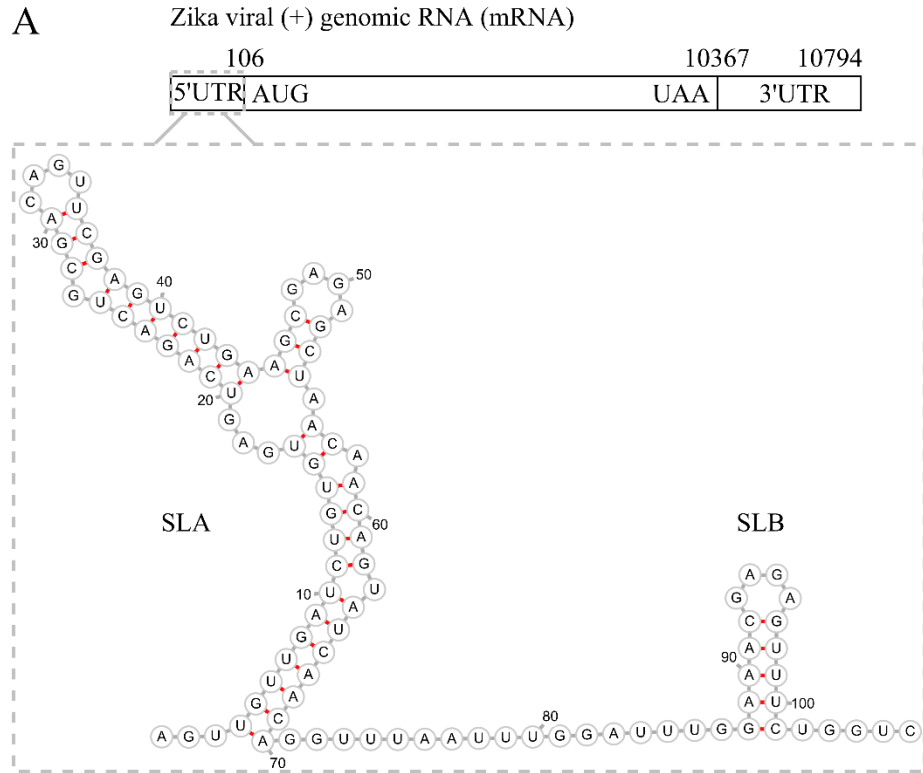


Figure 1.8 The secondary structure models of viral RNAs determined through enzymatic probing. (A) Zika 5' UTR, (B) SARS-Cov-2 5' UTR. Adapted from Li *et al.*⁸² and Miao *et al.*⁸³

1.5.3 Cap-independent and IRES-dependent translation pathway

For viruses utilizing cap-independent mode of translation initiation, *cis*-acting RNA elements called internal ribosome entry sites (IRESs) directly recruit ribosome independent of 5' end. IRESs enable viruses to continue viral protein translation even when the canonical cap-dependent protein translation pathway is repressed upon infection. IRESs are ~150-200 nucleotides long and contain several RNA pseudoknots folded into a compact structure (**Figure 1.9A-B**). IRES directly binds to the 40S ribosome subunits and eIF3 with multiple contacts (**Figure 1.9C-D**). The architecture of IRES-40S-eIF3 complex is not known; however, a cryo-electron microscopy (cryo-EM) structure of IRES with 40S shows how IRES engages with 40S⁸⁴. Upon viral infection, protein kinase R (PKR) is upregulated by interferon signaling pathway, PKR phosphorylates eukaryotic translation initiation factor eIF2 α , through which the mRNA translation is inhibited. However, eIF2 α phosphorylation stimulates the IRES-mediated translation⁸⁵. Therefore, a major host antiviral mechanism through activation of PKR and subsequent phosphorylation of eIF2 α can also be taken advantage of by viruses that are capable of sustaining viral protein translation through IRES-dependent pathway even though a global suppression of mRNA translation including antiviral effector proteins.

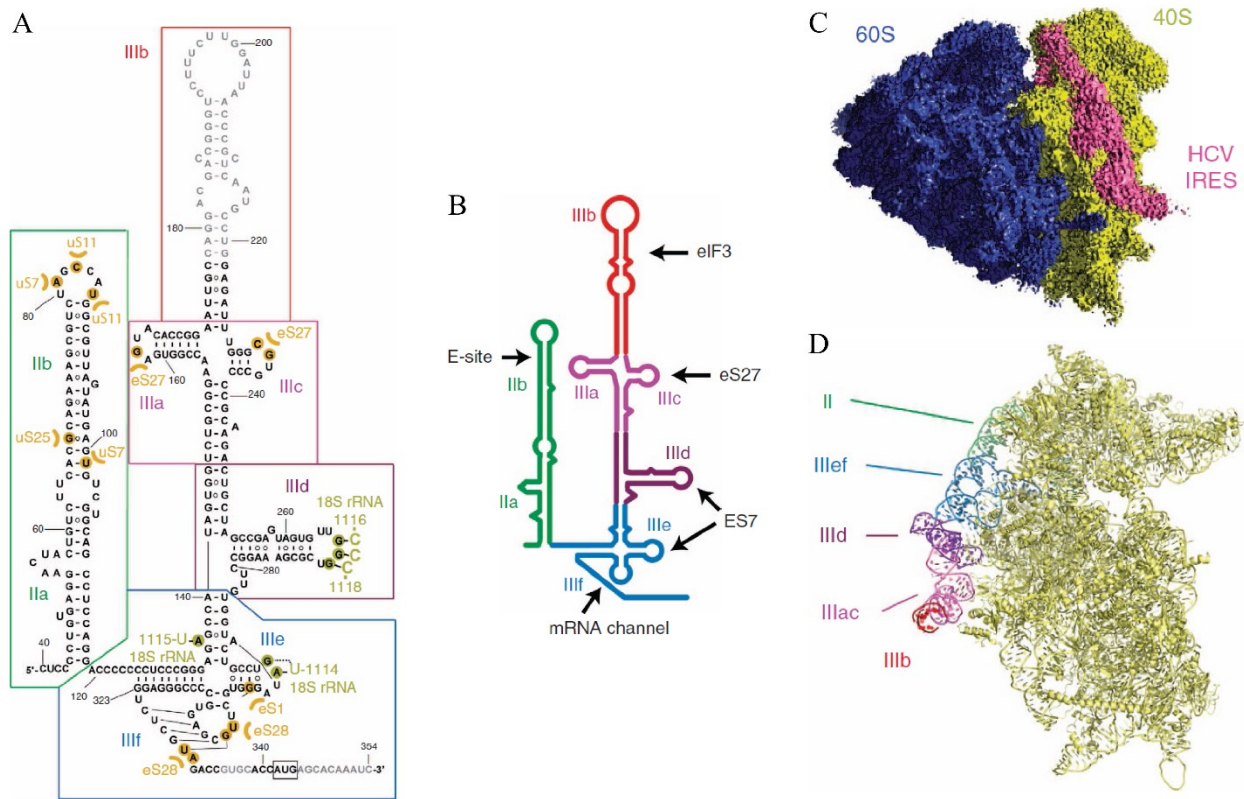


Figure 1.9 HCV IRES binds to human 40S ribosome. (A) Diagram of HCV IRES secondary structure, color labeled on different domains. The canonical base pairs are marked with lines and the non-standard base pairs are marked with circles. (B) Cartoon of HCV IRES, arrows indicate interaction sites on 40S ribosome. (C) The cryo-EM density map of HCV IRES bound to human 80S ribosome, showing that IRES only interacts with the 40S subunit. (D) The structure model of HCV IRES bound 40S ribosome. Figure used with permission from Quade *et al.*⁸⁴

1.5.4 Viral nsp1 mediated inhibition on host mRNA

SARS-Cov-2 evades host defense by global protein translation shutdown of the host cells. The non-structural protein 1 (nsp1) associates with ribosome to block the entry channel of mRNAs. The nsp1 mediated host translation inhibition acts in two aspects: nsp1 stalls canonical mRNA translation by binding to the 40S subunit of ribosome; nsp1 involves RNA cleavage to trigger the

endo-nucleolytic degradation of host mRNAs, including the innate immune effectors that would facilitate the clearance of the infection⁸⁶ (**Figure 1.10A-C**). The molecular basis of nsp1 mediated translation inhibition explains in part why the observed aberrant transcription induction failed to launch a robust expression of interferon responsive ISGs⁸⁷. In both SARS-Cov-2 and Influenza A virus infected cells, due to the shutdown of host translation interrupted by nsp1, the overall transcription level of ISGs, including the most upregulated IFITs, is substantially lower than cells induced with interferons or infected with mutant influenza virus lacking nsp1⁸⁷ (**Figure 1.10D**). Nevertheless, it is still unclear how SARS-Cov-2 overcomes the nsp1-mediated translation inhibition to produce its own viral proteins for propagation in host cells. The complex structure features in the 5'UTR of SARS-Cov-2 mRNA may provide clues to circumvent the ribosome entry blockage and to prevent the degradation of its own mRNAs. One possibility is that the IRES-dependent translation initiation pathway could rescue SARS-Cov-2 mRNA from the cap-dependent translation pathway blocked by nsp1.

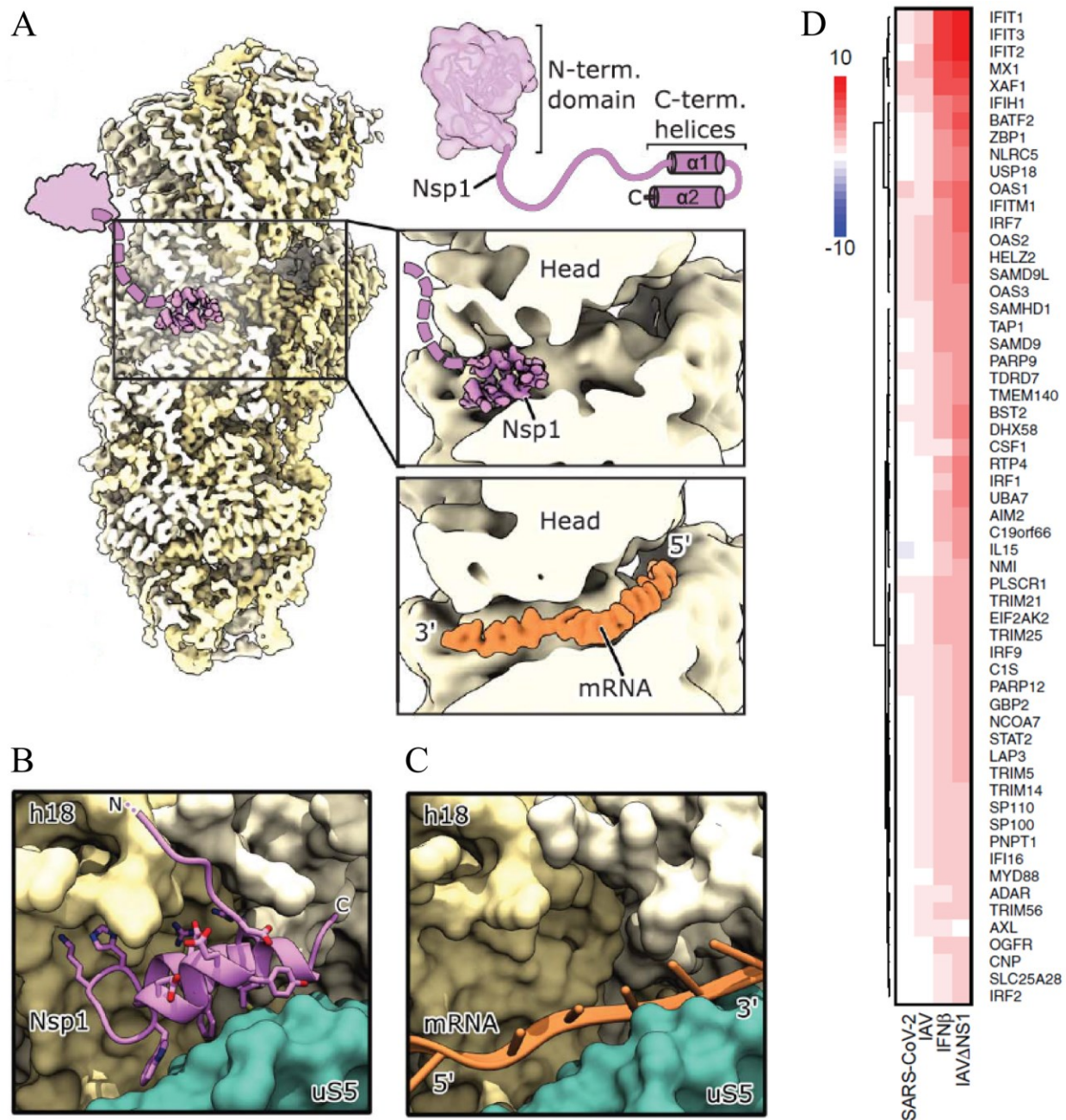


Figure 1.10 SARS-CoV-2 nsp1 mediated translation inhibition. (A) The cryo-EM structure of Nsp1-40S, with nsp1 shown in purple, rRNA and proteins shown in yellow; the putative position of the N-terminal domain of nsp1 is schematically indicated. (B) Nsp1 interacts with ribosome RNA helix h18 and the ribosome domain uS5. (C) The mRNA entry site based on a previous structure (PDB 6Y0G⁸⁸), located at the same pocket where nsp1 is accommodated⁸⁶. (D) Host

transcriptional response to SARS-CoV-2 infection, in contrast to treatment with IFN-I alone or infection with wild-type IAV or a mutant IAV lacking nsp1 (IAV Δ NS1), in primary human bronchial epithelial cells. Heatmap indicating the expression levels of differentially expressed genes involved in IFN-I responses⁸⁷. Figures used with permission from Thoms et al.⁸⁶ and Blanco-Melo et al.⁸⁷

1.5.5 microRNA induced protection

Viruses take advantage of host microRNAs (miRNAs) to evade host innate immune response and protect from degradation⁸⁹. One well studied example is the miR-122 interacting with 5' end of the HCV genome to produce 3' overhanging extension. The 3' overhang inhibits the recognition of cytoplasmic surveillance sensor RIG-I, and thus, subverts the interferon signaling. The 5' end of HCV RNA from HCV infected cells contains a 5' monophosphate, but the binding of miR122 protects HCV RNA from the degradation by 5' exonuclease Xrn-1 and also prevents recognition of the naked 5' end by innate immune sensors and effectors⁹⁰.

1.5.6 Target IFITs

Viruses have been discovered to subvert host defense by directly targeting IFITs. For example, Hepatitis E virus (HEV) RdRp binds to IFIT1 and therefore rescues HEV mRNA translation from IFIT1-mediated inhibition⁹¹. A vaccinia viral protein C9 mediates the proteasomal degradation of IFITs⁹². Influenza virus converts IFIT2 from an antiviral fighter to a pro-viral effector to increase viral protein production⁹³.

1.6 Remaining questions to be answered

The arms race between host and virus is dynamic and complex. IFITs, as one of the major players in host innate immune defense, have been studied to a significant extent over the past three decades since their discovery. However, there are still a few key questions that await to be answered: what is the antiviral mechanism of IFIT2-induced viral inhibition? What is the impact of viral Cap1 methylation and 5'UTR secondary structures on IFIT1 recognition? What is the molecular basis for the synergistic viral repression exerted by IFIT multi-protein complex? Based on these unanswered questions, this dissertation aims to extend our understanding of IFIT-mediated defense mechanisms.

1.7 Outline of the thesis

The work presented in this thesis attempts to answer the questions above. I arrange the project into three main chapters consisting of a fundamental method driving the project and new structure and function discoveries of IFITs as described below.

In Chapter 2, I developed a simple and effective method to separate heterogeneous short RNA transcripts at single nucleotide resolution by quaternary anion exchange chromatography. In order to make homogeneous RNA with sufficient amount for structure studies either by crystallization or cryo-electron microscopy (cryo-EM) approaches, two enzymes are indispensable: T7 RNA polymerase and Capping enzyme. I therefore established detailed purification methods to generate high-quality homemade enzymes as workhorses for RNA production. The short RNA production and purification method is the driving force for making short uncapped and capped RNA used for the biochemical and structure studies in the following two chapters.

In Chapter 3, I determined x-ray crystal structures and a cryo-EM model of IFIT2 bound to different RNAs, which clarified a long-standing false statement in the field: IFIT2 does not bind to ds AU-rich RNA as previously believed but binds to ssRNA with preference to AU-rich sequences. The structure findings show that IFIT2 can sequester short RNA without 5' cap moiety but also can clamp on the mRNA chain when binding to longer capped RNA. Thus, IFIT2 shows a distinct RNA binding mode from the 5' end sequestration mode seen in IFIT1 and IFIT5. This unique RNA binding mode supports a previous study about IFIT2-induced translation enhancement of bound transcripts. Unlike IFIT1 selectively targeting viral messenger RNA (mRNA), IFIT2 seems to act like a double-edged sword. Depending on the abundance of host mRNA versus viral mRNA in IFIT2 surrounding environment, IFIT2 either dominantly binds to host antiviral effector mRNAs and enhances their translation, therefore, indirectly inhibiting viral replication, or IFIT2 dominantly binds to viral mRNAs and enhances the translation of viral proteins, thus, directly promoting viral replication.

In chapter 4, I determined the first x-ray crystal structure of Cap1-RNA bound IFIT1. The structure analysis in comparison to Cap0-RNA bound IFIT1 structure shows that in the absence of secondary structure elements at RNA 5' end, Cap1 methylation alone cannot antagonize IFIT1 recognition, which is in line with the respective ITC affinity measurement of Cap0-, Cap1-, and Cap2-RNA interacting with IFIT1. In reality, viral mRNAs often carry very complex secondary structure elements at 5' end, such as stem-loops and hairpins, in addition to Cap1-methylation. In the presence of secondary structure elements at RNA 5' end, Cap1 methylation antagonizes IFIT1 recognition as seen on the gel binding assay with Zika Cap1-RNA whose 5' untranslated region (UTR) has two stem-loops as barriers to block IFIT1 access. Nevertheless, IFITs also exist in multiprotein complex form, including IFIT1/2/3 trimer and IFIT1/3 tetramer. The gel binding

assay with Zika 5' UTR Cap1-RNA showed that IFIT complexes can overcome viral 5' end stem-loop steric hindrance and Cap1 methylation, compared with IFIT1 alone. I further determined a cryo-EM structure of influenza RNA bound to IFIT1/2/3 complex, which reveals the synergistic collaboration among IFITs for strengthened RNA interactions.

Finally, I give some thoughts on future directions that can be explored to answer some remaining meaningful questions regarding other unknown functions of IFITs and the destiny of IFITs bound viral RNA. The further we explore the unknown world, the broader our horizon become. Facing the broadly unknown world, we realize that what we know is very limited and what we don't know is more than we thought. Discovery won't cease as long as there is passionate and persevered mind.

1.8 Bridge to Chapter 2

Once the objectives are set out, I need to make IFIT proteins and RNA in order to study their interactions. Previous studies from our lab have established purification methods for individual IFIT protein. However, the methods on RNA *in vitro* transcription and subsequent purification and modifications such as capping and methylation are not well established. The commercial capping enzyme is costly for long term usage. Therefore, the first challenge I faced is to generate homemade high-quality enzyme for making sufficient amount of capped RNA used in structure studies. When I dived into this challenging task, the unplanned purification method for short RNA was born along the way.

Chapter 2 A method to isolate short RNA at single nucleotide resolution using quaternary anion exchange chromatography

Abstract

We describe here a simple and effective method, free of acrylamide contamination and without using UV radiation, to separate heterogeneous short RNA transcripts at single nucleotide resolution by quaternary anion exchange chromatography. This method is designed especially for structure study purposes but can also be utilized for other applications requiring highly homogenous short RNA. The quality of short RNA isolated through this method is validated by gel electrophoresis, mass spectrometry, and the crystallization of a short RNA-bound protein complex.

2.1 Introduction

RNA plays a central role in biology, involving gene transcription and protein synthesis⁹⁴, catalysis⁹⁵, gene regulation⁹⁶, cell differentiation as well as tissue and organ development⁹⁷. RNA often associates with another key player in biology—protein—to carry out most of its functions. The main clues for understanding the functions of many ribonucleoprotein machineries lie in the elucidation of RNA-protein interaction at molecular level⁹⁸. X-ray crystallography and cryo-electron microscopy are the two main approaches to reveal molecular architectures of

biomacromolecules⁹⁹. Both techniques require high homogeneity of purified samples. Engineered fusion tags in recombinant proteins along with the well-established affinity, ion-exchange and size exclusion chromatography methods are routinely used to isolate target proteins with high purity¹⁰⁰. For RNA purification, denaturing polyacrylamide gel electrophoresis (PAGE) is the most widely used method to isolate RNA transcript¹⁰¹. However, gel extraction for RNA purification has some limitations, namely: the inevitable contaminant acrylamide; the challenge to separate full-length transcript from aborted or run-off transcripts with few nucleotides difference; the loss of RNA stuck in the gel; and perhaps too often ignored harm—potential skin damage from UV radiation. Despite these shortcomings, gel extraction is a robust method to obtain relatively homogeneous RNA.

Several studies have explored RNA purification using anion exchange chromatography. Easton *et al.* reported the purification of long RNA by diethyl-aminoethyl (DEAE) anion exchange chromatography, however, they admitted slight contamination of abortive transcripts in the purified product and incapability to separate 3' end heterogeneous transcripts¹⁰². Koubek *et al.* tested strong anion exchange Mono-Q column for purifying long transfer RNAs (tRNAs). They pointed out that the strong anion exchange column resolved nucleotides at higher resolution than weak anion exchange column such as DEAE column. Nevertheless, run-off transcripts were still observed in the purified full-length transcript; no convincing evidence to prove the homogeneity level of purified product, since 12% PAGE cannot resolve single nucleotide difference, especially given the length of RNA oligos ranging from 28 to 114 nucleotides¹⁰³. Karlsson *et al.* combined reverse phase ion-pairing and denaturing ion-exchange chromatography and purified four RNA samples ranging from 22 to 82 nucleotides, but the final fractionated RNA elutions were lacking homogeneity based on multiple bands observed on their gel analysis¹⁰⁴. The size-based separation

method is worse than the charge-based method for RNA purification. For instance, size exclusion chromatography can at most separate transcripts with different oligomerization states, but it cannot separate heterogeneous transcripts differed by a few nucleotides that do not significantly change the apparent size resulting from folding¹⁰².

The lack of an effective chromatography method for RNA isolation led us to develop this purification method using quaternary (Q) anion exchange column. The proposed method isolates short RNA (less than ~15 nucleotides) that fits the need for structural studies in terms of both quality at single nucleotide resolution level and quantity at mg level.

2.2 Results

2.2.1 Separation of *in vitro* transcribed short RNA by Q anion exchange chromatography

The optimized 500 µl *in vitro* transcription (IVT) reaction of 5 nucleotides (nt) GUAUA RNA was passed to a strong anion exchanger Q column. As seen on the chromatogram, a huge peak extending from ~98 ml to 114 ml eluted between ~251 and 272 mM NaCl during the gradient salt elution. Fractions were identified on a 20% polyacrylamide 7 M urea gel, the high percentage denaturing PAGE can resolve single nucleotide difference for short RNA. The fractions before the main peak are 4-nt abortive transcripts alone or mixed with 5-nt target transcripts. The fractions of the main peak are 5-nt transcripts GUAUA, with major 5' triphosphate (ppp) ends and very minor 5' diphosphate and monophosphate ends as confirmed by mass spectrometry (**Figure 2.1** and **Supplementary Figure 2.1**).

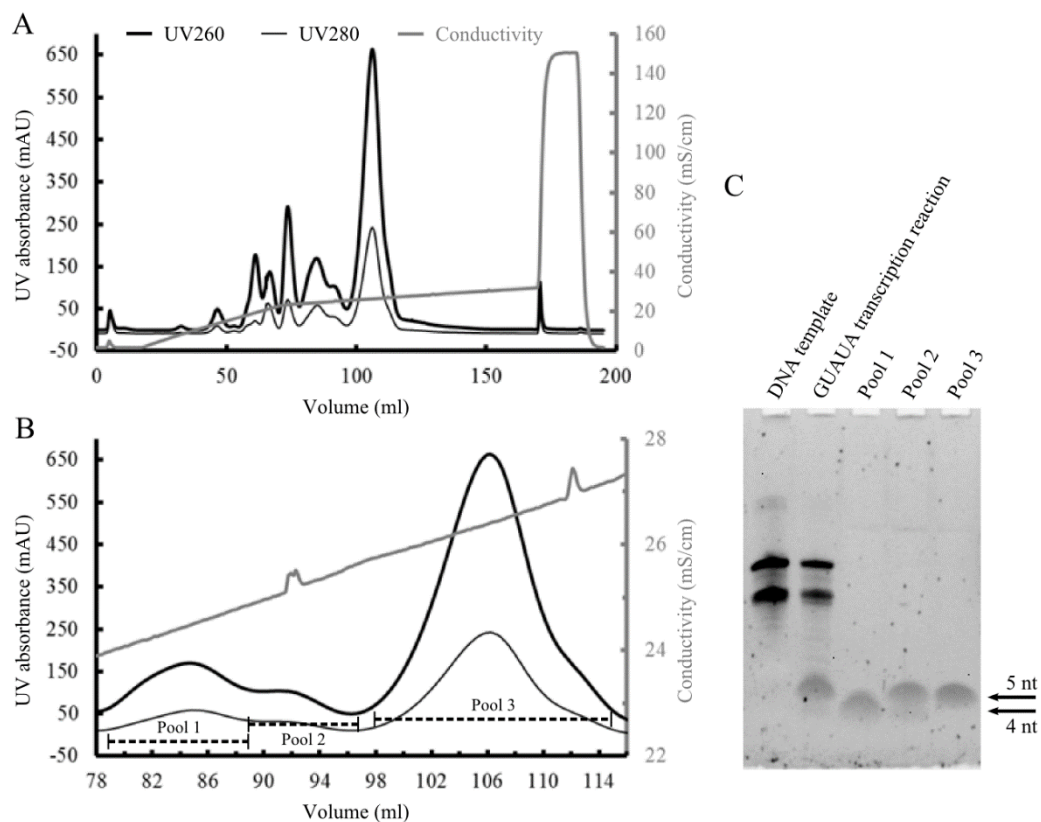


Figure 2.1 Purification of 5'PPP-GUAUA by Q column chromatography. (A) The chromatogram of sample application, gradient washing, gradient elution, and column washing; (B) a zoom-in view of (A) in the volume range of 78 to 116 ml; (C) gel analysis of chromatogram peaks with three pools of fractionations indicated in (B), 7 M urea, 20% denaturing PAGE at 200V for 80 min, SyBr Gold staining.

We also tested Q column purification of 10-nt GGUAGAAUAU RNA. The optimized 200 μ l IVT reaction of 10-nt RNA was loaded to Q column. As shown on the chromatogram, a huge peak extending from ~128 ml to 148 ml eluted between ~340 and 366 mM NaCl during the gradient salt elution. Based on the mass spectrometry quantification and analytical 20% polyacrylamide 7 M urea gel, the small peak right before the main peak corresponds to the 10-nt target transcripts;

the fractions of the main peak are the 11-nt sliding transcripts with an extra U added to the target transcripts (**Figure 2.2** and **Supplementary Figure 2.2**).

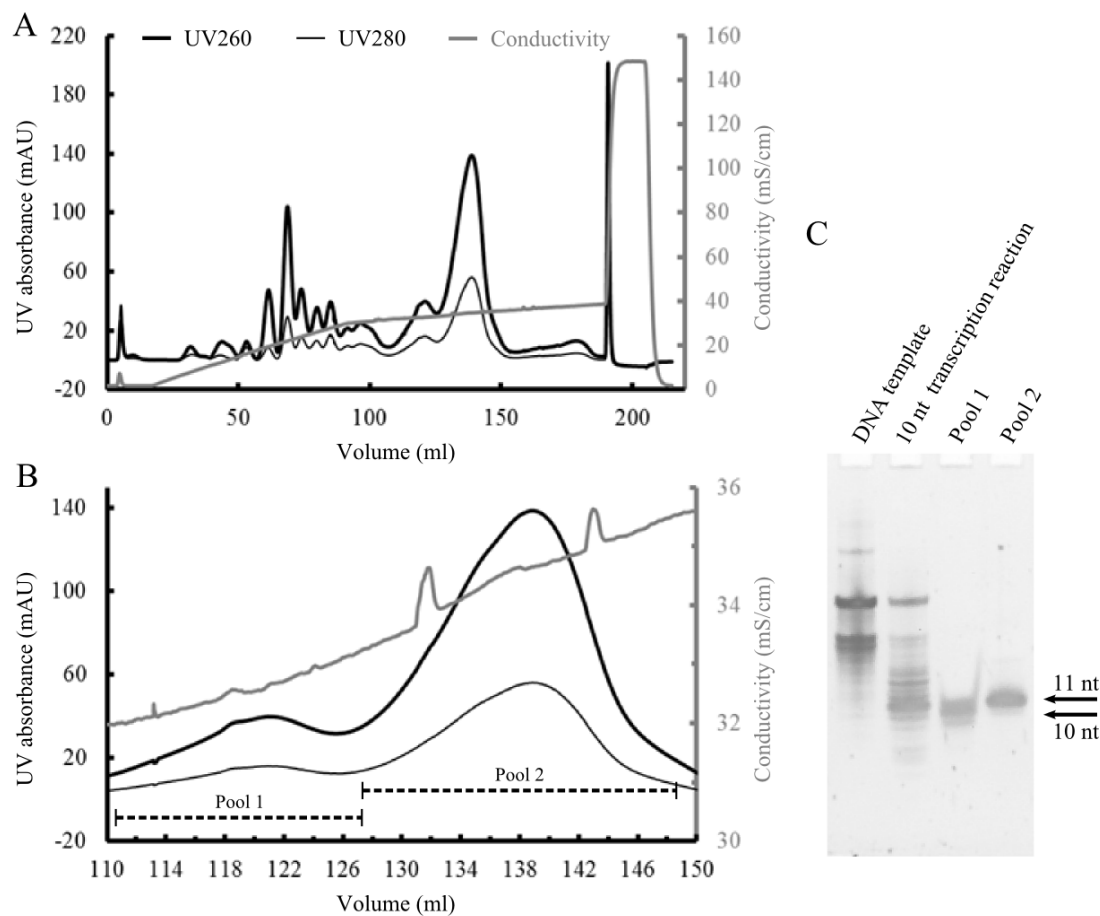


Figure 2.2 Purification of 10-nt 5'PPP-GGUAGAAUAU by Q column chromatography. (A)

The chromatogram of sample application, gradient washing, gradient elution, and column washing; (B) a zoom-in view of (A) in the volume range of 110 to 150 ml; (C) gel analysis of chromatogram peaks with two pools of fractionations indicated in (B), 7 M urea, 20% denaturing PAGE at 200V for 80 min, SyBr Gold staining.

2.2.2 Separation of capped RNA from uncapped RNA

Upon the separation of *in vitro* transcribed short RNA at single nucleotide resolution by Q column chromatography, we further tested the method to isolate capped short RNA from uncapped RNA differentiated by 7-methylguanosine (7mG) cap. We designed a 300 μ l capping reaction using ~184 μ g (~100 nmol) of the 5'ppp-GUAUA RNA purified from pool 3 in Figure 2.1, with our home-made capping enzyme to RNA molar ratio of 1 to 2000 and 1 to 100. On the chromatograms, a single peak spanning the volume from ~85 ml to 100 ml eluted between ~240 and 255 mM NaCl during the gradient salt elution at both molar ratios; for the molar ratio of ~1 to 2000, a second peak spanning the volume from ~110 ml to 120 ml eluted between ~264 and ~278 mM NaCl during the gradient salt elution (**Figure 2.3A-B**). According to the 20% polyacrylamide 7 M urea gel, fractions of the main elution peak were capped RNA transcripts of 7mG-ppp-GUAUA (Cap0-5nt), which was confirmed by mass spectrometry (**Figure 2.3C** and **Supplementary Figure 2.3**). The capping efficiency of the purified (His)₁₀-tagged capping enzyme was close to complete at the capping enzyme to RNA molar ratio of 1 to 100.

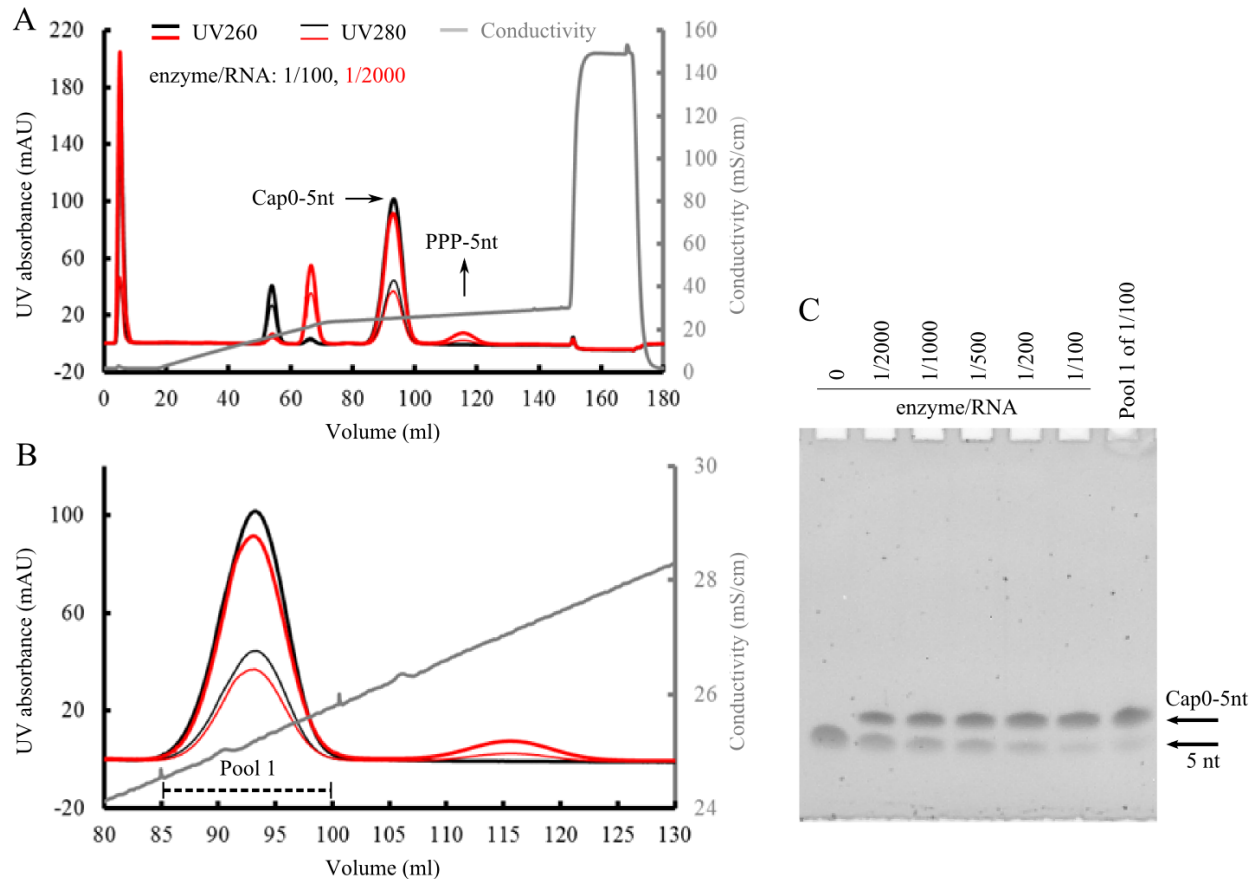


Figure 2.3 Purification of 5'PPP-GUAUA capping reaction by Q column chromatography.

(A) Q column chromatograms of capping reactions at capping enzyme to RNA molar ratio of 1 to 2000 and 1 to 100, respectively; (B) the enlarged view of chromatogram in the volume range of 80 to 130 ml; (C) gel analysis of capping reactions with various purified (His)₁₀-tagged vaccinia capping enzyme to RNA molar ratios, 7 M urea, 20% denaturing PAGE at 200V for 80 min, SyBr Gold staining.

2.2.3 Crystallization of purified capped RNA with a protein complex

To confirm the quality of the capped RNA produced above, we crystallized and solved its structure bound to human IFIT1 protein, for which a crystal structure was previously determined. The purified capped RNA was mixed with IFIT1 in a ratio of 1.5:1 (RNA:protein). Crystallization

using the published conditions resulted in a crystal that diffracted to ~ 2 Å resolution. A Fo–Fc map without inclusion of an RNA model revealed clear difference electron density for the 7mGTP moiety and the first 4 nucleotides of the 5-nt RNA, consistent with its sequence (**Figure 2.4**). The presence of the 5th nucleotide was confirmed by mass spectrometry (**Supplementary Figure 2.3**) but was not visible in the density, presumably due to flexibility, as reported in the initial structural study⁴⁹. The RNA model sequenced as 7mG-GUAU well fits into the electron density, suggesting that the quality of capped short RNA purified by Q chromatography is suitable for structural studies.

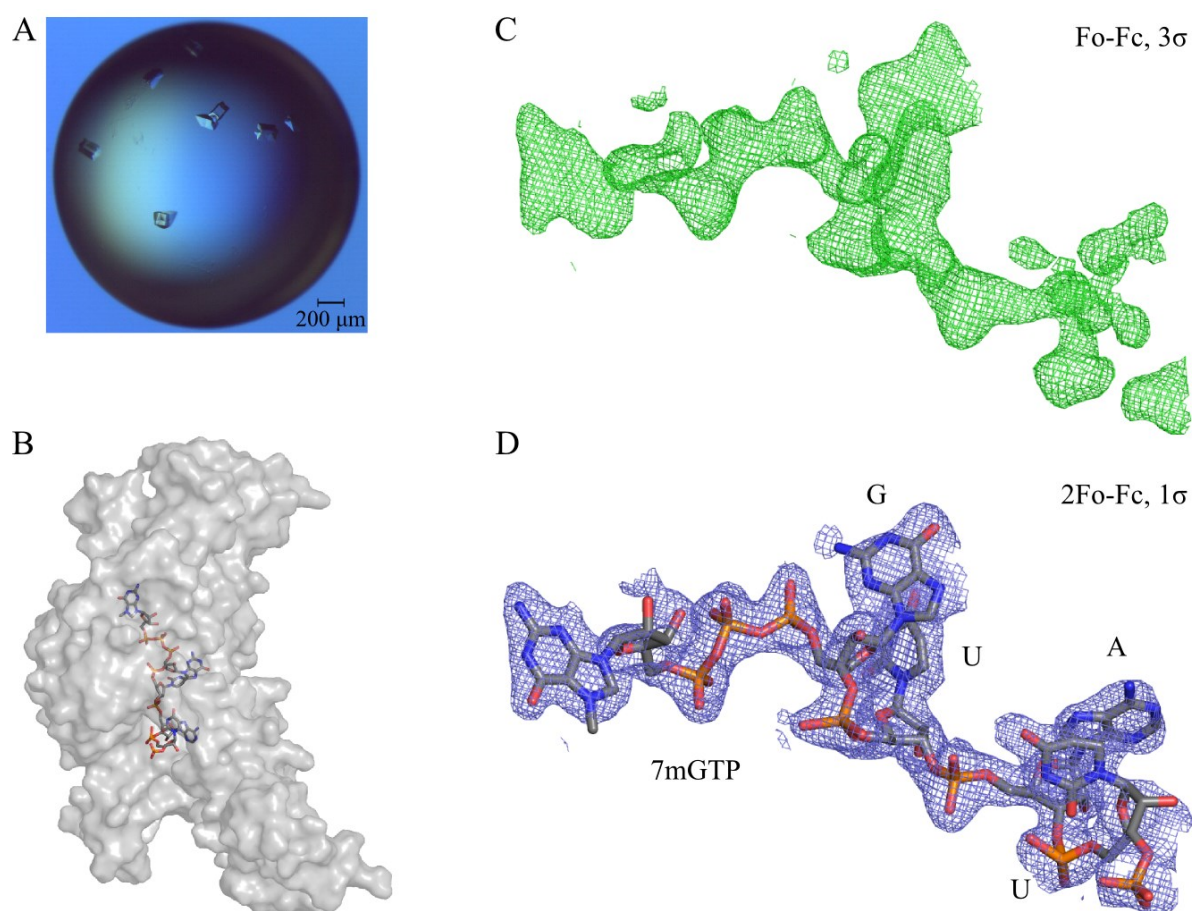


Figure 2.4 Crystallization test using 7mG-GUAUA isolated by Q column chromatography.

(A) A hanging drop of ~ 5 mg/ml IFIT1-RNA complex with crystals; (B) the refined 3D model of

the IFIT1-RNA complex with the capped short RNA sequestered in the protein binding pocket; (C) the Fo–Fc map contoured at 3σ before inclusion of RNA; (D) the refined 2Fo–Fc map contoured at 1σ with the modeled 7mGpppGpUpApUp RNA.

2.3 Discussion

In vitro transcription using bacteriophage T7 RNA polymerase has greatly expanded the capacity to study RNA *in vitro*. However, T7 RNA polymerase has long been known to introduce heterogeneous transcripts at either the 5' end near its initializing domain or the 3' end near the end of the template¹⁰⁵. Early termination of transcription generates abortive transcripts. On the other hand, slide of polymerase off the DNA template generates run-off transcripts. Attempt to generate the full-length transcript of designed sequence, starting at the exact starting point of coding DNA and ending at the finishing point of coding sequence, is rather nontrivial. Therefore, various ways have been introduced to address the heterogeneity issue from the perspectives of sequence design or purification. Self-cleaving ribozymes, including a 5' hammerhead and 3' hepatitis delta virus ribozymes, can be integrated into the termini of desired RNA sequence to flank the target transcript¹⁰⁶. This smart design is not spotless, given the fact that ribozyme cleavage is incomplete and the length of target transcript should be a few nucleotides differing from the length of ribozymes in order to achieve a good separation for gel extraction purification. Moreover, for short RNA (<10 nt), the ribozyme involved in sequence design is rather uneconomical, because only a very small portion of building blocks contribute to making the target RNA whereas the majority of ribonucleoside triphosphates (rNTPs) are wasted to make the long by-product ribozyme transcripts. Additionally, the traditional purification method—gel extraction—has been a faithful way to isolate target RNA. But it cannot resolve long heterogeneous transcripts that are close in length even if a high percentage gel is used. The handling of neurotoxic acrylamide and UV lamp

introduce potential health risks to an experimentalist using the gel extraction method. Thus, a more economic ribozyme-free design and hazard-minimized purification method (such as minimized usage of acrylamide, phenol/chloroform and UV radiation) is required. The method we present meets both requirements and is suitable for making and isolating short RNA with less than 10 nucleotides.

The sequence design contains non-coding stalling nucleotides (**Supplementary Table 2.1**). The stalling nucleotides “force” the polymerase to stop due to the absence of the specific type of rNTP coded by the stalling nucleotide. By doing so, it minimizes or even eliminates run-off transcripts. As seen in our cases, the stalling nucleotides work well to eliminate sliding transcripts for the 5-nt IVT reaction; however, sliding transcripts with an extra U are still prevalent for the 10-nt IVT reaction even with stalling nucleotides. How well the stalling nucleotides stop polymerase from producing run-off transcripts might be dependent on DNA template sequence, and the population of run-off transcripts by adding stalling nucleotides varies with RNA sequence. The stalling nucleotide is an accessory; depending on the sequence of the target transcript, it may contain all four types of rNTPs, making the addition of stalling nucleotides impossible. Quaternary anion exchange chromatography at single nucleotide resolution can separate target short RNA from all other transcripts, regardless of the sequence design strategies.

It is noteworthy that the unincorporated rNTPs eluted from Q column at the early stage (~140-195 mM) of gradient washing step, with volume ranging from ~45 to 70 ml. Another feature differentiating rNTPs from RNA transcript is the ratio of UV260/UV280: the ratio for rNTPs is above 3, but it is around 2.5 for RNA transcript (**Figure 2.1, 2.2 and 2.3**). It's possible to recycle unincorporated rNTPs for an RNA sequence with almost equal portion of each type of rNTPs.

The limitation of this method is that the single nucleotide resolution can only be achieved for short transcripts (less than ~10 nucleotides). Part of the reason could be that the affinity to Q column increases nonlinearly as the length of RNA extends. For RNA within 10 nucleotides, by applying salt gradient at a controlled rate, RNA elutes orderly with the overall charge differing by a single nucleotide. However, for longer transcripts, less controllable scenarios rise. A very refined salt gradient (e.g., 1 mM/ml) will loosen the binding in a “draining” manner so that upon reaching a certain salt concentration threshold, transcripts slowly elute from Q column without yielding a noticeable elution peak. A steeper salt gradient (e.g., 4-20 mM/ml) will enable the appearance of noticeable elution peaks, but each peak actually contains a mixture of transcripts close in length identified by denaturing PAGE and mass spectrometry. Due to stronger affinity of longer RNA (>10 nt) binding to Q column, salt gradient can no longer selectively elute RNA off the column charge-wise differed by single nucleotide, instead, it elutes RNA of varied lengths collectively. Therefore, for longer RNA purification, one has to pursue the traditional way using gel extraction, along with the inclusion of self-cleaving ribozymes in RNA sequence design.

We provide a purification method based on quaternary anion exchange chromatography to isolate highly homogeneous short RNA, 5' end triphosphorylated or capped, suitable for structural and other biochemical applications. This efficient method enables one to start a large-scale *in vitro* transcription reaction and end up with single nucleotide separation of target short RNA from other by-product transcripts in one day.

2.4 Methods

2.4.1 Purification of bacteriophage T7 RNA polymerase (RNAP)

The plasmid pAR1219 containing His-tagged bacteriophage T7 RNA polymerase gene was transformed into *E. coli* strain BL21 cells. A colony from ampicillin containing LB-agar plate was transferred into autoclaved LB media with 100 µg/ml ampicillin to grow overnight. The next morning 25 ml of overnight starter culture was inoculated into each 1 L autoclaved LB media with 100 µg/ml ampicillin. Cells were grown at 37 °C until optical density reaching ~0.6 and then induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and further grown for 3 hours. The following purification procedures were adapted from a previous study on T7 RNAP¹⁰⁷. Cells were harvested and resuspended in Ni binding buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol (βME)) supplemented with Roche protease inhibitor cocktail tablet and 0.1% Triton. The cells were lysed by using French homogenizer, and cell lysate was centrifuged at 50000 g for 30 min. The supernatant was loaded via AKTA purifier sample pump on self-packed Ni-NTA column (~2 ml bed volume of clean beads per litter cells) equilibrated in Ni binding buffer. The column was washed with Ni binding buffer until UV was stabilized, and a further washing was performed with 10% Ni elution buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 200 mM imidazole, 2 mM βME) to remove loosely bound contaminants until UV was stabilized. The His-tagged T7 RNAP was eluted from Ni-NTA column by applying a linear gradient of Ni elution buffer. The eluted protein was dialyzed overnight in dialysis buffer containing 50 mM Tris-HCl pH 8.0, 100 mM NaCl, and 2 mM βME. The next day, the dialyzed sample was passed to a 5 ml HiTrap SP HP column (GE Healthcare) equilibrated with SP buffer A (50 mM HEPES pH 7.5 and 2 mM dithiothreitol (DTT)), T7 RNAP was eluted using a linear gradient of SP buffer B (50 mM HEPES pH 7.5, 1 M NaCl, and 2 mM DTT). Two pools of

fractions were concentrated for the final polishing step of size exclusion chromatography. Each concentrated sample was injected to AKTA pure onto a Superdex 75 Increase 10/300 column (GE Healthcare) equilibrated in gel filtration buffer containing 50 mM Tris-HCl pH 8.0, 200 mM NaCl, and 2 mM DTT. The fractions corresponding to T7 RNAP were pooled and concentrated (**Supplementary Figure 2.4**). The final concentrated sample was spun down at 12-13 krpm for ~7 min before aliquoting into 2-4 mg/ml in 50% glycerol. The aliquots were ready to use and were stored at -20 °C (short term) or -80 °C (long term). This procedure produced ~2 mg highly purified T7 RNAP per liter culture.

2.4.2 *In vitro* transcription

We ordered synthetic DNA oligonucleotides and used them as templates for *in vitro* transcription (IVT). One strand contains a consensus 17-nt T7 promoter sequence plus an initializing nucleotide G. The other strand contains the complementary region to the T7 promoter, the adjacent region that encodes RNA of desired sequence, and ends up with a stalling nucleotide followed by one more random nucleotide (**Supplementary Table 2.1**). The two DNA oligos were respectively resuspended in RNase-free water for 100 µM stock, then an equal volume of each strand resuspension was transferred to a clean tube, annealed at 95 °C for ~3 minutes, and gradually cooled down to room temperature. The concentration of annealed double-stranded DNA template was determined on NanoVue (GE Healthcare) with default factor set as 50, and the concentration was further diluted to 1000 or 500 ng/µl for the convenience of template concentration optimization. The following ingredients were used for IVT reaction: buffer (autoclaved), either 1 M Tris-HCl, pH 8.1 at 37 °C or 1 M HEPES-KOH, pH 7.5 at 37 °C, at a final concentration of ~50 mM; 100 mM Spermidine at a final concentration of 2 mM; 1 M DTT at a final concentration of 20 mM; 100 mM each rNTP at a final concentration of 2-10 mM depending on the composition of

RNA sequence; double-stranded DNA template at a final concentration of 10-100 ng/μl; 2-4 mg/ml T7 RNA polymerase at a final concentration of 50-200 μg/ml; 1 M MgCl₂ (autoclaved) at a final concentration of 20-50 mM; RNase free water. A series of small scale (10 μl) reactions were required to find suitable buffer and optimized concentrations for rNTP, DNA template, T7 RNA polymerase, as well as Mg²⁺. Afterward, the IVT reaction was ready to scale up. The reaction was setup at room temperature by adding room temperature RNase-free water first and adding T7 RNA polymerase last. The reaction was kept at 37 °C incubator for 3-4 hours. Usually white fluffy pyrophosphate·Mg²⁺ precipitate was visible after 2 hours, indicating the working transcription. Extension of reaction time beyond 4 hours could cause RNA degradation outweighing a slight increase of yield. By adding 0.5 M EDTA pH 8.0 to the final concentration of ~50 mM, the reaction was quenched, and the pyrophosphate·Mg²⁺ precipitate was dissolved. The reaction solution was then spun down at 12-13 krpm in a benchtop microcentrifuge for ~7 minutes before injecting into the Q anion exchange chromatography.

2.4.3 Q column chromatography to isolate *in vitro* transcribed short

RNA

The AKTA pure fast-performance liquid chromatography machine without column was flushed by Milli-Q water until any water-soluble junk was washed away indicated by a minimized and stabilized UV signal. A 5 ml HiTrap Q HP column (GE Healthcare) was connected to the chromatography machine and further cleaned by 0.5 M NaOH to remove any trace of nuclease and protein aggregate until a minimized and stabilized UV signal was reached, followed by Milli-Q water washing to remove NaOH. The Q column was further cleaned with buffer B (20 mM Tris-HCl pH 7.8 at 0°C, 2 M NaCl, filtered and autoclaved) to remove any trace RNA bound to the column, then equilibrated with buffer A (20 mM Tris-HCl pH 7.8 at 0°C, filtered and autoclaved).

The ready-to-inject IVT reaction was injected into the chromatography machine. The chromatogram was programmed to run through the following four steps: sample application with buffer A only; gradient buffer B wash at ~4 mM/ml; gradient elution at ~1 mM/ml; and column washing with buffer B and buffer A, respectively. Afterward, the column was further cleaned with 0.5 mM NaOH and water for the next usage. The pool of fractions was then subjected to ethanol precipitation in anhydrous alcohol with 0.3 M NaOAc pH 5.2 (autoclaved) at -20 °C overnight. The RNA was pelleted by centrifugation, washed with 70% ethanol, air-dried, and then resuspended in RNase-free water for subsequent applications.

2.4.4 Mass spectrometry

The mass of RNA was determined by LC-MS using a Dionex Ultimate 3000 coupled to a Bruker Maxis Impact QTOF in negative ESI mode. 10-20 µl of 20 µM sample was separated on an AgilentAdvanceBio C18 column (particle size 2.7 µM; pore size 120 Å; diameter × length 2.1 × 50 mm). A programmed run with a gradient of 98% mobile phase A (100 mM HFIP and 5 mM TEA in H₂O) and 2% mobile phase B (MeOH) to 40% mobile phase A and 60% mobile phase B in 10 minutes was then performed. The data was processed using the Bruker DataAnalysis software v4.2.

2.4.5 Vaccinia capping enzyme Purification

We received the plasmid for expressing vaccinia capping enzyme as a gift from Dr. Remco Sprangers¹⁰⁸. However, this construct gave very unclean Ni elution despite using stepwise or gradient elution. The (His)₆-tagged D1 subunit was easily eluted at ~30 mM imidazole, and some bacterial proteins also came off together that could not be separated by subsequent steps, resulting in impure capping enzyme product which was also seen from several contaminant bands observed

on the SEC purified product¹⁰⁸, and consequently, nonoptimal capping efficiency (**Supplementary Figure 2.5**). The Shuman group who solved the crystal structure of vaccinia capping enzyme used (His)10-tagged D1 construct which endured 60 mM imidazole wash to make Ni elution much cleaner¹⁰⁹. We therefore re-engineered the (His)6-tagged construct by adding four extra histidines for tighter binding to the Ni-NTA column. The insertion mutagenesis primer design was based on the method developed by Liu¹¹⁰. The inserted extra histidines to the original enzyme construct were verified by sequencing. For the expression trials of the engineered (His)10-tagged construct, we tested five bacterial strains: BL21 (DE3), BL21 (DE3) pLyss, Rosetta 2, Rosetta 2 pLyss, and Arctic Express RIL. We found Rosetta 2 cells gave the highest yield of target protein and the lowest expression of background bacterial proteins. Therefore, the (His)10-tagged capping enzyme was expressed using Rosetta 2 cells in TB media. The purification of the capping enzyme was adapted according to its crystal structure paper¹⁰⁹. The cells were grown at 37 °C until OD reached ~1.2, then the temperature was reduced to 18 °C, and cells were induced with 1 mM IPTG and further grown overnight. The harvested cells were resuspended in Ni binding buffer (40 mM Tris-HCl pH 8.0, 200 mM NaCl, 20 mM imidazole, 2 mM β ME), supplemented with Roche protease inhibitor cocktail tablet, 0.1% Triton, and 100 μ g/ml lysozyme. The cells were lysed by using French homogenizer, and cell lysate was centrifuged at 50000 g for 30 min. The supernatant was loaded via AKTA purifier sample pump on to a self-packed Ni-NTA column (~2 ml bed volume of clean beads per litter cells) equilibrated in Ni binding buffer. The column was washed with Ni binding buffer until UV was stabilized, a further washing step was performed with 10% Ni elution buffer (40 mM Tris-HCl pH 8.0, 200 mM NaCl, 250 mM imidazole, 2 mM β ME) to remove loosely bound contaminants until UV was stabilized. The (His)10-tagged vaccinia capping enzyme was eluted from Ni-NTA column by applying a linear gradient of Ni elution buffer. The eluted

protein was dialyzed overnight in dialysis buffer (40 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM β ME). The next day, the dialyzed sample was passed to a 5 ml HiTrap Heparin HP affinity column (GE Healthcare) equilibrated with heparin buffer A (40 mM Tris-HCl pH 8.0 and 2 mM DTT). The capping enzyme was eluted using a linear gradient of heparin buffer B (40 mM Tris-HCl pH 8.0, 1 M NaCl, and 2 mM DTT). Two pools of fractions were concentrated for the final polishing step of size exclusion chromatography. Each concentrated sample was injected to AKTA pure on to a Superdex 75 Increase 10/300 column (GE Healthcare) equilibrated in gel filtration buffer (40 mM Tris-HCl pH 8.0, 400 mM NaCl, and 2 mM DTT). The fractions corresponding to the capping enzyme were pooled and concentrated. The final concentrated sample was spun down at 12-13 krpm for ~7 min before aliquoting into ~4 mg/ml in 50% glycerol. The aliquots were ready to use and were stored at -20°C (short term) or -80°C (long term). This procedure produced ~1 mg well purified vaccinia capping enzyme per liter culture (**Supplementary Figure 2.6**).

2.4.6 Capping reaction

The 10x capping buffer (500 mM Tris pH 8.0, 50 mM KCl, 10 mM MgCl_2) was prepared, autoclaved, aliquoted, and stored at -20°C . The capping reaction was set up with 1x capping buffer, 32 mM S-adenosyl-methionine (New England Biolabs) at a final concentration of ~0.2-0.5 mM, and 100 mM Guanosine triphosphate (GTP) at a final concentration of ~0.5 mM, 5'ppp-GUAUA, purified capping enzyme, 1 mM DTT, and RNase-free water. Small-scale 10 μl capping reactions were performed with varied enzyme to RNA ratios to find optimized capping efficiency. (**Figure 2.3C**) Then a larger scale capping reaction was set up: ~ 0.18 mg (~100 nmol) 5'ppp-GUAUA with capping enzyme to RNA molar ratio of 1:100. The reaction was incubated at 37°C for 2 hours. Afterward, the reaction was spun down at ~12-13 krpm for ~7 min to pellet

pyrophosphate·Mg²⁺ precipitate, and the supernatant was directly injected into the Q column chromatography.

2.4.7 Q column chromatography to isolate short capped RNA

The AKTA pure and the 5 ml Q column were cleaned in the same way as aforementioned with the same buffer A and B. The chromatogram was programmed to run through the following four steps: sample application with buffer A only; gradient wash at 4 mM/ml; gradient elution at 1 mM/ml; and column washing with buffer B and A, respectively. The ethanol precipitation and quantification of the capped RNA were the same as the treatment for *in vitro* transcribed RNA.

2.4.8 Crystallization

The expression and purification of IFIT1 protein (L457E/L464E mutant) as well as crystallization condition were adapted accordingly⁴⁹. The protein-RNA complex was incubated on ice for ~30 min after mixing each individually purified component. The complex was then set up for crystallization screening at a 24-well plate in three protein concentrations: 5, 7.5, and 10 mg/ml, with RNA to protein molar ratio of ~1.5:1. All dilutions were made with gel filtration buffer (20 mM Tris pH 7.8 at 0°C, 150 mM NaCl, supplemented with 1 mM DTT). The best crystal hit condition was at 4 °C, with 200 mM CaCl₂, 23-27% PEG 200, 100 mM Tris pH 8.1, at the protein concentration of ~5 mg/ml.

2.5 Acknowledgement

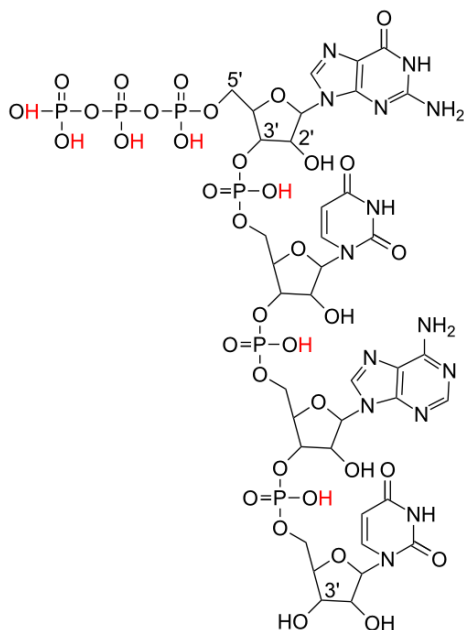
We thank Dr. Remco Sprangers (Universität Regensburg) for kindly providing the gift of 6xHis tagged vaccinia capping enzyme plasmid, and Dr. Alexander S. Wahba for RNA LC-MS (McGill University).

2.6 Supporting information

5'PPP-GUAU

$C_{38}H_{50}O_{36}N_{14}P_6$

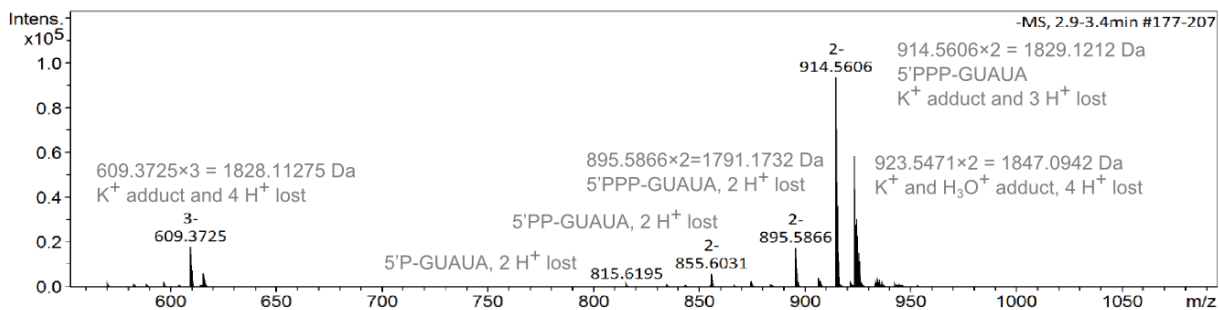
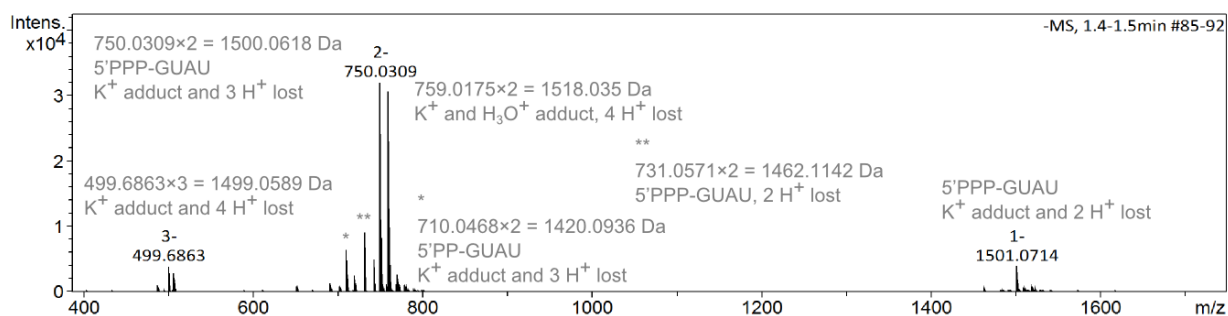
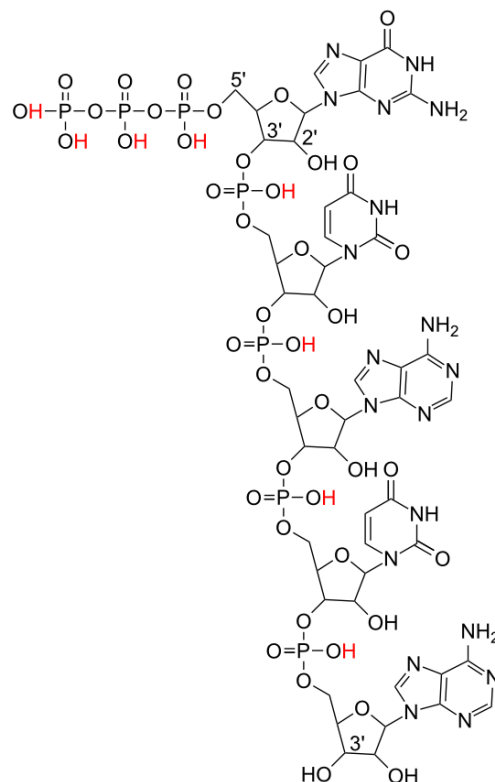
Complete molecular mass: 1464.724 g/mol (Da)



5'PPP-GUAUA

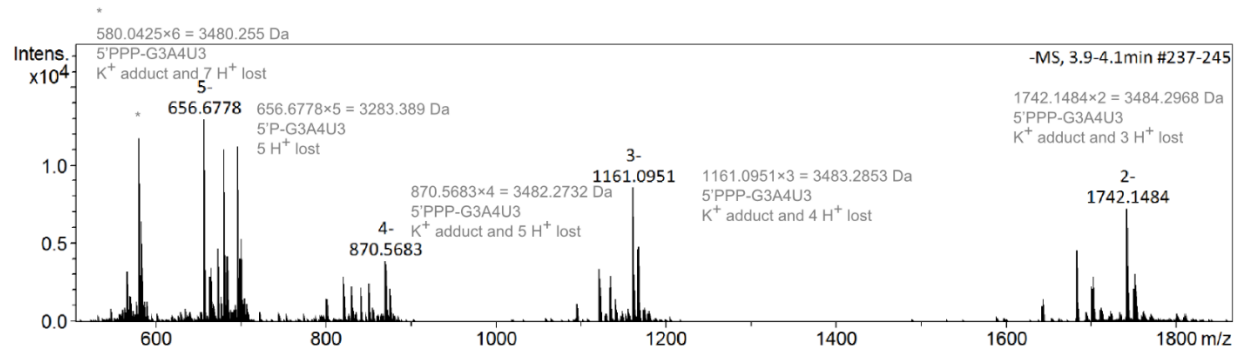
$C_{48}H_{62}O_{42}N_{19}P_7$

Complete molecular mass: 1793.9243 g/mol (Da)

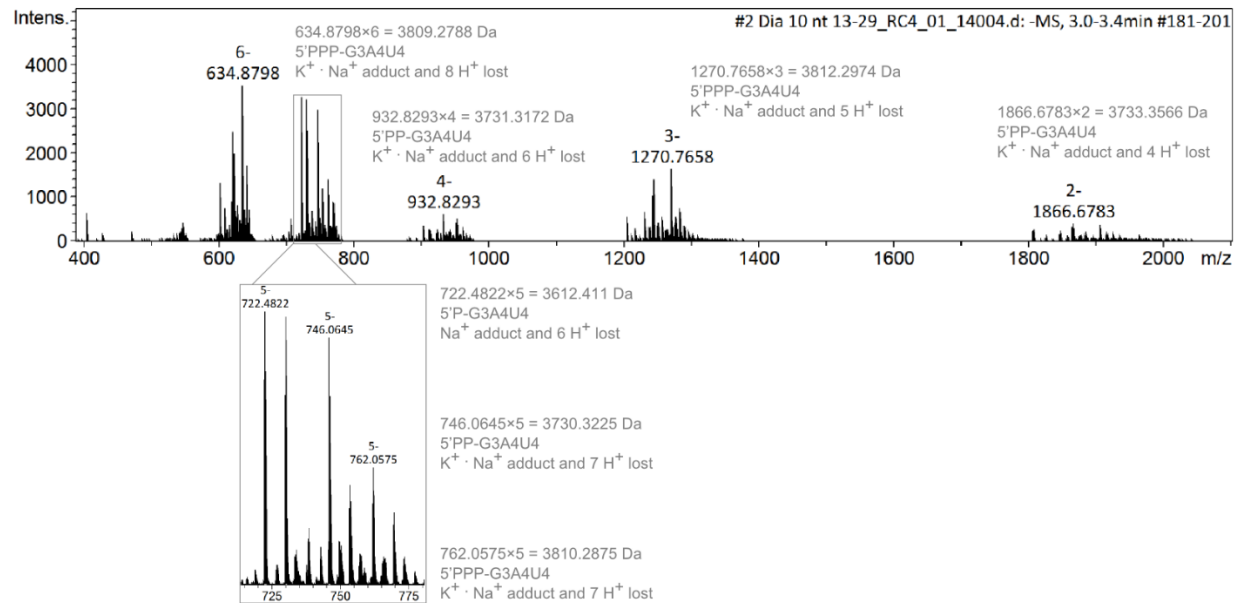
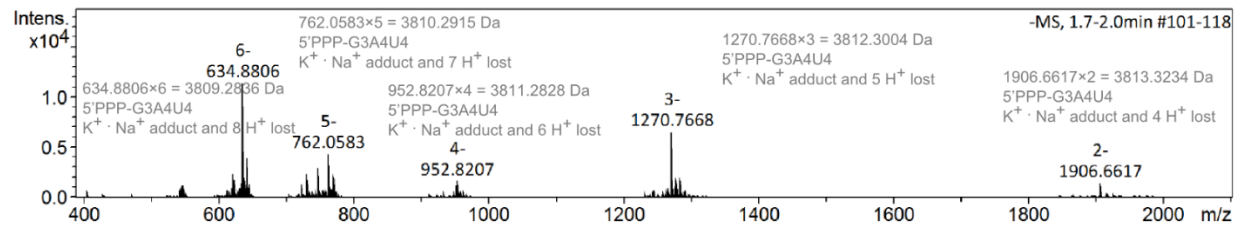


Supplementary Figure 2.1 The Q column purified 5'PPP-GUAU and 5'PPP-GUAUA confirmed by mass spectrometry. The chemical formula of 5'PPP 4-nt and 5-nt transcripts; the mass spectra of Q column fraction pools 1 and 3 in Figure 2.1.

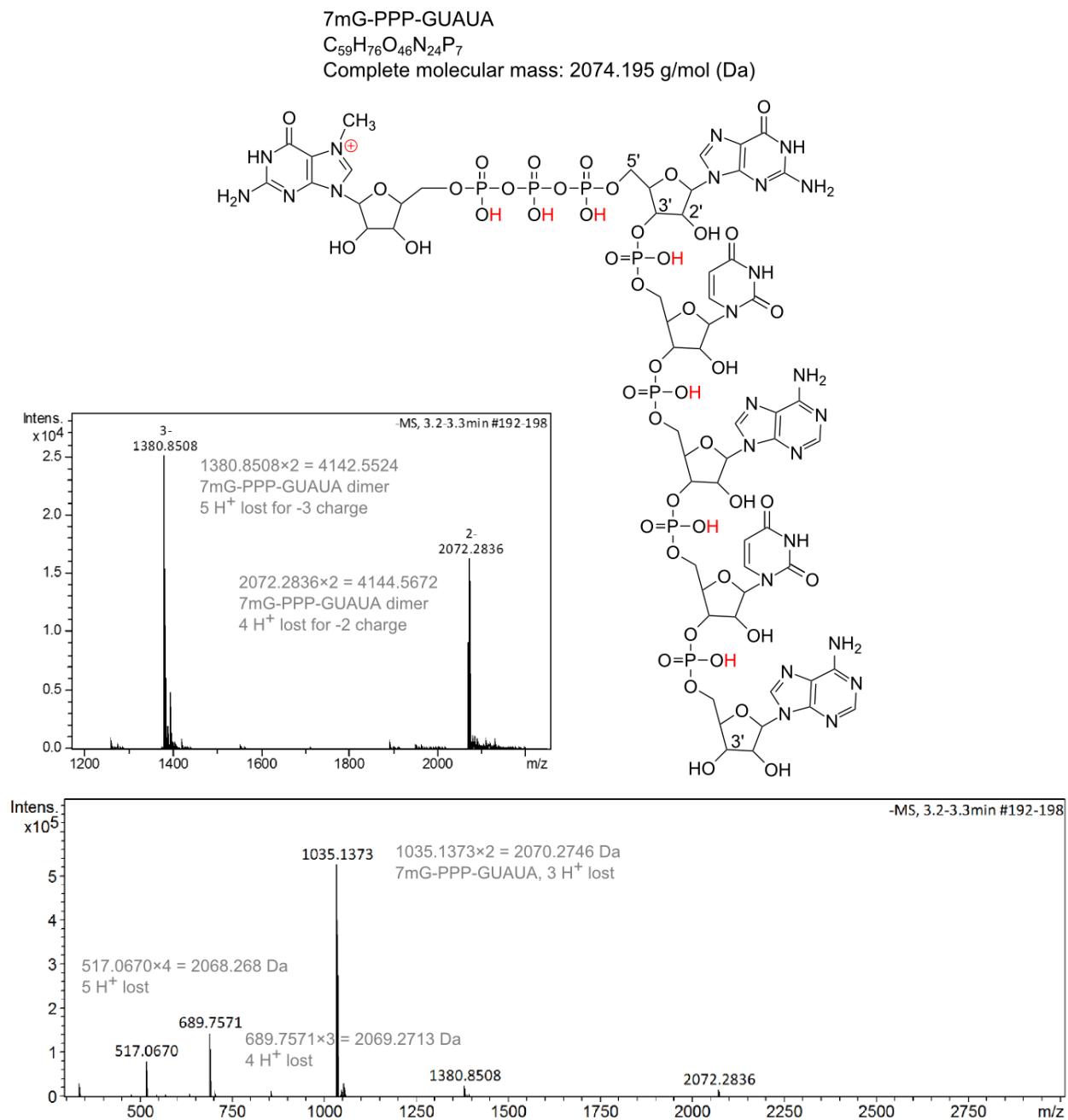
5'PPP-GGUAGAAUUAU (5'PPP-G3A4U3)
Complete molecular mass: 3448.934 g/mol (Da)



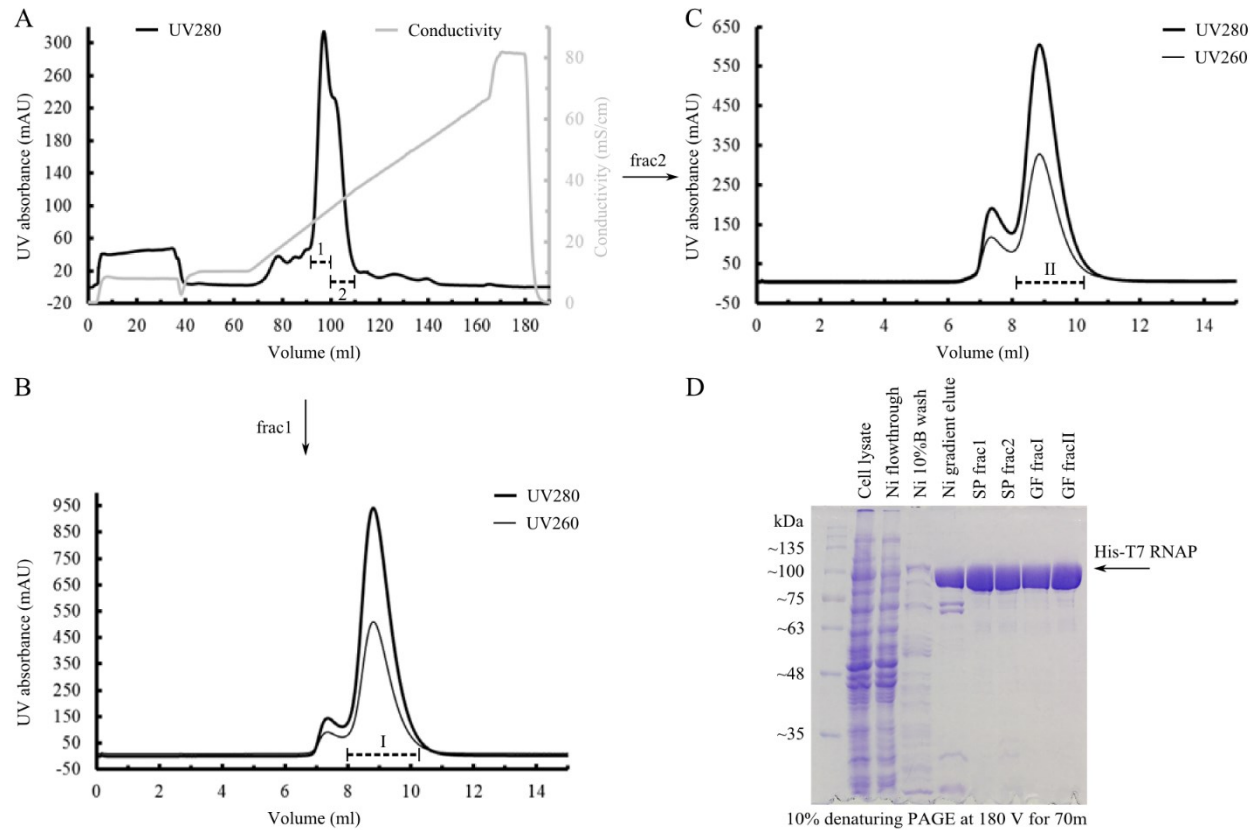
5'PPP-GGUAGAAUUAU+U (5'PPP-G3A4U4, sliding transcript)
Complete molecular mass: 3755.101 g/mol (Da)



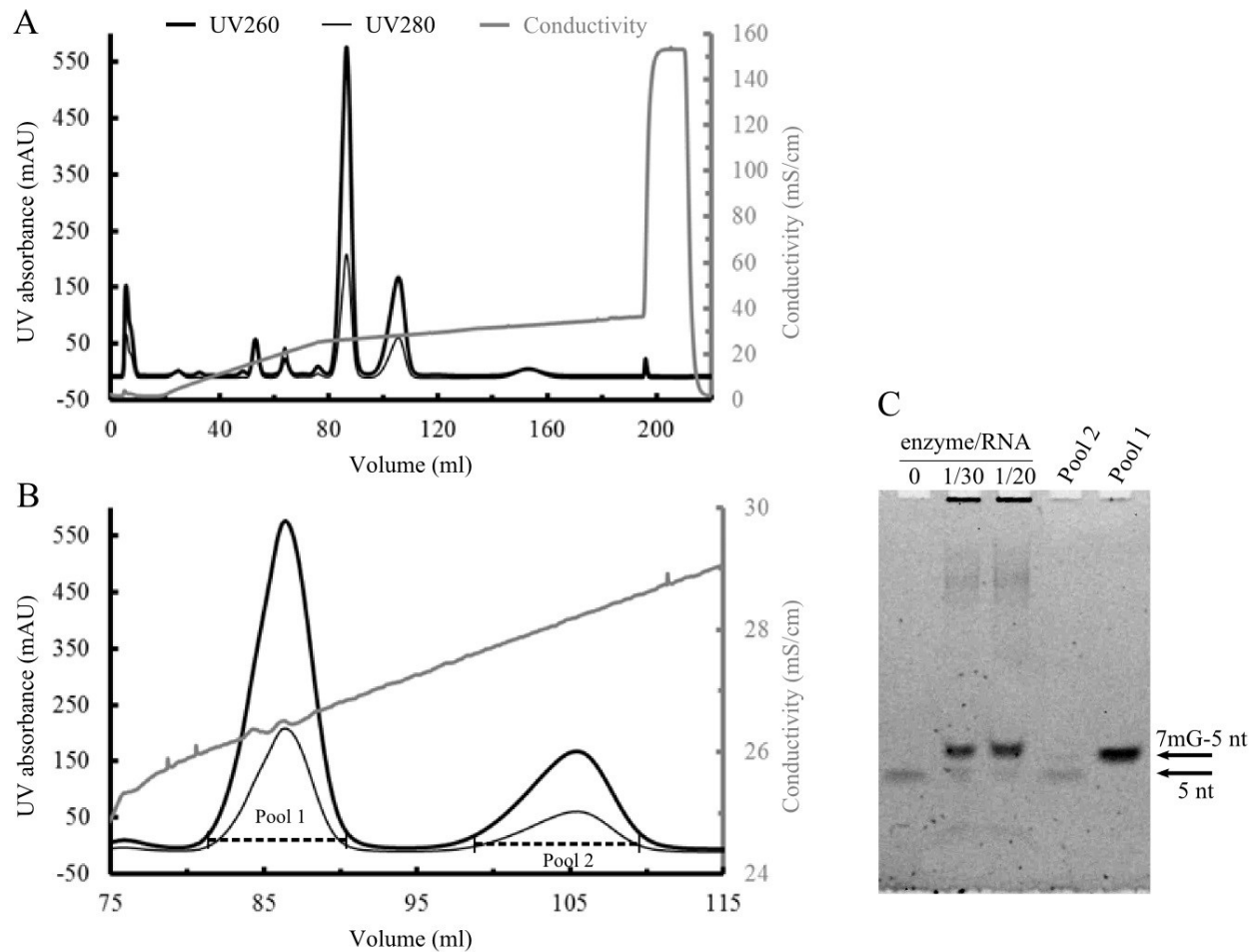
Supplementary Figure 2.2 The mass spectra of Q column fraction pool 1 (10-nt target transcripts) and pool 2 (11-nt sliding transcripts) in Figure 2.2.



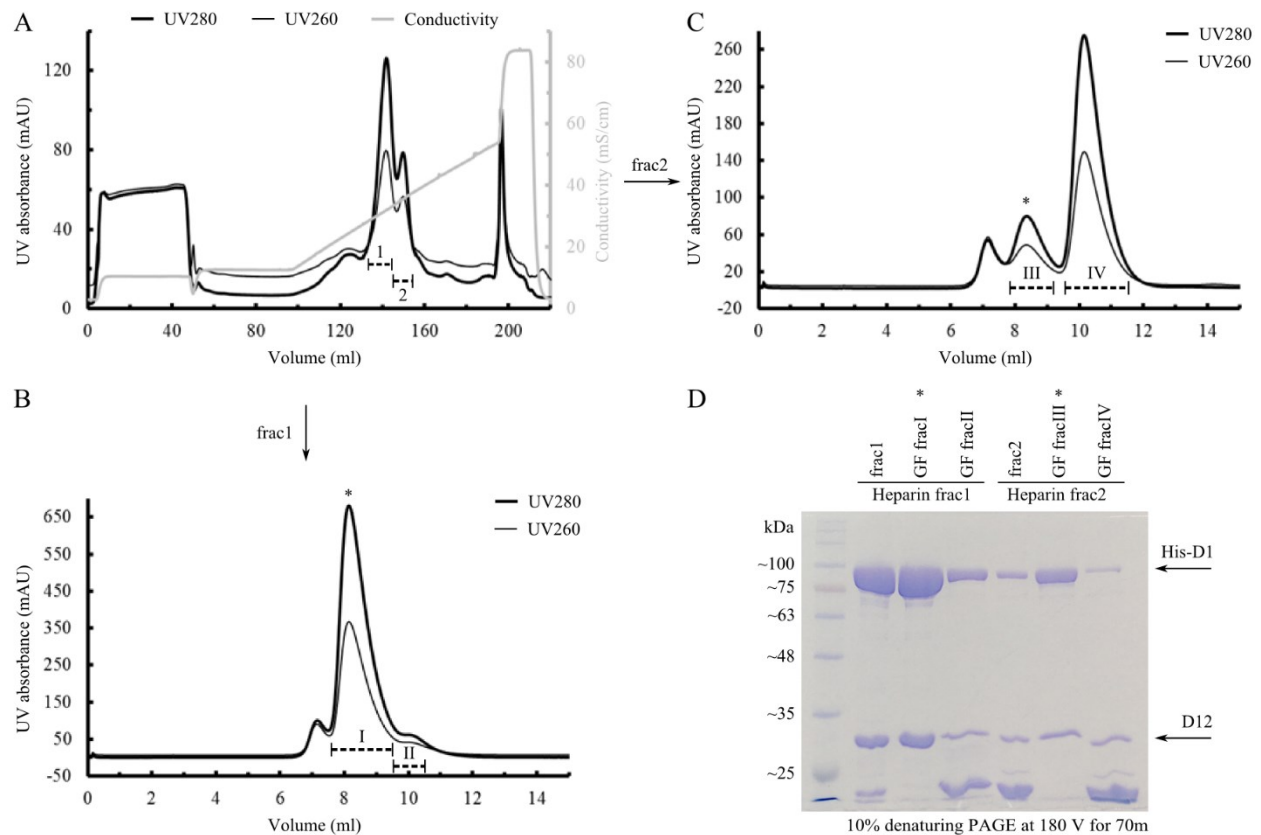
Supplementary Figure 2.3 The chemical formula of 7mG-PPP-GUAUA; the mass spectrum of Q column fraction pool 1 in Figure 2.3, with a zoom-in window.



Supplementary Figure 2.4 The purification of T7 RNAP. (A) Chromatogram of SP column run, (B) gel filtration (GF) chromatogram of SP pool of fractions 1, (C) gel filtration (GF) chromatogram of SP pool of fractions 2, and (D) sample purity at each step shown by SDS-PAGE.



Supplementary Figure 2.5 Purification of 1 ml 5'PPP-GUAUA capping reaction by Q column chromatography using impure capping enzyme from the original capping enzyme construct without tag modification. (A) The chromatogram of sample application, gradient washing, gradient elution, and column washing; (B) a zoom-in view of (A) in the volume range 75-115 ml; (C) a gel of capping reactions with three capping enzyme to RNA molar ratios, and 2 pools of fractionations indicated in (B).



Supplementary Figure 2.6 The purification of engineered (His)10-tagged vaccinia capping enzyme. (A) Chromatogram of heparin column run; GF chromatogram of heparin pool of fractions 1 (B) and 2 (C), peak corresponding to vaccinia capping enzyme complex marked with *; and (D) sample purity at heparin and GF steps shown by SDS-PAGE.

Supplementary Table S1. The 5' to 3' DNA sequences used for 5-nt and 10-nt IVT reactions.

5-nt non-coding strand	TAA TAC GAC TCA CTA TA G
5-nt template strand	AG TATAC TA TAG TGA GTC GTA TTA
10-nt non-coding strand	GAA AT TAA TAC GAC TCA CTA TA G
10-nt template strand	GG ATATTCTACC TA TAG TGA GTC GTA TTA AT TTC

T7 Class III promoter sequence is in bold. Color coding: **five additional nucleotides** for promoting T7 RNA polymerase binding (optional), **RNA coding sequence**, **stalling nucleotides** to stop T7 RNA polymerase from sliding (optional).

2.7 Bridge to Chapter 3

I established the methods for RNA in vitro transcription and subsequent purification as well as downstream capping and methylation in Chapter 2, which give me freedom to study IFIT2-RNA interaction from all possible perspectives available to us in Chapter 3.

Chapter 3 Human IFIT2 shows distinct RNA binding from IFIT1 and IFIT5

Abstract

IFIT2 has been known to inhibit viral infection, but the mechanism for IFIT2-exerted viral inhibition is unclear. Whether IFIT2 can capture viral RNA as two other family members IFIT1 and IFIT5 has remained unanswered over the past decade, due to the lack of RNA bound IFIT2 structure. In this study, we present two x-ray crystal structures and one cryo-EM model to unveil the molecular mechanisms of IFIT2 interacting with different RNAs. Unlike the sequestration of RNA 5' end observed in IFIT1 and IFIT5, IFIT2 appears distinct RNA binding modes to interact with various RNAs based on 5' end, sequences and secondary structures. The flexible RNA binding modes of IFIT2 expand our understanding about the non-redundant defense mechanisms of IFITs against viruses.

3.1 Introduction

IFIT2 is one of the members of a protein family called interferon-induced protein with tetratricopeptide repeats (IFITs) in humans, along with three other members that have been studied to-some-extent: IFIT1, IFIT3 and IFIT5¹¹¹. IFITs are generally known as innate immune response effector proteins. The basal expression level is barely detectable under normal conditions whereas massive production (~1-3 million copies of protein per cell) is rapidly triggered by viral infection⁴⁸.

The signaling cascades are interferon-dependent in most cases. Pathogen associated molecular patterns (PAMPs), such as viral nucleic acids, are detected by pattern recognition receptors (PRRs) present in cytosol and endosomes, including toll-like receptors, retinoic acid-inducible gene I (RIG-I)-like receptors and nucleotide-binding oligomerization domain (NOD)-like receptors. The recognition of PAMPs by PRRs activates interferon regulatory factors (IRF) and leads to the expression of type I interferons (IFN)³⁷. IFNs are secreted in paracrine and autocrine manners and bind to IFN receptors on the cell surface to alert surrounding uninfected cells, through the JAK-STAT signaling pathway, resulting in the upregulation of hundreds of IFN stimulated genes (ISGs). The antiviral signaling can also be interferon-independent especially at the very beginning of infection with low numbers of virus particles: IRF3 can directly trigger the expression of a small subset of ISGs in a more efficient way to prime the immune system and block viral replication without activating an elaborate immune response³⁹.

IFITs are among the most prevalent ISGs⁴⁰. As key players of the down-stream immune effector proteins, IFITs inhibit viral infections mainly through directly engaging in viral nucleic acids either individually or collectively⁴⁸. IFIT5 was discovered to sequester 5' triphosphate (PPP-) end of single-stranded (ss) viral RNA⁴⁶. IFIT1 was found to bind predominantly Cap0 ss viral RNA⁴⁹. IFIT3 does not bind to any RNA. Mouse ifit2, the homolog of human IFIT2, has shown an inhibitory effect on viral replication of positive-sense^{50, 51} and negative-sense⁵²⁻⁵⁴ ssRNA viruses. Nevertheless, the reason behind the observed viral inhibition remains unclear. Would IFIT2 repress viral replication also by directly binding to viral RNA, like the mechanism used by IFIT5 and IFIT1? Yang *et al.* solved the *apo* IFIT2 structure, which revealed a domain-swapped parallel dimer⁵⁵. Their biochemical studies suggested that IFIT2 potentially interact with ds AU-rich RNAs; however, the evidence supporting IFIT2-RNA interaction was controversial due to the excessive

amounts of RNA used in their gel-shift binding assays (~20 μ M vs. typically under 100 nM) and the absence of binding competing agent necessary for binding specificity.

Very interestingly, a recent study revealed that IFIT2 was converted from a traditional antiviral effector to a pro-viral enhancer during influenza viral infection. Tran *et al.* discovered that IFIT2 interacts with AU-rich regions of both host and influenza viral mRNA using UV crosslinking and immunoprecipitation together with sequencing analysis. Among IFIT2 bound RNA sites, ~20% of them contained a consensus sequence motif **UAGnnUAU**, n refers to any nucleotides. IFIT2 enhances the translation efficiency of bound transcripts by reducing ribosome pausing, influenza virus maximizes viral transcripts bound by IFIT2 over host transcripts and therefore enhances viral protein production to favor viral replication⁹³. Based on previous studies about immunological functions of IFIT2, we present here the structure basis of IFIT2-RNA interactions with two crystal structures at the resolution of 3.2 and 2.7 Å as well as a ~4 Å cryo-EM model. Together with biochemical studies and functional studies in cells, our findings suggest the following: i) IFIT2 can not bind to dsRNA but binds to ssRNA, ii) AU-rich sequences favor IFIT2-RNA binding, iii) IFIT2 sequesters short RNA without 5' cap, iv) the domain-swapped dimerization of IFIT2 does play a functional role on binding to host or viral mRNA.

3.2 Results

3.2.1 IFIT2 does not bind to dsRNA

Since the previous study suggested that IFIT2 interacted with ds AU-rich RNA⁹³, we tested IFIT2 interaction with a dsRNA. The 24-nt ssRNA has the sequence GG(AU)10CC, so it can self-anneal with another strand in an anti-parallel manner to form 48-nt ds RNA (**Figure 3.1A**), denoted as ds48. At the regular concentration of 50 nM for ds48, even in the absence of heparin, IFIT2 did

not show a noticeable shift with ds48, neither did IFIT5 as a negative control for it cannot accommodate dsRNA⁴⁶. However, at the very high concentration of 500 nM beyond the typical concentration of lower than ~100 nM, without the binding competing agent heparin, IFIT2 showed a very faint shifted band. With heparin, the shifted band disappeared, indicating that IFIT2 could form non-specific interactions with dsRNA of excessive amount without any binding competing agent (**Figure 3.1A**). We found that IFIT2 appears not to interact with ds AU-rich RNA and may bind to some RNA in a non-specific fashion, which can be verified by using binding competing agents.

3.2.2 IFIT2 specifically binds to ssRNA

Next we tested IFIT2 interaction with a RNA used in our previous studies⁴⁹, namely ss44. According to the secondary structure prediction by M-fold server¹¹²: the 44-nt RNA ss44 starts with a 22-nt linear region, followed by a stem loop, and ends with an 8-nt linear 3' overhang (**Figure 3.1B**). We qualitatively assessed the binding affinity of ss44 with IFIT2 and IFIT5 by titration in the presence of 0.01 mg/ml (~740 nM) heparin. The ss44 concentration was fixed at 50 nM and the protein concentration was gradually increased from 100 to 250 nM in increment of 50 nM. The estimated ss44 binding K_d is ~200-250 nM for IFIT2 and ~150-200 nM for IFIT5 (**Figure 3.1B**).

In order to locate the binding region in ss44, we “cut” ss44 into two equal-length 22-nt segments, namely ss1-22 and ss23-44 (**Figure 3.1C**), by *in vitro* transcription with tailored templates, respectively. Because ss1-22 and ss23-44 contained 50% less nucleotides compared to ss44, the signal for SyBr Gold staining was roughly 50% weaker. Therefore, we increased the concentrations of ss1-22 and ss23-44 twice as ss44 concentration to achieve a comparable staining with the ss44 for binding assay. When 100 nM of ss1-22 or ss23-44 and 50 nM of ss44 were incubated with IFIT2 and IFIT5, respectively, and were loaded on the gel shift assay with protein to RNA ratio of 4 to 1, ss1-22 and ss44 both specifically interacted with IFIT2 and IFIT5. However, ss23-44 showed no interaction with either IFIT2 or IFIT5 (**Figure 3.1C**). IFIT5, as a negative control, is known to specifically bind ss 5'ppp-RNA; ss23-44 forms stem loop at 5' end and thus prohibits IFIT5 from binding. Consistent with the observation of IFIT2 to ds48, IFIT2 does not interact with ss23-44 due to the hairpin at 5' end. These observations suggest that IFIT2 specifically binds to ssRNA.

3.2.3 AU-rich sequences enhance IFIT2 binding to ssRNA

Using size exclusion chromatography, we tested if IFIT2 could form a stable complex in solution with a 10-nt AU-rich RNA UGUAGAAUAU containing the consensus motif sequence⁹³. We incubated the 10-nt AU-rich RNA with IFIT2 dimer at a molar ratio of ~2.2:1 and injected the mixture on a size exclusion chromatography column. Clearly, a stable IFIT2-RNA complex was formed with the size larger than either the protein or RNA alone. We also observed a small elution peak corresponding to excessive RNA (**Figure 3.2A**). This suggests that IFIT2 dimer binds to two copies of RNA molecules, with one copy of RNA bound by each monomer.

Next, we used isothermal titration calorimetry to quantify and compare the binding affinity of IFIT2 dimer with two 11-nt RNAs: one AU-rich 5'HO-UGUAGAAUAUU and the other non-AU-rich 5'HO-GGGAGAGAGAG. For the AU-rich 11-nt ssRNA 5'HO-UGUAGAAUAUU with IFIT2 dimer, the measured K_d was $\sim 189.4 \pm 28.7$ nM (**Figure 3.2B**). In comparison, for the 11-nt non-AU-rich ssRNA 5'HO-GGGAGAGAGAG with IFIT2 dimer, the quantified K_d was $\sim 709 \pm 22$ nM (**Figure 3.2C**). For the same form RNA (5' end and length), the AU-rich sequence resulted in a tighter binding affinity with IFIT2 dimer approximately three-fold stronger than that of the non-AU-rich sequence. Therefore, the IFIT2-RNA interactions were strengthened by AU-rich sequences.

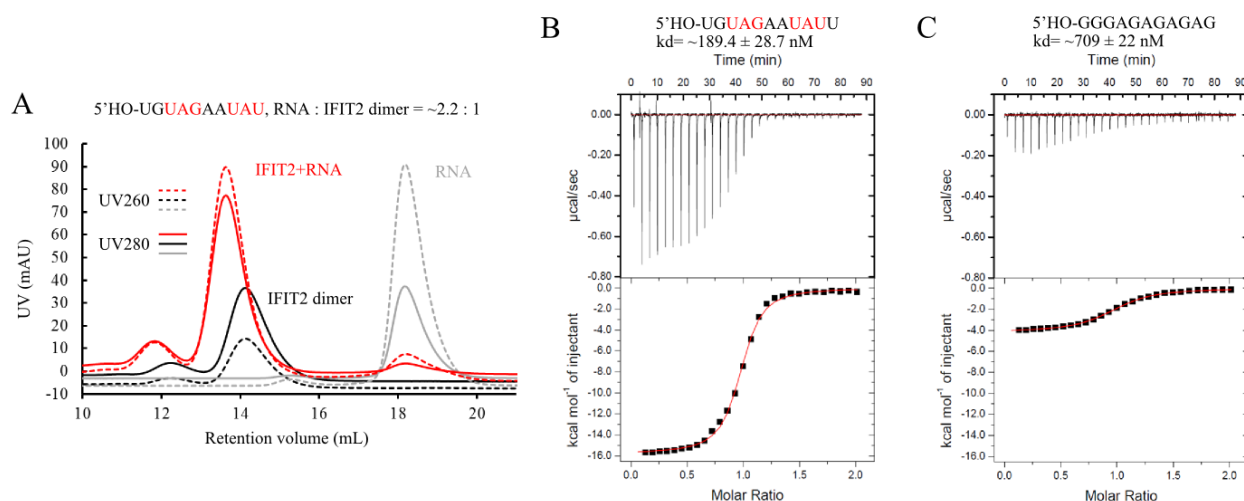


Figure 3.2 The qualitative and quantitative determination of the IFIT2-RNA interactions.

(A) Size-exclusion chromatography binding assay of IFIT2 with an AU-rich RNA; isothermal titration calorimetry measurement on Kd of IFIT2 with (B) an AU-rich RNA and (C) a non-AU-rich RNA.

3.2.4 The IFIT2-RNA crystal structure

The initial crystallization trials were performed with ss1-22 RNA, which was motivated by the observation of specific binding with IFIT2. Presumably due to weaker affinity than AU-rich RNA and RNA flexibility inside the IFIT2 binding tunnel, the electron density for the non-AU rich ss1-22 RNA was very discontinuous. Following the study by Tran *et al.*⁹³, we carried out co-crystallization of IFIT2 with an AU-rich RNA 5'-HO-UGUAGAAUAU based on the consensus sequence motif. After laborious screening of many ice-like fragile IFIT2-RNA crystals, we found one crystal diffracting at 3.2 Å that enables us to see a continual electron density for up to 8 nucleotides of the RNA (**Supplementary Figure 3.1**). In agreement with the stoichiometry observed on size-exclusion chromatography binding assay, each monomer of IFIT2 dimer bound to one copy of RNA (**Supplementary Figure 3.2A**). Overall, the bound RNAs make the C-termini

of IFIT2 less compact when superposing the RNA-bound IFIT2 onto the *apo* IFIT2 structure (**Supplementary Figure 3.2B**).

To address the RNA register ambiguity (5' end or 3' end entry into binding tunnel) of model building due to the relatively low resolution, we performed another co-crystallization screening of IFIT2 with 5'HO 5-Bromo-deoxyuridine AU rich RNA Br-dU-UAGUAUAUU (Integrated DNA Technologies, Inc.) using bromine anomalous signal as a marker to locate the 5' end of bound RNA in IFIT2. Similar to IFIT1 and IFIT5, the 5' end acted like a "head" entering the protein binding site followed by the 3' end "tail" (**Supplementary Figure 3.3**). The non-Bromo RNA bound IFIT2 crystal contained 1 copy of IFIT2 dimer in the asymmetric unit (ASU) and its symmetry belonged to the same space group, P212121, as the *apo* IFIT2 crystal structure. The best quality Bromo-RNA bound IFIT2 crystals were obtained under a different condition from that of the non-Bromo RNA bound IFIT2, resulting in 2 copies of IFIT2 dimers in each ASU and a lower symmetry corresponding to the space group of P1211 (**Figure 3.3**). The resolution of Bromo-RNA bound IFIT2 crystal structure was improved to ~2.7 Å as seen by a more continual and resolved RNA electron density map (**Supplementary Figure 3.3-3.4**). The molecular interactions between IFIT2 and RNA are very alike when comparing the two crystal structures. Therefore, the following structure analysis is carried out based on RNA chain E with IFIT2 chain A of 5'HO 5-Bromo-deoxyuridine AU rich RNA bound IFIT2 crystal structure.

Starting from 3' end of the bound RNA, the two uracil bases (9U and 8U) stacked on top of Tyr 383 via π - π interactions. The phosphate backbone ahead of the adenine 7A appeared very plastic to form a U-turn such that the uridine 6U was stacked on top of 8U despite the two uridines being separated by the adenine 7A. Gln 384 formed an H-bond with the uridine 8U base, and Asn 379 formed an H-bond with the phosphate backbone. Arg 376 and Arg 292 stabilized the phosphate

backbone through electrostatic interactions (**Figure 3.3**). Tyr 383, Gln 384, Asn 379, Arg 376, and Arg 292 are located at α -helix 19 of the C-terminal (**Supplementary Figure 3.2B**). On the other side of α -helix 19, the phosphate backbone was stabilized by Lys 417, 410, 414, and Arg 406 through electrostatic interactions, and the adenine 7A base formed H-bonding interaction with Glu 407 (**Figure 3.3**). Lys 417, 410, 414, Arg 406, and Glu 407 are located at α -helix 21 of the C-terminal (**Supplementary Figure 3.2B**). Moving further into the binding tunnel, the adenine 5A base formed π - π interactions with Phe 296 at helix 15, which served as a foundation for stacking with the uridine 4U and the guanine 3G. Two adjacent arginines 258 and 259 clipped the phosphate backbone through electrostatic interactions, acting like a molecular clipper. Arg 258-259 are located at helix 13. The uridine 1U near the 5' end stacked with Trp 188 (chain B) located at the loop connecting the swapped helix 9 and helix 10 (**Supplementary Figure 3.2B**). Lys 255 and Arg 251 from helix 13, Arg 41 from helix 2, as well as Arg184 (chain B) from the swapped helix 9 stabilized the phosphate backbone through electrostatic interactions (**Figure 3.3** & **Supplementary Figure 3.2B**).

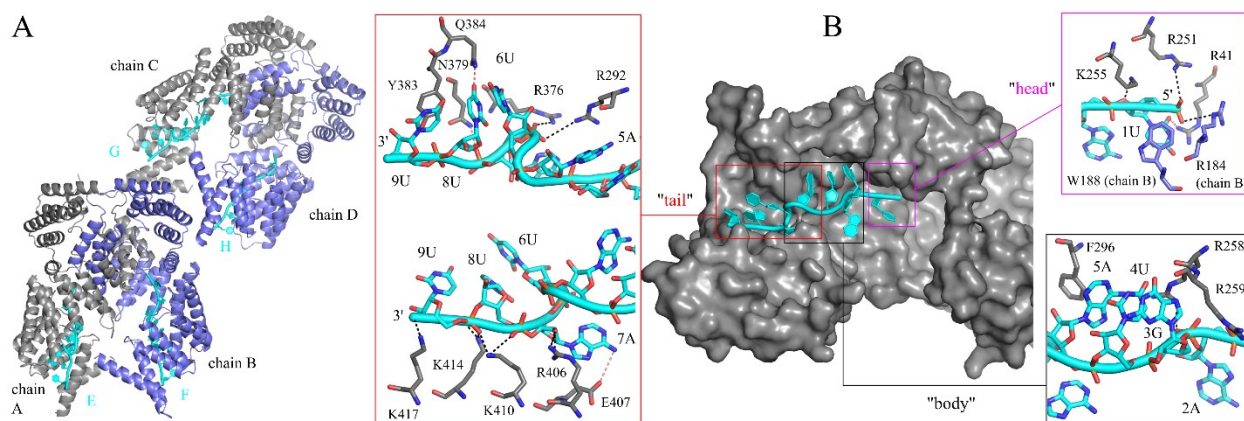


Figure 3.3 The crystal structure of RNA-bound IFIT2. (A) Two copies of IFIT2 dimers in the crystal asymmetric unit, with proteins labelled in chain A, B, C, D and RNA in chain E, F, G, H. (B) The molecular interactions between IFIT2 and RNA displayed in three regions based on

RNA: the 3' end "tail", "body", and the 5' end "head". Dashed lines indicate electrostatic (black) or H-bonding interactions (yellow). For clarity, only the protein chain A is shown in surface mode, protein residues involved in RNA binding are shown in stick, with chain A in gray and chain B in blue.

3.2.5 IFIT2 shows distinguishable RNA binding patterns from IFIT1 and IFIT5

Previous studies have revealed the sequestration binding mode of IFIT1 mainly targeting Cap0-ssRNA⁴⁹ and of IFIT5 mainly targeting 5'PPP-ssRNA⁴⁶. In order to examine the effect of RNA 5' end on IFITs binding, we performed isothermal titration calorimetry (ITC) measurements to quantify three RNA-binding IFITs, IFIT1, IFIT5 and IFIT2, with 11-nt RNAs, namely: Cap0-GGUAGAAUAUU, 5'PPP- GGUAGAAUAUU, 5'P-UGUAGAAUAUU, and 5'HO-UGUAGAAUAUU. As expected, IFIT1 showed the highest affinity to Cap0-ssRNA and the lowest affinity to 5'HO-ssRNA. Very interestingly, IFIT5 also showed the same trend as IFIT1: the binding affinity with Cap0-ssRNA was even higher than 5'PPP-ssRNA which was thought to be the main target of IFIT5. Surprisingly, the 5'HO-ssRNA still retained very tight binding to IFIT5. In line with a previous study on the broad and adaptable RNA binding properties of IFIT5, their affinity measurements based on gel binding assay also showed comparable nanomolar affinity tight binding for IFIT5 with 5'P-, 5'PPP-, and Cap0-RNA¹¹³. The crystal structure of 5'PPP-RNA bound IFIT5 likely revealed only one type of varied RNA targets. Future work is required to understand the adaptable RNA binding function of IFIT5, especially since human IFIT5 is neither stimulated by interferon treatment nor upregulated in common viral infections, including Influenza and corona viruses⁸⁷. IFIT2, on the other hand, showed an average binding affinity of ~200-300 nM to AU-rich RNAs (**Figure 3.4**) The higher than average affinity observed

for IFIT2 toward Cap0-GGUAGAAUAU might result from additional interaction with the cap due to its proximity to AU-rich sequences.

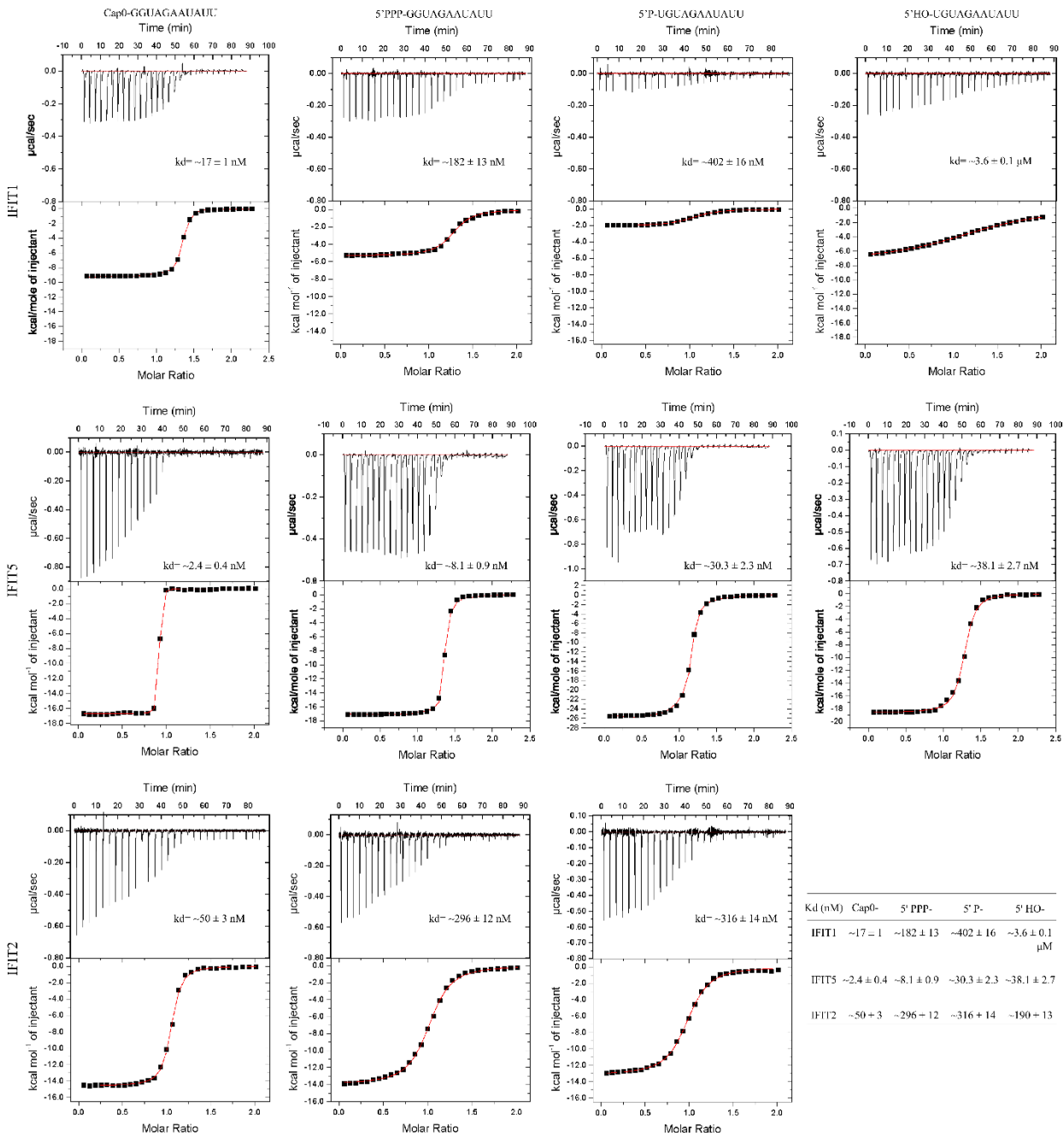


Figure 3.4 The effect of RNA 5’end on IFITs-RNA binding.Isothermal titration calorimetry binding affinity measurements of RNAs with varied 5’ ends (Cap0-, PPP-, P-, HO-) interacting

with IFIT1 (top row), IFIT5 (middle row) and IFIT2 (bottom row). K_d for IFIT2 with HO-RNA is taken from Figure 3.2B, thus, its ITC curve is not repeated here. Furthermore, we tested the interactions of IFIT2 with two 20-nt 5'HO-ssRNAs containing AU-rich and non-AU rich segments in reverse order: UA10-GA10 and GA10-UA10. Using gel binding assay, we observed that UA10-GA10 was completely shifted by IFIT2 at the concentration above K_d but GA10-UA10 showed weaker binding than UA10-GA10 to IFIT2 (**Figure 3.5A**), suggesting IFIT2 binds to short linear RNA in a 5' to 3' direction. The affinity difference can be seen more clearly when the domain-swapped IFIT2/3 heterodimer was used to test binding with the two RNAs. Despite IFIT3 does not bind to RNA, the IFIT2/3 heterodimer showed comparable binding affinity with IFIT2 homodimer for UA10-GA10 but weaker binding affinity than IFIT2 homodimer for GA10-UA10 (**Figure 3.5A**). UA10-GA10 showed higher binding affinity to IFIT2 due to the AU-rich sequences present at 5' end and GA10-UA10 showed weaker binding affinity to IFIT2 due to the non-AU rich sequences present at 5' end (**Figure 3.5B**).

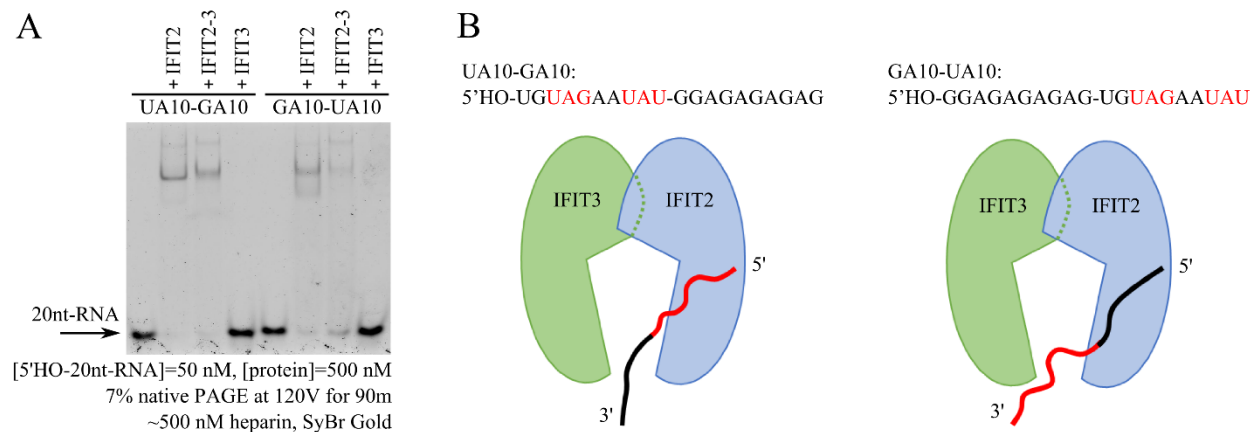


Figure 3.5 IFIT2 sequesters 5' end of short RNA. (A) The gel shift assay of IFIT2, IFIT2/3 and IFIT3 with two 20-nt ssRNAs containing AU-rich and non-AU rich segments in reverse order; (B) cartoon illustration of the sequestration binding mode of IFIT2 with 5' end of short RNA (AU-rich sequences in red and non-AU rich sequences in black).

IFIT2 shares ~46% sequence identity with IFIT1 and IFIT5, the majority of RNA binding residues are conserved amongst IFIT1, IFIT5 and IFIT2. Several key binding residues are unique to IFIT2, including Arginine 41/376/406 and Lysine 414 (**Supplementary Figure 3.6**). Although IFIT3 does not bind to RNA, it plays an important role in enhancing host antiviral responses. Knockdown of IFIT3 significantly increased virion titer, conversely, exogenous expression of IFIT3 effectively suppressed the infection of Sendai virus, Vesicular stomatitis virus, and Newcastle disease virus in HEK293 cells¹¹⁴. IFIT2 and IFIT3 both form domain-swapped dimers in solution, and they can also form domain-swapped heterodimers under mild denaturing condition or through co-expression. Furthermore, through IFIT1-IFIT3 interactions⁵⁷, IFIT1 and IFIT2/3 heterodimer can form stable trimeric IFIT1/2/3 complex in solution. In addition to the formation of the IFIT complex, IFIT3 was known to act like a scaffold protein for bridging interactions with other proteins to form protein complexes. For example, IFIT3 bridged the interaction between mitochondrial antiviral signaling (MAVS) complex and tumor necrosis factor receptor (TNFR)-associated factor family member-associated NF- κ B activator-binding kinase 1 (TBK1) to modulate the MAVS antiviral response¹¹⁴. Moreover, IFIT3 is associated with STAT1 and STAT2 to promote the heterodimerization of STAT1-STAT2 for strengthening the IFN signaling pathway¹¹⁵.

IFIT3 has ~57% sequence identity as IFIT2; interestingly, about half of the key RNA binding residues in IFIT2 were conserved in IFIT3. However, Arginine 406 and Lysine 410/414/417 located at the C-terminal of the RNA binding tunnel in IFIT2 as well as Arginine 41/184 located at the N-terminal of the binding tunnel were not conserved in IFIT3 (**Supplementary Figure 3.6**), which may significantly contribute to the incapability of RNA binding for IFIT3. Another reason may lie in the folding of IFIT3, IFIT3 is eighteen amino acids longer than IFIT2 (**Supplementary**

Figure 3.6). The folding of IFIT3 and the extra segment may block the RNA binding by some conserved RNA binding residues and make RNA inaccessible for IFIT3. The clear reason behind IFIT3 not binding to RNA requires determination of its molecular structure.

3.2.6 Biological function roles of IFIT2 dimerization

Daffis *et al.* reported that Murine Ifit2 restrained the accumulation of positive- (genomic) and negative-strand viral RNA in cells infected by West Nile Virus (WNV) mutant strain lacking 2' O-methyltransferase activity¹¹⁶. They suggested that mouse Ifit2 inhibits infection at or before the synthesis of the intermediate negative-strand viral RNA. We wondered whether the reported IFIT2-induced inhibition comes from direct binding to WNV genomic (+) strand RNA. Therefore, we *in vitro* transcribed the 79-nt 5' untranslated region (UTR) of WNV (+) strand RNA and tested its binding with IFIT2 on gel binding assay: IFIT2 specifically bound to 5' UTR of WNV (+) strand RNA. Very interestingly, the binding to WNV 5'UTR required the involvement of both monomers of the domain-swapped dimeric IFIT2 (**Figure 3.6A**). After observing specific interactions on gel binding assay in the presence of heparin, in order to quantify the affinity toward WNV (+) strand 5' UTR, we titrated IFIT2 binding to 5' PPP- and Cap1-WNV UTR without heparin: the binding affinities of IFIT2 to 5' PPP- and Cap1-WNV UTR were similar ~40-60 nM (**Figure 3.6B**). Unlike the sequestration mode of IFIT2 binding to the short ssRNA, each monomer bound to one copy of RNA independently; even the heterodimeric IFIT2/3 was still able to bind to one copy of ssRNA (**Figure 3.5A**). Only IFIT2 homodimer was able to interact with WNV 5' UTR but not for the IFIT2/3 heterodimer. According to the secondary structure prediction by M-fold server, the 5' UTR of WNV (+) strand RNA started with a short 3-nt overhang at 5' end, continued with two adjacent stem-loops, and ended with a 5-nt overhang at 3' end (**Figure 3.6C**). The WNV (+) 5' UTR unraveled another binding mode of IFIT2 utilizing both monomers.

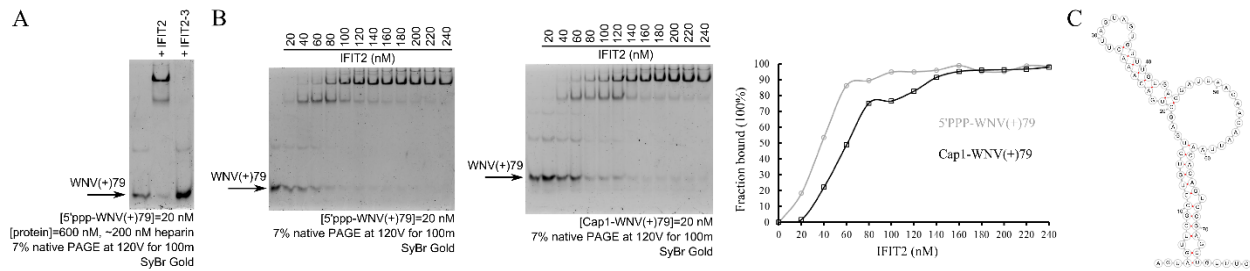


Figure 3.6 The interaction of IFIT2 with WNV 5' UTR. (A) The gel shift binding assay of homodimeric IFIT2 and heterodimeric IFIT2/3 with 79-nt WNV 5' UTR; (B) the titration of IFIT2 binding to PPP- and Cap1-WNV 5' UTR as well as the binding curve based on titration for the determination of K_d; (C) secondary structure of WNV 5' UTR predicted by the *M-fold* server¹¹².

To reveal the binding mechanism of IFIT2 with WNV 5' UTR, we performed cryo-EM studies on Cap1-WNV 5' UTR bound IFIT2. The ~4 Å cryo-EM map revealed that both monomers were involved in the binding, proving what we observed on the gel binding assay. Stunningly, the cryo-EM model unveils a remarkable binding pattern of IFIT2 acting on mRNA for the first time. IFIT2 dimer binds to the mRNA chain in a sophisticated manner to clamp on the chain. One monomer of IFIT2 dimer "senses" an AU-rich element on the mRNA chain, the monomer initiates a "grab" at the AU-rich site, "drags" the chain into the binding pocket as shown in the crystal structure, the chain gets bent downwards and turns toward the other monomer. The second monomer anchors the chain and guides the RNA chain to bend in order to be accommodated in the binding pocket of the second monomer. The significant conformation adjustment of the mRNA chain fully enables the mRNA segment to be accommodated in the RNA binding pockets of both monomers, by doing so, ensuring a certain affinity required for translation enhancement with other unknown cellular factors. The RNA electron density map in one monomer is slightly better than the other, likely due to the enhanced binding affinity from AU-rich sequences. Along the linker between the two

monomers, the RNA density map appears discontinuous, reflecting the flexibility of the RNA chain in this range where there is no interaction with protein residues (**Figure 3.7A**).

The higher-resolution crystal structure provides guidance for the cryo-EM model building. Although the resolution could not reach the level of accurately modeling individual nucleotides (below 2 Å), it is still feasible to map detailed interactions between the IFIT2 dimer and the 37-nt (29U to 65A) WNV 5' UTR segment. Starting from the initial "grabbing" site, Tyr 383 formed π - π stacking interactions with the uridine bases of 29U and 30U. Arg 376 and Arg 292 stabilized the phosphate backbone through electrostatic interactions. Arg 184 (chain B) and Arg 251 electrostatically stabilized the phosphate backbone in addition to the "molecular clipper" acted by Arg 258-259. Trp 188 (chain B) formed stacking interactions with 38U. Lys 221, Lys 406, and Lys 414 electrostatically stabilized the phosphate backbone. Then, the RNA chain extended toward the other monomer. Along the linker, RNA was stacked with its bases, without any supporting stabilization from IFIT2. The other monomer anchored the phosphate backbone by using Lys 410 and Arg 406. Moving upward, the "molecular clipper" Arg 258-259 and Lys 255 electrostatically stabilized the phosphate backbone. Then the RNA chain was bent downward via electrostatic interactions with Arg 41 and π - π stacking interactions between 56A and Trp 188 (chain A). Toward the end, Arg 302 electrostatically stabilized 63A which was stacked on Tyr 383. The IFIT2 dimer completed binding to the mRNA chain through the cooperation of both monomers (**Figure 3.7B**).

Compared with the crystal structure, the 5' UTR of WNV mRNA appeared dramatical conformation adjustment to be bound by the IFIT2 dimer: the 5' cap and 3' tail are beyond the IFIT2 binding site. Consequently, both 5' and 3' ends are not visible because they freely move beyond the "control" of IFIT2. The cryo-EM model likely represents the real picture of how IFIT2 binds to mRNA in cells. The crystal structure rather provides a partial picture of the RNA binding

mechanism of IFIT2, owing to the short RNA used for crystallization trials; the high flexibility of the long RNA, e.g., WNV 5' UTR, prevented the formation of IFIT2-RNA crystals. Nevertheless, many residues involved in RNA binding in the crystal structure are also responsible in the cryo-EM model. The monomers act collaboratively to clamp on the RNA chain. Therefore, we named it as “clamp” mode to differ from the “sequester” mode seen in IFIT5 and IFIT1.

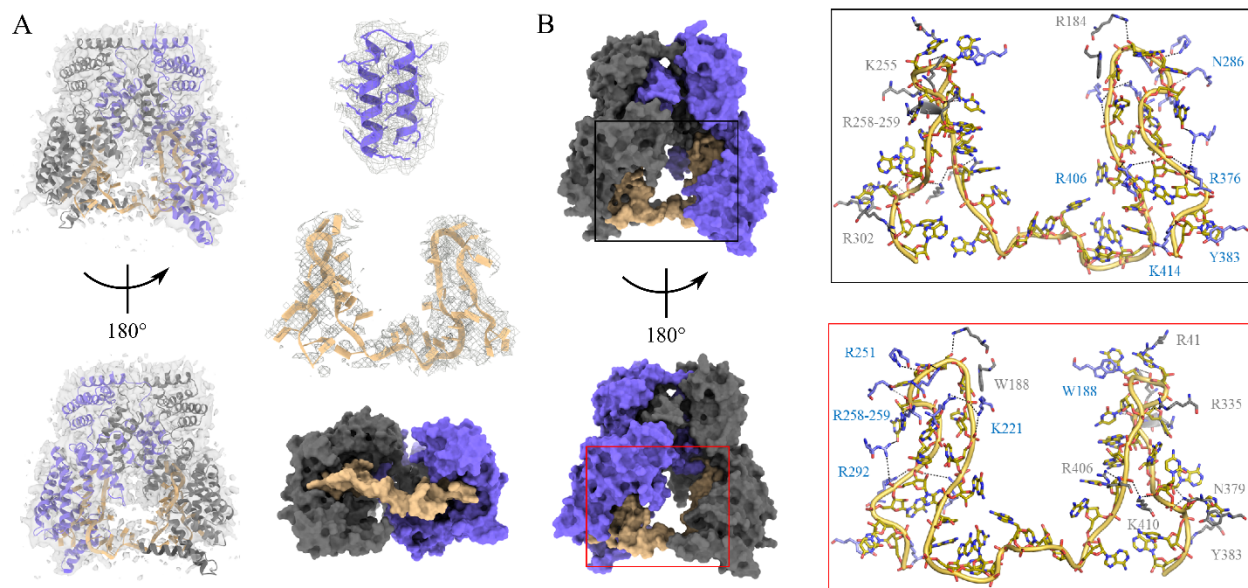


Figure 3.7 The cryo-EM model of an RNA bound by both monomers in IFIT2. (A) The ~ 4 Å cryoEM map with the model fit in the map with front and back views. The electron density maps of protein and bound RNA. A top view of the model showing both monomers bind to the RNA. The IFIT2 monomer of chain A is colored in blue, chain B in gray, and RNA is colored in yellow. (B) The molecular interactions between IFIT2 and RNA corresponding to the highlighted region of the protein-RNA complex in surface mode. Dashed lines indicate electrostatic (black) or H-bonding interactions (gray).

IFIT1 forms elongated dimers in solution through antiparallel C-termini protein-protein interactions¹¹⁷. Each monomer acts independently to interact with RNA based on the stoichiometry

IFIT1 monomer to RNA ratio of 1:1 in ITC (**Figure 3.4**). IFIT5 exists as monomer form in solution. In contrast, IFIT2 and 3 form domain-swapped dimers in solution and only IFIT2 binds to RNA. The biological function of the domain-swapped dimerization architecture in IFIT2 is clearly seen in the cryo-EM model of IFIT2 dimer binding to a viral mRNA 5' UTR, both monomers in IFIT2 cooperate for clamping on the RNA chain, which cannot be accomplished by either monomer alone (**Figure 3.6A, 3.7**).

3.2.7 Structure guided mutation analysis and translation reporter assay

Based on the IFIT2-RNA crystal structure, we generated six mutants: mutant 1, 41E/R184E/R251E; mutant 2, K255E/R258E/R259E; mutant 3, R292E/R376E/R406E; mutant 4, K410E/K414E/K417E; mutant 5, Y383V/Q384A; and mutant 6, Loop45-54 replacement (replace IFIT2 loop 45-54 by IFIT1 loop 46-55). We first tested the wild-type (WT) IFIT2 and mutants in the “go-through” mode using 20-nt 5'HO-RNAs on gel binding assay and observed the following: mutant 2-4 impaired RNA binding, mutant 5 weakened RNA binding, and mutant 1 and 6 showed no effect on RNA binding (**Supplementary Figure 3.7**). We then tested the WT IFIT2 and mutants in the “clip” mode using the 79-nt 5'PPP- and Cap1-WNV 5' UTR on gel binding assay and observed the following: mutant 2-4, mutant 5, and mutant 6 appeared to have similar effects to the “go-through” mode; surprisingly, mutant 1 had no effect on binding to the 20-nt 5'HO-RNAs but impaired binding to the 79-nt 5'PPP- and Cap1-WNV 5' UTR RNAs (**Supplementary Figure 3.8**). Structure alignment of RNA bound IFIT1, IFIT5 and IFIT2 showed that the loop region (aa 46-55 in IFIT1, aa 45-54 in IFIT5 and IFIT2) of IFIT1 and IFIT5 kept a similar position but it shifted away substantially in IFIT2. Despite replacing the loop region in IFIT2 by IFIT1, the loop45-54 mutant showed no effect on RNA binding in both binding modes, suggesting the displacement of this loop region in IFIT2 was trivial for RNA binding.

Whether the IFIT2-induced translation enhancement results from RNA binding or not remains questionable. In order to verify the structure models and confirm whether the translation stimulation effect comes from RNA binding, we performed a similar reporter assay as in the study by Tran *et al.*⁹³. Briefly, we co-transfected IFIT2 with two reporter genes in 293T cells, respectively. One reporter gene contains the WT IAV nucleoprotein (NP) sequence with AU-rich regions, the other reporter gene contains juggled IAV NP (JUG) sequence in which AU-rich regions are removed without changing the NP protein sequence. Then we compared the translation level of the two genes influenced by IFIT2.

We first titrated the IFIT2 concentration to find out the lowest IFIT2 transfection concentration that gives the highest difference in translation level between the control and IFIT2 transfected cells. We chose 100 ng of IFIT2 as the optimal transfection amount (**Supplementary Figure 3.9A**). Then we tested WT IFIT2 and the six mutants above. However, none of the mutants showed significant reduction of translation level compared with WT IFIT2 (**Supplementary Figure 3.9B, C**), even the effective mutants that abolish IFIT2-RNA binding *in vitro*. The RNA binding involves with ~16 residues of IFIT2. The mRNAs in cells are neither short nor linear, similar to the WNV(+)₇₉ RNA. Therefore, IFIT2 mostly likely adopts the “clip” mode to interact with cellular mRNAs, the binding involves both monomers, resulting to ~32 RNA binding residues. Thus, double or triple mutations out of a large number of RNA binding residues likely have negligible effect on RNA binding in cells. We designed two long mutants: one segment with 9 amino acids, namely, 9aa mutant: R41E, R184E, W188V, R251E, K255E, R258E, R259E, R292E, F296A; the other segment with 7 amino acids, namely, 7aa mutant: R376E, Y383V, Q384A, R406E, K410E, K414E, K417E. The WT IFIT2 showed approximately two-fold higher translation level than either

9aa or 7aa mutant (**Figure 3.8A-B**), indicating that RNA binding of IFIT2 is important for the translation enhancement in cells.

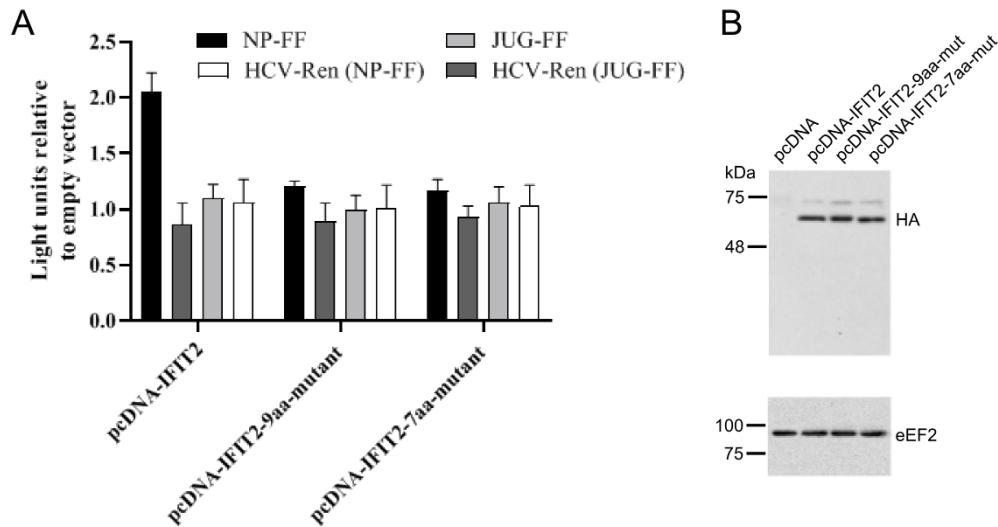


Figure 3.8 A firefly luciferase reporter assay showing the effect of WT IFIT2 and mutants on translation stimulation. IFIT2 9aa or 7aa mutants fail to stimulate expression of a reporter mRNA bearing the NP sequence transfected into cells. (A) Relative translation of FireFly mRNA bearing either NP or JUG sequence in 293T cells transfected with 100 ng of IFIT2 vector or the two mutants. Values are set relative to empty vector $\pm$ standard deviation. (N=6). HCV-Ren mRNA was used as transfection control. (B) Expression levels of IFIT2 and mutants in transfected 293T cells. Eukaryotic translation elongation factor 2 (eEF2) is used as loading control.

3.3 Discussion

IFIT2 has been known to inhibit viral infection; however, the mechanism for IFIT2-exerted viral inhibition is unclear. Because the structure motif tetratricopeptide repeats are well known for mediating protein-protein interactions⁴⁵, early studies on the antiviral function of IFIT1 and IFIT2 proposed that IFIT1 bound to one subunit of the eukaryotic translation initiation factor 3 (eIF3) to

block the formation of stable eIF3-ternary complex and IFIT2 bound to two subunits of eIF3 to destabilize the translation ternary complex and to hinder the formation of 48S pre-initiation complex⁴⁷. Therefore, IFIT1 and IFIT2 restrained viral replication through the inhibition on viral protein synthesis. Later, Pichlmair et al. discovered that IFITs, conventionally thought to bind protein partners, directly engaged viral RNA⁴⁸. In the subsequent studies, the crystal structures of RNA-bound IFIT5 and IFIT1 were determined, revealing how IFITs utilized TPR repeats to form positively charged binding pocket for sequestering viral RNA⁴⁶. Especially for IFIT1, the biological function on Cap0-RNA binding became clearer: to compete with eIF4E for binding to 2'O-nonmethylated viral mRNA and therefore to shut down viral protein production¹¹⁸. With ~46% sequence identity as two other RNA binding family members, IFIT2, not surprisingly, might exert viral defense actions similarly through direct interaction with viral RNA. However, the RNA binding ability of IFIT2 has remained questionable over the past decade due to the lack of RNA bound IFIT2 structure.

In this study, we determined two x-ray crystal structures and one cryo-EM structure to reveal the mechanism of IFIT2 interacting with different RNAs. IFIT2 seems able to adopt various modes of interacting with different RNAs based on 5' end, sequences and consequent folding (secondary structures). The binding tunnel of IFIT2 can only accommodate ssRNA, preferentially AU-rich sequences. For short linear RNA, each monomer of IFIT2 acts as an independent machinery to sequester the 5' end of the RNA chain. For RNA with complex secondary structures, such as stem-loops, each monomer of IFIT2 is no longer an independent binder but rather acts collectively to clamp the accessible RNA region using both monomers. The two binding modes complement so that IFIT2 acts on a variety of mRNA transcripts, including both host and viral mRNAs as reported by Tran *et al.*⁹³

IFIT1 and IFIT5 inhibit the translation of bound transcripts by sequestering RNA 5' end. IFIT2, however, clips along the mRNA chain and enhances the translation efficiency of bound transcripts. The exact mechanism of the enhanced translation efficiency resulting from reduced ribosome pausing induced by IFIT2 is undefined. Our data using an *in vitro* translation system showed that IFIT2 alone cannot enhance translation efficiency (**Supplementary Figure 2.9D**), suggesting IFIT2 likely cooperates with other factors in cells to enhance translation efficiency of bound transcripts. Prior studies in mice consistently showed ifit2-induced inhibition on viral replication⁵⁰⁻⁵⁴. Influenza virus was the first one that co-opted IFIT2 in favor of virus survival⁹³. At the time of writing this manuscript, a new study reported that IFIT2 inhibited influenza A viral RNA synthesis in HEK293 cells¹¹⁹, conflicting with the pro-viral role of IFIT2 found by Tran et al.⁹³ The contradictory results might be due to different methods used for altering gene expression and viral infection assays¹¹⁹. The inhibition was observed by overexpression of IFIT1 and IFIT2 for 48 h followed by infection for another 48 h¹¹⁹, clearly, viral RNA inhibition occurs at IFITs abundant condition. On the other hand, the viral mRNA enhancement was observed by comparing WT and CRISPR-knockout cells upon infection⁹³. The viral RNA enhancement was observed in normal cells versus IFITs knockout cells, so there is a possibility that viruses replicate not as well in the CRISPR treated “defective” cells as in normal cells. At the early stage of infection, many ISGs are induced rapidly and host antiviral effector mRNAs are more abundant than viral mRNA in cytoplasm. IFIT2 perhaps mainly binds to self mRNAs and enhances the translation of antiviral effectors. Consequently, viruses cannot be sustained, the antiviral state is lifted up, viral proteins and mRNAs are cleared out from infected cells, and life inside infected cells resumes normal state. However, if the defense battle is lost at the starting point, the IFN signaling pathway is hijacked by invading viruses due to comprised immune system or other pre-existing health conditions. Viral

mRNAs are more abundant than self mRNAs, and IFIT2 perhaps predominantly binds to viral mRNA and enhances the translation of viral proteins. Devastatingly, self-defense effectors cannot be sustained, the dysfunctional antiviral state maintained, viral proteins and mRNAs occupy infected cells, numerous new viral particles are assembled and released to invade surrounding cells, and the infection continues to be more severe. This is followed by the collapse of innate immune defense, and depending on the fight by adaptive immune system, the infection can be either conquered or intensified to become life threatening.

IFIT2 binds to a wide range of RNAs with varied affinities, AU-rich sequences are present both in host and viral mRNAs; thus, IFIT2 binds to RNA in a sequence non-specific manner, i.e., IFIT2 cannot selectively bind to host mRNA over viral mRNA, and vice versa. Therefore, depending on the enriched transcripts in cytoplasm upon viral infection, IFIT2, without betraying its designated role, can enhance either self or viral mRNA translation by binding to surrounding available transcripts.

3.4 Materials and Methods

3.4.1 IFIT2 cloning, expression and purification

The DNA sequence of human IFIT2 (UniProtKB accession number: P09913) was cloned into a pSMT3 vector (pET-28a backbone) between restriction enzyme BamHI and NotI sites. Fusion proteins carry an N-terminal His6 small ubiquitin-like modifier protein (SUMO) tag which is cleavable by Ubiquitin-like-specific protease 1 (Ulp1). Recombinant IFIT2 was expressed using BL21 (DE3) cells in Luria Broth culturing media. Cells were grown at 30 °C initially until the optical density reached ~0.7, then the temperature was reduced to 18 °C. Cells were induced with 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG), further grown overnight, and harvested next

morning. The cell pellet was resuspended in Ni binding buffer (20 mM Tris pH 7.8 at 0 °C, 500 mM NaCl, 10 mM imidazole, 5-10% glycerol (volume by volume, v/v)), supplemented with 2 mM 2-Mercaptoethanol (β ME), 1 mM phenylmethylsulfonyl fluoride (PMSF), and 0.1% Triton X-100 (v/v). The resuspended cells were lysed by using a French homogenizer, followed by centrifugation at 50,000 g for 30 min. The supernatant was loaded via AKTA purifier sample pump on a self-packed Ni-NTA column (~2.5 ml bed volume clean beads per 1 L culture) equilibrated in Ni binding buffer. After loading the supernatant, the Ni-NTA column was washed with Ni washing buffer (20 mM Tris pH 7.8 at 0 °C, 1 M NaCl, 10 mM imidazole, 5-10% glycerol, 2 mM β ME) until the UV absorbance was stabilized. For gradient elution, the AKTA system pump A and B were equilibrated in Ni binding buffer and Ni elution buffer (20 mM Tris pH 7.8 at 0 °C, 500 mM NaCl, 300 mM imidazole, 5-10% glycerol, 2 mM β ME), respectively. Further washing was performed with 10% Ni elution buffer to remove loosely bound contaminants until the UV absorbance was stabilized. The His6-SUMO-tagged IFIT2 was eluted off the Ni-NTA column by applying a linear gradient of Ni elution buffer. The first round Ni eluted tagged protein was cleaved by adding home-purified SUMO protease Ulp during dialysis at 4 °C overnight in dialysis buffer (20 mM Tris pH 7.8 at 0 °C, 100 mM NaCl, 2 mM β ME). The next day, the dialyzed sample was reloaded to the Ni-NTA column cleaned with Ni elution buffer and Ni washing buffer and equilibrated with dialysis buffer. The tag-cleaved IFIT2 protein flew through the column, and the collected second round Ni flowthrough sample was centrifuged to remove any precipitation before passing to an anion-exchange HiTrap Q HP column (GE Healthcare). The 5 ml Q column was washed with Q-B buffer (20 mM Tris pH 7.8 at 0 °C, 1 M NaCl, 1 mM DTT) and then equilibrated with Q-A buffer (20 mM Tris pH 7.8 at 0 °C, 1 mM DTT). The IFIT2 protein was gradually eluted during a salt gradient ranging from ~15%-45% of Q-B buffer. The pool of Q column elution with

protein of interest was concentrated and further polished by a Superdex 200 10/300 GL column (GE Healthcare) using gel filtration (GF) chromatography in GF buffer (20 mM Tris pH 7.8 at 0 °C, 150 mM NaCl, 1 mM DTT). The pool of final polished protein was concentrated and stored at –80 °C.

3.4.2 IFIT2 mutagenesis

Primers for IFIT2 mutants were designed based on this method¹¹⁰, and mutation sites were verified by sequencing. All the mutants were purified following the same procedures as WT IFIT2. For cell-based assay, the pCDNA3-NP-2A-GFP plasmid (a gift from Dr. Andrew Mehle) was double digested with KpnI and XhoI, the IFIT2 sequence with an N-terminal HA tag flanked by the two cloning sites above was amplified by PCR reaction using the bacterial IFIT2 plasmid as template, the HA-IFIT2 insert was subcloned into the digested pCDNA plasmid, and the pCDNA3-HA-IFIT2 plasmid was verified by sequencing.

3.4.3 IFIT1, IFIT3, IFIT5, and IFIT2/3 cloning, expression and purification

The DNA sequence of human IFIT1 (UniProtKB accession number: P09914), IFIT3 (UniProtKB accession number: O14879), and IFIT5 (UniProtKB accession number: Q13325) was cloned into a pSMT3 vector (pET-28a backbone) between restriction enzyme BamHI and NotI sites, respectively. Fusion proteins carry an N-terminal His6-SUMO tag which is cleavable by Ulp1 protease.

The recombinant IFIT1 purification procedure was the same as that of IFIT2 at Ni-NTA step. Afterward, the 2Ni flowthrough sample was passed to a cation-exchange HiTrap SP HP column (GE Healthcare) using the same ion-exchange buffer as Q column. The IFIT1 protein was

gradually eluted during a salt gradient ranging from ~15%-35% of buffer B. Recombinant IFIT3 purification was the same as IFIT2 purification except a salt gradient ranging from ~15%-30% of Q-B buffer was used in the Q column step. Recombinant IFIT5 purification was the same as IFIT2 purification except for a salt gradient ranging from ~10%-40% of Q-B buffer was used in the Q column step. Both IFIT5 and IFIT1 were further polished by size-exclusion chromatography.

For the subcloning of IFIT2/3, the DNA sequence of human IFIT2 (UniProtKB accession number: P09913) was cloned into the multiple cloning site (MCS) 2 of a pCDFDuet-1 vector between restriction enzyme EcoRV and XhoI sites, with a N-terminal SUMO tag. The DNA sequence of human IFIT3 (UniProtKB accession number: O14879) was cloned into MCS1 of the same vector between restriction enzyme NcoI and NotI sites, with a N-terminal Hisx6-SUMO tag. The expression and purification procedures of IFIT2/3 were the same as those of IFIT2, except Heparin column was used as the intermediate step between Ni-NTA and size-exclusion chromatography.

3.4.4 In vitro transcription, RNA capping and 2'O-methylation, and

RNA mass spectrometry

For the 10-nt GGUAGAAUAU RNA, the *in vitro* transcription and purification were followed with the method developed in Chapter 2. The ss44 and WNV(+)⁷⁹ in vitro transcription reactions were set up in the same way as the 10-nt RNA, except a self-cleaving hepatitis delta virus ribozyme sequence¹²⁰ was added after the desired RNA sequence at the 3' end of the DNA template. The dsDNA template was generated by using the recursive-PCR method¹²¹. The RNA capping reaction was set up in the same way as described in Chapter 2. For the one-step 5'PPP to Cap1 reaction, mRNA Cap 2'-O-Methyltransferase (New England Biolabs) was added to the reaction in addition to the homemade vaccinia capping enzyme. The capping reaction was left at 37°C for ~2 hours,

afterward, 98% formamide denaturing loading buffer was added to the reaction, spun down, and loaded onto the 7M urea TBE polyacrylamide denaturing gel for RNA extraction. The RNA recovered from ethanol precipitation was resuspended in RNase-free water and stored at -20°C before use. The quality of RNA was verified by mass spectrometry with the protocol reported in Chapter 2.

3.4.5 Electrophoretic mobility shift assay

Purified protein was stored in protein storage buffer (20 mM Tris pH 7.6, 150 mM NaCl, 50% glycerol, 1 mM EDTA, 5 mM β ME, 0.5 mM PMSF). Protein and RNA were set up in a 10 μ l reaction, using 10x binding buffer (500 mM Tris pH 7.6, 1.25 M NaCl, 10 mM EDTA) and incubated at room temperature for ~10 min. For the binding specificity test, heparin sodium salt was included. The sample was loaded on a freshly casted native TBE gel and run the gel with 0.5x TBE buffer in the cold room. For quantification purpose, the gel bands were quantified by using BioRad Image Lab program.

3.4.6 Size exclusion chromatography binding assay

The superdex 200 10/300 GL column was equilibrated in GF buffer (20 mM Tris pH 7.8 at 0 °C, 150 mM NaCl, 1 mM DTT). The amount of ~0.5 mg (~4.5 nmol) purified IFIT2 dimer was injected into the column. For the second run, ~10 nmol of 11-nt RNA 5'HO-UGUAGAAUAU (Integrated DNA Technologies, Inc.) was injected into the same column. For the last run, ~0.5 mg purified IFIT2 dimer was incubated with ~10 nmol of the 11-nt RNA on ice for ~15 min and injected into the same column.

3.4.7 Crystallization of IFIT2 with RNA

The purified IFIT2 in GF buffer was incubated with molar excess 5'HO-UGUAGAAUAUU on ice for ~30 min before setting up crystallization. The crystals were obtained in 5-10 mg/ml protein drops mixed with an equal volume of reservoir solution containing 0.1 M Bis Tris propane pH 6.5, 200 mM NaNO₃, 20% (w/v) PEG3350 using the hanging-drop vapor diffusion method at 4 °C. The crystals were harvested and transferred into the cryoprotectant drop with 30% PEG200 and 70% reservoir solution, and then flash frozen in liquid nitrogen. The X-ray diffraction data was collected at the American Photon Source beamline 24-IDE and processed using DIALS¹²². The structure was solved using PHASER implemented in CCP4i2¹²³ with an ensemble of the *apo* IFIT2 dimer structures⁵⁵ as a search model and refined using Refmac¹²⁴ and Coot¹²⁵.

The purified IFIT2 was mixed with molar excess 5'HO-5-Br-dU-UAGUAUAUU and set up hanging drops at the condition of 0.1 M HEPES pH 7.2, 150 mM NaSCN, 22% (w/v) PEG3350 at 4 °C. The crystals were harvested and cryoprotected in the drop of 30% PEG400 and 70% reservoir solution, and flash frozen in liquid nitrogen. The X-ray diffraction data were collected at the Canadian Light Source beamline CMCF-BM. The downstream processing proceeded in the same way as above.

3.4.8 Isothermal titration calorimetry

The IFIT1 or IFIT2 protein of 20 μM (calculated in monomer form) in 50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM Tris (2-carboxyethyl) phosphine (TCEP) was loaded in the cell. RNA of 200 μM in the same buffer as protein was loaded in the syringe. The injection and mixing were run in a MicroCal iTC200 machine (GE Healthcare). The IFIT5 protein of 10 μM in the same buffer above was loaded in the cell and RNA of 100 μM in the same buffer as the protein was loaded in

the syringe. Measurements for IFIT1 with Cap0-RNA, IFIT5 with 5'PPP-, 5'P- and 5'HO-RNA were carried out in the VP-ITC MicroCalorimeter (MicroCal). The ITC data were acquired with a reference power of 10 μ cal/s and processed by fitting to a one-site binding model to determine n (number of binding sites) and K_d (dissociation constant) using the accompanied Origin 7.0 software.

3.4.9 In vitro transcription and translation for firefly luciferase reporter assay

The pcDNA3-NP-2A-FFLUC-PEST and pcDNA3-jugNP-2A-FFLUC-PEST vectors were linearized with ApaI and used for in vitro transcription using the Hiscribe T7 High Yield RNA Synthesis Kit from NEB following manufacturer's recommendations. RNA was then purified by phenol:chloroform extraction, followed by a G50 sephadex column purification and ethanol precipitation. RNA was then capped and methylated using vaccinia capping enzyme and mRNA Cap 2'-O-methyltransferase (New England Biolabs). mRNA was translated in Krebs-II extracts as described previously using 10 ng/ μ l in presence of increasing amounts of IFIT2¹²⁶. HCV-Ren mRNA was prepared as described¹²⁷ and used in translation at 10 ng/ μ l.

3.4.10 Cell culture

HEK293T cells were grown at 37°C in DMEM supplemented with 10% FBS and Glut/Pen/Strep. Prior to transfection, 1 million cells were seeded in each well of a six-well plate. The following day, cells were transfected with the indicated amounts of each construct using PEI¹²⁸ for a total of 2 μ g of DNA per well. pcDNA3 empty vector was supplemented to make up to 2 μ g DNA in the conditions where less than 2 μ g of IFIT2 expression vector was used. The next morning transfected cells were trypsinized, resuspended in 3 ml of DMEM, and 400 μ l of cells were re-plated into four

wells of a 24-well plate. Six hours later, 100 ng of either the NP-2A-FFLUC-PEST or jugNP-2A-FFluc-PEST mRNA reporters, along with 100 ng of HCV-Ren mRNA, was transfected in 200 μ l of OptiMEM using 1 μ l of DMRIE-C reagent according to the manufacturer's recommendations (Thermo Fisher Scientific). Cells were lysed 16 h later using Passive Lysis Buffer (Promega) and FF and Ren luciferase activity determined on a Fluostar 96-well plate reader (BMG Labtech) using homemade reagents¹²⁹.

3.4.11 Cryo-EM sample preparation, data collection and processing

The purified IFIT2 complex stored at -80°C was thawed on ice, spun down, and injected on to a Superdex 200 10/300 GL column (GE Healthcare) equilibrated in Tris gel filtration buffer (20 mM Tris pH 7.6 at 0°C , 150 mM NaCl, 1 mM DTT) for removal of aggregates. The peak fractions of IFIT2 protein were concentrated and incubated with Cap1-WNV(+)₇₉ at RNA to dimeric protein molar ratio of $\sim 2:1$. Tween20 was added to the sample at a final concentration of 0.005% for cryo-EM sample freezing. 3.5 μ l of ~ 2 mg/ml Cap1-WNV(+)₇₉ bound IFIT2 was applied to a glow-discharged C-flat 300 mesh 2/1 copper holey carbon grid. Then the grid was blotted at 4°C and 100% humidity using a Vitrobot IV (FEI) and plunge-frozen into liquid ethane. Microscopy data were collected at the McGill Facility for EM Research using an FEI Titan Krios TEM operating at 300 kV with a Gatan K3 direct electron detection camera and a Gatan GIF BioQuantum LS imaging filter. Movies were collected in counting mode using SerialEM, with a total dose of 60 $\text{e}/\text{\AA}^2$ over 40 frames and a set defocus range of -1 to -2 μm at a nominal magnification of 130,000, resulting in a pixel size of 0.675 $\text{\AA}$. Micrographs were motion corrected using cryoSPARC (v4.2)¹³⁰, followed by contrast transfer function (CTF) estimation. Micrographs with CTF estimated resolution beyond 5 $\text{\AA}$ were excluded. A manual picking was performed on a few selected micrographs, the picked particles were trained using Topaz¹³¹, the Topaz trained model

was used to pick particles for all the manually curated micrographs. The Topaz picked particles were subject to 2D classification and ab-initio reconstruction to remove junk particles. The genuine particles were used as the input to train Topaz for picking. Several iterations of Topaz picking and classifications were performed until no further improvement on the number of genuine particles. Once particle picking was done, multiple rounds of ab-initio reconstruction were performed to select a highly homogeneous set of particles. The final homogeneous particles were refined using homogeneous refinement followed by non-uniform refinement in cryoSPARC.

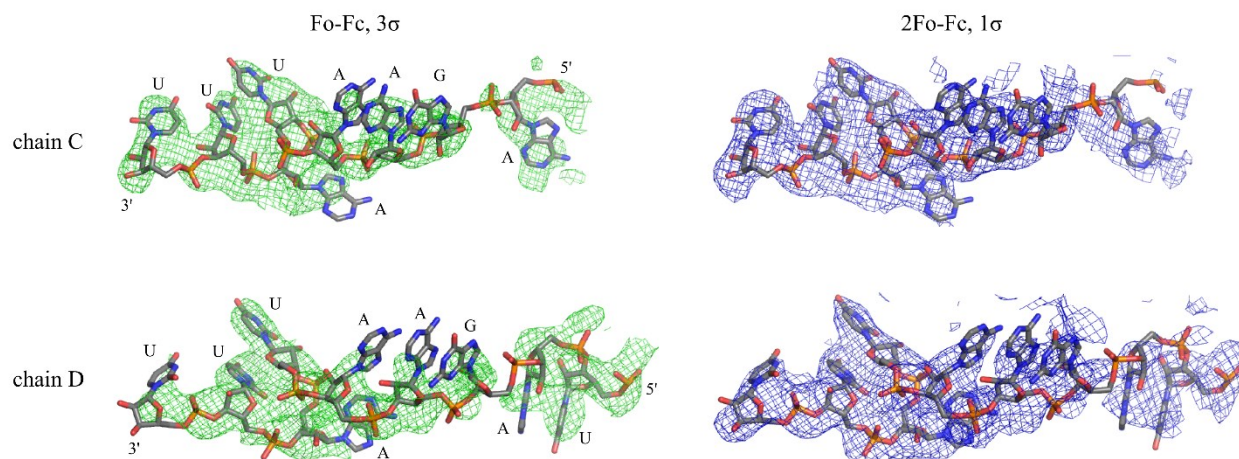
3.4.12 Cryo-EM Modeling and refinement

IFIT2 from the RNA bound crystal structure was fit into the non-uniform refined map by using ChimeraX¹³². The fitted IFIT2 model was further relaxed in both cartesian and torsion spaces by using Rosetta electron density scoring function^{133, 134}. The output relaxed model and the map were loaded in Coot¹²⁵ for RNA model building. Once the RNA model was completed and refined in Coot, the built RNA bound IFIT2 was input to Rosetta again for relaxation and refinement guided by the electron density map. The output relaxed model was input to Phenix¹³⁵ for real space refinement.

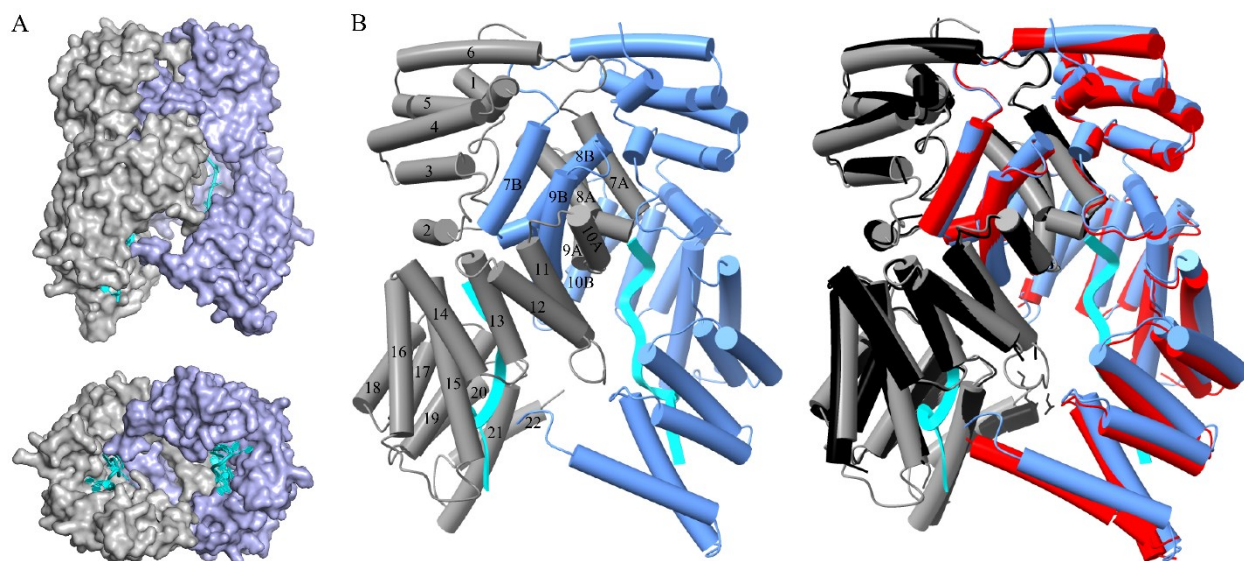
3.5 Acknowledgement

We thank the American Photon Sources and the Canadian Light Sources for X-ray diffraction experiments. We also thank the cryo-EM facility at McGill University for screening and data collection. We are grateful to Dr. Andrew Mehle (University of Wisconsin, Madison) for kindly providing the gift of Influenza nucleoprotein native and juggled pCDNA plasmids with GFP reporters or firefly luciferase, and Dr. Alexander S. Wahba for RNA LC-MS (McGill University).

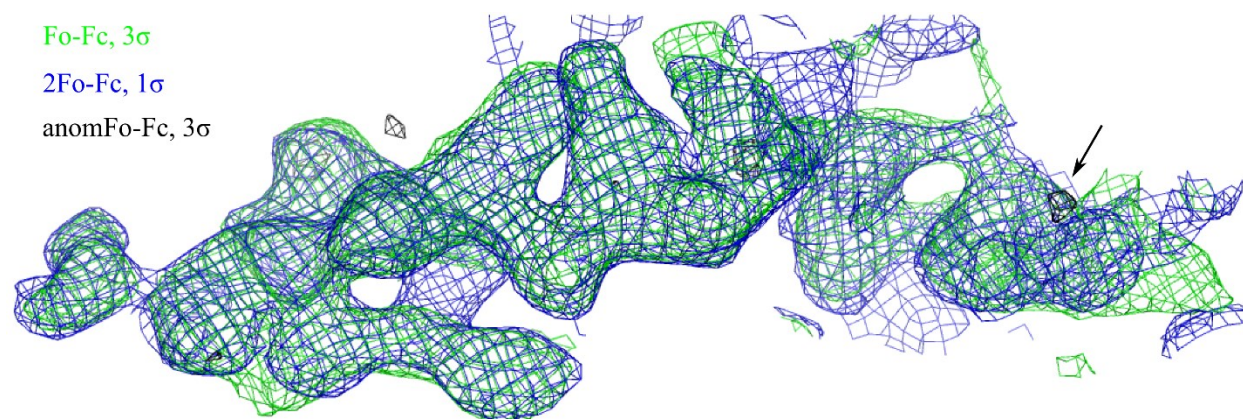
3.6 Supporting information



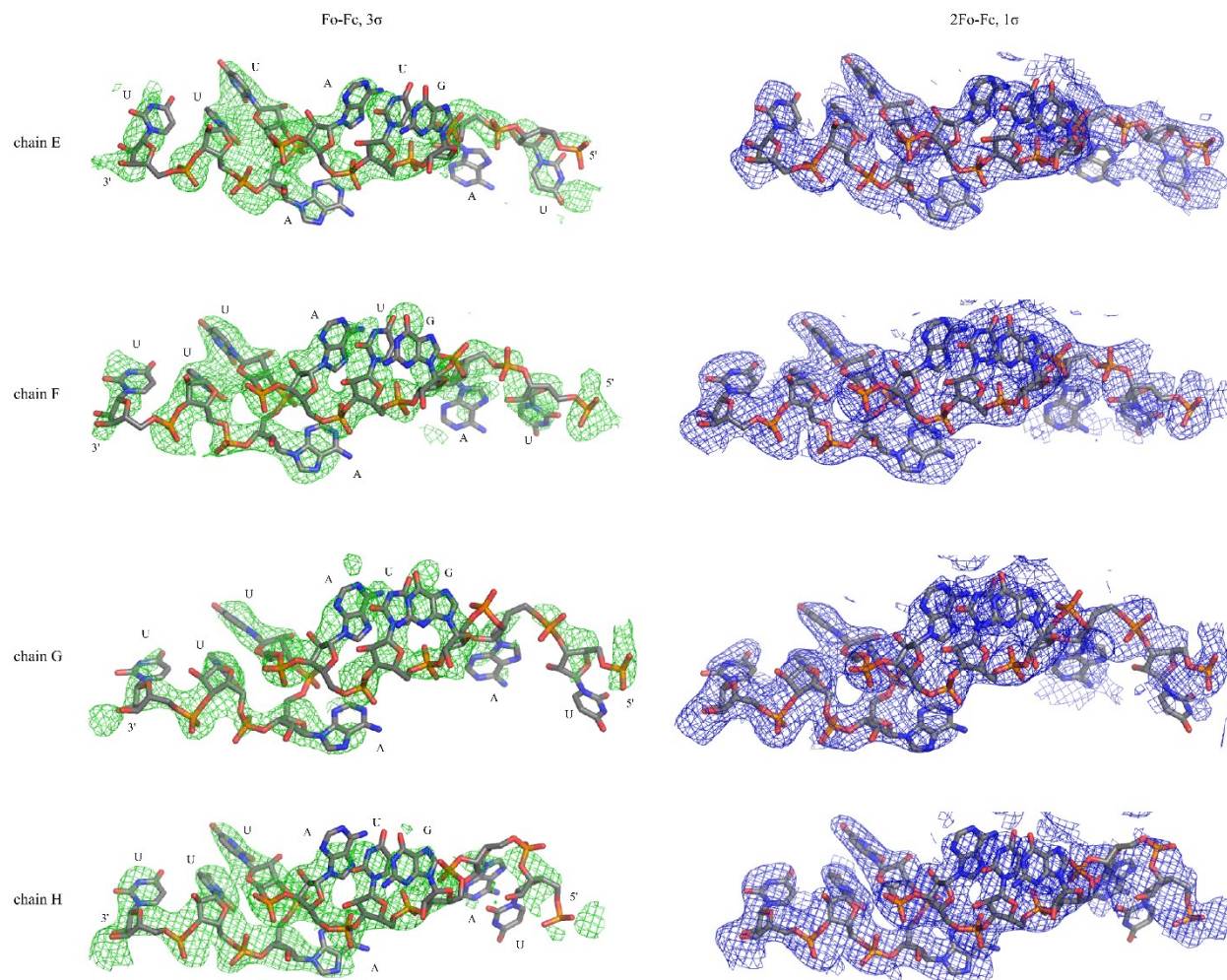
Supplementary Figure 3.1 The difference electron density maps before model building and the all featured electron density maps after building RNA models, calculated from the diffraction data of IFIT2 with 5'HO-UGUAGAAUAUU.



Supplementary Figure 3.2 (A) The overall view of RNA bound IFIT2 crystal structure, (B) the RNA bound IFIT2 structure (grey and blue) superimposed to the *apo* IFIT2 crystal structure (black and red).



Supplementary Figure 3.3 The difference electron density map (green), the all featured electron density map (blue), and the anomalous difference peaks (black) before model building, calculated from the diffraction data of IFIT2 with 5' 5-Br-dU-UAGUAUAUU. The anomalous signal of Bromide, indicated by the arrow, was enhanced by changing the x-ray wavelength from 0.954 Å (13.0 keV) to 0.9 Å (13.776 keV).



Supplementary Figure 3.4 The difference electron density maps before model building and the all featured electron density maps after building RNA models, calculated from the diffraction data of IFIT2 with 5' 5-Br-dU-UAGUAUAUU.

```

IFIT1: 1  MSTNGDDHQVKDSLEQLRCHFTWELSIDDDMPDLENRVLDQIEFLDTKYSVGIHNLAYVKHLKGQNEEA 71
IFIT5: 1  -MSEIRKDTLKAILLELECHFTWNLLKEDIDLFEVEDTIGQQLFLLTKSRLALYNLLAYVKHLKGQNKDA 70
IFIT2: 1  -MSENNKNSLESSLRQLKCHFTWNLMEGENSLDDFEDKVFYRTEFQNRFEKATMCNLLAYLKHLKGQNEAA 70
                                         41

IFIT1: 72  LKSLKEAENLMQEEHDNQANVRSVLTWGNFAWYYHMGRLAEAQTYLDKVENICKLSNPFYRMECPEID 142
IFIT5: 71  LECLEQAEEIIQQEHSDKEEVRSVLTWGNYAWVYYHMDQLEEAQYTGKIGNVCKKLSSPSNYKLECPETD 141
IFIT2: 71  LECLRKAEEIIQQEHADQAEIRSVLTWGNYAWVYYHMGRLSDVQIYVDKVKHVCEKFSSP--YRIESPELD 139

IFIT1: 143 CEEGWALLKCGGKNYERAKACFEKVLEVDPENPESAGYAISAYRLDGFKLATKNHKPFSLLPLRQAVRLN 213
IFIT5: 142 CEKGWALLKFGGKYYQKAKAAFEKALEVEPDNPEFNIGYAITVYRLDDSDRE-GSVKSFSGLPLRKAVTLN 211
IFIT2: 140 CEEGWTRLKCGGNQNERAKVCFEKALEKKPKNPEFTSGLAIASYRLDNWPPSQ-----NAIDPLRQAIRLN 205

IFIT1: 214 PDNGYIKVLLALKLQDEG----QEAEGEKYIEEALANMSSQTYVFRYAAKFYRKGSVDKALELLKKALQE 280
IFIT5: 212 PDNSYIKVFLALKLQDVH----AEAEGEKYIEEILDQISSQPYVLYAAKFYRKNSWNKALELLKKALEV 278
IFIT2: 206 PDNQYLKVLLALKLHKMREEGEEEGEGEKLVEEAELEKAPGVTDVLSAAKFYRRDEPDKAIELLKKALEY 276

IFIT1: 281 TPTSVLLHHQIGLCYKAQMIQIKEATKGQPRGQNREKLDKMIRSAIFHFESAVEKPTFEVAHLDLARMYI 351
IFIT5: 279 TPTSSFLHHQMGLCYRAQMIQIKKATHNRPKGDKLKVDELISSAIFHFKAAMERDSMFAYTDLANMYA 349
IFIT2: 277 IPNNAYLHCQIGCCYRAKVFQVMNLRENGMYGKR--KLELIGHAVAHLKKADEANDNLFVCSILASLHA 345

IFIT1: 352 EAGNHRKAEENFQKLLCMKPVEETMQDIHFHYGRFQEFQKKSVDVNAIIHYLKAIKIEQASLTRDKSINSL 422
IFIT5: 350 EGGQYSNAEDIFRKALRLENITDDHKHQIHYHYGRFQEFHRKSENTAIHHYLEALKVKDRSPLRTKLTSAL 420
IFIT2: 346 LADQYEEAEYYFQKEFS-KELTPVAKQLLHLRYGNFQLYQMKCEDKAIHHFIEGVKINQKSEKEKMKDKL 415
                                         376                               406       414

IFIT1: 423 KKLVLRKLRKALDLESLSLLGFVYKLEGNMNEALEYYERALRLAADFENSVRQGP----- 478
IFIT5: 421 KKLSTKRLCHNALDVQSLSALGFVYKLEGEKRQAAEYYEKAQKIDPENAEFLTALCELRLSI 482
IFIT2: 416 QKIAKMRLSKNGADSEALHVLAFLLQELNEKMQQADEDSERGLESGSLIPSASSWNGE----- 472

```

Supplementary Figure 3.5 The sequence alignment among IFIT1, IFIT5 and IFIT2. Key RNA binding residues in IFIT1, IFIT5, and IFIT2 are colored in blue, magenta, and red, respectively.

```

                                41
IFIT2: 1  MSENKNSLESSLRQLKCHFTWNLMEGENSLDDFEDKVFY41RTEFQNREFKATMCNLLAYLKHLKGQNEAA 70
IFIT3: 1  MSEVTKNSLEKILPQLKCHFTWNLFKEDSVSRDLEDRCVNCIEFLNTEFKATMYNLLAYIKHLDGNNEAA 70

IFIT2: 71  LECLRKAEEELIQQEHADQAEIRSLVTWGNyawvyyhmgRLSDVQIYVDKVKHVCEKFSSPYRIESPELDC 140
IFIT3: 71  LECLRKAEEELIQQEHADQAEIRSLVTWGNyawvyyhLGRLSDAQIYVDKVKQTCKKFSNPYSIEYSELDC 140

                                184
IFIT2: 141 EEGWTRLKCGGNQNERAKVCFEKALEKKPKNPEFTSGLAIASY184RLDNWPPSQNAIDPLRQAIRLNPQNQY 210
IFIT3: 141 EEGWTQLKCGR--NERAKVCFEKALEEKPNNEPFSGLAIAMYHLDNHPEKQFSTDVLKQAIELSPDNQY 208

IFIT2: 211 LKVLLALKLHKMREEGEEGEGEKLVEEAELEKAPGVTDVLRSAAKFYRRKDEPDKAIELLKKALEYIPNN 280
IFIT3: 209 VKVLLGLKLQKMNKE---AEGEQFVEEAELEKSPCQTDVLRSAAKFYRRKGDLDKAIELFQRVLESTPNN 274

IFIT2: 281 AYLHCQIGCCY281RAKVFQVMNLRENGMYGKRKLELIGHAVAHLKKADEANDNLFRCVCSILASLHALADQY 350
IFIT3: 275 GYLYHQIGCCYKAKVRQMONTGESEASGNKEMIEALKQYAMDYSN----KALEKGLNPLNAYSDLAEFL 339

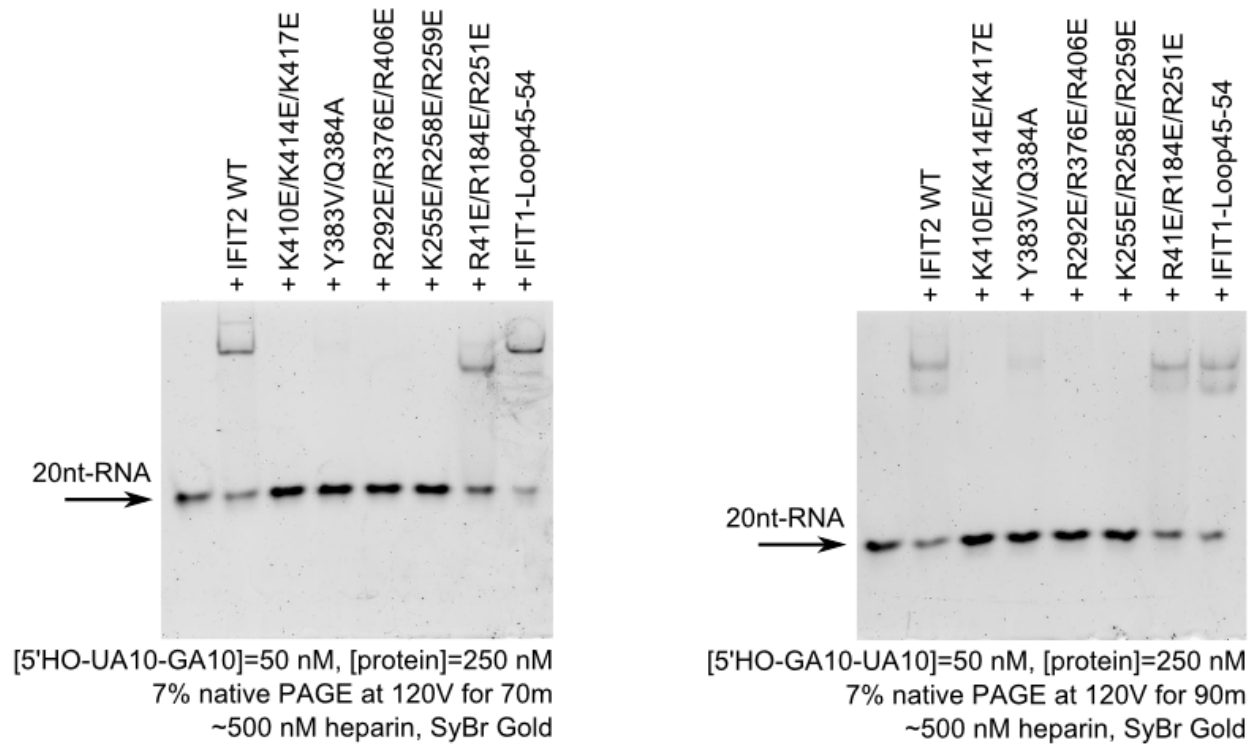
                                406 410 414 417
IFIT2: 351 EEAEEYFQKEFSKELTPVAKQLLHL406RYGNFQL410YQMKCEDKAIHHFIEGVKINQKS414REKEK417MKDKLQKIAK 420
IFIT3: 340 E--TECYQTPFNKEVPDAEKQQSHQRYCNLQKYNKSEDTAVQHGLEGLSISKKSTDKEEIKDQPQNVSE 407

IFIT2: 421 MRLSKNGADSEALHVLAFLLQELNEKMQQADE-----DSERGLESGSLIPSASSWNGE----- 472
IFIT3: 408 NLLPQNAPNYWYLQGL--IHKQNGDLLQAAKCYEKELGRLLRDAPSGI--GSIFLSASELEDGSEEMGQG 473

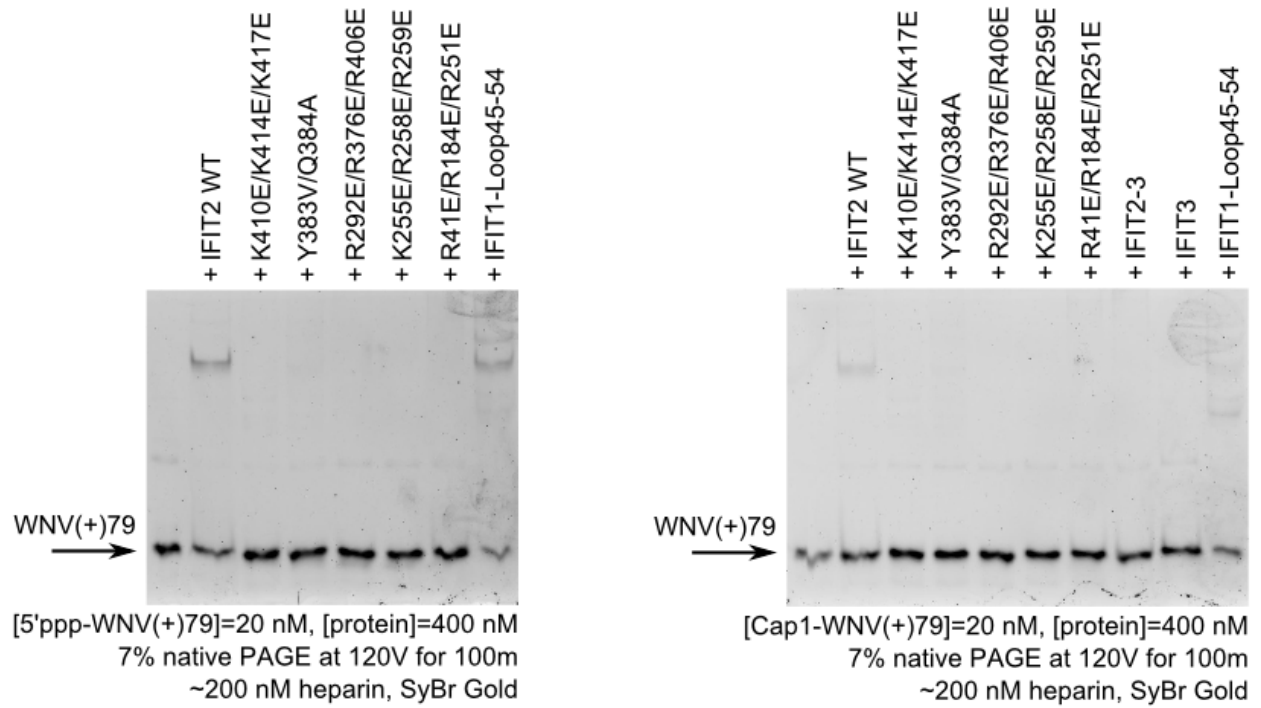
IFIT2: -----
IFIT3: 474 AVSSSPRELLSNSEQLN 490

```

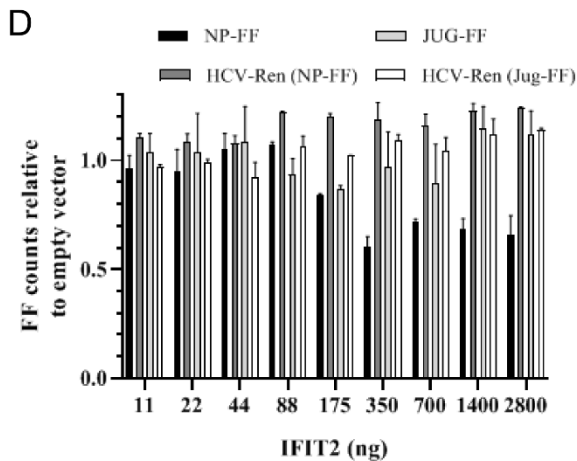
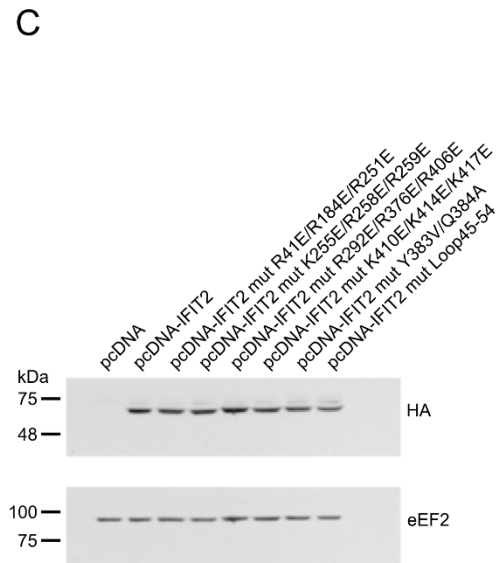
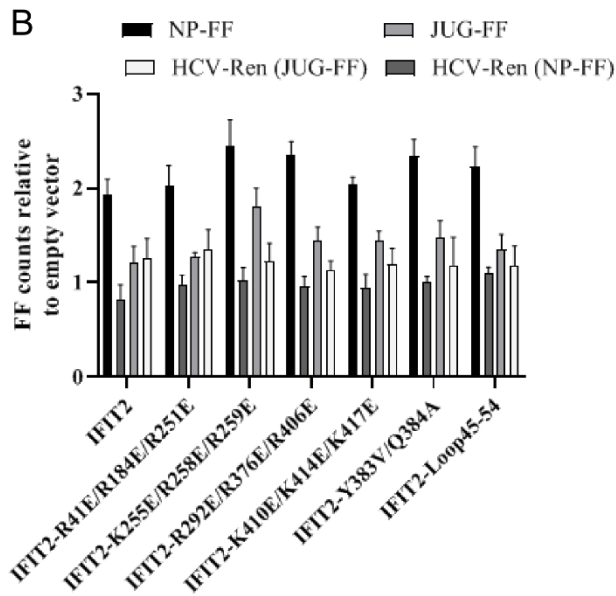
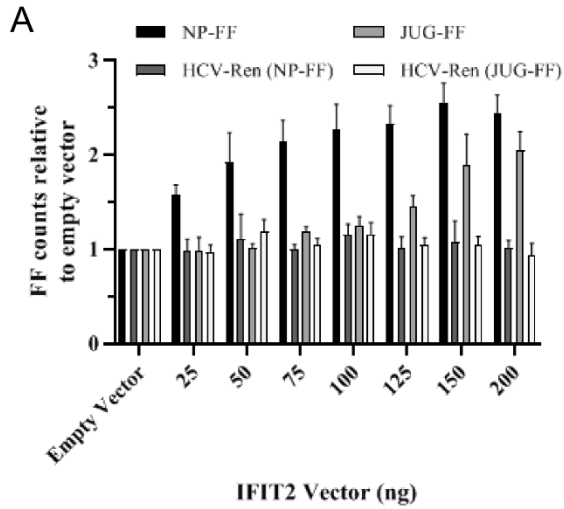
Supplementary Figure 3.6 The sequence alignment between IFIT2 and IFIT3. Key RNA binding residues in IFIT2 are colored in red.



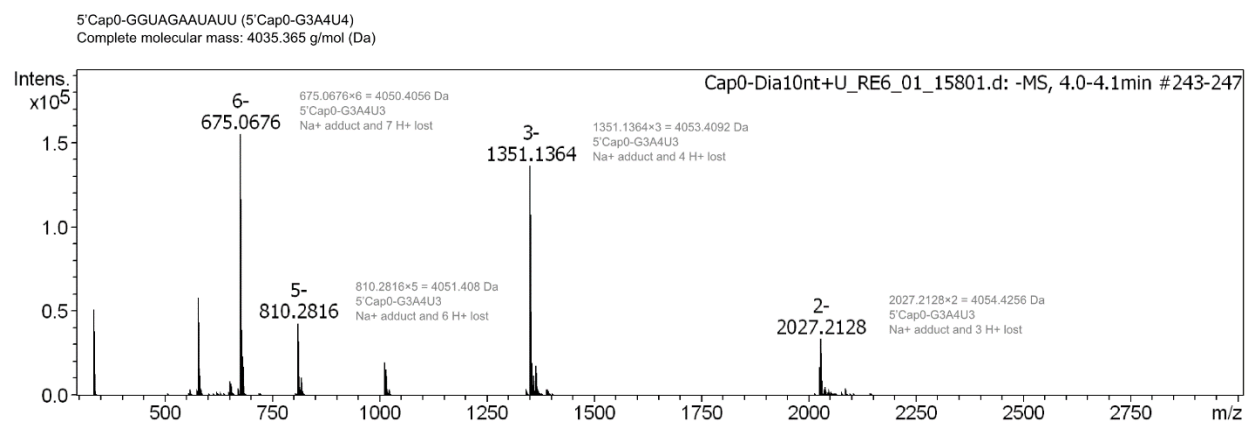
Supplementary Figure 3.7 The gel shift binding assay of WT IFIT2 as well as mutants with 20-nt 5'HO-RNAs containing AU-rich and non-AU rich segments in reverse order (sequences seen in **Figure 3.5A**).



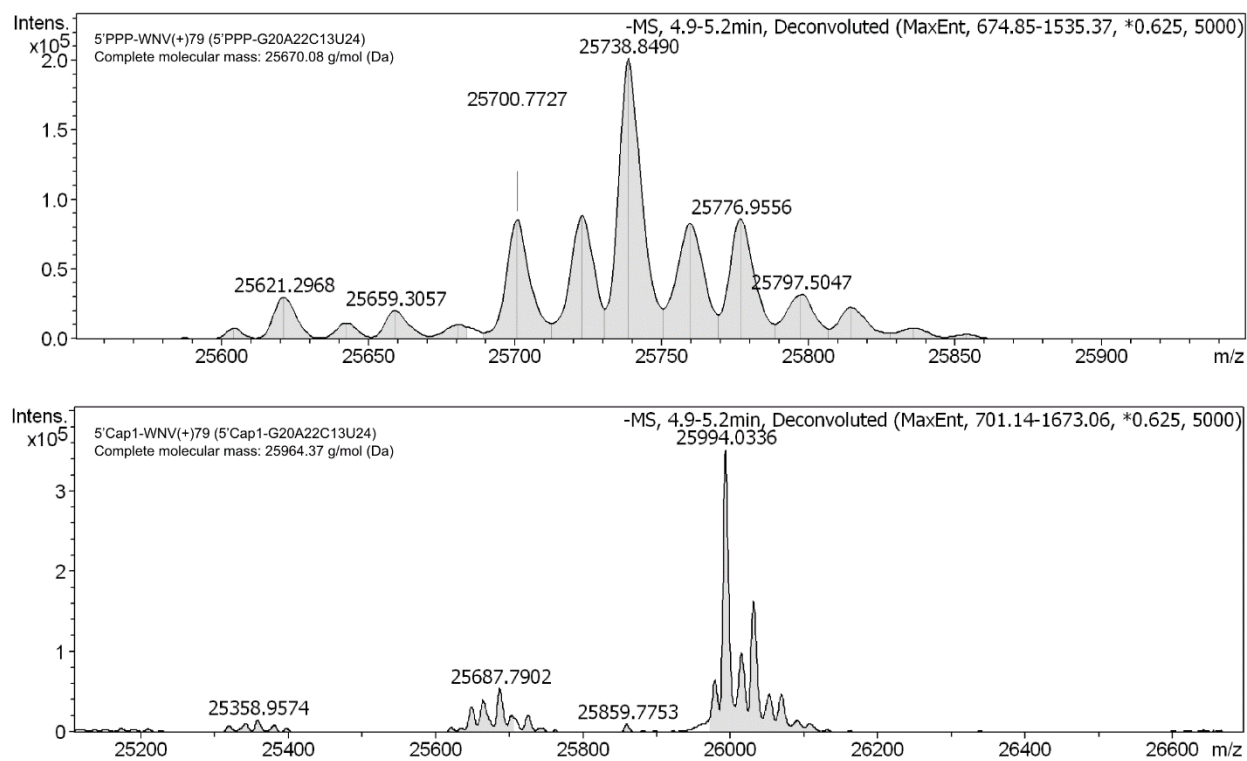
Supplementary Figure 3.8 The gel shift binding assay of WT IFIT2 as well as mutants with 79-nt 5'PPP- and Cap1-WNV 5' UTR. The hetero-dimeric IFIT2-3 and IFIT3 showed no interaction with WNV 5' UTR.



Supplementary Figure 3.9 Translation stimulation of NP and Juggled NP by IFIT2 and mutants in 293T cells. (A) Relative translation of FF mRNA bearing either the NP or the JUG sequence in 293T cells in presence of increasing amounts of IFIT2. Values are set relative to empty vector $\pm$ s.d. HCV-Ren mRNA was used as transfection control (N=6). (B) Relative translation of FF mRNA bearing either the NP or the JUG sequence in 293T cells transfected with 100 ng of IFIT2 vector or the various mutants. Values are set relative to empty vector $\pm$ standard deviation (N=6). HCV-Ren mRNA was used as transfection control. (C) Expression levels of IFIT2 and mutants in transfected 293T cells. eEF2 is used as loading control. (D) Relative in vitro translation of a FF reporter mRNA bearing either the NP or the Jug sequence in Krebs extract in presence of increasing amounts of purified recombinant IFIT2. Values are set relative to buffer $\pm$ standard deviation (N=2). HCV-Ren mRNA was used as translation control.



Supplementary Figure 3.10 The mass spectrometry analysis of the Cap0-RNA used for the ITC experiments in **Figure 3.4**. The mass spectrometry analysis of the 5'PPP-RNA was reported in chapter 2 (**Supplementary Figure 2.1**).



Supplementary Figure 3.11 The mass spectrometry analysis of the 5'PPP- and Cap1-WNV(+)-79 used for the quantitative binding assay in **Figure 3.6**.

	IFIT2 with 5'HO UGUAGAAUAUU	IFIT2 with 5'HO Br-dU- UAGUAUAUU
Data collection		
Space group	P212121	P1211
Cell dimensions		
a, b, c (Å)	78.7, 94.1, 165.5	79.4, 95.0, 154.7
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 92.7, 90.0
Resolution (Å)	48.76-2.81	154.58-2.70
I/ σ I	5.8 (0.5)	7.7 (0.7)
CC1/2	0.325	0.410
Completeness (%)	97.9% (91.3%)	99.9% (99.0%)
R _{meas}	0.636	0.242
R _{pim}	0.126	0.091
Multiplicity	25.0	7.1
Refinement		
Resolution (Å)	48.81-3.20	154.58-2.70
No. reflections	20742	63198
Rwork/Rfree	0.246/0.310	0.213/0.274
No. atoms		
Protein chain A	3629	3582
Protein chain B	3653	3608
Protein chain C		3576
Protein chain D		3636
RNA chain C	171	189
RNA chain D	191	189
RNA chain E		189
RNA chain F		189
Solvent		12
B-factors		

Protein chain A	102.8	56.0
Protein chain B	112.0	55.3
Protein chain C		63.7
Protein chain D		58.7
RNA chain C	174.4	90,4
RNA chain D	212.6	59.7
RNA chain E		71.6
RNA chain F		63.0
Solvent		54.7
RMS Bond lengths (Å)	0.0106	0.0136
RMS Bond angles (°)	2.60	2.78
Ramachandran outliers	0%	0%
Ramachandran favoured	97.41%	96.88%
Rotamer outliers	3.98%	4.01%
Clashscore	1.59	5.66
Molprobity score	1.48	1.95
Deposition		
PDB ID		

Supplementary Table 1 X-ray data collection and refinement statistics.

3.7 Bridge to Chapter 4

How IFIT2 binds to RNA and the functional role of IFIT2 associated with RNA binding become clearer after our work presented in Chapter 3. It also taught me an important lesson: the short RNA I learnt to make, purify, and modify in Chapter 2 may not be the best option to study the biological functions of RNA binding proteins. For instance, the biological function of IFIT2 dimerization would not be revealed if we stick to the short artificial RNA. We realized its function by using a piece of viral RNA that is similar to mRNA in cells. When our understanding on RNA binding of IFIT1 and IFIT2 grows from pieces into a small picture, we start to ponder the purpose for the existence of IFIT1/2/3 complex, which leads us to elucidate the functional roles of IFIT1 and IFIT1/2/3 complex in a context that is close to the real encounter of viral RNAs in cells, by utilizing virus derived RNA sequences for biochemical and structural studies.

Chapter 4 The effect of methylation and secondary structure elements on viral RNA binding by IFIT1 protein and IFIT complexes

Abstract

The crystal structure of short Cap0-RNA-bound human IFIT1 has revealed the molecular basis of IFIT1 acting on 5' end of non-self RNA. Viruses commonly counteract IFIT1 recognition by Cap1 methylation since most viral mRNAs are in Cap1 form. Moreover, the secondary structure elements present in the 5' untranslated region of viral RNA can further subvert IFIT1 action. The effect of Cap1 methylation and secondary structure elements present in viral mRNA on IFIT1 RNA binding is not clear. The effect of these self-protective measures on IFIT complexes is also unknown. We use biophysical and biochemical binding assays and structure approaches to investigate the effect of methylation and secondary structure elements at 5' end on viral RNA binding by IFIT1 protein and IFIT complexes.

4.1 Introduction

IFIT1, as one of the major antiviral defense effectors, has been shown to sequester 2'O unmethylated viral mRNA from binding to eIF4E, and consequently, inhibiting viral protein translation⁴⁹. Facing the potent viral inhibition exerted by IFIT1, viruses are driven to counteract to survive and thrive. One of the counteracting measures viruses take is to mimic host mRNA caps.

In reality, many viruses are capable of producing 2'O methylated mRNA (Cap1 form). For example, the family of dsDNA Hepadnaviridae and ss(+)RNA Retroviridae use the host capping machinery in the nucleus to generate Cap1 viral mRNA. The family of dsDNA Poxviridae, dsRNA Reoviridae, ss(+)RNA Flaviviridae and Coronaviridae, encode their own capping enzymes to generate Cap1 viral mRNA in cytoplasm⁷⁸. Recent studies showed that SARS-Cov-2 nsp16, the viral 2'O methyltransferase, is a key player to evade the innate immune detection by melanoma differentiation-associated protein 5 (MAD5)¹³⁶ and the restriction by IFIT1 and IFIT3¹³⁷. Apart from mimicking host mRNA caps, viral RNAs are highly structured. Complex secondary structure elements are present at 5' UTR of viral mRNA. For instance, the 5'UTR of Zika has two stem-loops⁸² and the 5' UTR of SARS-Cov-2 contains several stem-loops and pseudoknots⁸³. Such higher-order structure elements not only act as barriers to block the access by host sensors and effectors but also plays import role in viral replication, pathogenesis, and transmissibility¹³⁸. Evidence about the effect of Cap1 methylation and secondary structure elements on IFIT1-RNA interaction is still lacking.

Stawowczyk et al. performed sedimentation through glycerol gradient and Western blot analysis on HeLa cells treated with IFNs, and they detected the existence of a trimer-like IFIT1/2/3 complex with molecular weight in the range of ~150 to 200 kDa⁵⁸. The precise role of this complex in inhibiting viral replication is unknown but the binding of IFIT3 to IFIT1 has been shown to enhance its RNA binding affinity^{56,57}.

Using in vitro viral RNA binding assays combined with structure analysis, our work showed how viruses utilize methylation and secondary structure at 5'UTR of genomic or messenger RNA to circumvent IFIT1 defense. However, IFIT1, IFIT2, and IFIT3 cooperate to overcome the 5' end

blocking strategies developed by viruses. Therefore, synergistic RNA binding makes IFIT1/2/3 or IFIT1/3 complex superior to IFIT1 alone.

4.2 Results

4.2.1 The methylation effect on RNA binding affinity of IFIT1

In order to examine the effect of 2'O methylation on IFIT1 binding, we performed isothermal titration calorimetry (ITC) measurements to quantify IFIT1 with three capped 11-nt RNAs: Cap0-GGUAGAAUAUU, Cap1-GGUAGAAUAUU and Cap2-GGUAGAAUAUU. The different levels of methylation status were confirmed by mass spectrometry (**Supplementary Figure 4.1**). IFIT1 showed the highest affinity to Cap0-GGUAGAAUAUU of ~4 nM, a medium affinity to Cap1-GGUAGAAUAUU of ~21 nM, and the lowest affinity to Cap2-GGUAGAAUAUU of ~176 nM (**Figure 4.1**). From Cap0 to Cap1 methylation, the affinity dropped by ~5 fold, and from Cap1 to Cap2, the affinity dropped by ~8.5 fold, suggesting a significant weakening effect on IFIT1-RNA recognition by Cap1 methylation and a detrimental effect on IFIT1-RNA binding by Cap2 methylation. Our results are in line with a previous study on the kinetic analysis of IFIT1-RNA binding for which their affinity measurements, based on biolayer interferometry, also showed comparable nanomolar affinity tight binding for IFIT1 with Cap0-RNA in the range of ~6-18 nM, with Cap1-RNA in the range of ~21-44 nM¹³⁹. The difference between our ITC result and their result may be due to different RNA sequences and different technique.

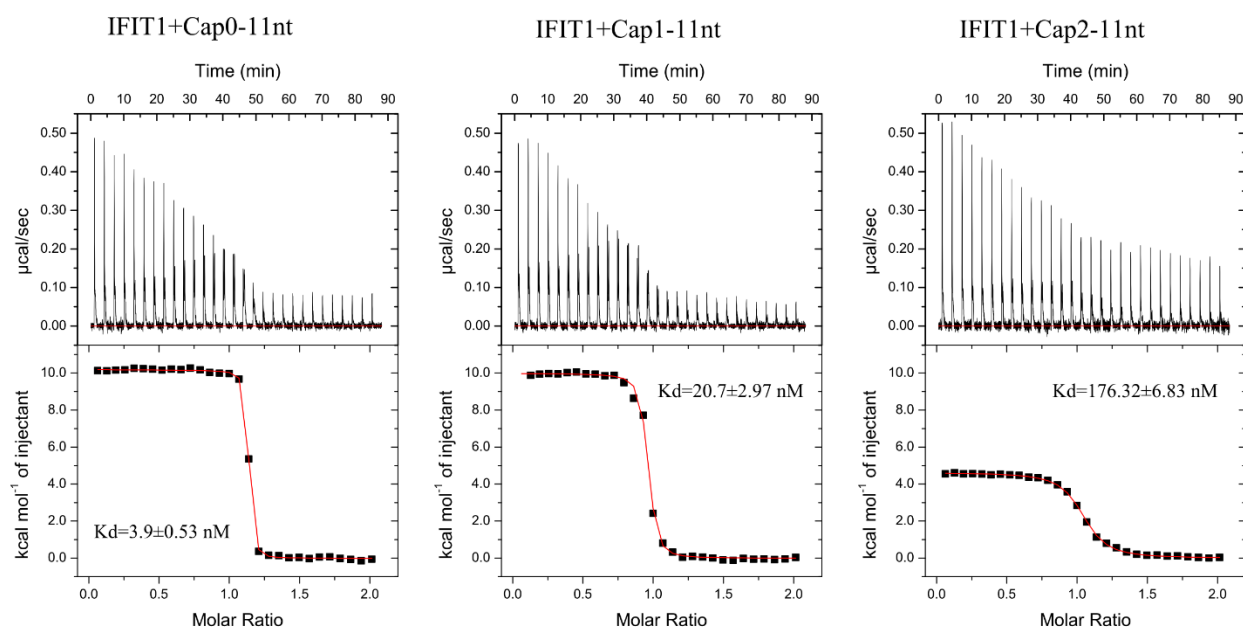


Figure 4.1 The 2'O methylation effect on IFIT1-RNA binding. The isothermal titration calorimetry measurements of K_d between IFIT1 and different capped 11-nt RNAs with methylation status of Cap0, Cap1, and Cap2, respectively.

4.2.2 Structure basis of 2'O-methylation on IFIT1-Cap1-RNA interactions

The ITC affinity measurement showed that upon adding one methyl group to the 2'O ribose position of the first nucleotide adjacent to the 7mG cap, the IFIT1-RNA binding affinity decreased by ~5 fold. Therefore we sought clues of the Cap1 methylation effect by structure determination of IFIT1 bound to different capped RNAs. Our previous study on Cap0-AAAAA bound IFIT1 study obtained high resolution crystal structure by using a IFIT1 double mutant (DM) that destroyed the dimerization but not affect its RNA binding⁴⁹. Therefore, we used IFIT1DM for the crystallization trials, which ended with crystal formation in IFIT1 DM with Cap0- and Cap1-RNA, respectively. The IFIT1DM with Cap2-RNA was not crystallized, likely due to the weak binding

and flexibility. The Cap0-RNA bound IFIT1DM and Cap1-RNA bound IFIT1DM both were crystallized in the same crystal form, with the same space group and unit cell (**Supplementary Table 4.1**). The unbiased difference electron density clearly showed the absence and presence of Cap1-methylation (**Supplementary Figure 4.2**). The overall structure alignment of protein and RNA showed no major changes (**Figure 4.2A**). Then we focused on the residues surrounding the 2'O ribose position. Three residues, Tyr157, Arg187, and Phe191, as well as one water molecule showed subtle displacement when aligned (**Figure 4.2B**). The methyl carbon weakened the IFIT1-RNA interactions in three aspects: 1) the methyl carbon caused more clashes with Tyr157, Arg187, and Phe191, as well as with the water molecule mediating hydrogen bonds, which can be seen through the shorter interatomic distances; 2) the subtle displacement shifted R187 further from the 5'-5' triphosphate linker, resulting in weaker electrostatic interactions, which can be seen through the longer interatomic distances; 3) in addition to clash, the water mediated H-bonds between ribose 2'O and R187 carboxylate O were weakened. (**Figure 4.2C-D**)

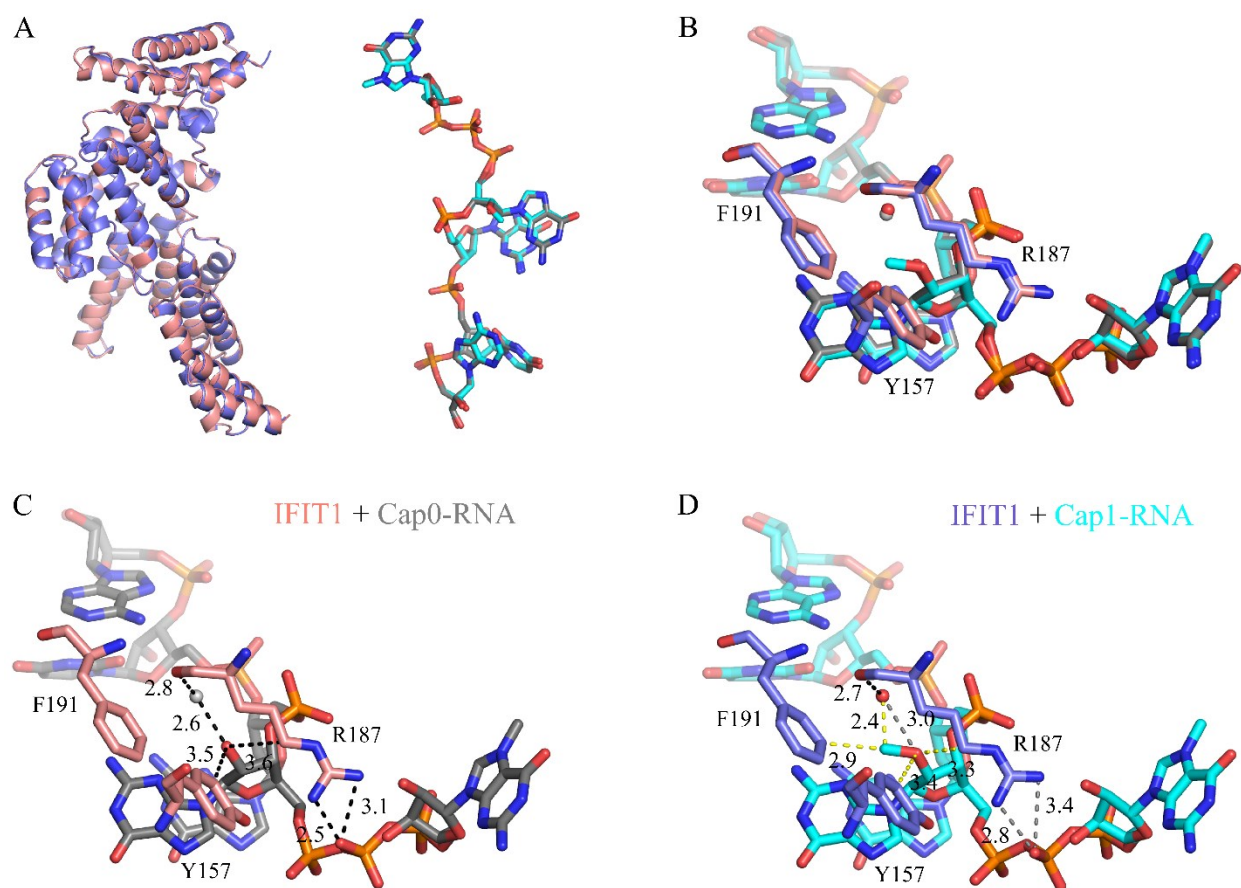


Figure 4.2 The molecular basis of 2'O methyl group weakening IFIT1-RNA interactions. (A) Structure alignment of protein and RNA, (B) three residues and one water molecule (sphere) surrounding the 2'O position of the first nucleotide adjacent to the cap, (C) interactions with 2' hydroxyl in IFIT1-Cap0-RNA structure, and (D) interactions with 2'-O-methyl in IFIT1-Cap1-RNA structure. Stable interactions in black, weakened interactions in grey, van der Waals repulsions in yellow.

4.2.3 The secondary structure effect on IFIT1-RNA binding

Viral RNAs are highly structured, such as the complex secondary structure elements in 5'UTR of Zika⁸² and SARS-Cov-2 genomic RNA⁸³ experimentally mapped by using nuclease digestion coupled with sequencing. The 5'UTR of Zika contains two stem-loops (SL): SLA and SLB. The

5'UTR of SARS-Cov-2 contains five stem-loops, SL1-5 and SL5 itself forms a four-way junction, as well as two pseudo-knots (PK). Zika and SARS-Cov-2 both belong to (+) ssRNA virus, their mRNAs are in Cap1 form by their own capping enzyme and methyltransferase. Therefore, we next sought to study the effect on IFIT1 recognition by viral mRNA 5' end with Cap1 status and secondary structure elements. However, because the in vitro transcription yield for SARS-Cov-2 5'UTR of SARS-Cov-2 was extremely low, presumably due to the very complex secondary structures, we use Zika 5'UTR to demonstrate the effect of Cap1 methylation in the presence of secondary structure on IFITs action. We generated IFIT1/3 complex based on the intrinsically strong protein-protein interactions between IFIT1 and IFIT3 (**Supplementary Figure 4.3A-B**). We made IFIT2/3 complex through co-expression strategy, then incubated purified IFIT2/3 complex with purified IFIT1, and ran size exclusion chromatography to obtain purified IFIT1/2/3 complex (**Supplementary Figure 4.3C-D**).

Zika genome is a positive-sense RNA with approximately 10.8 kilo bases (kb). The Zika 5' UTR has only three nucleotides overhang (**Figure 4.3**), resulting in limited accessibility for interaction with IFIT1. When the highly structured RNA is in uncapped form, only the complex IFIT1/2/3 and IFIT1/3 showed limited binding to 5'PPP-Zika RNA. When it is in Cap0 form, IFIT1 showed interaction though not as strong as the complexes. When it is in the functional Cap1 form, IFIT1 alone no longer showed binding but IFIT complexes manifested limited binding. Our ITC results revealed the ~5 nM high affinity binding between IFIT1 and the short Cap0 linear RNA. The absence of secondary structure elements made the RNA fully accessible to IFIT1 RNA binding site. Zika viral RNA 5' end binding assay showed the significant hindrance for IFIT1 binding due to the cap-proximal huge stem-loop A. SLA acts as a big barrier to block IFIT1 from accessing the RNA, therefore, substantially reducing IFIT1 binding affinity. In the presence of secondary

structure elements, Cap1 methylation impaired IFIT1 binding. However, the IFIT1/2/3 and IFIT1/3 complexes showed stronger effect to overcome Zika 5' stem-loop steric hindrance than IFIT1 alone (**Figure 4.3**). To evaluate the binding affinity of Zika Cap0- and Cap1-5'UTR with IFIT1, IFIT1/3, IFIT1/2/3, we did titration binding assays. The estimated K_d for IFIT1 were ~100-200 nM for Cap0 binding and ~500-600 nM for Cap1 binding. The estimated K_d for IFIT1/3 were ~50 nM for Cap0 binding and ~120-180 nM for Cap1 binding. The estimated K_d for IFIT1/2/3 were ~60 nM for Cap0 binding and ~120-180 nM for Cap1 binding. (**Supplementary Figure 4.4**) The IFIT1/3 or IFIT1/2/3 complexes showed ~4 fold stronger binding affinity than IFIT1 alone to Zika 5'UTR.

Zika (+) genomic RNA (mRNA), NCBI Reference Sequence: NC_012532.1

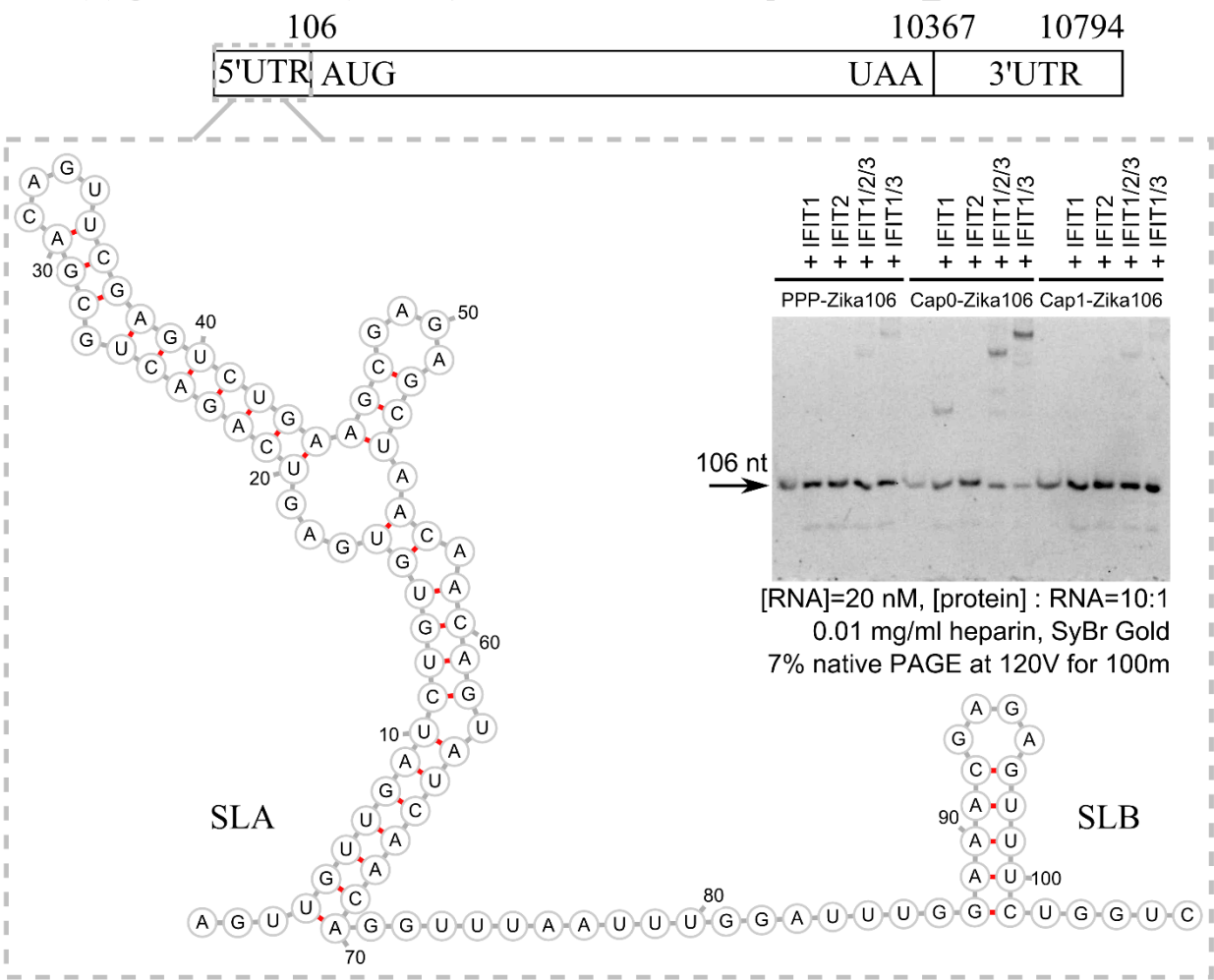


Figure 4.3 The 5'UTR of Zika mRNA and its interaction with IFITs. Zika genomic RNA map, the secondary structure of 5'UTR determined by enzymatic and chemical probing¹⁴⁰, and the gel binding assay of PPP-, Cap0- and Cap1-Zika 5' UTR with individual IFIT1 and IFIT2, as well as the IFIT1/2/3 and IFIT1/3 complexes.

4.2.4 IFIT complex synergistically binds to RNA

Influenza A virus is known to upregulate IFITs⁸⁷. Therefore, we tested IFIT binding to the 5'UTR of nucleoprotein (NP), one of the viral proteins produced at the early stage responsible for encapsidating viral genome for virion packaging. The 45 nucleotides (nt) 5'UTR of NP has a

single-stranded region followed by a stem-loop as predicted by M-fold server. (**Figure 4.4**). We tested its binding with individual IFIT and IFIT1/2/3 complex. The 5'PPP NP 5'UTR showed specific binding to IFIT1, IFIT2, and IFIT5, no binding was observed for IFIT3 as it is known not to bind to RNA. Interestingly, IFIT1 showed stronger binding than IFIT5, as IFIT1 specifically targets Cap0-RNA⁴⁹ and IFIT5 specifically targets 5'PPP-RNA⁴⁶. The single stranded region of NP 5'UTR does not contain enriched AU sequences, therefore, IFIT2 showed very weak binding. The IFIT1/2/3 complex completely shifted the RNA, demonstrating stronger RNA binding than IFIT1 or IFIT2 alone. To elucidate the synergy of IFIT1/2/3 complex in RNA binding, we performed structural studies to reveal the molecular basis for the observation of enhanced RNA binding.

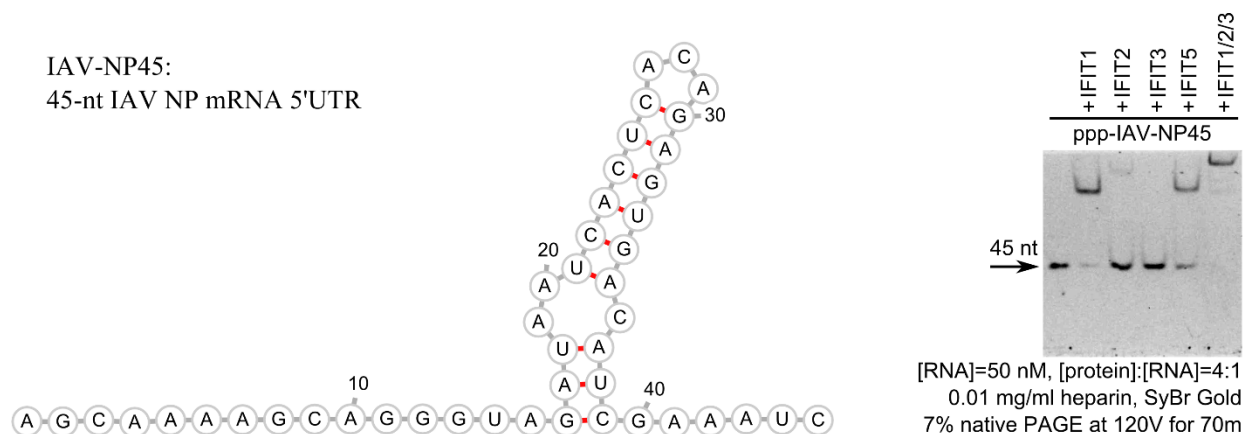


Figure 4.4 The 5'UTR of IAV NP mRNA and its interaction with IFITs. The secondary structure of 45-nt IAV NP 5'UTR predicted by the M-fold server¹¹², and the gel binding assay of 5'ppp-IAV-NP45 with individual IFIT1, IFIT2, IFIT3, IFIT5 and the IFIT1/2/3 complex.

4.2.5 RNA bound IFIT1/2/3 complex revealed by Cryo-EM

The extensive crystallization trials were unsuccessful for the IFIT1/2/3 complex without or with RNA, presumably due to the flexibility of this multi-protein complex. Then, we took the approach

of cryo-EM for the structure determination of the IFIT1/2/3 complex. Our pilot cryo-EM dataset of RNA-bound IFIT1/2/3 complex revealed a seriously freezing-induced protein denaturation problem: the complex fell apart during freezing. We then included the commonly used crosslinker glutaraldehyde in cryo-EM sample preparation of the 5'PPP-IAV-NP45 bound IFIT1/2/3 complex. Despite the slight preferred orientation issue, the crosslinked intact complex particles eventually resolved a ~ 3.9 Å cryo-EM map that reveals the assembly of the multi-protein IFIT complex and the elegant manner of capturing a 22-nt viral RNA segment by the IFIT complex (**Figure 4.5**).

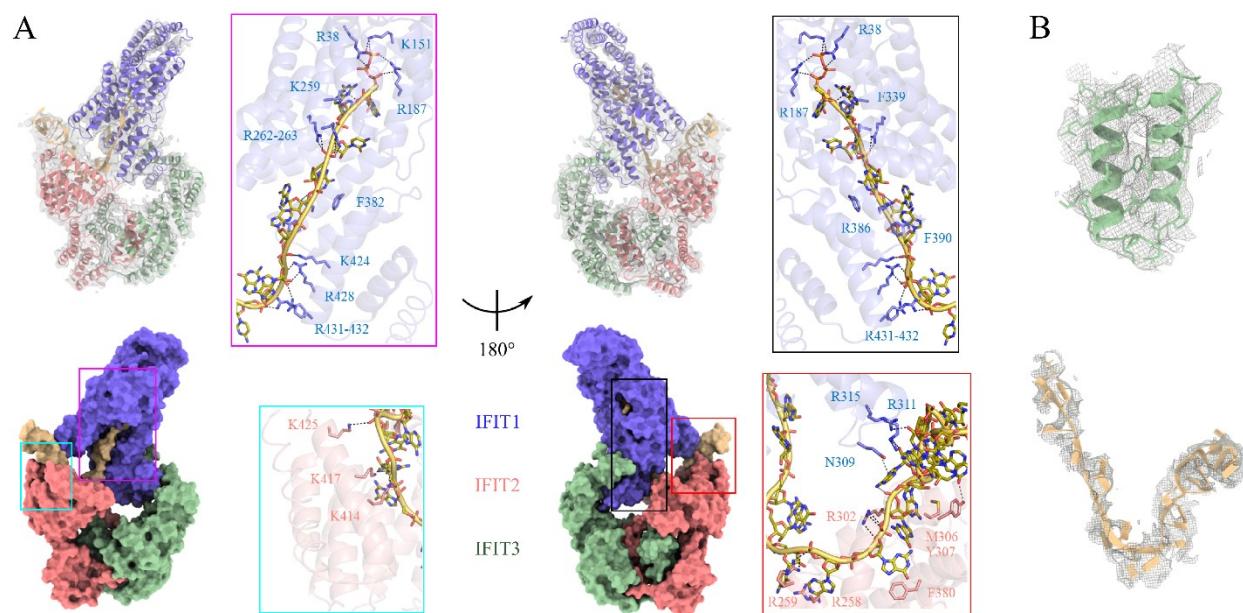


Figure 4.5 The cryo-EM model of RNA-bound IFIT1/2/3 complex at ~ 3.9 Å. (A) The overall cryo-EM model fit in the map and the assembly of the IFIT1/2/3 complex (front and back views), as well as the detailed molecular interactions between proteins and RNA. IFIT1 is colored in blue, IFIT2 in pink, IFIT3 in light green, and RNA in yellow. (B) The electron density map quality for protein and RNA.

As seen in the IFIT1-RNA crystal structure, not surprisingly, IFIT1 sequesters the 5' end of RNA when in complex with IFIT3 and IFIT2. The RNA stretches out of the IFIT1 binding pocket, the

chain is tethered by RNA binding helices located at the C-terminal of IFIT2. The RNA chain then bends and extends upward to be clamped by both IFIT2 and IFIT1. Beyond the "control" of IFIT1 and IFIT2, the RNA tail flaps around freely, resulting in invisible density to locate the flexible 3' end. As for interactions between proteins and RNA, starting from the 5' end, Arg 38, Lys 151, and Arg 187 interacted with the triphosphate. Phe 339 formed π - π stacking interactions with the guanine base of 2G. Lys 259 stabilized the phosphate backbone, in addition to the "molecular clipper" of Arg 262-263. Phe 382 stacked with the adenine base of 4A via π - π interactions, Phe 390 stacked with 6A. Arg 386, Lys 424, and Arg 428 along with another "molecular clipper" of Arg431-432 electrostatically interacted with the phosphate backbone near 7A, 8G, and 9C. The RNA was anchored by IFIT2 as soon as it stretched out of the IFIT1 binding pocket. The two adjacent arginines 258 and 259 acted differently from being as a "molecular clipper" as seen in Chapter 3: Arg 258 stacked with 10A and Arg 259 interacted with the phosphate backbone. Then, the RNA chain reached a site where IFIT1 and IFIT2 clamped on the chain. From IFIT2, Arg 302 formed electrostatic interactions with the phosphate backbone and Phe 380 stacked with the guanine base of 11G. Tyr 307 served as the base to stack with Met 306 which stacked on 13G. Three lysines 414, 417, and 425, as two of them seen in Chapter 3, acted together to stabilize the phosphate backbone. From IFIT1, Gln 309 formed an H-bond with the adenine base of 15A; Arg 311 and Arg 315 formed electrostatic interactions with 15A and 16G. Beyond the reach of protein residues, the nucleotides numbered 17A to 22C mostly formed base-stacking interactions (**Figure 4.5A**).

4.3 Discussion

Cap1 methylation has been thought as a protective measure utilized by host to prevent action by innate immune sensor such as RIG-I and MDA5 as well as effector such as IFIT1 to avoid

autoimmune attack. But the function of Cap2 methylation remains unclear except for an additional safety layer to protect host from innate immune response. Also, the ratio of Cap1 versus Cap2 host mRNA is unclear. A recent study provided important discoveries about the mechanisms of Cap2 methylation on host mRNA. Despic and Jaffrey developed a method, namely CircLigase-assisted mapping of caps by sequencing (CLAM-Cap-seq), to map transcriptome wide Cap2 methylation. Compared with the rapidly co-transcriptional events of 7mG and Cap1 methylation in nucleus, the cytoplasmic Cap2 methylation was rather slow and mRNAs were gradually enriched in Cap2 methylation over their lifetime. Slow Cap2 methylation provides a reaction time window for innate immune effectors to respond to rapidly replicating viral Cap1 mRNAs before viral mRNA acquires a high level of Cap2 methylation. Meanwhile, it suppresses the autoimmune response to host Cap1 mRNAs that are gradually modified to Cap2 RNA as they age⁸⁰. Our binding affinity and crystal structures show Cap2 methylation indeed effectively shields host mRNA from being acted on by IFIT1, supporting the milestone discovery of Cap2 methylation mechanism by Despic and Jaffrey⁸⁰. IFIT1 maintains high affinity binding to Cap1-RNA, which is necessary because the majority of viruses are capable of producing Cap1-mRNA either by their own encoded capping enzyme and 2'O methyltransferase or by hijacking the host's capping and methylation apparatus⁷⁸. However, the tight binding to Cap1-RNA inevitably makes the host mRNA in Cap1 form susceptible to IFIT1. The newly discovered Cap2 methylation mechanism loosens our worries on autoimmune action by IFIT1. Over the lifespan from translation to degradation, host mRNAs are gradually methylated to Cap2 status, minimizing the risk of being recognized by IFIT1. Therefore, Cap2 methylation is a fundamental signature of host mRNA distinguished from the viral counterpart.

These secondary structures not only act as barriers to block the recognition of innate sensors and effectors, therefore, downplaying the host defense, but they are also functional elements that are critical for virus infectivity and pathogenicity. For example, a long-range interaction between the Zika 5'UTR and viral Envelope protein coding region is significant for ribosome binding and scanning on 5' Cap1 mRNA⁸². The complex SL5 contains a four-way junction, which is present in all coronavirus; the start codon of nsp1 is located in SL5. The structure features in the 5' UTR of SARS-Cov-2 and the AUG start codon location downstream of a four-way junction make it reminiscent of internal ribosome entry site (IRES) found in HCV⁸³.

IFIT3 has been known to modulate and strengthen IFIT1 binding to RNA. Johnson et al. solved the crystal structure of Cap0-RNA bound IFIT1 in complex with C-terminal three helices of IFIT3⁵⁷. Their viral infectivity assay transfecting the C-terminal domain depleting IFIT3 mutant can no longer enhance IFIT1 RNA binding, suggesting that the TPR-mediated protein-protein interactions between C-termini of IFIT1 and IFIT3 are essential to strengthen IFIT1 binding to RNA. The synergistic IFIT1/3 complex RNA binding is most likely contributed by the C-terminal domain of IFIT3 as seen in the crystal structure⁵⁷. However, for the IFIT1/2/3 complex, apart from the modulation by IFIT3 C-terminal domains, nothing is known about the architecture of the three-protein complex as well as the role of IFIT2 in the complex. Our cryo-EM structure of the RNA-bound IFIT1/2/3 complex reveals the assembly of this multi-protein complex and the synergy of the complex binding to a viral RNA. IFIT1 sequesters the 5' end of the RNA, IFIT2 uses C-terminal RNA binding helices to tether the RNA chain outside the IFIT1 binding pocket to further stabilize the RNA binding, and IFIT1 and IFIT2 both are involved to clamp on the RNA chain as the final step of capturing the viral RNA. In the IFIT1/2/3 complex, IFIT1 acts as a primary binder and its RNA binding is enhanced through IFIT3 modulation⁵⁷; IFIT2 acts as a secondary binder to

stabilize the RNA chain stretching out of the IFIT1 binding pocket. The auxiliary role of IFIT2 in RNA binding cannot be revealed without a longer RNA used for the structural studies. Viruses commonly antagonize IFIT1 action by Cap1 methylation and secondary structure elements at the 5' end. Nevertheless, the IFIT1/2/3 complex could overcome the counteracting measure that some viral Cap1-mRNAs potentially subvert the combat by IFIT1 alone.

4.4 Material and Methods

4.4.1 IFIT1 cloning, expression and purification

The DNA sequence of human IFIT1 (UniProtKB accession number: P09914) was cloned into a pSMT3 vector (pET-28a backbone) between restriction enzyme BamHI and NotI sites. Fusion proteins carry an N-terminal His6-SUMO tag which is cleavable by Ulp1 protease. Recombinant IFIT1 was expressed using BL21 (DE3) cells in Luria Broth culturing media. Cells were grown at 30 °C initially until the optical density reached ~0.7, then the temperature was lowered to 18 °C. Cells were induced with 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG), further grown overnight, and harvested next morning. The cell pellet was resuspended in Ni binding buffer (20 mM Tris pH 7.8 at 0 °C, 500 mM NaCl, 10 mM imidazole, 5-10% glycerol (volume by volume, v/v)), supplemented with 2 mM 2-Mercaptoethanol (β ME), 1 mM phenylmethylsulfonyl fluoride (PMSF), and 0.1% Triton X-100 (v/v). The resuspended cells were lysed by using a French homogenizer, followed by centrifugation at 50,000 g for 30 min. The supernatant was loaded via AKTA purifier sample pump on a self-packed Ni-NTA column (~2 ml bed volume of clean beads per 1 L culture) equilibrated in Ni binding buffer. After loading the supernatant, the Ni-NTA column was washed with Ni washing buffer (20 mM Tris pH 7.8 at 0 °C, 1 M NaCl, 10 mM imidazole, 5-10% glycerol, 2 mM β ME) until the UV absorbance was stabilized. For gradient

elution, the AKTA system pump A and B were equilibrated in Ni binding buffer and Ni elution buffer (20 mM Tris pH 7.8 at 0 °C, 500 mM NaCl, 300 mM imidazole, 5-10% glycerol, 2 mM β ME), respectively. Further washing was performed with 10% Ni elution buffer to remove loosely bound contaminants until the UV absorbance was stabilized. The His6-SUMO-tagged IFIT1 was eluted off the Ni-NTA column by applying a linear gradient of Ni elution buffer. The first round Ni eluted tagged protein was cleaved by adding home-purified SUMO protease Ulp during dialysis at 4 °C overnight in dialysis buffer (20 mM Tris pH 7.8 at 0 °C, 100 mM NaCl, 2 mM β ME). The next day, the dialyzed sample was reloaded to the Ni-NTA column cleaned with Ni elution buffer and Ni washing buffer and equilibrated with dialysis buffer. The tag-cleaved IFIT1 protein flew through the column, the second round Ni flow-through sample was centrifuged to remove any precipitation and then passed to a cation-exchange HiTrap SP HP column (GE Healthcare). The 5 ml SP column was washed with buffer B (20 mM Tris pH 7.8 at 0 °C, 1 M NaCl, 1 mM DTT) and then equilibrated with buffer A (20 mM Tris pH 7.8 at 0 °C, 1 mM DTT). The IFIT1 protein was gradually eluted during a salt gradient ranging from ~15%-35% of buffer B. The pool of SP column elution was concentrated and further polished by a Superdex 200 10/300 GL column (GE Healthcare) using gel filtration (GF) chromatography in GF buffer (20 mM Tris pH 7.8 at 0 °C, 150 mM NaCl, 1 mM DTT). The pool of final polished protein was concentrated and stored at –80 °C.

The IFIT1 DM (L457E/L464E) was obtained from our previous study⁴⁹, and it was expressed and purified in the same way as IFIT1.

4.4.2 IFIT1/2/3 complex purification

The expression and purification method of IFIT2/3 mentioned in Chapter 3 were followed. The purified IFIT1 and IFIT2/3 were incubated with 1:1 molar ratio on ice for ~30 minutes, they were

then injected on to a Superdex 200 10/300 GL column (GE Healthcare) equilibrated in gel filtration buffer (20 mM Tris pH 7.8 at 0 °C, 150 mM NaCl, 1 mM DTT). The pool of final polished protein was concentrated and stored at –80 °C.

4.4.3 TbMTr2 purification

The plasmid for TbMTr2 with an N-terminal 6xHis tobacco etch virus (TEV) cleavable tag was obtained from a previous study⁴⁹. It was expressed in BL21 (DE3), the first round Ni-NTA followed the same path as IFIT1 except for using a prepacked HisTrap HP 5 ml column (Cytiva). The Ni eluted tagged protein was dialyzed at 4 °C overnight in dialysis buffer (20 mM Tris pH 7.8 at 0 °C, 100 mM NaCl, ~5% glycerol). The next day, the dialyzed sample was loaded to a HiTrap DEAE FF 1ml column (GE Healthcare) equilibrated with buffer A (20 mM Tris pH 7.8 at 0 °C, 1 mM DTT), TbMTr2 flew through DEAE column and was collected while other contaminants bound to the column. The flow-through of DEAE column was loaded to a HiTrap SP HP 5ml column (GE Healthcare). The SP column was washed with buffer B (20 mM Tris pH 7.8 at 0 °C, 1 M NaCl, 1 mM DTT) and then equilibrated with buffer A. The TbMTr2 protein was gradually eluted during a salt gradient ranging from ~15%-24.5% of buffer B. The pool of SP column elution was concentrated and further polished by a Superdex 200 10/300 GL column (GE Healthcare) in GF buffer (20 mM Tris pH 7.8 at 0 °C, 150 mM NaCl, 2 mM DTT). The pool of final polished protein was concentrated, aliquoted to 50% glycerol stock, and stored at –20 °C (short term) or –80 °C (long term).

4.4.4 Cap2 methylation and electrophoretic mobility shift assay

The in vitro transcription and purification of the short 11-nt RNA was based on the method developed in Chapter 2. The Cap1 methylation reaction was set in the same way as in Chapter 1.

For Cap2 methylation reaction, we set up a 300 μ l reaction in the following way: $\sim$ 30 nmol of Cap1-11nt, with 30 μ l of 10x capping buffer, $\sim$ 150 nmol of SAM, $\sim$ 4 mM DTT, $\sim$ 3 nmol of TbMTr2 with the enzyme to RNA ratio of $\sim$ 1:10, and added RNase-free water to the reaction volume. The well mixed reaction was kept at 27 °C for 3 hours. After the reaction, 98% formamide loading buffer was added and the reaction was loaded to 7M urea 20% acrylamide TBE gel for gel extraction. The RNA recovered from ethanol precipitation was resuspended in RNase free water and stored at -20°C before use. The quality of RNA was verified by mass spectrometry with the protocol reported in Chapter 2. The 5'UTR of Zika and IAV NP in vitro transcription reactions were set up in the same way as the short RNA, except a self-cleaving hepatitis delta virus ribozyme sequence¹²⁰ was added after the desired RNA sequence at the 3' end of the DNA template. The dsDNA template was generated by using recursive-PCR method¹²¹. The electrophoretic mobility shift assay was set up in the same way as mentioned in Chapter 3.

4.4.5 Isothermal titration calorimetry

The IFIT1 protein of 25 μ M (calculated in monomer form) in 50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM DTT was loaded in the cell; the Cap0-, Cap1-, and Cap2-RNA of 250 μ M in the same buffer as protein was loaded in the syringe, the injection and mixing were run in a MicroCal iTC200 machine (GE Healthcare) at 4 °C. The ITC data were acquired with a reference power of 10 μ cal/s and processed by fitting to a one-site binding model to determine n (number of binding sites) and Kd (dissociation constant) using the accompanied Origin 7.0 software.

4.4.6 Crystallization of IFIT1 with capped RNA

The purified IFIT1 DM (L457E/L464E) of 5 and 7.5 mg/ml in GF buffer was incubated with molar excess capped RNA on ice for $\sim$ 30 min before setting up crystallization. The molar ratio of Cap0-

GGUAGAAUAUU to IFIT1DM was 1.5:1, Cap1-GGUAGAAUAUU to IFIT1DM of 2:1 and Cap2-GGUAGAAUAUU to IFIT1DM of 4:1. No crystal was formed in drops with IFIT1DM and Cap2-RNA. The Cap0- or Cap1-RNA bound IFIT1DM crystals were obtained in 5-7.5 mg/ml protein drops mixed with equal volume of reservoir solution containing 0.1 M Tris pH 8.1, 200 mM CaCl₂, 21-23% (w/v) PEG200 using the hanging-drop vapor diffusion method at 4 °C, based on a previous study⁴⁹. The crystals were harvested and flash frozen in liquid nitrogen without additional cryoprotectant. The X-ray diffraction data was collected at the Canadian Light Source beamline CMCF-BM and processed using DIALS¹²². The structure was solved using PHASER implemented in CCP4i2¹²³ with an ensemble of the IFIT1 structure from PDB 5udi⁴⁹ after removing RNA as a search model and refined using Refmac¹²⁴ and Coot¹²⁵.

4.4.7 Size Exclusion Chromatography-coupled Multi-Angle Light Scattering (SEC-MALS)

The purified recombinant protein (IFIT1/3 and IFIT1/2/3), ~70 µl of ~3 mg/ml in the GF buffer (20 mM Tris pH 7.6, 150 mM NaCl, 1 mM DTT), was injected on to a Superdex 200 10/300 GL column with a flow rate of 0.4 ml/min and analyzed by a MiniDAWN TREOS light-scattering detector (Wyatt Technology Corporation) and a Optilab rEX (Wyatt) refractive index detector. The experiment was carried out at room temperature. The calibration using BSA (~70 µl of ~3 mg/ml) as a standard for molecular weight, the chromatographic and scattering profiles were analyzed using Astra software (Wyatt).

4.4.8 High-Performance Liquid Chromatography (HPLC) and Mass

Spectrometry

The purified IFIT1/2/3 protein complex, 16 μ l of 0.1 mg/ml in GF buffer (20 mM Tris pH 7.6, 150 mM NaCl, 1 mM DTT), was applied to a PLRP-S reverse-phase column (5 μ m bead size, 1000 Å pore size) at 80 °C pre-equilibrated in 95% mobile phase A (0.1% formic acid in water) and 5% mobile phase B (0.1% formic acid in 100% acetonitrile) on an Agilent Technologies 1260 Infinity HPLC system coupled to a Bruker amaZon speed ETD ion trap mass spectrometer. The sample was eluted with a linear gradient from 5% buffer A to 100% buffer B. The detailed running parameters were based on this study¹⁴¹. Data was analyzed using the Bruker DataAnalysis software.

4.4.9 Cryo-EM sample preparation, data collection and processing

The purified IFIT1/2/3 complex stored at –80 °C was thawed on ice, spun down, and incubated with molar excessive 5'PPP-IAV NP45. The protein-RNA mixture was injected on to a Superdex 200 10/300 GL column (GE Healthcare) equilibrated in HEPES gel filtration buffer (50 mM HEPES pH 7.5 at 0 °C, 150 mM NaCl, 1 mM DTT) for buffer exchange and removal of aggregates. The peak fraction of RNA-bound protein (~1 mg/ml, UV260/280 of ~1.38) was used for cryo-EM sample freezing. First, 1 μ l of ~0.175% glutaraldehyde was applied to a glow-discharged C-flat 300 mesh 2/1 copper holey carbon grid, then 3.5 μ l of ~0.5 mg/ml 5'PPP-IAV NP45 bound IFIT1/2/3 complex was added to the grid and resuspended with the tiny drop of glutaraldehyde. Then the grid was blotted at 4 °C and 100% humidity using a Vitrobot IV (FEI) and plunge-frozen into liquid ethane. Data were collected at the McGill Facility for EM Research using an FEI Titan Krios TEM operating at 300 kV with a Gatan K3 direct electron detection camera and a Gatan GIF BioQuantum LS imaging filter. Movies were collected in counting mode using SerialEM, with a

total dose of $80 \text{ e}/\text{\AA}^2$ over 40 frames and a set defocus range of -0.5 to $-2.25 \text{ }\mu\text{m}$ at a nominal magnification of 130,000, resulting in a pixel size of $0.675 \text{ }\text{\AA}$. Micrographs were motion corrected using cryoSPARC¹³⁰ (v4.2), followed by contrast transfer function (CTF) estimation. Micrographs with CTF estimated resolution beyond $5 \text{ }\text{\AA}$ were excluded. A manual picking was performed on a few selected micrographs, the picked particles were trained using Topaz¹³¹, then the Topaz trained model was used to pick particles for all the manually curated micrographs. The Topaz picked particles were subject to 2D classification and ab-initio reconstruction to remove junk particles. The genuine particles were used as the input to train Topaz for picking. Several iterations of Topaz picking and classifications were performed until no further improvement on the number of genuine particles. Once the particle picking was done, multiple rounds of ab-initio reconstruction were performed to select a highly homogeneous set of particles. The final homogeneous particles were refined using homogeneous refinement followed by non-uniform refinement in cryoSPARC.

4.4.10 Cryo-EM Modeling and refinement

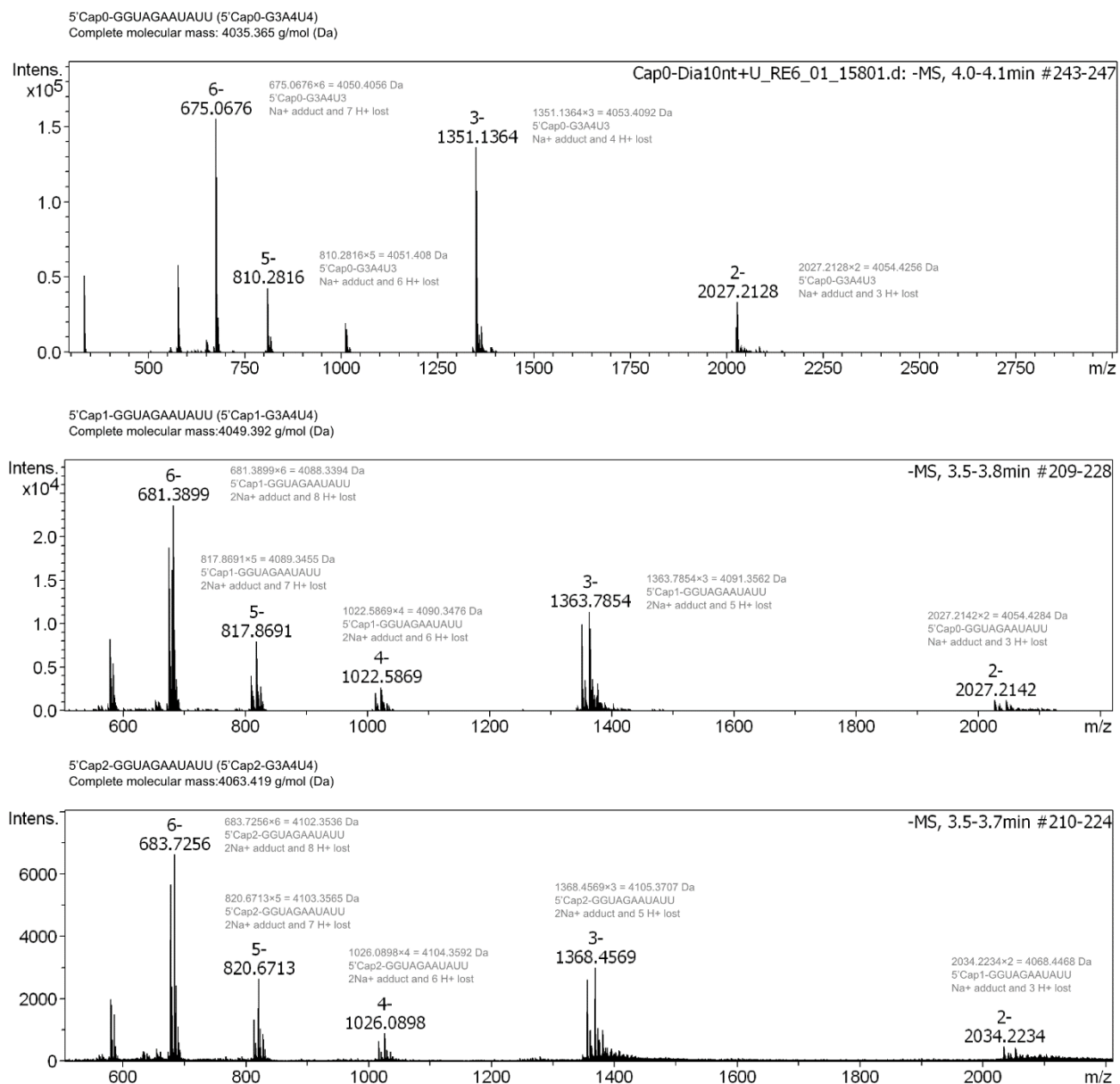
The two half maps from non-uniform refinement were input to DeepEMhancer¹⁴² for map sharpening. The sharpened map was used for model building. IFIT1 and C-terminal IFIT3 (PDB:6c6k⁵⁷), IFIT2 from the RNA bound crystal structure in Chapter 3, and IFIT3 from the AlphaFold predicted model (AF-Q5T765-F1) were divided into helix fragments (2-3 helices). The fragments were fit into the sharpened map based on secondary structure features by using ChimeraX¹³². Then, the loop regions in each fragment were removed, the initial model containing fitted helices and the sharpened map were input to the program AlphaFold_unmasked¹⁴³ for further model building. The output ranked solutions were carefully examined in ChimeraX, the solution that fit the map the best was used to model RNA in Coot¹²⁵. Once the RNA model was completed and refined in Coot, the built RNA bound IFIT complex model was further relaxed both

in cartesian and torsion space by using Rosetta electron density scoring function^{133, 134}. The output relaxed model and the non-uniform refined map were input to Phenix¹³⁵ for real space refinement.

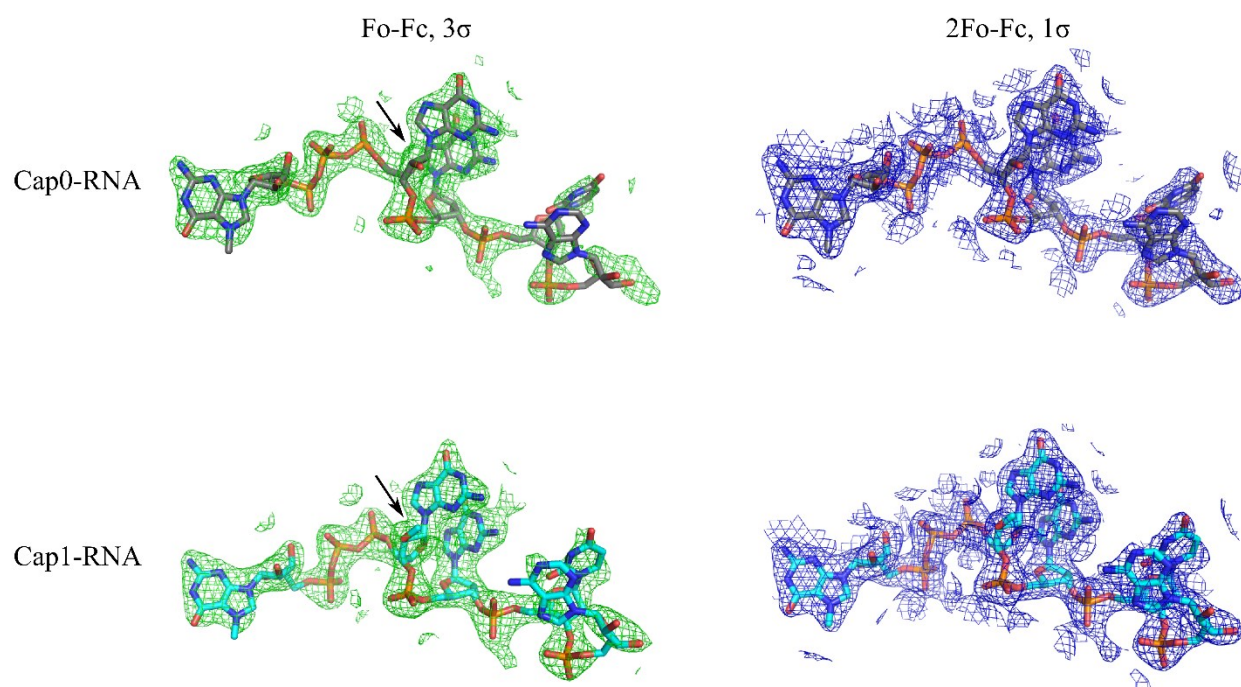
4.5 Acknowledgement

We thank the Canadian Light Sources for X-ray diffraction experiments. We also thank the cryo-EM facility at McGill University for screening and data collection. We are grateful to Dr. Alexander S. Wahba for RNA LC-MS (McGill University).

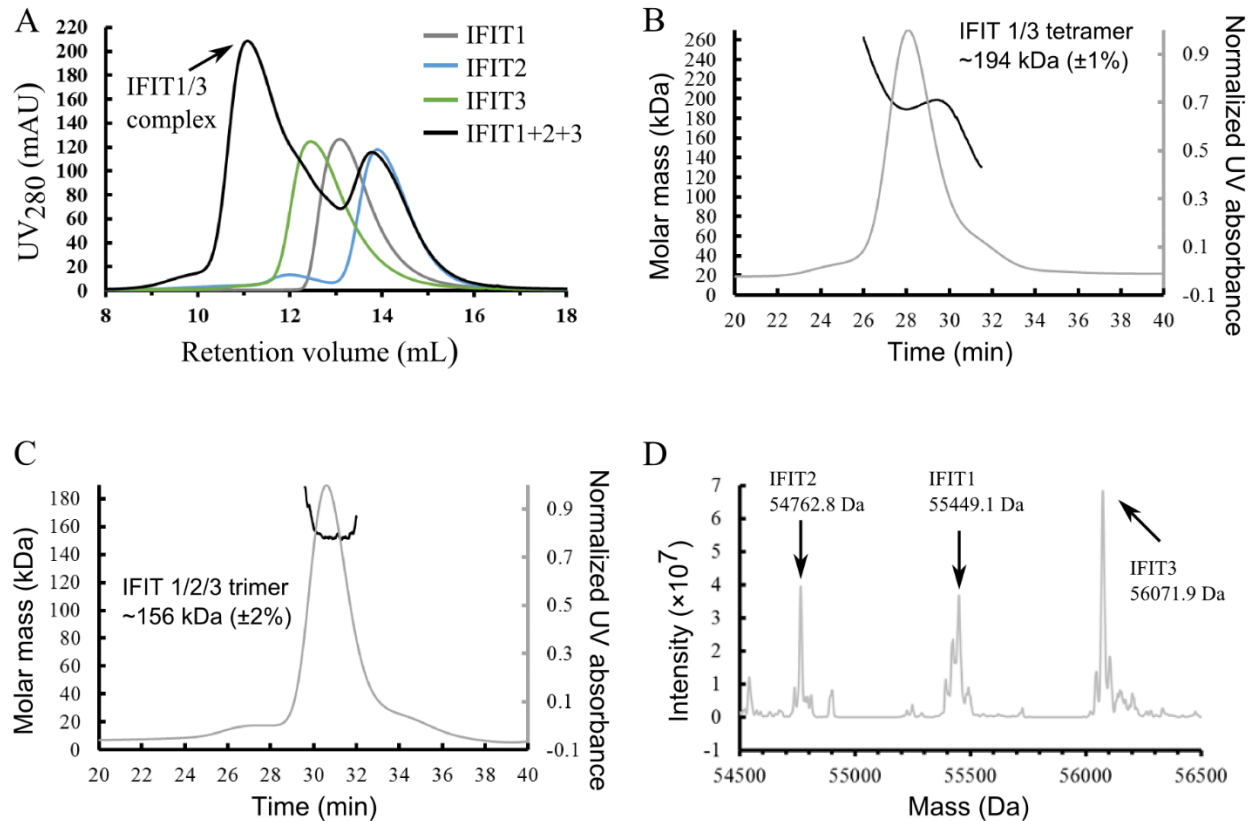
4.6 Supporting information



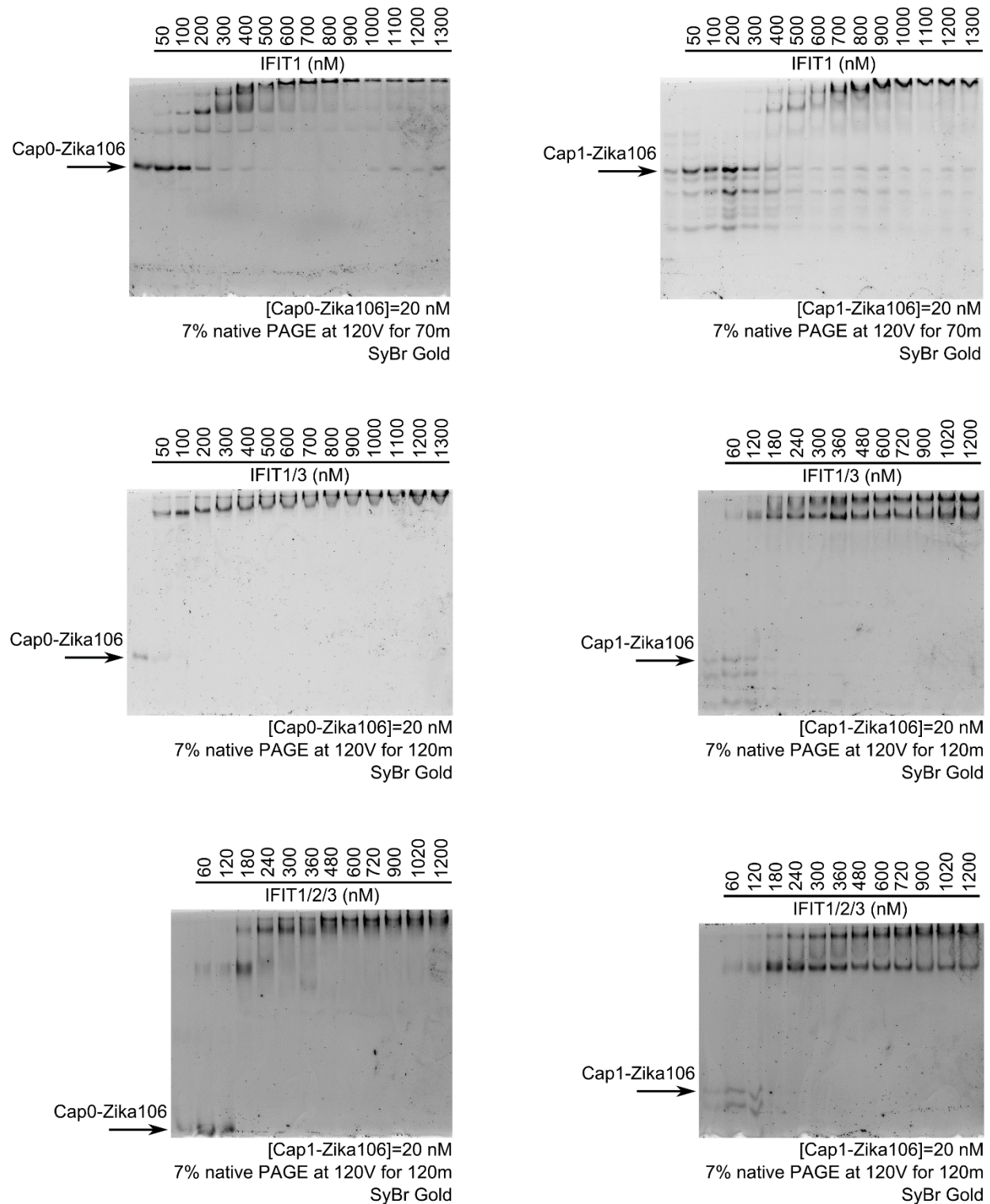
Supplementary Figure 4.1 The mass spectrometry analysis of the Cap0-, Cap1-, and Cap2-RNA used for the ITC experiments in **Figure 4.1**.



Supplementary Figure 4.2 The difference electron density maps before model building (left) and the all featured electron density maps after building RNA models (right), calculated from the diffraction data of IFIT1 DM with Cap0- and Cap1-RNA, respectively. Arrows indicating the position of 2'O and 2'O methylation.



Supplementary Figure 4.3 Characterization of IFIT1/3 and IFIT1/2/3 complex. (A) Size exclusion chromatography of individually purified IFIT1, IFIT2, and IFIT3, as well as the mixing of three protein together, indicating only IFIT1 and IFIT3 can form stable complex in solution. (B) Molecular weight measurement of IFIT1/3 complex by size exclusion chromatography coupled with multi-angle light scattering (SEC-MALS). (C) Molecular weight measurement of IFIT1/2/3 complex by SEC-MALS. (D) Mass spectrometry detected the presence of all three proteins in the IFIT1/2/3 complex.



Supplementary Figure 4.4 Titration the binding of Zika Cap0- and Cap1-5'UTR with IFIT1, IFIT1/3, and IFIT1/2/3, respectively.

	IFIT1 DM with Cap0- GGUAGAAUAUU	IFIT1 DM with Cap1- GGUAGAAUAUU
Data collection		
Space group	P4222	P4222
Cell dimensions		
a, b, c (Å)	112.1, 112.1, 92.8	111.2, 111.2, 92.8
α , β , γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	112.1-2.4	111.2-2.3
I/ σ I	8.0 (1.5)	9.5 (2.1)
CC1/2	0.804	0.844
Completeness (%)	100 (99.8)	98.4 (96.6)
R _{meas}	0.308	0.262
R _{pim}	0.061	0.051
Multiplicity	25.8	26.4
Refinement		
Resolution (Å)	112.1-2.4	111.2-2.3
No. reflections	23764	26172
Rwork/Rfree	0.198/0.247	0.190/0.249
No. atoms		
Protein chain A	7450	7456
RNA chain B	169	173
Solvent	58	51
B-factors		
Protein chain A	34.6	40.7
RNA chain B	23.7	34.2
Solvent	33.4	39.7
RMS Bond lengths (Å)	0.010	0.014
RMS Bond angles (°)	1.93	2.35

Ramachandran outliers	0	0
Ramachandran favoured	98.91%	98.47%
Rotamer outliers	4.96%	9.18%
Clashscore	2.21	2.99
Molprobity score	1.53	1.82
Deposition		
PDB ID		

Supplementary Table 4.1 X-ray data collection and refinement statistics.

Chapter 5 General discussion and outlook

5.1 Conclusions

The work presented in this dissertation centers on exploring the antiviral defense mechanisms of IFITs in human innate immunity. The thesis consists of three main chapters: Chapter 2 introduces a method developed for generating short and homogeneous 5'PPP- or Capped RNA required for structural studies; Chapter 3 and Chapter 4 extensively explain novel characterizations of the biological function of IFIT2 and IFIT complex in action toward viral RNA through biochemical, structural and functional analysis.

In Chapter 2, I developed a new purification method to obtain homogeneous short RNAs. For 5' monophosphate or hydroxyl RNA, chemical synthesis is routinely used to make up to ~200 nucleotides at an affordable cost. However, for 5'PPP-RNA, the synthetic method has not been successful due to poor yield, hazardous chemicals required for synthesis, and the degradation of the triphosphate group under the synthetic reaction conditions^{144, 145}. Therefore, to produce RNA with an intact 5'PPP structure, it is more appropriate to opt for in vitro transcription with T7 RNA polymerase. Nevertheless, the T7 bacteriophage RNA polymerase is well known to generate heterogeneous transcripts, either aborted or run-off transcripts. Traditionally, the incorporation of ribozyme sequence into the DNA template used for transcription or gel purification is the option to obtain the homogenous transcript. But for short RNA with less than ~15 nucleotides, the available methods are inefficient or unsatisfactory. The method developed in Chapter 1 is simple, efficient, and effective to obtain homogeneous short RNA. In addition, capped RNA (Cap0-, Cap1-, Cap2-) production is also challenging for chemical synthesis; Cap1- and Cap2-RNA have to be modified enzymatically using specific methyltransferase. Cap1 or Cap2 is an essential feature for

host mRNA to avoid autoimmune response or for mRNA vaccine to suppress the innate immune response. Even though capping enzyme is commercially available, the relatively high cost hinders extensive usage for large-scale production of capped RNA. Therefore, I re-engineered the tag for a vaccinia capping enzyme construct and established a detailed purification protocol for generating high-quality capping enzymes with close to complete capping efficiency. These tools described in Chapter 2, the improved capping enzyme purification protocol, the detailed purification protocol to make high-quality T7 RNA polymerase, and the short RNA purification method, are powerful workhorses for RNA synthesis, modification, and purification. Short RNAs are suitable for structural studies owing to less flexibility compared with longer RNAs; however, they may not reflect fully how RNA binding proteins act in cells due to the limited length, as seen from the example of an incomplete picture of the mechanism of IFIT2 by using short RNA in Chapter 3.

In Chapter 3, I first used biochemical assay to clarify a misleading assumption on the RNA binding property of IFIT2: IFIT2 does not bind to dsRNA but rather binds to ssRNA. Guided by the discovery of an AU-rich sequence motif interacted with IFIT2 by Tran and co-workers⁹³, I solved the crystal structure of AU-rich RNA-bound IFIT2. The structure provides insight into how each monomer in IFIT2 independently sequesters short uncapped RNA, but it cannot explain the biological function of IFIT2 domain-swapped dimerization since each monomer acts as an independent RNA binder. Later a virus-derived relatively long RNA unfolds the functional role of IFIT2 dimerization. It is seen in the cryo-EM model of IFIT2 dimer binding to a viral mRNA 5' UTR segment: both monomers in IFIT2 cooperate for clamping along the RNA chain, which cannot be accomplished by either monomer alone. The cryo-EM model most likely represents the real picture of how IFIT2 binds to mRNA in cells. The crystal structure rather provides an incomplete picture of the RNA binding mechanism of IFIT2. Unlike the sequestration RNA

binding mode seen in the RNA-bound IFIT1 and IFIT5 crystal structures, IFIT2 shows a distinct binding mode by utilizing both monomers to clamp on the mRNA. By doing so, IFIT2, along with some other unknown cellular factors, enhances the translation of bound mRNA. These studies suggest that IFIT2 acts like a double-edged sword: it can enhance the translation of host defense effectors to conquer the invading viral pathogens, conversely, viral intruders can also take advantage of IFIT2 to boost viral protein production to counteract host defense forces. Another question awaits to be answered regarding the viral inhibition effect observed on IFIT2, either in IFIT2 knockdown cells or *ifit2* knockout mice compared with the wild-type control during viral infection. Would the IFIT2-exerted viral inhibition directly come from sequestering the 5'PPP viral genome replication intermediates¹⁴⁶, or indirectly come from the enhanced translation of host effector proteins such as IFIT1, IFIT3, and even IFIT2 itself?

In Chapter 4, I first described the effect of 2'O methylation and secondary structure at viral RNA 5' end on IFIT1 binding. The ITC measurements give a very quantitative comparison of the RNA binding affinity to IFIT1 influenced only by the 2'O-methylation status (Cap0-, Cap1-, and Cap2-RNA). The IFIT1-Cap2-RNA binding affinity is ~9 fold weaker than that of IFIT1-Cap1-RNA, suggesting Cap2 methylation is a protection layer on host mRNA from autoimmune response by host innate immune sensors and effectors such as MAD5 and IFIT1; this result further supports the gradually enriched Cap2 mechanism on cellular mRNA discovered by Despici and Jaffrey⁸⁰. With the lesson learned in Chapter 3, short linear RNA *versus* long mRNA with complex secondary structure elements, the binding affinity estimated from the biochemical binding assay using viral mRNA segments, such as Zika 5'UTR and IAV NP 5'UTR, are more realistic than affinities quantified using short artificial RNA sequence. The gel binding assay using the Zika 5'UTR sequence shows how IFITs interact with a piece of viral RNA with different 5' end

modifications, demonstrating the protective means viruses utilize to evade IFIT1 action. The first crystal structure of Cap1-RNA bound IFIT1 shows that IFIT1 slightly adapts to interact with Cap1-RNA; a collective subtle changes in electrostatic, H-bonding, and van der Waals interactions eventually results in a significant decrease in binding affinity, by comparing with the crystal structure of the same RNA bound IFIT1 except in Cap0 form. Even though Cap1-viral RNA weakens IFIT1 binding, the IFIT complex synergistically binds to Cap1-viral RNA tighter than IFIT1 alone. The stronger RNA binding by IFIT complex observed on gel binding assay is clearly explained by our cryo-EM model of IAV NP 5'UTR bound IFIT1/2/3 complex. For the first time, the cryo-EM structure reveals the assembly of the multi-protein IFIT complex and the elegant manner of capturing a 22-nt viral RNA segment by the complex. IFIT1 sequesters the 5' end of the RNA, IFIT2 anchors the RNA chain outside the IFIT1 binding pocket to further stabilize the RNA binding, and IFIT1 and IFIT2 act together to clamp the RNA chain as the last step of capturing the viral RNA. In the IFIT1/2/3 complex, IFIT1 acts as a primary binder and its RNA binding is enhanced through IFIT3 modulation⁵⁷; IFIT2 acts as a secondary binder to stabilize the RNA chain stretching out of the IFIT1 binding pocket. Once again, if a short RNA is used in the cryo-EM study, we cannot fully understand the auxiliary role of IFIT2 in RNA binding. Together with Chapter 3, we see a paradoxical function role of IFIT2: the primary RNA binding role of IFIT2 in the absence of family members to enhance translation of bound transcript, the auxiliary RNA binding role of IFIT2 in the complex to inhibit translation of bound viral transcript. Viruses commonly antagonize IFIT1 action by Cap1 methylation and secondary structure elements at the 5' end, nevertheless, our binding assays and cryo-EM model affirm, to some extent, that IFIT1/2/3 complex could overcome the counteracting measure that some viral Cap1-mRNAs potentially

subvert the combat by IFIT1 alone. It is not the end of the story. What is the regulatory mechanism for cells to make individual IFITs or to assemble IFIT1/2/3 complex in response to a viral infection?

5.2 Future work

The findings made in this dissertation complement previous discoveries about the antiviral functions of the IFIT family and deepen our understanding of IFIT functions. However, there are still a few questions worthy to be explored in the future from the following aspects:

5.2.1 Function of IFIT5 needs to be redefined

The crystal structure of 5'PPP-RNA bound IFIT5 likely revealed only one type of varied RNA targets. Future work is required to understand the adaptable RNA binding function of IFIT5, especially that human IFIT5 is neither stimulated by interferon treatment nor upregulated in common viral infections, including Influenza and corona viruses⁸⁷. Therefore, it is inappropriate to classify IFIT5 as a member in ISG family, since it's not induced by interferon signaling pathway. As an RNA binding protein, with high affinity binding adapted to act on varied forms of RNA as seen in Chapter 3 of this thesis, Cap0/Cap1-, 5'PPP-, 5'P-, and 5'HO-RNA, how IFIT5 regulates the function of bound transcripts still awaits to be answered.

5.2.2 Molecular basis for the IFIT2 involved translation enhancement

Tran *et al.* showed that IFIT2 promotes translation by reducing ribosome pausing based on the ribosome profiling experiment⁹³. In agreement, another study showed that IFIT2 preferred binding to 3' UTR of endogenous mRNAs. The IFIT2-RNA interaction stabilized RNA and enhanced its translation under normal conditions without either interferon treatment or viral infection¹⁴⁷. Nevertheless, neither of the studies could explain deeper about how IFIT2 reduces ribosome

pausing or stabilizes the bound transcript by interacting with RNA. Our in vitro translation system data in Chapter 3 clearly showed that IFIT2 alone cannot promote the translation of the bound transcript, which suggests other co-factors are essential to stabilize the transcript and enhance its translation together with IFIT2. What are the other factors required to promote translation? Do they work collectively through the connection with RNA, or do they form a multiple-protein complex with IFIT2 to modulate IFIT2-RNA binding, or do they not only bind to RNA but also interact with IFIT2 as part of large ribonucleoprotein complex? Future work using immunoprecipitation and proteomics-based approaches is necessary to discover the multi-factor involved in translation enhancement machinery.

5.2.3 The destiny of IFIT-bound transcript

IFITs capture viral RNA either individually or collectively to stop translation of viral proteins. But what's the fate/destination of IFIT-bound RNA? The protein-RNA binding is a dynamic event, constantly in an association/dissociation equilibrium. As long as the viral RNA is not destroyed, it is always a potential threat: a chance to escape from IFITs and get translated by surrounding ribosomes. Our preliminary nuclease digestion assay showed IFIT1 and IFIT5 cannot protect the bound RNA from the 3'-5' exonuclease digestion and the 5'-3' exonuclease XRN-1 digestion, even though the 5' end is sequestered by IFIT1 and IFIT5. This result frees our concern that RNA binding by IFITs could rather become a protection for viral RNA from exonucleases in cells. However, an exception is IFIT2. IFIT2 protects RNA from exonuclease digestion, which confirms its translation-stimulated function and once again points to a distinguished RNA binding function from the other two family members. Despite exonucleases being present in cell cytoplasm to degrade unwanted RNA as part of RNA quality control and waste processing, IFIT2, somehow, protects the bound transcript and ensures the bound transcript to go for an efficient translation with

less pause. Investigation on degradation of IFIT1-bound viral RNA also requires further work to see whether direction interactions exist between IFIT1 and exonucleases, such that the IFIT1-bound viral RNA is efficiently released to exonuclease for degradation to free out IFIT1 for targeting remaining viral RNA.

We have just recovered from the Covid-19 pandemic and we don't know when and what will be the next global viral outbreak. Our innate immune system, as the front line of host defense, is critical to protect humans from infectious diseases. The deeper we understand the innate immune system, the better we are prepared for the next worldwide infectious disease to save lives especially for vulnerable people, through therapeutic interventions to reactivate the dysfunctional innate immune defense system.

References

1. Janeway, C. J.; Travers, P.; Walport, M.; Shlomchik, M., *Immunobiology: The Immune System in Health and Disease*. New York, Garland Science: 2001.
2. World Health Organization, *Schistosomiasis and soil-transmitted helminthiases: progress report 2021*.
3. World Health Organization, *World Malaria Report*; 2022.
4. WADMAN, M. *First malaria vaccine slashes early childhood mortality*; 2023.
5. WHO fungal priority pathogens list to guide research, development and public health action; 2022.
6. Zhao, Y.; Ye, L.; Zhao, F.; Zhang, L.; Lu, Z.; Chu, T.; Wang, S.; Liu, Z.; Sun, Y.; Chen, M.; Liao, G.; Ding, C.; Xu, Y.; Liao, W.; Wang, L., Cryptococcus neoformans, a global threat to human health. *Infectious Diseases of Poverty* **2023**, 12 (1), 20.
7. WHO fact sheets: tuberculosis; 2022.
8. Viboud, C.; Simonsen, L.; Fuentes, R.; Flores, J.; Miller, M. A.; Chowell, G., Global Mortality Impact of the 1957–1959 Influenza Pandemic. *The Journal of Infectious Diseases* **2016**, 213 (5), 738-745.
9. Peckham, R., Viral surveillance and the 1968 Hong Kong flu pandemic. *Journal of Global History* **2020**, 15 (3), 444-458.
10. Gregg, M.; Hinman, A.; Craven, R., The Russian flu. Its history and implications for this year's influenza season. *JAMA* **1978**, 240, 4.
11. Dawood, F. S.; Iuliano, A. D.; Reed, C.; Meltzer, M. I.; Shay, D. K.; Cheng, P.-Y.; Bandaranayake, D.; Breiman, R. F.; Brooks, W. A.; Buchy, P.; Feikin, D. R.; Fowler, K. B.; Gordon, A.; Hien, N. T.; Horby, P.; Huang, Q. S.; Katz, M. A.; Krishnan, A.; Lal, R.; Montgomery, J. M.; Mølbak, K.; Pebody, R.; Presanis, A. M.; Razuri, H.; Steens, A.; Tinoco, Y. O.; Wallinga, J.; Yu, H.; Vong, S.; Bresee, J.; Widdowson, M.-A., Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *The Lancet Infectious Diseases* **2012**, 12 (9), 687-695.
12. 2014-2016 Ebola Outbreak in West Africa; <https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.html>.
13. de Oliveira, W. K.; de França, G. V. A.; Carmo, E. H.; Duncan, B. B.; de Souza Kuchenbecker, R.; Schmidt, M. I., Infection-related microcephaly after the 2015 and 2016 Zika virus outbreaks in Brazil: a surveillance-based analysis. *The Lancet* **2017**, 390 (10097), 861-870.
14. Hu, B.; Guo, H.; Zhou, P.; Shi, Z.-L., Characteristics of SARS-CoV-2 and COVID-19. *Nature Reviews Microbiology* **2021**, 19 (3), 141-154.
15. Temmam, S.; Vongphayloth, K.; Baquero, E.; Munier, S.; Bonomi, M.; Regnault, B.; Douangboubpha, B.; Karami, Y.; Chrétien, D.; Sanamxay, D.; Xayaphet, V.; Paphaphanh, P.; Lacoste, V.; Somlor, S.; Lakeomany, K.; Phommavanh, N.; Pérot, P.; Dehan, O.; Amara, F.; Donati, F.; Bigot, T.; Nilges, M.; Rey, F. A.; van der Werf, S.; Brey, P. T.; Eloit, M., Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. *Nature* **2022**, 604 (7905), 330-336.
16. WHO Coronavirus (COVID-19) Dashboard; <https://covid19.who.int/>.
17. *Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses*; 2012.

18. dsDNA(RT) replication/transcription; <https://viralzone.expasy.org/1938>.
19. WHO Influenza (Seasonal); 2023.
20. The Baltimore classification clusters viruses into families depending on their type of genome.; <https://viralzone.expasy.org/254>.
21. Kaihatsu, K.; Yamabe, M.; Ebara, Y., Antiviral Mechanism of Action of Epigallocatechin-3-O-gallate and Its Fatty Acid Esters. *Proceedings of the National Academy of Sciences* **2018**, 23 (10), 2475.
22. Hepatitis B; <https://www.who.int/news-room/fact-sheets/detail/hepatitis-b>.
23. WHO HIV and AIDS; 2023.
24. National Cancer Institute HPV and Cancer. **2020**.
25. Du, Y.; Chen, C.; Zhang, X.; Yan, D.; Jiang, D.; Liu, X.; Yang, M.; Ding, C.; Lan, L.; Hecht, R.; Zhu, C.; Yang, S., Global burden and trends of rotavirus infection-associated deaths from 1990 to 2019: an observational trend study. *Virology Journal* **2022**, 19 (1), 166.
26. Why is Dengue a Global Issue? **Centers for Disease Control and Prevention, USA**.
27. Samji, T., Influenza A: Understanding the Viral Life Cycle. *Yale Journal of Biology and Medicine* **2009**, 82, 153–159.
28. Lamb, R. A.; Choppin, P. W., THE GENE STRUCTURE AND REPLICATION OF INFLUENZA VIRUS. *Annual Review of Biochemistry* **1983**, 52 (1), 467-506.
29. Malone, B.; Urakova, N.; Snijder, E. J.; Campbell, E. A., Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Reviews Molecular Cell Biology* **2022**, 23 (1), 21-39.
30. Hepatitis A; <https://www.who.int/news-room/fact-sheets/detail/hepatitis-a>.
31. Hepatitis C. <https://www.who.int/news-room/fact-sheets/detail/hepatitis-c>.
32. den Boon, J. A.; Diaz, A.; Ahlquist, P., Cytoplasmic Viral Replication Complexes. *Cell Host & Microbe* **2010**, 8 (1), 77-85.
33. Schwartz, M.; Chen, J.; Janda, M.; Sullivan, M.; den Boon, J.; Ahlquist, P., A Positive-Strand RNA Virus Replication Complex Parallels Form and Function of Retrovirus Capsids. *Molecular Cell* **2002**, 9 (3), 505-514.
34. Kopek, B. G.; Perkins, G.; Miller, D. J.; Ellisman, M. H.; Ahlquist, P., Three-Dimensional Analysis of a Viral RNA Replication Complex Reveals a Virus-Induced Mini-Organelle. *PLOS Biology* **2007**, 5 (9), e220.
35. Knoops, K.; Kikkert, M.; Worm, S. H. E. v. d.; Zevenhoven-Dobbe, J. C.; van der Meer, Y.; Koster, A. J.; Mommaas, A. M.; Snijder, E. J., SARS-Coronavirus Replication Is Supported by a Reticulovesicular Network of Modified Endoplasmic Reticulum. *PLOS Biology* **2008**, 6 (9), e226.
36. Wolff, G.; Limpens, R. W. A. L.; Zevenhoven-Dobbe, J. C.; Laugks, U.; Zheng, S.; de Jong, A. W. M.; Koning, R. I.; Agard, D. A.; Grünewald, K.; Koster, A. J.; Snijder, E. J.; Bárcena, M., A molecular pore spans the double membrane of the coronavirus replication organelle. *Science* **2020**, 369 (6509), 1395-1398.
37. Takeuchi, O.; Akira, S., Recognition of viruses by innate immunity. *Immunological Reviews* **2007**, 220 (1), 214-224.
38. Schoggins, J. W.; Rice, C. M., Interferon-stimulated genes and their antiviral effector functions. *Current Opinion in Virology* **2011**, 1 (6), 519-525.
39. Paladino, P.; Cummings, D. T.; Noyce, R. S.; Mossman, K. L., The IFN-Independent Response to Virus Particle Entry Provides a First Line of Antiviral Defense That Is

- Independent of TLRs and Retinoic Acid-Inducible Gene I1. *The Journal of Immunology* **2006**, *177* (11), 8008-8016.
40. Der, S. D.; Zhou, A.; Williams, B. R. G.; Silverman, R. H., Identification of genes differentially regulated by interferon α , β , or γ using oligonucleotide arrays. *Proceedings of the National Academy of Sciences* **1998**, *95* (26), 15623.
 41. Chebath, J.; Merlin, G.; Metz, R.; Benech, P.; Revel, M., Interferon-induced 56,000 Mr protein and its mRNA in human cells: molecular cloning and partial sequence of the cDNA. *Nucleic Acids Research* **1983**, *11* (5), 1213-1226.
 42. Levy, D.; Larner, A.; Chaudhuri, A.; Babiss, L. E.; Darnell, J. E., Interferon-stimulated transcription: isolation of an inducible gene and identification of its regulatory region. *Proceedings of the National Academy of Sciences* **1986**, *83* (23), 8929-8933.
 43. Yu, M.; Tong, J.-H.; Mao, M.; Kan, L.-X.; Liu, M.-M.; Sun, Y.-W.; Fu, G.; Jing, Y.-K.; Yu, L.; Lepaslier, D.; Lanotte, M.; Wang, Z.-Y.; Chen, Z.; Waxman, S.; Wang, Y.-X.; Tan, J.-Z.; Chen, S.-J., Cloning of a gene (RIG-G) associated with retinoic acid-induced differentiation of acute promyelocytic leukemia cells and representing a new member of a family of interferon-stimulated genes. *Proceedings of the National Academy of Sciences* **1997**, *94* (14), 7406-7411.
 44. Niikura, T.; Hirata, R.; Weil, S. C., A Novel Interferon-Inducible Gene Expressed during Myeloid Differentiation. *Blood Cells, Molecules, and Diseases* **1997**, *23* (3), 337-349.
 45. Perez-Riba, A.; Itzhaki, L. S., The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition. *Current Opinion in Structural Biology* **2019**, *54*, 43-49.
 46. Abbas, Y. M.; Pichlmair, A.; Górna, M. W.; Superti-Furga, G.; Nagar, B., Structural basis for viral 5'-PPP-RNA recognition by human IFIT proteins. *Nature* **2013**, *494* (7435), 60-64.
 47. Terenzi, F.; Hui, D. J.; Merrick, W. C.; Sen, G. C., Distinct Induction Patterns and Functions of Two Closely Related Interferon-inducible Human Genes, ISG54 and ISG56. *JOURNAL OF BIOLOGICAL CHEMISTRY* **2006**, *281* (45), 34064-34071.
 48. Pichlmair, A.; Lassnig, C.; Eberle, C.-A.; Górna, M. W.; Baumann, C. L.; Burkard, T. R.; Bürckstümmer, T.; Stefanovic, A.; Krieger, S.; Bennett, K. L.; Rülcke, T.; Weber, F.; Colinge, J.; Müller, M.; Superti-Furga, G., IFIT1 is an antiviral protein that recognizes 5'-triphosphate RNA. *Nature Immunology* **2011**, *12* (7), 624-630.
 49. Abbas, Y. M.; Laudénbach, B. T.; Martínez-Montero, S.; Cencic, R.; Habjan, M.; Pichlmair, A.; Damha, M. J.; Pelletier, J.; Nagar, B., Structure of human IFIT1 with capped RNA reveals adaptable mRNA binding and mechanisms for sensing N1 and N2 ribose 2'-O methylations. *Proceedings of the National Academy of Sciences* **2017**, *114* (11), E2106-E2115.
 50. Cho, H.; Shrestha, B.; Sen, G. C.; Diamond, M. S., A Role for Ifit2 in Restricting West Nile Virus Infection in the Brain. *Journal of Virology* **2013**, *87* (15), 8363-8371.
 51. Butchi, N. B.; Hinton, D. R.; Stohlman, S. A.; Kapil, P.; Fensterl, V.; Sen, G. C.; Bergmann, C. C., Ifit2 Deficiency Results in Uncontrolled Neurotropic Coronavirus Replication and Enhanced Encephalitis via Impaired Alpha/Beta Interferon Induction in Macrophages. *Journal of Virology* **2014**, *88* (2), 1051-1064.
 52. Fensterl, V.; Wetzel, J. L.; Ramachandran, S.; Ogino, T.; Stohlman, S. A.; Bergmann, C. C.; Diamond, M. S.; Virgin, H. W.; Sen, G. C., Interferon-Induced Ifit2/ISG54 Protects Mice from Lethal VSV Neuropathogenesis. *PLOS Pathogens* **2012**, *8* (5), e1002712.
 53. Wetzel, J. L.; Fensterl, V.; Sen, G. C., Sendai Virus Pathogenesis in Mice Is Prevented by Ifit2 and Exacerbated by Interferon. *Journal of Virology* **2014**, *88* (23), 13593-13601.

54. Davis, B. M.; Fensterl, V.; Lawrence, T. M.; Hudacek, A. W.; Sen, G. C.; Schnell, M. J., Ifit2 Is a Restriction Factor in Rabies Virus Pathogenicity. *Journal of Virology* **2017**, *91* (17), e00889-17.
55. Yang, Z.; Liang, H.; Zhou, Q.; Li, Y.; Chen, H.; Ye, W.; Chen, D.; Fleming, J.; Shu, H.; Liu, Y., Crystal structure of ISG54 reveals a novel RNA binding structure and potential functional mechanisms. *Cell Research* **2012**, *22* (9), 1328-1338.
56. Fleith, R. C.; Mears, H. V.; Leong, X. Y.; Sanford, T. J.; Emmott, E.; Graham, S. C.; Mansur, D. S.; Sweeney, T. R., IFIT3 and IFIT2/3 promote IFIT1-mediated translation inhibition by enhancing binding to non-self RNA. *Nucleic Acids Research* **2018**, *46* (10), 5269-5285.
57. Johnson, B.; VanBlargan, L. A.; Xu, W.; White, J. P.; Shan, C.; Shi, P.-Y.; Zhang, R.; Adhikari, J.; Gross, M. L.; Leung, D. W.; Diamond, M. S.; Amarasinghe, G. K., Human IFIT3 Modulates IFIT1 RNA Binding Specificity and Protein Stability. *Immunity* **2018**, *48* (3), 487-499.e5.
58. Stawowczyk, M.; Van Scoy, S.; Kumar, K. P.; Reich, N. C., The Interferon Stimulated Gene 54 Promotes Apoptosis. *Journal of Biological Chemistry* **2011**, *286* (9), 7257-7266.
59. Zhu, Y.; Chen, G.; Lv, F.; Wang, X.; Ji, X.; Xu, Y.; Sun, J.; Wu, L.; Zheng, Y.-T.; Gao, G., Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. *Proceedings of the National Academy of Sciences* **2011**, *108* (38), 15834-15839.
60. Takata, M. A.; Gonçalves-Carneiro, D.; Zang, T. M.; Soll, S. J.; York, A.; Blanco-Melo, D.; Bieniasz, P. D., CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature* **2017**, *550* (7674), 124-127.
61. Meagher, J. L.; Takata, M.; Gonçalves-Carneiro, D.; Keane, S. C.; Rebendenne, A.; Ong, H.; Orr, V. K.; MacDonald, M. R.; Stuckey, J. A.; Bieniasz, P. D.; Smith, J. L., Structure of the zinc-finger antiviral protein in complex with RNA reveals a mechanism for selective targeting of CG-rich viral sequences. *Proceedings of the National Academy of Sciences* **2019**, *116* (48), 24303-24309.
62. Garneau, N. L.; Wilusz, J.; Wilusz, C. J., The highways and byways of mRNA decay. *Nature Reviews Molecular Cell Biology* **2007**, *8* (2), 113-126.
63. Liu, Q.; Greimann, J. C.; Lima, C. D., Reconstitution, Activities, and Structure of the Eukaryotic RNA Exosome. *Cell* **2006**, *127* (6), 1223-1237.
64. Bonneau, F.; Basquin, J.; Ebert, J.; Lorentzen, E.; Conti, E., The Yeast Exosome Functions as a Macromolecular Cage to Channel RNA Substrates for Degradation. *Cell* **2009**, *139* (3), 547-559.
65. Lykke-Andersen, J., Identification of a Human Decapping Complex Associated with hUpf Proteins in Nonsense-Mediated Decay. *Molecular and Cellular Biology* **2002**, *22* (23), 8114-8121.
66. Stevens, A., 5'-Exoribonuclease 1: Xrn1. In *Methods in Enzymology*, Nicholson, A. W., Ed. Academic Press: 2001; Vol. 342, pp 251-259.
67. Staeheli, P.; Haller, O.; Boll, W.; Lindenmann, J.; Weissmann, C., Mx protein: Constitutive expression in 3T3 cells transformed with cloned Mx cDNA confers selective resistance to influenza virus. *Cell* **1986**, *44* (1), 147-158.
68. Schoggins, J. W., Interferon-Stimulated Genes: What Do They All Do? *Annual Review of Virology* **2019**, *6* (1), 567-584.

69. McDougal, M. B.; De Maria, A. M.; Ohlson, M. B.; Kumar, A.; Xing, C.; Schoggins, J. W., Interferon inhibits a model RNA virus via a limited set of inducible effector genes. *EMBO reports* **2023**, e56901.
70. Itell, H. L.; Humes, D.; Overbaugh, J., Several cell-intrinsic effectors drive type I interferon-mediated restriction of HIV-1 in primary CD4⁺ T cells. *Cell Reports* **2023**, 42 (6).
71. Human MxA Protein: An Interferon-Induced Dynamin-Like GTPase with Broad Antiviral Activity. **2011**, 31 (1), 79-87.
72. Evans, D. T.; Serra-Moreno, R.; Singh, R. K.; Guatelli, J. C., BST-2/tetherin: a new component of the innate immune response to enveloped viruses. *Trends in Microbiology* **2010**, 18 (9), 388-396.
73. Gizzi, A. S.; Grove, T. L.; Arnold, J. J.; Jose, J.; Jangra, R. K.; Garforth, S. J.; Du, Q.; Cahill, S. M.; Dulyaninova, N. G.; Love, J. D.; Chandran, K.; Bresnick, A. R.; Cameron, C. E.; Almo, S. C., A naturally occurring antiviral ribonucleotide encoded by the human genome. *Nature* **2018**, 558 (7711), 610-614.
74. Spence, J. S.; He, R.; Hoffmann, H.-H.; Das, T.; Thinon, E.; Rice, C. M.; Peng, T.; Chandran, K.; Hang, H. C., IFITM3 directly engages and shuttles incoming virus particles to lysosomes. *Nature Chemical Biology* **2019**, 15 (3), 259-268.
75. Deymier, S.; Louvat, C.; Fiorini, F.; Cimorelli, A., ISG20: an enigmatic antiviral RNase targeting multiple viruses. *FEBS openbio* **2022**, 12 (6), 1096-1111.
76. Richardson, R. B.; Ohlson, M. B.; Eitson, J. L.; Kumar, A.; McDougal, M. B.; Boys, I. N.; Mar, K. B.; De La Cruz-Rivera, P. C.; Douglas, C.; Konopka, G.; Xing, C.; Schoggins, J. W., A CRISPR screen identifies IFI6 as an ER-resident interferon effector that blocks flavivirus replication. *Nature Microbiology* **2018**, 3 (11), 1214-1223.
77. Leppek, K.; Das, R.; Barna, M., Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nature Reviews Molecular Cell Biology* **2018**, 19 (3), 158-174.
78. Decroly, E.; Ferron, F.; Lescar, J.; Canard, B., Conventional and unconventional mechanisms for capping viral mRNA. *Nature Reviews Microbiology* **2012**, 10 (1), 51-65.
79. Gebhardt, A.; Laudenbach, B. T.; Pichlmair, A., Discrimination of Self and Non-Self Ribonucleic Acids. *JOURNAL OF INTERFERON & CYTOKINE RESEARCH* **2017**, 37 (5), 184-197.
80. Despic, V.; Jaffrey, S. R., mRNA ageing shapes the Cap2 methylome in mammalian mRNA. *Nature* **2023**, 614 (7947), 358-366.
81. Hill, C. H.; Brierley, I., Structural and Functional Insights into Viral Programmed Ribosomal Frameshifting. *Annual Review of Virology* **2023**, 10, 217-242.
82. Li, P.; Wei, Y.; Mei, M.; Tang, L.; Sun, L.; Huang, W.; Zhou, J.; Zou, C.; Zhang, S.; Qin, C.-F.; Jiang, T.; Dai, J.; Tan, X.; Zhang, Q. C., Integrative Analysis of Zika Virus Genome RNA Structure Reveals Critical Determinants of Viral Infectivity. *Cell Host & Microbe* **2018**, 24 (6), 875-886.e5.
83. Miao, Z.; Tidu, A.; Eriani, G.; Martin, F., Secondary structure of the SARS-CoV-2 5'-UTR. *RNA Biology* **2020**, 1-10.
84. Quade, N.; Boehringer, D.; Leibundgut, M.; van den Heuvel, J.; Ban, N., Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-Å resolution. *Nature Communications* **2015**, 6 (1), 7646.

85. Fernandez, J.; Yaman, I.; Sarnow, P.; Snider, M. D.; Hatzoglou, M., Regulation of Internal Ribosomal Entry Site-mediated Translation by Phosphorylation of the Translation Initiation Factor eIF2. *Journal of Biological Chemistry* **2002**, 277 (21), 19198-19205.
86. Thoms, M.; Buschauer, R.; Ameismeier, M.; Koepke, L.; Denk, T.; Hirschenberger, M.; Kratzat, H.; Hayn, M.; Mackens-Kiani, T.; Cheng, J.; Straub, J. H.; Stürzel, C. M.; Fröhlich, T.; Berninghausen, O.; Becker, T.; Kirchhoff, F.; Sparrer, K. M. J.; Beckmann, R., Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* **2020**, 369 (6508), 1249-1255.
87. Blanco-Melo, D.; Nilsson-Payant, B. E.; Liu, W.-C.; Uhl, S.; Hoagland, D.; Möller, R.; Jordan, T. X.; Oishi, K.; Panis, M.; Sachs, D.; Wang, T. T.; Schwartz, R. E.; Lim, J. K.; Albrecht, R. A.; tenOever, B. R., Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell* **2020**, 181 (5), 1036-1045.e9.
88. Bhaskar, V.; Graff-Meyer, A.; Schenk, A. D.; Cavadini, S.; von Loeffelholz, O.; Natchiar, S. K.; Artus-Revel, C. G.; Hotz, H.-R.; Bretones, G.; Klaholz, B. P.; Chao, J. A., Dynamics of uS19 C-Terminal Tail during the Translation Elongation Cycle in Human Ribosomes. *Cell Reports* **2020**, 31 (1).
89. Cullen, B. R., MicroRNAs as mediators of viral evasion of the immune system. *Nature Immunology* **2013**, 14 (3), 205-210.
90. Machlin, E. S.; Sarnow, P.; Sagan, S. M., Masking the 5' terminal nucleotides of the hepatitis C virus genome by an unconventional microRNA-target RNA complex. *Proceedings of the National Academy of Sciences* **2011**, 108 (8), 3193-3198.
91. Pingale, K. D.; Kanade, G. D.; Karpe, Y. A., Hepatitis E virus polymerase binds to IFIT1 to protect the viral RNA from IFIT1-mediated translation inhibition. *Journal of General Virology* **2019**, 100 (3), 471-483.
92. Liu, R.; Moss, B., Vaccinia Virus C9 Ankyrin Repeat/F-Box Protein Is a Newly Identified Antagonist of the Type I Interferon-Induced Antiviral State. *Journal of Virology* **2018**, 92 (9), e00053-18.
93. Tran, V.; Ledwith, M. P.; Thamamongood, T.; Higgins, C. A.; Tripathi, S.; Chang, M. W.; Benner, C.; García-Sastre, A.; Schwemmle, M.; Boon, A. C. M.; Diamond, M. S.; Mehle, A., Influenza virus repurposes the antiviral protein IFIT2 to promote translation of viral mRNAs. *Nature Microbiology* **2020**, 5 (12), 1490-1503.
94. Hoagland, M. B.; Stephenson, M. L.; Scott, J. F.; Hecht, L. I.; Zamecnik, P. C., A SOLUBLE RIBONUCLEIC ACID INTERMEDIATE IN PROTEIN SYNTHESIS. *Journal of Biological Chemistry* **1958**, 231 (1), 241-257.
95. Serganov, A.; Patel, D. J., Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nature Reviews Genetics* **2007**, 8 (10), 776-790.
96. Morris, K. V.; Mattick, J. S., The rise of regulatory RNA. *Nature Reviews Genetics* **2014**, 15 (6), 423-437.
97. Fatica, A.; Bozzoni, I., Long non-coding RNAs: new players in cell differentiation and development. *Nature Reviews Genetics* **2014**, 15 (1), 7-21.
98. Nagai, K., RNA—protein complexes. *Current Opinion in Structural Biology* **1996**, 6 (1), 53-61.
99. Rossmann, M. G.; Morais, M. C.; Leiman, P. G.; Zhang, W., Combining X-Ray Crystallography and Electron Microscopy. *Structure* **2005**, 13 (3), 355-362.

100. Young, C. L.; Britton, Z. T.; Robinson, A. S., Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnology Journal* **2012**, 7 (5), 620-634.
101. Petrov, A.; Wu, T.; Puglisi, E. V.; Puglisi, J. D., Chapter Seventeen - RNA Purification by Preparative Polyacrylamide Gel Electrophoresis. In *Methods in Enzymology*, Lorsch, J., Ed. Academic Press: 2013; Vol. 530, pp 315-330.
102. Easton, L. E.; Shibata, Y.; Lukavsky, P. J., Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA* **2010**, 16 (3), 647-653.
103. Koubek, J.; Lin, K. F.; Chen, Y. R.; Cheng, R. P.; Huang, J. J. T., Strong anion-exchange fast performance liquid chromatography as a versatile tool for preparation and purification of RNA produced by in vitro transcription. *RNA* **2013**, 19 (10), 1449-1459.
104. Karlsson, H.; Baronti, L.; Petzold, K., A robust and versatile method for production and purification of large-scale RNA samples for structural biology. *RNA* **2020**, 26 (8), 1023-1037.
105. Martin, C. T.; Muller, D. K.; Coleman, J. E., Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry* **1988**, 27 (11), 3966-3974.
106. Walker, S. C.; Avis, J. M.; Conn, G. L., General plasmids for producing RNA in vitro transcripts with homogeneous ends. *Nucleic Acids Research* **2003**, 31 (15), e82-e82.
107. Durniak, K. J.; Bailey, S.; Steitz, T. A., The Structure of a Transcribing T7 RNA Polymerase in Transition from Initiation to Elongation. *Science* **2008**, 322 (5901), 553-557.
108. Fuchs, A.-L.; Neu, A.; Sprangers, R., A general method for rapid and cost-efficient large-scale production of 5' capped RNA. *RNA* **2016**, 22 (9), 1454-1466.
109. Kyrieleis, Otto J. P.; Chang, J.; de la Peña, M.; Shuman, S.; Cusack, S., Crystal Structure of Vaccinia Virus mRNA Capping Enzyme Provides Insights into the Mechanism and Evolution of the Capping Apparatus. *Structure* **2014**, 22 (3), 452-465.
110. Liu, H.; Naismith, J. H., An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology* **2008**, 8 (1), 91.
111. Diamond, M. S.; Farzan, M., The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nature Reviews Immunology* **2013**, 13 (1), 46-57.
112. Zuker, M., Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **2003**, 31 (13), 3406-3415.
113. Katibah, G. E.; Qin, Y.; Sidote, D. J.; Yao, J.; Lambowitz, A. M.; Collins, K., Broad and adaptable RNA structure recognition by the human interferon-induced tetratricopeptide repeat protein IFIT5. *Proceedings of the National Academy of Sciences* **2014**, 111 (33), 12025-12030.
114. Liu, X.-Y.; Chen, W.; Wei, B.; Shan, Y.-F.; Wang, C., IFN-Induced TPR Protein IFIT3 Potentiates Antiviral Signaling by Bridging MAVS and TBK1. *The Journal of Immunology* **2011**, 187 (5), 2559-2568.
115. Yang, Y.; Zhou, Y.; Hou, J.; Bai, C.; Li, Z.; Fan, J.; Ng, I. O. L.; Zhou, W.; Sun, H.; Dong, Q.; Lee, J. M. F.; Lo, C. M.; Man, K.; Yang, Y.; Li, N.; Ding, G.; Yu, Y.; Cao, X., Hepatic IFIT3 predicts interferon- α therapeutic response in patients of hepatocellular carcinoma. *HEPATOLOGY* **2017**, 66 (1), 152-166.
116. Daffis, S.; Szretter, K. J.; Schriewer, J.; Li, J.; Youn, S.; Errett, J.; Lin, T.-Y.; Schneller, S.; Zust, R.; Dong, H.; Thiel, V.; Sen, G. C.; Fensterl, V.; Klimstra, W. B.; Pierson, T. C.; Buller, R. M.; Gale Jr, M.; Shi, P.-Y.; Diamond, M. S., 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature* **2010**, 468 (7322), 452-456.

117. Abbas, Y. M.; Martínez-Montero, S.; Cencic, R.; Pelletier, J.; Pawelek, P. D.; Damha, M. J.; Nagar, B., A conserved homo-dimerization interface in human IFIT1 provides insights into IFIT interactome assembly. *bioRxiv* **2017**, 152850.
118. Habjan, M.; Hubel, P.; Lacerda, L.; Benda, C.; Holze, C.; Eberl, C. H.; Mann, A.; Kindler, E.; Gil-Cruz, C.; Ziebuhr, J.; Thiel, V.; Pichlmair, A., Sequestration by IFIT1 Impairs Translation of 2'O-unmethylated Capped RNA. *PLOS Pathogens* **2013**, *9* (10), e1003663.
119. Zhu, Z.; Yang, X.; Huang, C.; Liu, L., The Interferon-Induced Protein with Tetratricopeptide Repeats Repress Influenza Virus Infection by Inhibiting Viral RNA Synthesis. *Viruses* **2023**, *15* (7), 1412.
120. Schürer, H.; Lang, K.; Schuster, J.; Mörl, M., A universal method to produce in vitro transcripts with homogeneous 3' ends. *Nucleic Acids Research* **2002**, *30* (12), e56-e56.
121. Bowman, J. C.; Azizi, B.; Lenz, T. K.; Roy, P.; Williams, L. D., Preparation of Long Templates for RNA In Vitro Transcription by Recursive PCR. In *Recombinant and In Vitro RNA Synthesis: Methods and Protocols*, Conn, G. L., Ed. Springer Protocols: 2012; Vol. 941, pp 19-41.
122. Winter, G.; Waterman, D. G.; Parkhurst, J. M.; Brewster, A. S.; Gildea, R. J.; Gerstel, M.; Fuentes-Montero, L.; Vollmar, M.; Michels-Clark, T.; Young, I. D.; Sauter, N. K.; Evans, G., DIALS: implementation and evaluation of a new integration package. *Acta Crystallographica Section D* **2018**, *74* (2), 85-97.
123. Potterton, L.; Agirre, J.; Ballard, C.; Cowtan, K.; Dodson, E.; Evans, P. R.; Jenkins, H. T.; Keegan, R.; Krissinel, E.; Stevenson, K.; Lebedev, A.; McNicholas, S. J.; Nicholls, R. A.; Noble, M.; Pannu, N. S.; Roth, C.; Sheldrick, G.; Skubak, P.; Turkenburg, J.; Uski, V.; von Delft, F.; Waterman, D.; Wilson, K.; Winn, M.; Wojdyr, M., CCP4i2: the new graphical user interface to the CCP4 program suite. *Acta Crystallographica Section D* **2018**, *74* (2), 68-84.
124. Murshudov, G. N.; Skubak, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A., REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D* **2011**, *67* (4), 355-367.
125. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D* **2004**, *60* (12 Part 1), 2126-2132.
126. Novac, O.; Guenier, A. S.; Pelletier, J., Inhibitors of protein synthesis identified by a high throughput multiplexed translation screen. *Nucleic Acids Research* **2004**, *32* (3), 902-915.
127. Robert, F.; Cencic, R.; Cai, R.; Schmeing, T. M.; Pelletier, J., RNA-tethering assay and eIF4G:eIF4A obligate dimer design uncovers multiple eIF4F functional complexes. *Nucleic Acids Research* **2020**, *48* (15), 8562-8575.
128. Fukumoto, Y.; Obata, Y.; Ishibashi, K.; Tamura, N.; Kikuchi, I.; Aoyama, K.; Hattori, Y.; Tsuda, K.; Nakayama, Y.; Yamaguchi, N., Cost-effective gene transfection by DNA compaction at pH 4.0 using acidified, long shelf-life polyethylenimine. *Cytotechnology* **2010**, *62* (1), 73-82.
129. Dyer, B. W.; Ferrer, F. A.; Klinedinst, D. K.; Rodriguez, R., A Noncommercial Dual Luciferase Enzyme Assay System for Reporter Gene Analysis. *Analytical Biochemistry* **2000**, *282* (1), 158-161.
130. Punjani, A.; Rubinstein, J. L.; Fleet, D. J.; Brubaker, M. A., cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* **2017**, *14* (3), 290-296.

131. Bepler, T.; Morin, A.; Rapp, M.; Brasch, J.; Shapiro, L.; Noble, A. J.; Berger, B., Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature Methods* **2019**, *16* (11), 1153-1160.
132. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S.; Croll, T. I.; Morris, J. H.; Ferrin, T. E., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **2021**, *30* (1), 70-82.
133. Khatib, F.; Cooper, S.; Tyka, M. D.; Xu, K.; Makedon, I.; Popović, Z.; Baker, D.; Players, F., Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences* **2011**, *108* (47), 18949-18953.
134. Maguire, J. B.; Haddox, H. K.; Strickland, D.; Halabiya, S. F.; Coventry, B.; Griffin, J. R.; Pulavarti, S. V. S. R. K.; Cummins, M.; Thieker, D. F.; Klavins, E.; Szyperski, T.; DiMaio, F.; Baker, D.; Kuhlman, B., Perturbing the energy landscape for improved packing during computational protein design. *Proteins* **2021**, *89* (4), 436-449.
135. Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkoczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L.-W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D., Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallographica Section D* **2019**, *75* (10), 861-877.
136. Russ, A.; Wittmann, S.; Tsukamoto, Y.; Herrmann, A.; Deutschmann, J.; Lagisquet, J.; Ensser, A.; Kato, H.; Gramberg, T., Nsp16 shields SARS-CoV-2 from efficient MDA5 sensing and IFIT1-mediated restriction. *EMBO reports* **2022**, *23* (12), e55648.
137. Schindewolf, C.; Lokugamage, K.; Vu, M. N.; Johnson, B. A.; Scharon, D.; Plante, J. A.; Kalveram, B.; Crocquet-Valdes, P. A.; Sotcheff, S.; Jaworski, E.; Alvarado, R. E.; Debbink, K.; Daugherty, M. D.; Weaver, S. C.; Routh, A. L.; Walker, D. H.; Plante, K. S.; Menachery, V. D., SARS-CoV-2 Uses Nonstructural Protein 16 To Evade Restriction by IFIT1 and IFIT3. *Journal of Virology* **2023**, *97* (2), e01532-22.
138. Bassett, M.; Salemi, M.; Magalis, B. R., Lessons Learned and Yet-to-Be Learned on the Importance of RNA Structure in SARS-CoV-2 Replication. *Microbiology and Molecular Biology Reviews* **2022**, *86* (3), e00057-21.
139. Miedziak, B.; Dobieżyńska, A.; Darzynkiewicz, Z. M.; Bartkowska, J.; Miskiewicz, J.; Kowalska, J.; Warminski, M.; Tyras, M.; Trylska, J.; Jemielity, J.; Darzynkiewicz, E.; Grzela, R., Kinetic analysis of IFIT1 and IFIT5 interactions with different native and engineered RNAs and its consequences for designing mRNA-based therapeutics. *RNA* **2020**, *26* (1), 58-68.
140. Lodeiro, M. F.; Filomatori, C. V.; Gamarnik, A. V., Structural and Functional Studies of the Promoter Element for Dengue Virus RNA Replication. *Journal of Virology* **2009**, *83* (2), 993-1008.
141. Huguenin-Dezot, N.; Alonzo, D. A.; Heberlig, G. W.; Mahesh, M.; Nguyen, D. P.; Dornan, M. H.; Boddy, C. N.; Schmeing, T. M.; Chin, J. W., Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid. *Nature* **2019**, *565* (7737), 112-117.
142. Sanchez-Garcia, R.; Gomez-Blanco, J.; Cuervo, A.; Carazo, J. M.; Sorzano, C. O. S.; Vargas, J., DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Communications Biology* **2021**, *4* (1), 874.

143. Mirabello, C.; Wallner, B.; Nystedt, B.; Azinas, S.; Carroni, M., Unmasking AlphaFold: integration of experiments and predictions in multimeric complexes. *bioRxiv* **2023**, 2023.09.20.558579.
144. Zlatev, I.; Lackey, J. G.; Zhang, L.; Dell, A.; McRae, K.; Shaikh, S.; Duncan, R. G.; Rajeev, K. G.; Manoharan, M., Automated parallel synthesis of 5'-triphosphate oligonucleotides and preparation of chemically modified 5'-triphosphate small interfering RNA. *Bioorganic & Medicinal Chemistry* **2013**, *21* (3), 722-732.
145. Sarac, I.; Meier, C., Efficient Automated Solid-Phase Synthesis of DNA and RNA 5'-Triphosphates. *Chemistry – A European Journal* **2015**, *21* (46), 16421-16426.
146. Nikonov, A.; Mölder, T.; Sikut, R.; Kiiver, K.; Männik, A.; Toots, U.; Lulla, A.; Lulla, V.; Utt, A.; Merits, A.; Ustav, M., RIG-I and MDA-5 Detection of Viral RNA-dependent RNA Polymerase Activity Restricts Positive-Strand RNA Virus Replication. *PLOS Pathogens* **2013**, *9* (9), e1003610.
147. Luo, E.-C.; Nathanson, J. L.; Tan, F. E.; Schwartz, J. L.; Schmok, J. C.; Shankar, A.; Markmiller, S.; Yee, B. A.; Sathe, S.; Pratt, G. A.; Scaletta, D. B.; Ha, Y.; Hill, D. E.; Aigner, S.; Yeo, G. W., Large-scale tethered function assays identify factors that regulate mRNA stability and translation. *Nature Structural & Molecular Biology* **2020**, *27* (10), 989-1000.