

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


Host Determinants of Viral Entry and Replication

Citation

Tartell, Michael Armen. 2022. Host Determinants of Viral Entry and Replication. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Link

https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37372131

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

HARVARD UNIVERSITY
Graduate School of Arts and Sciences



DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Division of Medical Sciences

Committee on Virology

have examined a dissertation entitled

Host Determinants of Viral Entry and Replication

presented by Michael Armen Tartell

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature: Lee Gehrke

Typed Name: Dr. Lee Gehrke

Signature: jonathan kagan (May 6, 2022 18:22 EDT)

Typed Name: Dr. Jonathan Kagan

Signature: Silvia Rouskin
Silvia Rouskin (May 6, 2022 15:02 EDT)

Typed Name: Dr. Silvi Rouskin

Signature: Stacy M Horner
Stacy M Horner (May 6, 2022 13:19 EDT)

Typed Name: Dr. Stacy Horner

Date: May 05, 2022

Host Determinants of Viral Entry and Replication

A dissertation presented

By

Michael Armen Tartell

To

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Virology

Harvard University

Cambridge, Massachusetts

May 2022

©2022 Michael Armen Tartell

All Rights Reserved.

Dissertation Advisor: Dr. Sean P. J. Whelan Michael Armen Tartell

Host Determinants of Viral Entry and Replication

ABSTRACT

As obligate intracellular parasites, viruses rely on the biological machinery of their hosts to replicate. Each step of the viral life cycle, from entry, to protein translation and genome replication, to egress, involves host factors and their associated mechanisms to function. The presence of these factors, and the virus's ability to exploit them by mimicking the host, can determine the success of viral infection. Cells in turn have evolved mechanisms to defeat viral subversion, including self/non-self recognition by the innate immune system, and cell-intrinsic responses induced by interferons. Using vesicular stomatitis virus (VSV), we found that the host N6-adenosine messenger RNA (mRNA) cap methyltransferase, phosphorylated C-terminal domain interacting factor 1 (PCIF1), attenuates the antiviral response to interferon-beta (IFN-β). PCIF1 N6-methylates the cap-proximal nucleotide of VSV mRNA to $m^7Gpppm^6A_m$, and we define substrate requirements for this modification. PCIF1-dependent N6-methylation had no impact on mRNA stability, translation, or viral infectivity. However, cells lacking PCIF1 or expressing a catalytically inactive PCIF1 exhibit an augmented inhibition of viral replication and gene expression following IFN-β pretreatment. We further demonstrate that mRNA cap structures of rabies and measles viruses are also modified by PCIF1. This identifies a function of PCIF1 and cap-proximal m^6A_m in attenuation of the host response to viral infection. Host factors, specifically host proteases, are

required to trigger membrane fusion and entry of coronaviruses into cells, including severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The serine protease TMPRSS2 fulfills this function in lung cells, and is an attractive therapeutic target for antiviral drugs against COVID-19. Using rational structure-based drug design, substrate specificity screening of TMPRSS2, and entry inhibition screening using infectious chimeric VSV-SARS-CoV-2, we discovered covalent small-molecule ketobenzothiazole (kbt) TMPRSS2 inhibitors which are structurally distinct from and have improved activity over existing inhibitors Camostat and Nafamostat. Lead compound MM3122 has a half-maximal inhibitory concentration (IC_{50}) of 430 pM in blocking entry of VSV-SARS-CoV-2 into Calu-3 human lung epithelial cells, and a similar IC_{50} of 870 pM against a chimeric VSV Middle East respiratory syndrome coronavirus (VSV-MERS-CoV). Excellent metabolic stability, safety, and pharmacokinetics in mice make it a promising drug candidate for COVID-19 treatment.

ACKNOWLEDGMENTS

I have been exceptionally fortunate in my life to have been supported and encouraged by a multitude of people to whom I am immensely grateful. First, I must thank my family. My parents and grandparents nurtured in me a love of science from a very early age, and their guidance and teachings paved the way for my career. I would not be the person, nor the scientist, I am today without them. Second, I've had exceptional scientific mentors over many years: Dr. Martha Russell (Wilton High School), Dr. Robert Means (Yale University), Dr. Glenn Randall (University of Chicago), Dr. Sara Cherry (University of Pennsylvania), and Dr. Sean Whelan (Harvard University and WashU). Each of you played an essential role in my development at each stage of my career, from enthusiastic high schooler to rigorous research scientist, and your mentorship gave me the tools necessary to succeed and advance to each stage in my career. Thank you in particular to Sean, my mentor and adviser for these many years of my PhD. I am immensely grateful for your advice, patience, and guidance of my evolution as a scientist. You have taught me much both about virology and being a scientist, and I can only hope one day to repay your mentorship and generosity to others in kind. Thank you as well to the virology and microbiology communities at Harvard and WashU. Every scientist is a product of their environment, and I have been fortunate to work in and be supported by many people of extraordinary talent who advised and helped me over many years. Finally, a thank you to my friends, from high school to college, many bell ringing towers, and my fellow virology students. You made the journey as important as the destination, and the time we have shared together made the bad times bearable, and the good times great. Thanks to you all!

DEDICATION

To my parents and grandparents, whose love, support, and encouragement made this work possible.

TABLE OF CONTENTS

TITLE PAGE……..……i

COPYRIGHT PAGE……ii

ABSTRACT……..……iii

ACKNOWLEDGMENTS……..…v

DEDICATION……...vi

TABLE OF CONTENTS……..…vii

LIST OF FIGURES……..ix

LIST OF TABLES……xi

CHAPTER 1: INTRODUCTION……..……..……1

Vesicular Stomatitis Virus……2
Host Capping Enzymes and PCIF1……17
Internal RNA Modifications in Cellular and Viral mRNA……...25
Coronavirus Viral Entry……..36
Addressing Questions in Host Determinants of Viral Entry and Replication…….41

CHAPTER 2: Viral messenger RNA cap methylation by PCIF1 attenuates the antiviral activity of interferon-β……..……43

Author Affiliations and Publication Information……..44
Abstract……45
Significance……46
Introduction……..47
Results……50
Discussion……...72
Materials and Methods……..76
Acknowledgments……..83

CHAPTER 3: A novel class of TMPRSS2 inhibitors potently block SARS-CoV-2 and MERS-CoV viral entry and protect human epithelial lung cells……..84

Author Affiliations and Publication Information……..…85
Abstract……87
Significance……88
Introduction……..89
Results……92

Discussion……116
Materials and Methods……119
Acknowledgments……127

CHAPTER 4: DISCUSSION……...……...129

Summary of Results……130
Open Questions in Coronavirus Entry……..130
Open Questions in PCIF1 and m^6A_m……133
A Model for PCIF1 and m^6A_m in Viral Infection…….. 162

REFERENCES……163

APPENDIX 1: Supplementary Information to Chapter 2……...202

APPENDIX 2: Supplementary Information to Chapter 3……...228

APPENDIX 3: Supplementary Information to Chapter 4……...273

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Figure 1.1: The VSV Life Cycle……..3
Figure 1.2: Structure and Functions of large protein (L), the RNA-dependent RNA-polymerase (RdRp) of VSV……..……6
Figure 1.3: Schematic of enzymatic steps in mRNA capping and methylation……11
Figure 1.4: Structure of SARS-CoV-2 spike protein and entry pathways……..39

CHAPTER 2: Viral messenger RNA cap methylation by PCIF1 attenuates the antiviral activity of interferon-β

Figure 2.1: VSV mRNAs contain a 5' $m^7Gpppm^6A_m$ cap-structure……...…51
Figure 2.2: PCIF1 is the cap-proximal N6-methyltransferase……....54
Figure 2.3: Effect of mRNA cap methylation on PCIF1 modification of VSV mRNA…...57
Figure 2.4: Effect of m^6A_m on viral mRNA translation and stability……60
Figure 2.5: Effect of IFN-β pretreatment of cells on viral infection……63
Figure 2.6: PCIF1 does not regulate VSV mRNA sensing or IFIT1/3 binding……..67
Figure 2.7: Effect of PCIF1 on other non-segmented negative-stranded RNA viruses..70

CHAPTER 3: A novel class of TMPRSS2 inhibitors potently block SARS-CoV-2 and MERS-CoV viral entry and protect human epithelial lung cells

Figure 3.1: Structures of the initial TMPRS2 inhibitors discovered by screening existing HGFA, matriptase, and hepsin serine protease inhibitors……..……...94
Figure 3.2: Inhibition of SARS-CoV-2 cell entry into Calu-3 lung epithelial cells……..…95
Figure 3.3: Structures of compounds and molecular docking model to TMRPSS2…….97
Figure 3.4: Combined PS-SCL for substrate specificity of TMPRSS2, matriptase, Hepsin and HGFA……..…….100
Figure 3.5: Amino acid, turnover, substrate specificity information of TMPRSS2…….102
Figure 3.6: Structures of rationally designed covalent reversible kbt inhibitors of TMPRSS2…..……..106
Figure 3.7: Activity of compounds against SARS-CoV-2 and cell viability..……112
Figure 3.8: Mouse pharmacokinetics of compounds……..114

CHAPTER 4: Discussion

Figure 4.1: No change in bulk IFN-β induced transcripts……..136
Figure 4.2: Identification of N6-methylation specific RNA binding proteins……139
Figure 4.3: Cap-proximal nucleotides synthesized during IFN-β treatment……144
Figure 4.4: Localization of PCIF1 during VSV infection……146
Figure 4.5: Viral determinants of N6-methylation……148
Figure 4.6: PCIF1 knockout mice have no detectable m^6A_m……152
Figure 4.7: Infection of mice with VSV……..154

APPENDIX 1: Supplementary Information to Chapter 2

Figure A1.S1: VSV mRNAs in HeLa cells contain a 5' $m^7Gpppm^6A_m$ cap-structure….204
Figure A1.S2: PCIF1 methylates viral mRNA……..206
Figure A1.S3: Cap-methylation requirements for *in vitro* methylation of VSV mRNA...208
Figure A1.S4: Effect of m^6A_m on VSV mRNA stability……210
Figure A1.S5: Effect of m^6A_m on translation of VSV mRNA reporters……212
Figure A1.S6: Effect of IFN-β pretreatment of cells on viral infection in a second PCIF1-addback clone……..215
Figure A1.S7: Effect of IFN-β pretreatment of cells on viral infection in multiple single-cell clones of PCIF1 KO cells…….217
Figure A1.S8: VSV mRNA in A549 cells contain m^6A_m……..219
Figure A1.S9: Effect of IFN-β pretreatment and PCIF1 knockout on rates of viral translation……..221
Figure A1.S10: PCIF1 does not regulate VSV mRNA sensing late in infection…….223
Figure A1.S11: Effect of PCIF1 on viral-encoded eGFP expression……...226

APPENDIX 2: Supplementary Information to Chapter 3

Figure A2.S1: Compounds were tested against the protease domain of Human Recombinant TMPRSS2……..……230
Scheme A1.S1: Synthesis of acyclic P4-P1 tetrapeptide ketobenzothiazoles (kbts)…234
Scheme A2.S2: Cycloamide synthesis (2, 19, and 20)…….234
Scheme A2.S3: Synthesis of cyclic peptide aryl ether 21…….235
Figure A2.S2: Km curve for Boc-QAR-AMC using full-length TMPRSS2……...267
Figure A2.S3: IC50 inhibition curves of full-length TMPRSS2/Boc-QAR-AMC……..268
Figure A2.S4: Cell-based TMPRSS2 enzyme inhibition data of Z FH7116 (1) and VD2173 (2) in HEK-293 cells…….270
Figure A2.S5: MM3122 Acute Toxicity Study……..270
Figure A2.S6: Inhibition of SARS-CoV-2 cell entry into Calu-3 lung epithelial cells By ZFH7116 (1) and VD2173 (2) using VSV-SARS-CoV2-Spike protein Pseudotypes……..271
Figure A2.S7: Inhibition of SARS-CoV-2 cell entry into Calu-3 lung epithelial cells by ZFH7116 (1) and VD2173 (2) using VSV-SARS-CoV2-Spike protein Chimeras…….271
Figure A2.S8: Bacterial periplamic expression and purification of TMPRSS2-protease domain……..272

APPENDIX 3: Supplementary Information to Chapter 4

Figure A3.1: RNA-binding protein pulldown assay is specific for capped oligos.....….274

LIST OF TABLES

CHAPTER 3: A novel class of TMPRSS2 inhibitors potently block SARS-CoV-2 and MERS-CoV viral entry and protect human epithelial lung cells

Table 3.1: Structures and inhibition data of compounds ……..104
Table 3.2: Activity of lead compounds against VSV pseudotype MERS viral entry into human Calu- 3 lung epithelial cells…..……..……107

CHAPTER 4: Discussion

Table 4.1: Determining the 50% lethal dose in male and female wild type and PCIF1 knockout mice challenged with VSV……...…….153

APPENDIX 1: Supplementary Information to Chapter 2

Table A1.1: Oligonucleotides used for qPCR or mRNA selection…….227

APPENDIX 2: Supplementary Information to Chapter 3

Table A2.S1. Protease selectivity of ZFH7116 and VD2173……231
Table A2.S2. Full protease selectivity data of ZFH7116 and VD2173……232

APPENDIX 3: Supplementary Information to Chapter 4

Table A3.1: Top 20 proteins bound to the indicated RNA oligos in the indicated cell lysate……..275

Chapter 1:

Introduction

Michael Tartell

Vesicular stomatitis virus

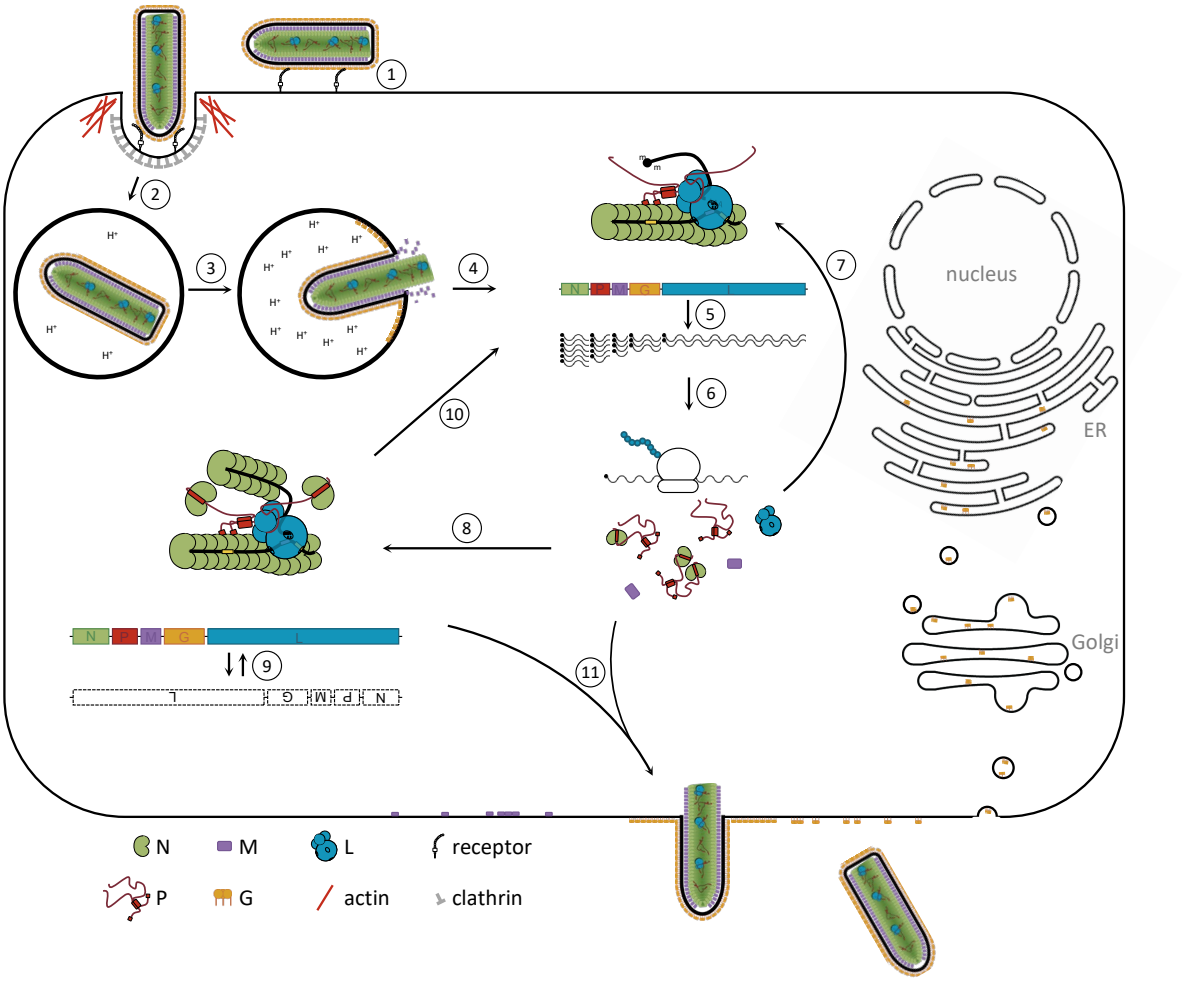
Introduction to VSV and its replication cycle

Vesciular stomatitis virus (VSV) is a nonsegmented negative-sense (NNS) RNA virus, and the prototype member of the *Rhabdoviridae* family of viruses [1] . The single-stranded viral genome encodes for five proteins. In order of their position in the negative-sense genome (3'->5'): the nucleocapsid protein (N) encapsidates the genomic RNA, the phosphoprotein (P) acts as a critical co-factor for the viral polymerase, the matrix protein (M) has multiple roles in blocking nuclear export of nascent host RNAs (including those enacting an innate immune response) as well as viral assembly, the glycoprotein (G) carries out entry into the cell by fusion between viral and cellular membranes, and the large protein (L) is a polyfunctional RNA dependent RNA polymerase (RdRp), and capping enzyme [2-7].

The replicative cycle of VSV is as follows **(Fig 1.1)**. VSV enters cells by clathrin mediated endocytosis [8-11]. Endosomal acidification drives conformational changes in G, that fuse the viral membrane with that of the endosome [12]. The viral core (N, P, L) is sufficient for mRNA synthesis and genome replication, all of which take place in the cytoplasm [13-17]. First, primary mRNA transcription is carried out from the genome (encapisdated in N) by L in consort with P, synthesizing one positive-sense mRNA strand for each gene [5, 7, 18-20]. These mRNAs are then translated by host ribosomes. G alone is translated by membrane-bound ribosomes and co-translationally inserted into the endoplasmic reticulum, and trafficked through the secretory pathway to the plasma membrane; the other viral proteins are translated by soluble ribosomes in the cytosol [21-25]. Following primary mRNA transcription, L synthesizes full-length

Figure 1.1: The viral life cycle. **(1)** The virion comes into contact with a host cell and the glycoproteins bind to their receptors. **(2)** The virion enters the cell through endocytosis using a clathrin- and actin-depend mechanism. **(3)** The pH decreases inside the endosome which triggers conformation changes of the glycoproteins and fusion of the viral and cellular membranes. **(4)** The content of the virion is released in the cytoplasm and matrix proteins dissociate from the nucleocapsid. **(5)** Primary transcription: the encapsidated genome is transcribed by the polymerase complexes present in the virion. **(6)** Viral mRNAs are translated by the host cell machinery. **(7)** Newly made polymerases transcribe the genome and increase primary transcription. **(8)** Once the N protein concentration reaches a certain threshold, the polymerase replicates the genome and uses N0-P complexes to encapsidate the nascent RNA. **(9)** The antigenomes are used as a template to generate viral genomes. **(10)** Secondary transcription: newly made genomes are used as a template for transcription. **(11)** Nucleocapsids are transported to the cytoplasmic membrane where they assemble with M and G proteins into viral particles. Courtesy of Louis-Marie Bloyet (submitted to Field's Virology).

Figure 1.1 (continued):



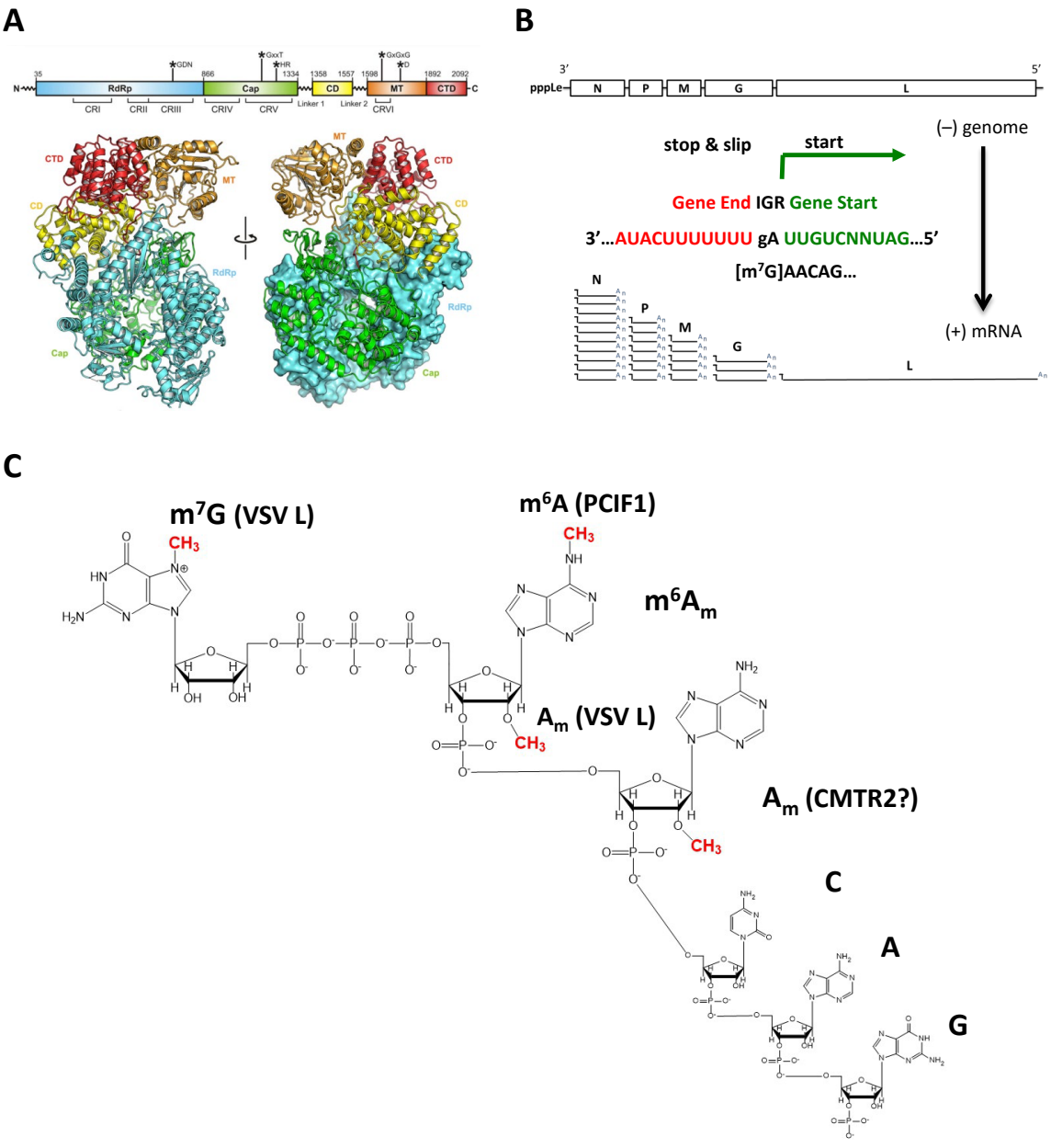
postive-sense (complementary) anti-genome, which acts as a template for genome replication [26-28]. This process depends on available N to encapsidate the newly synthesized genomic RNA [29]. These additional genomes can them be used as templates for secondary transcription in consort with newly synthesized viral proteins, greatly amplifying viral mRNA levels [30, 31]; by 6 hours post infection, VSV transcripts constitute over 65% of the mRNA of HeLa cells [32]. Viral assembly occurs at the plasma membrane, where viral cores (encapsidated N-RNA genomes with P and L) are condensed by M and new viral particles bud through G-containing microdomains in a manner that requires the late domain of M, possibly using the cellular endosomal sorting complex required for transport (ESCRT) machinery [33-36].

VSV Transcription

The VSV RdRp (L) performs all steps of viral transcription and genome replication, including replicating functions normally done by a multitude of host enzymes [2, 15, 37]. Early work showed that, while the RNA itself of VSV was not infectious (unlike positive-sense RNA viruses), virions added to a reaction containing NTPs, $MgCl_2$, and detergent were sufficient for RNA synthesis [2] . Later work clarified which viral proteins were required, and found that L and P were required to reconstitute transcription from a template of RNA encapsidated with N [13]. Most recently, cryo-electron microscopy studies have allowed determination of the structure of L, including in complex with P [3, 38]. VSV L contains 5 functional domains: an RNA-dependent RNA-polymerase domain (RdRp), capping domain (Cap), connector domain (CD), methyltransferase domain (MT), and C-terminal domain (CTD) **(Fig 1.2A)** [3]. The CD and CTD likely play organization roles in the overall structure and positioning of the

Figure 1.2: Structure and Functions of large protein (L), the RNA-dependent RNA-polymerase (RdRp) of VSV. **(A)** Schematic and structure of VSV L, indicating the RdRp domain, capping domain, connector domain, methyltransferase domain, and C-terminal domains. Adapted from [3]. **(B)** Schematic of the VSV genome and mRNAs, including conserved gene stop (red) and gene start (green) sequences in the (-)RNA genome which allow the transcription of each (+) mRNA gene sequence (black). The five viral genes are transcribed in a cascade of abundance related to their position in the viral genome: nucleoprotein (N) > phosphoprotein (P) > matrix protein (M) > glycoprotein (G) > large protein (L). **(C)** Structure of the conserved 5' cap and first 5 transcribed nucleotides of the VSV mRNA 5' UTR, including methylation sites (red) made by the indicated viral or cellular methyltransferases.

Figure 1.2 (continued):



various enzymatic domains [3]. The RdRp domain has a “fingers-palm-thumb” structure common to all template dependent polynucleotide polymerases [3]. The capping domain co-transcriptionally adds a 5’ cap structure onto nascent mRNA through an unconventional mechanism whereby a monophosphate RNA is transferred onto a GDP acceptor through a covalent L-pRNA intermediate [3, 39]. This polyribonucleotidyltransferase activity requires conserved residues within the capping domain including a GxxT and HR motif [40]. The methyltransferase domain catalyzes transfer of a methyl group from s-adenosyl-methionine (SAM) onto the 5’ cap at the ribose 2’O and guanosine N7 positions [41], and is structurally related to ribose-2’-O methyltransferases [3]. P makes contacts with L in multiple domains, specifically the CTD, between the CTD and RdRp, and in the groove between the CTD and the connector domain [38]. These interactions keep L in a closed, compact, and initiation-competent conformation [38].

All these enzymatic functions of L performed together ensure that not only is viral RNA synthesized, but that mRNAs mimic the structures of host mRNA. This allows viral mRNA to gain access to critical cellular machinery required for replication (e.g. host ribosomes for translation) and blocks self/non-self recognition mechanisms of the innate immune system (e.g. sensing of uncapped triphosphate mRNAs by binding of retinoic acid inducible gene I – RIG-I) [42, 43]. Viruses deficient in capping or methylation are attenuated both in cell culture and in pathogenesis during infection of mice [44]. The specific enzymatic processes used by L to synthesize mRNA, and how they differ from those used for cellular mRNA, significantly affect how they are processed by and function in the cell.

L initiates transcription from a single site on the 3' end of the viral genomic RNA, and synthesizes a short "leader" RNA with a 5' terminal triphosphate [45-48]. Following that, L synthesizes 5 discrete mRNAs, each coding for a single gene from the non-segmented genome, by recognizing conserved start and stop sequences in the genome **(Fig 1.2B)** [5-7, 49]. Each gene begins with the sequence UUGUCNNUAG in the genomic RNA, and thus each viral mRNA is synthesized with the 5' positive-sense sequence AACAGNNAUC. Similarly, each gene ends with the sequence AUACUUUUUUU in the genomic RNA [50-53]. Upon reaching this sequence, L synthesizes a poly(A) tail of approximately 150 nt through reiterative copying of the poly(U) sequence, then terminates transcription [51, 52, 54-59]. It then reinitiates synthesis at the subsequent gene-start sequence, though this process is inefficient, resulting in a ~30% decrease in abundance following each transcript and creating a cascade of abundance from N (first to be synthesized and most abundant) subsequently to P, M, G, and finally L (last to be synthesized and least abundant) [18, 19].

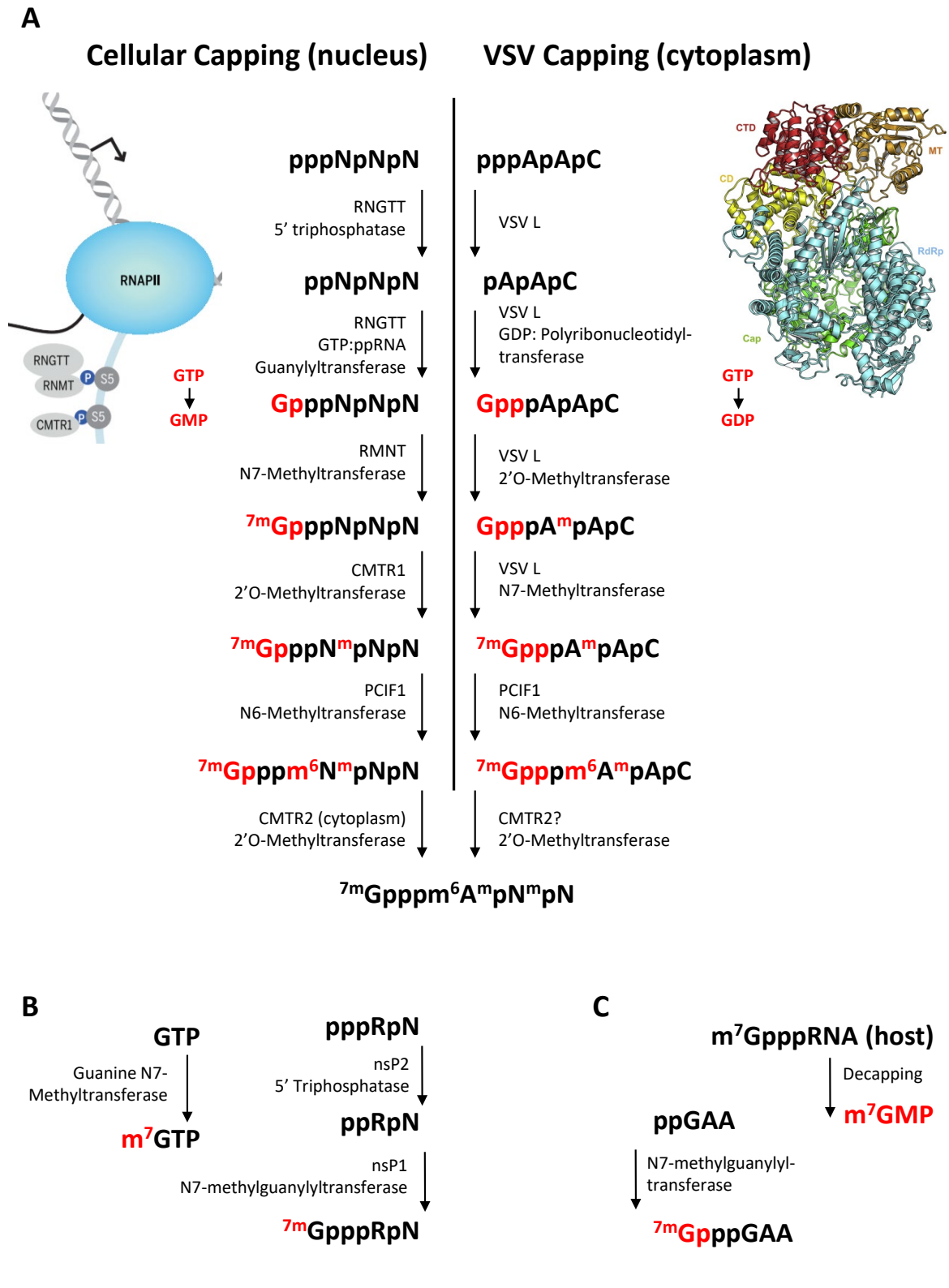
In addition to the formation of a poly(A) tail, VSV polymerase caps and methylates its mRNAs to mimic cellular mRNAs, though the enzymatic process used for their synthesis differs **(Fig 1.3A)** [37]. Specifically, cells and viruses both synthesize a 5' terminal N7-methyl-guanosine, connected to the first transcribed nucleotide by a triphosphate linker **(Fig 1.2C)**. The first transcribed nucleotide (an adenosine, corresponding to the first nucleotide of the transcriptional start sequence) is further methylated at the 2'O position [60-62]. Addition of the cap-structure by L requires a nascent transcript length of 31-nucleotides [63]. The 5' triphosphatase activity of L is assumed to first digest GTP to GDP. Then, the GDP polyribonucleotidyltransferase

(PRNTase) transfers the 5' pRNA of the nascent stand to the GDP acceptor through an unusual covalent enzyme-RNA intermediate (L-pRNA) at a histidine residue conserved within the PRNTase domain [40, 64]. This activity is specific for the transcriptional start sequence for VSV mRNA, but not the leader RNA start sequence, and is disrupted by mutations in the transcriptional start site [53]. The capped mRNAs are then sequentially methylated at first the 2'O, then the N7 positions by the methyltransferase domain of L [41, 60-62, 65, 66]. Though the final structure formed is identical to those found on eukaryotic mRNAs (m^7GpppN_m, commonly referred to as "cap1"), the enzymatic activities used in eukaryotes differ both in enzymatic function (the use of guanylyltransferase activity for capping instead of a PRNTase, and sequential order of N7- followed by 2'O-methylation), location (viral replication in the cytoplasm versus transcription in the nucleus), and number of enzymes (L, versus multiple enzymes associated with the phosphorylated C-terminal domain of RNA polymerase II) **(Fig 1.2C, 1.3A)** [37]. These host functions will be reviewed in detail in a subsequent section.

Although all the above functions of the viral polymerase can be recapitulated in vitro, studies of VSV observed key differences between viral mRNAs synthesized in vitro and in cells. VSV mRNA caps were N6-methylated at the cap-proximal (first transcribed) nucleotide, producing N6, 2'O- dimethyladenosine (m^6A_m) specifically at that position **(Fig 1.2C)** [60, 61]. This modification was detected in in the majority of all viral mRNAs, with abundance ranging from ~85% in HeLa cells to ~90-100% in BHK cells [60]. An additional 2'O methylation was also present at the second transcribed nucleotide (A_m) with an abundance ranging from 40-50% in HeLa cells to 90%+ in BHK cells [60]. Thus, the VSV mRNA 5' UTR synthesized by viral proteins alone forms the

Figure 1.3: Schematic of enzymatic steps in mRNA capping and methylation. (A) Comparison of the enzymes and reactions performed for cellular mRNA capping in the nucleus by enzymes associated with the phosphorylated C-terminal domain of RNA polymerase II, or by VSV polymerase in the cytoplasm. **(B)** Capping reactions performed by viruses using a guanine N7-methyltransferase. **(C)** Capping reactions performed by viruses which acquire N7-methylguanosine monophosphate from a host transcript and use an an N7-methylguanylyltransferase. Images adapted from [3] and [67].

Figure 1.3 (continued):



structure $m^7GpppA_mpApCpApG$-, while the 5' UTR synthesized during infection of a cell is additionally modified to $m^7Gppp(m^6)A_mpA(_m)pCpApG$- with varying abundances for each additional methylation [60-62]. It was presumed that these modifications were made by a cellular enzyme, since modifications at these methylation sites and positions had been characterized in eukaryotic mRNA caps, and could not be recapitulated using only viral proteins [60, 68, 69]. However, the cellular enzymes which synthesized them had not yet been characterized, and thus whether these mystery enzymes also modified viral mRNAs could not yet be determined. Later, VSV was found to be modified with internal monomethyl-m^6A modifications (N6-methyladenosine) as well [70, 71].

Translation during VSV Infection

VSV are translated by host ribosomes in the cytosol, and require host initiation factors for efficient translation [32, 72-74] . In the canonical cap-dependent initiation pathway of the cell, the 5' mRNA cap structure is first recognized and bound by the cap binding protein eIF4E, which specifically recognizes the 7-methylguanosine 5' cap, the helicase eIF4A, and a scaffolding protein eIF4G [75-77]. Together, these proteins form a complex referred to as eIF4F [78]. eIF4G then interacts with polyA binding protein (PABP), which binds the poly(A) tail of at the 3' end of mRNAs, thus forming a looped structure with translation complexes binding the 5' and 3' mRNA ends [76, 79, 80]. Following formation of the eIF4F structure, the 40S ribosomal subunit is recruited, as well as eIF1, eIF1A, eIF3, eIF5, and eIF2 in complex with the initiating methionine-bound tRNA and GTP [76, 80]. eIF3 itself is made of multiple subunits, 13 (eIF3a-m) in humans and most mammals [81]. This forms the 43S pre-initiation complex, which then begins the process of start-codon recognition by scanning down the 5' UTR mediated

primarily by eIF1 and eIF1A [82]. Upon start codon recognition by the 40S ribosome, which recognizes complementarity to the anticodon of the methionine-bound tRNA, eIF5 hydrolyzes the GTP bound to eIF2 to GDP and catalyzes binding of the 60S ribosome subunit, forming an 80S ribosome ready to translate the bound mRNA [76, 80, 83].

As VSV mRNAs are made with the 5' 7-methylguanosine cap structure, early studies hypothesized that VSV used canonical cap-dependent translation [72, 84-86]. In both wheat germ or rabbit reticulocyte in vitro translation systems, a significant decrease in VSV translation was observed with mRNAs decapped by β-elimination, or in competition with other cap binding molecules (pm^7G alone or cap analogs) [72]. VSV mRNAs also do not contain a known internal ribosome entry site (IRES), unlike other viruses which contain these conserved structures that can substitute for cellular translation factors [87]. Indeed, the 5' UTRs (untranslated region) of VSV mRNAs are much shorter than most cellular 5' UTRs [87]. M is the longest at 42 nt, while N, P, and L each have a 5' UTR of of 13 nt or fewer [50].

However, multiple lines of evidence suggest a more complicated mechanism controlling VSV translation. VSV translation is highly sensitive to depletion of rpl40, a component of the large ribosomal subunit [74]. Depletion of rpl40 had minimal effect on bulk mRNA translation, polysome association, and rRNA processing in yeast and mammalian cells [74]. However, VSV gene expression was sensitive to rpl40 depletion both in mammalian and yeast cells, as well as a subset of mRNAs in yeast cells [74]. Though no conserved sequence element could be identified in this set of mRNAs, a reporter mRNA containing the VSV 5' UTR was also sensitive to rpl40 in yeast extracts,

indicating a cis-acting element present in the viral RNA [74]. The precise mechanism behind the rpl40-dependence of VSV translation remains largely unknown.

Furthermore, VSV protein synthesis is not rate-limited by eIF4E, as is thought to be the case for cellular mRNA [88, 89]. VSV infection results in strong inhibition of host gene expression starting at 2 hours post infection (hpi), while viral protein synthesis continues robustly [32, 90]. Two main mechanisms have been identified to explain this effect. First, infection results in the sequestration of eIF4E by eIF4E-binding proteins (4EBP1, 2, or 3), a known mechanism of translational regulation by the cell [91]. 4EBP can be phosphorylated by mammalian target of rapamycin (mTOR) at multiple sites (threonines 37, 46; serine 65, threonine 70) [92, 93]. Phosphorylation of 4EBP causes it to disassociate from eIF4E, thus allowing cap-dependent translation to proceed. This allows for activation of translation through mTOR, integrating signaling pathways for growth factors or amino acid availability [92-96]. Conversely, infections with many viruses have been shown to downregulate mTOR signaling and cause hypophosphorylation of 4EBP, thus sequestering eIF4E and inhibiting 4E-dependent translation, while other viruses have been known to activate mTOR signaling [77, 97, 98]. RNAs with a 5' terminal oligopyrimidine (TOP) motif were hypersensitive to this effect [94] VSV infection was shown to cause hypophosphorylation of 4EBP and cap-dependent translational shutdown during infection [91]. However, if VSV translation is truly cap-dependent, this process should affect both the host and VSV. On the contrary, VSV replicates robustly under conditions of 4E-dependent translational shutdown, including treatment with rapamycin (which causes mTOR inhibition and eIF4E hypophosphorylation) [91], depletion of eIF4E [32], cleavage of eIF4G by the poliovirus

2A protease [99], and under conditions of hypoxia [100]. Furthermore, a mutant VSV whose polymerase is deficient in N7-methylation of the 5' mRNA cap ($GpppA_m$ structure, resulting from a point mutation G1670A in the s-adenosyl-L-methionine (SAM) binding domain of L) replicates robustly in cells and its mRNA is efficiently translated despite lacking the key moiety for eIF4E recognition [32].

Second, the viral matrix protein (M) disrupts the normal functioning of multiple activities in the nucleus [101-104]. M binds to the Rae1-Nup98 nuclear pore complex and prevents newly synthesized mRNAs from being exported from the nucleus to the cytoplasm, thus preventing their translation by cytoplasmic machinery [105, 106]. Critically, this includes mRNAs encoding an antiviral response (e.g. interferon stimulated genes – ISGs), effectively blocking a significant part of the innate immune response to viral infection [103, 104, 107]. Other studies have identified inhibition of cellular transcription by VSV in a manner involving M and TFIID, a cellular transcription factor [101]. Many other nuclear factors re-localize from the nucleus to cytoplasm during VSV infection, including hnRNPA1 [102]. The impact of these proteins and their re-localization on the virus is unclear, though they may act as regulators of apoptotic signaling [102]. A single point mutation in M (M51R) renders it deficient in these functions [104, 105, 107, 108].

How VSV replicates robustly under these conditions of translational shutdown remains unclear. One recent study proposed that translation of VSV mRNA at early time points proceeds through an eIF4E dependent mechanism, and then switches to an unknown 4E-independent mechanism later during infection [32]. Furthermore, rather than gaining a specific mechanical advantage over cellular transcripts, VSV overwhelms

the cell with an abundance of viral RNA [32]. How this 4E-independent mechanism functions is unknown. However, recent work has identified eIF3d as a cap-binding component of the eIF3 complex which can drive cap-dependent but eIF4E-independent translation of transcripts under conditions of 4E-shutdown [109-111]. eIF3 and its subunits have also been implicated in cap-independent translational mechanisms by binding to monomethyl m^6A sites in the 5' UTR of specific cellular transcripts (several nucleotides downstream of the 5' cap) [112]. VSV translation is strongly sensitive to depletion of eIF3 subunits, though the reduction in the rate of viral translation was similar to cellular translation, indicating VSV transcripts may not be specifically sensitive to eIF3 depletion [32].

Host Capping Enzymes and PCIF1

As detailed above, how a virus interfaces with and hijacks its host endures as a fertile ground for studying biological mechanisms. Indeed, the discovery and characterization of the 5' cap in eukaryotic RNA is inexorably intertwined with the study of viruses.

Discovery of the 5' mRNA Cap Structure

Before the first identification and characterization of the 5' mRNA cap in 1974, studies of 5' RNA structures were primarily performed using *E. Coli* or bacteriophage RNA [113, 114]. mRNAs in these systems contained 5' triphosphate ends, while rRNA and tRNA contained 5' monophosphate ends with abundant modified nucleotides [113, 114]. Studies of viruses were critical to discovering the 5' cap present in eukaryotic mRNAs, with key contributions by Kin-ichiro Miura, Aaron Shatkin, and Yasuhiro Furuichi with reovirus and cytoplasmic polyhedrosis virus (CPV), both dsRNA viruses in

the *Reoviridae* family [113-119]. Miura and Furuichi initially performed a classical sequencing experiment with CPV by labeling the 3'-terminal nucleotides on both dsRNA strands with ^{3}H-borohydride following periodate-mediated oxidation, digesting the labeled RNA with nucleases, and identifying the labeled nucleosides by two-dimensional paper chromatography [118]. To their surprise, 30% of their signal was incorporated into a "non-nucleoside material" or NNM [116, 118]. Meanwhile, while attempting to identify the 5' nucleotides through a methodology involving oxidation and β-elimination of the 5' end, followed by dephosphorylation and labeling with $^{[32]}$P-γ-ATP, the 5' nucleotide was determined to be a form of adenosine that did not quite match their marker for authentic adenosine [117]. Further chemical analysis revealed this to be 2'O-methyladenosine, providing the first evidence for modified nucleotides outside of tRNA and rRNA [115, 120]. When developing an *in vitro* transcription system for CPV, Furuichi further discovered that not only did the addition of the methyl donor SAM greatly enhance viral transcription, but that the RNA produced a "blocked" structure resistant to alkaline phosphatase [121].

In the same year, other groups began identifying methylated nucleotides in mRNA as well. Fritz Rottman's laboratory identified all 2' O-methylated bases in mRNA, as well as m^6A and m^7G in the mRNA of Novikoff hepatoma cells [122]. After meeting and discussing their findings at a 1974 Gordon Research Conference, Rottmann, Shatkin, and Perry published a paper predicting an m^7GppN_m cap structure in mammalian and some viral RNA, nearly accurate except for the prediction of a di-phosphate instead of a tri-phosphate linker connecting m^7G and the first transcribed 2'O-methylated nucleotide [123]. In 1975, Furuichi and Shatkin were the first to publish

the correct cap structure for CPV and reovirus mRNAs, m^7GpppA_m and m^7GpppG_m, observing that the long elusive NNM incorporated ^{32}P from $^{[32]}P$-α-GTP, $^{[3]}H$ from $^{[3]}H$-SAM, and the beta phosphate from $^{[32]}P$-β-ATP [117, 119]. These structures were also found in the mRNAs of HeLa cells and vaccinia virus by multiple groups, with Bernard Moss's laboratory characterizing an additional methylated nucleotide at the cap-proximal position, m^6A_m [68, 69, 124]

Around this same time, Amiya Banerjee's laboratory made the aforementioned discovery that VSV mRNAs contained a cap themselves, and were modified to m^6A_m in cells [60, 62]. Though the capping and methyltransferase machinery that made these modifications had not yet been found, mRNA caps could be eliminated chemically by beta-elimination, and their requirement for efficient translation *in vitro* was soon discovered [72]. Surprisingly, the GDP moiety of GTP was incorporated into these caps, in contrast to the GMP moiety for cellular caps, hinting at the unusual process of VSV cap synthesis [125]. However, the presence or absence of m^6A_m had no impact on ribosome association in either wheat germ or reticulocyte ribosomes [126]. Other modified nucleotides present internally in mRNA were also discovered, and will be detailed in subsequent sections.

Host mRNA Capping

Unlike VSV, which uses one multifunctional enzyme to cap its mRNAs in the cytoplasm, eukaryotic mRNAs are capped using a series of enzymes associated with the phosphorylated C-terminal domain of RNA polymerase II in the nucleus [37, 127]. The cap0 structure (N7-methylguanosine connected to the first transcribed nucleotide by a triphosphate linker) is conserved in eukaryotes at a minimum (m^7GpppN),

including, yeast and amoebas [113, 114, 128, 129]. As eukaryotes evolved, cap structures became more complex. Some eukaryotes also contain 2'O, but not N6-methylation, at the cap-proximal and second nucleotide ($m^7GpppN_mN_m$), including brine shrimp, trout, and insects (e.g. *D. melanogaster and A. albopictus*) [113, 114]. Higher eukaryotes, including humans, mice, rats, rabbits, chickens, and zebrafish, also contain m^6A_m ($m^7GpppN_mN_m$, or if N=A, $m^7Gpppm^6A_mN_m$) [67, 113, 114, 130, 131].

The processes of mRNA capping were originally studied in vaccinia virus and yeast [128]. Mammalian capping processes were found to proceed through a similar mechanism, which is as follows **(Fig 1.3A)** [37]. The mRNA cap is synthesized co-transcriptionally on the 5' end of nascent mRNAs when they reach a length of 22-25 nucleotides [127, 132, 133]. The 5' triphosphate of the nascent transcript is first digested by an RNA triphosphatase to a 5' diphosphate. Next, an mRNA guanylyltransferase transfers GMP derived from GTP to the 5' end of the RNA, generating a GpppN capped structure. These two functions are carried out by the enzyme RNGTT. The mRNA is subsequently methylated by an RNA guanine-N7-methyltransferase (RNMT) at the N7 position, and then methylated by CMTR1 at the N1 2'O position (m^7GpppN_m) [37, 113, 127, 128, 132, 134-140]. Though all these enzymes are associated with the phosphorylated CTD of RNA pol II [134, 136], the mechanism of 2'O methylation by the methyltransferase CMTR2 ($m^7GpppN_mN_m$) is less clear [141, 142]. Methylation at N2 is less common than at N1, with only 50% methylation at N2 being reported [142]. Interestingly, though CMTR1 localizes exclusively to the nucleus, and 2'O methyltransferase activity at N1 was primarily reported in nuclear extracts, CMTR2 localizes to both the nucleus and the cytoplasm, and 2'O methyltransferase

activity at N2 was reported exclusively in cytoplasmic extracts [140-142]. Why methylation at N2 seems to occur spatially and temporally separate from methylation at N1 is unknown.

Viral mRNA Capping

Viruses have evolved their own mechanisms of acquiring a 5' cap structure, thus mimicking host mRNAs to gain access to cellular translation machinery, or blocking access to innate immune mechanisms meant to inhibit their function. These generally fall into one of four main categories: cap synthesizing viruses, cap snatching viruses, pol II transcribed viruses, and cap independent viruses [37, 143].

Cap synthesizing viruses use their viral-encoded replication machinery to synthesize their own mRNA cap. This has been extensively reviewed recently, and will be summarized here [37, 143]. Viruses which synthesize their own caps fall into four separate subcategories depending on their specific mechanism of synthesis. The first is "conventional" synthesis, which follows the same general enzymatic process as eukaryotes described earlier **(Fig 1.3A)**. Large DNA viruses fall into this category, including *Poxvirdae* (vaccinia virus), baculovirus, and mimivirus, as well as dsRNA viruses of higher eukayotes (reovirus, cytoplasmic polyhedrosis virus, Bluetongue virus) and positive-sense flaviviruses (west nile virus). The second varies from "conventional" synthesis in the order and mechanism of N7-methylation. Here, a GTP-N7-methyltransferase methylates GTP to m^7GTP, which is then transferred by an N7-methylguanylyltransferase to the previously hydrolyzed 5'-diphosphorylated RNA to generate m^7GpppN-RNA (cap0) **(Fig1.3B)**. This occurs in positive-sense RNA viruses, including alphaviruses (semliki forest virus, sindbis viruses, Venezuelan equine

encephalitis virus) and potexviruses (bamboo mosaic virus). These viruses also lack 2'O methyltransferases, and instead contained a structured element in their 5' UTR capable of avoiding IFIT1-mediated restriction. The third category involves an enzymatically unusual form of cap-snatching used by yeast dsRNA totiviruses. Instead of using m^7GTP as a substrate for capping, they transfer m^7GMP obtained from a full length host mRNA to cap their own viral RNAs with cap0 **(Fig 1.3C)**. Finally, the fourth category includes negative-sense RNA viruses, including VSV and rabies virus (RABV), which use the unusual PRNTase mechanism described previously in this review **(Fig 1.3A)** [37, 143]. Other viruses in the order *Mononegavirales* (e.g. filoviridae such as ebola virus and paramyxoviridae such as measles virus) are inferred to synthesize their caps by a similar mechanism, as a PRNTase domain is present in the polymerases of these viruses [40]

Cap snatching viruses use cellular mRNAs to prime synthesis of their viral RNA including *Orthomyxoviridae* [144-146], *Arenaviridae* [147], and *Bunyaviridae* [143, 148]. The replication machinery of these viruses binds to cellular mRNAs and uses cleaves them using endonuclease activity a short distance downstream of their 5' cap. The virus then transcribes its own transcript downstream of the "snatched" transcript. This is carried out by multiple subunits of the IAV polymerase [144-146], or contained within the N terminal of bunyaviral or arenaviral L protein [147, 148]. Whether cap snatching is preferred on differentially methylated transcripts is unknown. Studies with rift valley fever virus and hantavirus suggest these viruses cap snatch transcripts from P bodies [97, 149], and thus RNA sequences currently targeted for degradation end up being

preferentially snatched, though this may simply opportunistic based on the RNAs spatially available [97].

Pol II transcribed viruses consist of most DNA viruses (excluding *Poxviridae*) or *Retroviridae* (which integrate into the cellular genome). Instead of encoding their own transcription machinery, these viruses are able to replicate and express their mRNAs through transcription by host RNA polymerase II. As such, they also gain access to all the capping and methylation machinery used by Pol II for host transcripts [4, 143]. Such structures ($m^7GpppN_mN_m$ or $m^7Gpppm^6A_mN_m$) have been characterized in adenovirus, simian virus 40, herpes simplex virus 1, and human polyomaviruses [150-154].

Cap independent viruses, such as *Picornaviridae, Astroviridae, and Calciviridae* do not contain a cap structure at the 5' end of their mRNAs, and have evolved different mechanisms to accomplish their goals. For example, they have a protein (e.g. VPg) covalently attached to the 5' end of their mRNA and/or an IRES which recruits translation machinery independent of the presence of a cap, and have diverse mechanisms for inhibiting the innate immune response they might elicit [143, 155].

Discovery of Host Methyltransferases

Host methyltransferases in mammalian cells were first isolated in the 1970s and 1980s by fractionation and purification of cell lysates, resulting in proteins containing a known enzymatic activity. Using these experiments, protein fractions containing the activities of an N7-methyltransferase [156], 2'O methyltransferase at N1 [142], 2'O methyltransferase at N2 [142], and N6-methyltransferase at cap-proximal adenosines were found and isolated [157]. However, these methods could only determine the size of the protein enriched in each fraction, not the identity of the gene encoding them.

Many of these genes were first identified in other eukaryotes such as yeast or *D. Melanogaster* [113, 128], and their human counterparts later identified by searching for homologous sequences. However, though this strategy was successful for methytransferases conserved between these organisms as early as the 1990s [158], the lack of m^6A_m in these model organisms delayed the discovery of its methyltransferase, and by extension its function, for decades.

The methyltransferase responsible for synthesizing m^6A_m was first published in 2019 by several independent groups [67, 130, 131, 159]. The protein in question had been previously identified as a protein which interacted with the phosphorylated CTD of RNA pol II, hence the name phosphorylated CTD-interacting protein 1 (*PCIF1*) [160, 161]. The first group to publish renamed it as cap-specific adenosine-N6-methyltransferase (CAPAM) [67], though this was not universally adopted [130, 131, 159]. Each group found that PCIF1 was indeed an N6-methytransferase specific for adenosine at the cap-proximal position [67, 130, 131, 159]. Biochemical data on the preference for an N7- and 2'O-methylated substrate, supported by a crystal structure of the protein and its associated with Pol II, led to the hypothesis that PCIF1 co-transcriptionally methylates mRNA synthesized by Pol II in the nucleus after it has already been capped and methylated at the N7- and 2'O- positions [67]. Though now armed with the ability to make *PCIF1*-knockout cells to probe its function, each group came to a different conclusion about its role in RNA stability and translation. Their specific experiments and data argued for either a small increase [67, 131, 162], decrease [130], or no change [163] in RNA stability or translation. Earlier biochemical studies showed that m^6A_m inhibits decapping by the enzyme DCP2, a key process in

the 5' -> 3' RNA decay pathway [162]. However, while this would seemingly impact RNA stability, a subsequent study in *PCIF1*-knockout cells showed stabilization of only a small set of transcripts, indicating N6-methylation was not a causative effect of the stability found in m^6A_m-annotated transcripts [131]. The reasons for contradictions between these studies are not immediately clear. One possibility is that each group used a slightly different method for annotating m^6A_m positive transcripts before determining their relative stability or translation efficiency. Broadly, each group used a method involving immunoprecipitation of transcripts using an antibody that targets m^6A_m and m^6A indiscriminately, followed by high-throughput sequencing to determine the identity of a transcript [67, 130, 131]. However, each group differed in the experimental and computational strategy used to separate true m^6A_m-positive transcripts from contaminating m^6A-positive but m^6A_m-negative transcripts. Thus, despite each paper agreeing that m^6A_m is present in the majority of mRNA caps initiating on A (m^6A_m to A_m comparison), a clear function for it remained unknown.

Internal RNA Modifications in Cellular and Viral mRNA

The cell contains many other modified nucleotides not found in the 5' cap which have become intense topics of study over the last decade. The presence of these nucleotides has diverse impacts on the stability and translatability of the transcripts which contain them. These modifications have also been found in viral mRNA and can impact their life cycles [164-169]. Their specific role in the cellular response to viral infection and innate immunity will be discussed in detail in the next section. A summary of their roles more broadly in the cell and viral life cycle follows, including relevant "writer" (enzyme which installs the modification), "eraser" (enzyme which removes the

modification) and “reader” (protein which specifically binds the modification) proteins for mRNA. Many of these modifications (and others) are also present in ribosomal RNA (rRNA) and transfer RNA (tRNA), which will not be discussed here.

N6-methyladenosine (m^6A)

m^6A is the most abundant mRNA modification [165]. It is made by an RNA methyltransferase complex involving methyltransferase-like 3 (METTL3 – the only catalytically active portion of the complex), METTL14, VIRMA (Virilizer Like m6A methyltransferase associated), HAKAI (an E3 ubiquitin ligase), RBM15 (RNA binding motif protein 15), WTAP (Wilm’s tumor 1-associating protein), and ZC3H13 (zinc finger CCCH domain-containing protein 13) [167]. m^6A is erased by the demethylases FTO (fat mass and obesity-associated protein) and ALKBH5 (alkB homologue 5) [167]. FTO also can demethylate m^6A_m, and whether FTO prefers m^6A_m as its substrate is debated [162, 163, 170, 171]. m^6A is bound by many known reader proteins, each of which mediates a different function in the cell. m^6A is present throughout an mRNA, but is primarily found near the stop codon around the 3’ UTR, and secondarily found at the 5’ UTR, with low levels present in the coding sequence [172]. When present in the 5’ UTR, m^6A can bind eIF3 directly to drive cap-independent translation, a process favored under conditions of cellular stress such as heat-shock [112, 173]. Heat shock also induces the addition of m^6A at 5’ UTR sites, enabling the usage of this pathway [112, 173]. When in the 3’ UTR, m^6A recruitment of other reader proteins produces a variety of different (and sometimes opposing) effects on RNA stability and translation, many mediated by YTH-domain family (YTHDF) and YTH-domain containing (YTHDC) proteins [166, 171]. YTHDF1 recruits translation complexes to positively regulate

translation of transcripts, YTHDF2 promotes decay of transcripts, and YTHDF3 has both functions [174, 175]. YTHDC1 has been shown to regulate splicing of RNA in the nucleus, while YTHDC2 also positively regulates translation [176-178]. m^6A also plays a role in nuclear export by binding hnRNPA2B1 [179, 180].

m^6A is synthesized in the genomic or mRNA of many viruses in a METTL3-dependent manner. This includes *Flaviviridae* (HCV, DENV, WNV, YFV, ZIKV), *Coronaviridae* (SARS-CoV-2), *Picornaviridae* (EV71), *Paramyxoviridae* (RSV), *Orthomyxoviridae* (IAV), *Pneumovirdae* (hMPV), *Retroviridae* (HIV, MLV), *Polyomaviridae* (SV40), *Adenoviridae* (Ad2), *Herpesviridae* (HSV-1, KSHV, EBV), and *Hepadnaviridae* (HBV) [165]. The effects of m^6A on these viruses are as diverse as their life cycles. It promotes gene expression for some viruses in a manner dependent on YTHDF-reader proteins (e.g. enhances HIV translation), though in *Flavivirdae* the binding of those reader proteins to viral RNA inhibits viral assembly [168, 181-185]. For viruses which replicate in the nucleus it can promote nuclear export of viral RNA and/or splicing [183, 186].

N1-methyladenosine (m^1A)

The presence of m^1A in mRNA is controversial, with multiple studies identifying it in thousands or very few mRNA transcripts [187-190]. One study reported that m^1A was primarily located in the 5' UTR of mRNA and correlated with increased translation [187]. A complex of TRNA methyltransferase 6 non-catalytic subunit (TRM6) and TRNA methyltransferase 61A (TRM61) has been shown to synthesize m^1A on mRNA, though other writers may yet be undiscovered [187, 188]. m^1A can be demethylated by ALKBH3 [191], and has only been described in the cap-snatching virus IAV [192].

5-Methylcytidine (m^5C)

Though there are many methyltransferases which synthesize m^5C, only NOP2/Sun RNA methyltransferase 2 (NSUN2) has been demonstrated to modify mRNAs [166, 193]. The extent of m^5C methylation in mRNA is unclear, with studies ranging from over 8000 to a select few mRNAs [194, 195]. A more recent mapping method estimated the number of m^5C at several hundred [196]. m^5C has been implicated in mRNA export via the protein Aly/REF export factor (ALYREF) [193]. m^5C has been proposed to positively regulate translation of select transcripts [197].

m^5C is incorporated into multiple viruses, including Sindbis virus, HIV, adenovirus 2, and murine leukemia virus (MLV) [165, 198-201]. For HIV and MLV, the host methyltransferase NSUN2 makes this modification [198, 199]. Loss of NSUN2 inhibited HIV mRNA translation and splicing [199].

N4-Acetylcytidine (ac^4C)

The N-acetyltransferase 10 (NAT10) can acetylate cytosine in mRNA at the N4 position to form ac^4C [202]. This has been reported to increase mRNA stability and translation efficiency both in cellular RNA and in HIV RNA [202, 203].

Inosine (A to I editing)

Inosine is a modified base synthesized by adenosine deaminase acting on RNA (ADAR1 or ADAR2) by the deamination of adenosine, a process referred to as A-to-I editing [165]. This causes several changes to the mRNA containing it. The major function of inosine is thought to be modification of the secondary structure of mRNA to reduce dsRNA formation and recognition by the innate immune system (this will be discussed further later) [166]. However, because inosine can be resolved as

guanosine, adenosine, or rarely uracil, it can also cause changes in protein translation [166]. ADAR1-mediated A-to-I editing has been implicated in HIV, VSV, measles virus, and hepatitis delta virus (HDV) [204].

Pseudouridine (ψ)

Pseudouridine, the C5-glycoside isomer of uridine, is the most abundant RNA modification found in total RNA, and is found extensively in mRNA [166]. It can be synthesized by multiple different pseudouridine synthase (PUS) or tRNA pseudouridine synthase (TRUB) proteins, and there are no known erasers [205]. Though each protein appears to have a substrate preference, the exact enzymes which contribute to ψ in mRNA have not been specifically defined [205]. One study defined TRUB1, TRUB2, PUS1, PUS7, and RPUSD2 as enzymes which could synthesize different proportions of ψ in mRNA [206]. ψ sites defined in cells were present equally throughout the 5' UTR, 3' UTR, and coding sequence of cells, though this could vary depending on the study and method of mapping [205, 207, 208]. Addition of ψ is thought to stabilize transcripts and promote translation, though many of these studies are based on *in vitro* addition of ψ to transcripts rather than studying its abrogation in the cell [205, 209]. A recent study also described effects of ψ on pre-mRNA alternative splicing [210]. Ψ has been described in many viruses, including zika virus (ZIKV), DENV, HCV, HIV, and poliovirus (PV), though its effect on these viruses is unknown [165, 169, 211].

Host Response to Viral Infection

The innate immune system of the cell serves as the first line of defense against infection by pathogens, including viruses. These systems have evolved to detect when a cell has been infected, and then upregulate expression of specific genes and/or

signaling pathways to mount an antiviral response. Modified nucleotides in RNA, including at the 5' cap, play key roles at all steps in this process for the cell [43, 212]. In turn, the evolution of a virus to acquire these modifications can allow it to hijack and co-opt the very mechanisms meant to defend against it [43, 212]. This section will focus on RNA modifications and how they function in these innate immune pathways.

Self/Non-Self Recognition

Mechanisms in the cell have evolved to discriminate viral from cellular processes and allow for selective targeting against the virus without collateral damage to the cell. The first step in this process is detection of pathogen-associated molecular patterns (PAMPs) by pattern-recognition receptors (PRRs) [213]. PRRs are proteins which bind to the presence (or absence) of a particular chemical moiety or molecule not normally present in the cell and associated with a pathogen (PAMP), allowing them to act as sensors of viral infection [165]. One key example of this is double-stranded RNA (dsRNA) which is made by many viruses (e.g. *Reoviridae*) but is not made by the cell, which uses single-stranded RNA (ssRNA) in its mRNA [214]. dsRNA is bound by a variety of PRRs (e.g retinoic acid inducible gene 1 (RIG-I) and protein kinase R (PKR)), which then cause a downstream antiviral response (transcriptional regulation of interferons and interferon-stimulated genes, or inhibition of eIF2α, respectively) [214]. RNA modifications made by the cell can inhibit binding by PRRs, thus allowing the cell to discriminate self from non-self [212]. Conversely, viruses have evolved in response to synthesize or steal these modifications for themselves, thus blunting the impact of an antiviral response [43, 164, 165, 168, 212-214]. A summary of key RNA sensors which are known to discriminate modified nucleotides follows.

Toll-Like Receptors (TLRs):

TLRs are a family of noncatalytic PRRs present both intracellularly and at the cell-surface. Though TLRs as a family have many roles in the specific recognition of viruses or bacteria, three are specific for RNA [215-217]. All of these three are present in the endosome, positioning them to detect nucleic acids from viruses which enter by endocytosis. TLR3 binds dsRNA [218-223], while TLR7 and TLR8 are specific for ssRNA containing either a 5' triphosphate or cap0 structure (i.e. RNA lacking m^7G or 2'O methylation as self markers) [224-228]. TLR7 or 8 binding induces downstream signaling through myeloid differentiation primary response gene 88 (MyD88) which results in the activation of NF_kB and interferon signaling [164, 165]. TLR3 does the same, but signals through Toll/IL-1R homology domain-containing adaptor protein inducing IFN-β (TRIF) instead of MyD88 [164, 165]. *In vitro* modification of oligos with m^6A potently limited detection by TLR3, while modification with m^6A, m^5C or ψ limited detection by TLR7 and TLR8 (this effect was also observed with 2-thiouridine (s^2U) and 5-methyluridine (m^5U) which occur in tRNA) [229-231]. However, this has not yet been shown to occur in a viral infection system.

RIG-I-Like Receptors (RLRs):

RLRs are a family of three cytosolic proteins, specifically RIG-I, melanoma differentiation-associated antigen 5 (MDA5), and laboratory of genetics and physiology 2 (LGP2), which modulates the activity of RIG-I or MDA5 [143, 164, 165, 232]. RIG-I is thought to primarily recognize the 5' end of short dsRNA or ssRNA with a terminal triphosphate [42, 232]. Additional methylations further inhibit RIG-I binding (e.g. cap1 more strongly inhibits binding than cap0) [43, 233]. The presence of internal m^6A

methylations have been shown inhibit RIG-I binding and signaling with many negative-sense RNA viruses, including human metapneumovirus (hMPV), SeV, MeV, and VSV, as well as HIV and hepatitis B virus (HBV) [70, 71, 234, 235]. Synthetic addition of ψ or m^6A also prevented RIG-I binding, though the effect with ψ has only been demonstrated *in vitro* thus far [236]. MDA5 primarily binds to long dsRNAs, but can also sense long ssRNAs [232]. Despite preferring a distinct type of ligand from RIG-I, MDA5 is also inhibited by many of the same RNA self-markers. MDA5 preferentially binds to RNA lacking a 2'O methylated cap1 structure [237]. MDA5, and to a lesser extent RIG-I, are inhibited by the presence of A-to-I modifications made by ADAR1 [204, 238-240]. These modifications are used by the cell to reduce dsRNA structures formed in certain transcripts, thereby ensuring they would be recognized as self rather than non-self. Though many viruses are modified by ADAR1, studies on this effect have focused on the PKR-dependent impact of this modification [204].

Other RNA sensors

TLRs and RLRs are considered to be the primary sensor proteins for viral RNA which then upregulate a downstream transcriptional response and synthesis of effector proteins (e.g.ISGs) [164, 165, 212]. However, two viral RNA sensors straddle the line between sensor/signaling protein and effector. PKR binds to dsRNA and phosphorylates eIF2α, thus inhibiting cellular translation [241-243]. OAS proteins (oligoadenylate synthase 1, 2, 3) are also dsRNA sensors which synthesize 2'-5' oligoadenylate, a nucleotide signaling molecule that binds to and activates RNASEL, which in turn degrades both viral and cellular ssRNA (mRNA, tRNA, and rRNA) as a mechanism to inhibit viral replication [243, 244]. A-to-I editing by ADAR1 has been

shown to reduce dsRNA structures in the cell to inhibit binding by both PKR and OAS, including for viral mRNA [204, 238-240, 244, 245].

Interferon-Stimulated Genes

Upon sensing viral nucleic acids, downstream signaling pathways upregulate the transcription of antiviral genes, including interferons and interferon-stimulated genes (ISGs) [246-248]. ISGs have diverse functions in the cell and restrict all parts of the viral life cycle, including entry, replication, assembly, and egress [246]. Some ISGs specifically target RNAs lacking self markers in order to specifically target viral RNAs, including IFN-induced proteins with tetratricopeptide repeats (IFIT proteins) [43, 249-251], and interferon stimulate gene 20 (ISG20) [252-256].

There are four IFIT proteins present in humans, IFIT1, IFIT2, IFIT3, and IFIT5. These proteins can act alone or in complex with each other to sense the 5' end of mRNAs [43, 249-251, 257, 258]. Generally, they bind to the 5' end of mRNAs lacking certain self markers, which inhibits ribosome binding and decreases translation of these transcripts [259]. IFIT1 and IFIT3 form a complex which specifically binds and represses mRNAs which contain cap0 (m^7GpppA) but not cap1 (m^7GpppA_m) [250]. Viruses which lack a functional 2'O methyltransferase are strongly inhibited, including west nile virus (WNV), Japanese encephalitis virus (JEV), DENV, multiple coronaviruses, vaccinia virus, Venezuelan equine encephalitis virus, sindbis virus, and HCV (reviewed in [260]). *In vitro* assays and structural studies have shown that 2'O methylation at the second nucleotide (N2) also inhibits binding and translational repression by IFIT1, though this effect has not been demonstrated during viral infection [251]. It should be noted that 2'O methylation does not totally repress IFIT1 binding,

and IFIT1 still restricts several viruses (including VSV) which contain 2'O methylation [250]. IFIT5 recognizes uncapped mRNAs with a 5' triphosphate [251, 261]. In contrast, IFIT2 appears to bind to the 3' UTR of mRNAs and appears to prefer AU rich sequences [262]. Though IFIT2 is antiviral against many viruses, including VSV, it is strangely proviral for IAV, for which it increases translational efficiency by preventing ribosome pausing [263]. This study also observed a similar effect in many cellular mRNAs, including those involved in the antiviral response, raising the possibility that IAV was hijacking a cellular mechanism [263]. Murine IFIT proteins differ from their human counterparts. Murine Ifit1 is not an ortholog of human IFIT1, but an evolutionarily divergent paralog [257]. Ifit1 cannot bind 2'O methylated RNAs at all (as opposed to human IFIT1), and other mouse Ifit proteins (Ifit1b and Ifit1c) instead bind to cap1 RNAs strongly and restrict both host and viral mRNAs containing cap1[257, 264]. Another RNA-binding protein whose mechanism remains somewhat enigmatic is ISG20 [252-256]. ISG20 inhibits a wide variety of viruses from many families, from VSV to DENV, HIV, and HBV (reviewed in [252]). Its mechanism of action is debated, from direct RNA degradation as an exonuclease to a translational repressor [252]. It is reported by multiple studies to discriminate between viral and cellular mRNAs, though the specific mechanism behind this is unclear [253-256]. One study reported binding of ISG20 to YTH-domain family 2 (YTHDF2) protein, an m^6A-binding protein, and dependence of the antiviral effect against HBV on the presence of a specific m^6A in a specific stem-loop structure [253] .

Self/non-self discrimination in ISGs need not only occur with the additional of RNA modifications, but can also be sequence based. The antiviral ISG zinc finger

antiviral protein (ZAP) targets 5' CG 3' dinucleotides suppressed in vertebrate genomes, but present in a mutant HIV genome [265].

Host Regulation of the Interferon Response

The interferon response uses self/non-self recognition and RNA modifications not only as mechanisms to target viral mRNAs, but to regulate host gene expression as well [164, 165, 167, 168, 266]. Multiple studies have demonstrated that m^6A promotes the nuclear export of key mRNAs involved in innate immune signaling. One showed that the mRNAs of key signaling proteins in the interferon response (MAVS, TRAF3, and TRAF6) are all modified with m^6A, that they become demethylated during VSV infection through the activity of the helicase DDX46 and the demethylase ALKBH5, and this causes the retention of these transcripts in the nucleus [267]. Another study showed that hnRNPA2B1 bound to m^6A modified transcripts of cGAS IFI16, and STING and promoted their nuclear export, thus enhancing cytoplasmic antiviral signaling [180]. In contrast, another study showed that though the presence of m^6A also promoted the nuclear export of DHX58, p65, and IKKγ, this caused them to be targeted by YTHDF2 for degradation instead of promoting their expression [268]. Similiarly, the mRNA of interferon-β is modified by m^6A, as also targeted for YTHDF2-dependent degradation in the cytoplasm [269, 270]. How different m^6A transcripts are sorted and bound by different RNA-binding proteins to up- or down-regulate their expression is unknown. Many ISGs are also modified with m^6A, and this modification enhances their translation in a YTHDF1-dependent manner [168, 271-273]. Depletion of the internal N6-methyltransferase METTL3 caused a dampened innate immune response against VSV, though this effect was independent of cap-proximal methylation by PCIF1 [272]. A

recent preprint reported that NSUN2, a m^5C methyltransferase, negatively regulates the stability of IRF3 mRNA [274]. 2'O methylation by the cellular methyltransferase CMTR1 also allows ISG expression to continue while IFIT1 is active, and its inhibition causes a dampened innate immune response [266]. The effect of CMTR2 and methylation at the second nucleotide remains understudied.

Coronavirus Viral Entry

In December 2019, there was an outbreak of respiratory illness and fever in Wuhan, China, the cause of which was identified as a newly emerged beta coronavirus (2019-nCOV). By March 2020, this virus (soon termed severe acute respiratory syndrome coronavirus 2, or SARS-CoV-2) had spread throughout the world and was declared a pandemic by the World Health Organization. The resulting global health crisis and lack of therapeutic agents triggered an immediate and vigorous study of SARS-CoV-2 and the development of tools for those studies. The necessity of contributing to this work caused many, including our lab, to shift our focus away from our main projects towards studying SARS-CoV-2.

<u>VSV Chimeras as Tools to Study Viral Entry</u>

A robust reverse genetics system allows the insertion of foreign genes into the VSV genome, as well as the modification or deletion of viral genes or sequences [275]. As the viral glycoprotein (G) is the sole determinant of viral entry, it can be replaced with the entry protein of a different virus [276]. These proteins can become incorporated into the recombinant virion and create an infectious, chimeric VSV that adopts the entry mechanism and tropism of the inserted glycoprotein [276-286]. A similar chimeric VSV was constructed using the viral spike protein (S) of SARS-CoV-2, which mediates all

steps of coronavirus entry [287]. This virus (VSV-SARS-CoV-2) mimicked the response of a clinical isolate of SARS-CoV-2 to inhibition of viral entry and/or antibody neutralization tests, allowing for studies at biosafety level 2 (BSL2) of the mechanisms, receptors, and inhibitors of SARS-CoV-2 viral entry [287].

Coronavirus Entry

Coronaviruses are enveloped, positive-sense, single-stranded RNA viruses in the order *Nidovirales* [288] . SARS-CoV-2 is a member of the subfamily *orthocoronavirinae* and genus *betacoronavirus*, which also includes the human pathogens severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and human coronavirus (HCoV)-OC43; viral entry is mediated by the glycoprotein spike [288, 289]. Though the specific entry mechanism of each of these viruses is slightly different, they all follow several conserved elements. Each S protein has two structurally distinct conformations, prefusion and postfusion, and when triggered, the transition between these two states completes the process of membrane fusion [289, 290]. The process of triggering occurs through a combination of binding to a host entry receptor and cleavage by host proteases at multiple possible locations in the viral life cycle (proprotein convertases during virion packaging or production, extracellular proteases after the completion of egress, cell surface proteases after attachment to a new cell, or endosomal/lysosomal proteases after endocytosis during entry) [289-292]. Which host factors are used varies not only by virus, but also by cell type for the same virus [289, 293, 294].

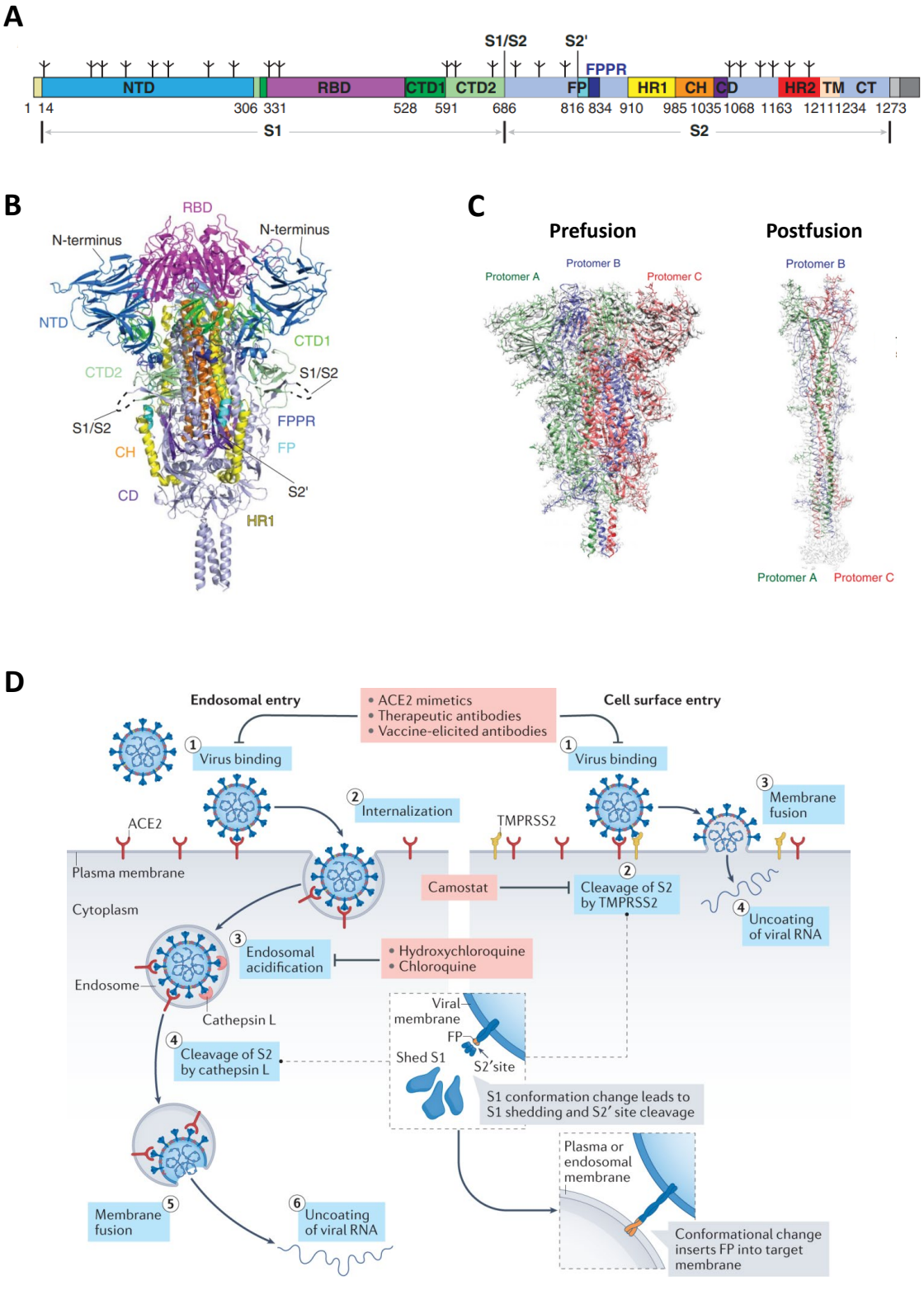
The structure of the SARS-CoV-2 S protein has been solved in multiple conformations by multiple groups [295-300], as have SARS-CoV and MERS-CoV S

proteins [301]. Like SARS-CoV and MERS-CoV, SARS-CoV-2 S has two non-covalently associated subunits, termed S1 and S2 **(Fig 1.4A-B)** [289, 290]. S1 contains the receptor binding domain (RBD), which binds angiotensin-converting enzyme 2 (ACE2), the host receptor for SARS-CoV-2 and SARS-CoV (as opposed to MERS, which uses dipeptidyl-peptidase 4 – DPP4); the receptor binding motif (RBM) in this domain makes the contacts necessary for binding with ACE2 [295, 296, 302, 303]. The RBD is the primary target of neutralizing antibodies produced during infection or as a result of vaccination. S1 also contains an N-terminal domain, and two C-terminal domains, all of which can also be targeted by neutralizing antibodies [304-306]. S1 wraps around prefusion S2, which contains the fusion peptide [299]. Upon triggering, S2 undergoes a significant rearrangement into the postfusion conformation, and then exposes and inserts the fusion peptide into the target cell membrane to cause fusion pore formation, thus completing the process of entry **(Fig 1.4C)** [299].

In order to effect triggering, two proteolytic cleavage steps must occur in the SARS-CoV-2 spike by host proteases [289]. The first is at the S1-S2 boundary, and for SARS-CoV-2 is performed by the host protease furin at a multibasic site in the S1-S2 junction [294, 307-309]. This step is thought to occur during virion production in an infected cell before the completion of the egress process, and distinguishes SARS-CoV-2 from SARS-CoV, which neither has this multibasic site nor is cleaved during maturation by furin [294, 307, 309]. Interestingly, the furin site appears to be specifically required for SARS-CoV-2 entry into lung cells and for transmission in animal models, and has little effect or is lost upon passage in other cell lines (e.g. vero cells, a monkey kidney cell line) [294, 308, 310, 311].

Figure 1.4: Structure of SARS-CoV-2 spike protein and entry pathways. (A) Schematic of the subdomains of spike. Tree-like symbols indicate known glycans. NTD – N-terminal domain, RBD – receptor binding domain, CTD – C-terminal domain (1 or 2), FP – fusion peptide, FPPR – fusion-peptide proximal region, HR – heptad repeat (1 or 2), CH – central helix, CD – connector domain, TM – transmembrane anchor, CT – cytoplasmic tail. Adapted from [299]. **(B)** Cryo-electron microscopy structure of trimeric spike in the prefusion conformation, with the domains labeled. Adapted from [299]. **(C)** Cryo-electron microscopy structures of trimeric spike in the prefusion and postfusion conformations, with each protomer of the trimer labeled. Adapted from [299]. **(D)** Schematic of SARS-CoV-2 entry pathways. **(1)** Extra cellular virus binds to ACE2, inducing a conformational change in the S1 subunit and exposure of the S2' cleavage site. These steps may be inhibited by ACE2 mimetics or antibody neutralization, which binds the virus and blocks attachment to cellular ACE2 at the plasma membrane. Viral entry diverges here dependent on which proteases are present to cleave S2'. **Left: (2)** In the absence of TMPRSS2, virus is internalized into an endosome, which then **(3)** becomes an acidified endosome in a process blocked by chloroquine and hydroxychloroquine. **(4)** Cathepsins (most notably cathepsin L), which require this acidified environment, then cleave spike at S2', which **(5)** catalyzes membrane fusion and finally **(6)** uncoating of viral RNA. **Right: (2)** In the presence of TMPRSS2, S2' is likely cleaved at the cell surface by TMPRSS2, which is inhibited by camostat. **(3)** This triggers membrane fusion at the plasma membrane and **(4)** results in uncoating of viral RNA. Adapted from [289].

Figure 1.4 (continued):



After completion of S1-S2 cleavage, a site in S2 named S2' must also be cleaved to fully complete triggering and membrane fusion **(Fig 1.4D)** [293, 307]. The location and host proteases responsible for this event vary, occurring either through an endosomal or cell surface entry process depending on the proteases expressed in a particular cell. At the cell surface, S2' is cleaved by a serine protease, primarily TMPRSS2, but can also be cleaved by TMPRSS4, TMPRSS11A, and others [293, 307, 312]. TMPRSS2, highly expressed in the gastrointestinal and respiratory epithelium, has been well characterized as a host protease for multiple coronaviruses and influenza viruses [313]. Alternatively, SARS-CoV-2 can enter through an endosomal-dependent entry pathway. Under these conditions, the S2' site is cleaved by cathepsin proteases, most notably cathepsin L, and to a far lesser extent cathepsin B [293, 294, 314]. Cathepsins play critical roles in entry of other viruses as well, not only other coronaviruses like SARS-CoV, but also reoviruses and ebola virus [278, 315, 316]. Whether SARS-CoV-2 uses a TMPRSS2- or cathepsin-dependent entry pathway seems highly dependent on the cells it infects, with furin cleavage during maturation enhancing entry in a TMPRSS2-dependent manner [293, 294]. For example, entry into vero or 293T cells is cathepsin-dependent, while entry into several lung cell types (caco-2, calu-3, primary human airway epithelial cells) is TMPRSS2-dependent [294]. Given the requirement of host proteases for entry of SARS-CoV-2, as well as broadly for coronaviruses and others, they present an attractive target for drug development [291].

Addressing Questions in Host Determinants of Viral Entry and Replication

The role of cap-proximal m^6A_m, both in the cell and in viral infection, remains poorly understood. We report a novel role for PCIF1-mediated N6-methylation in the

type I interferon response to VSV infection. PCIF1 methylates the cap proximal adenosine, previously 2'O methylated by VSV L, to m^6A_m. In contrast to cellular mRNA, viral mRNA did not require prior guanine N7-methylation for efficient PCIF1-dependent methylation. PCIF1 had no impact on viral mRNA stability, mRNA translation, or replication under basal conditions. However, loss of PCIF1 enhanced the antiviral effect of IFN-β pretreatment in a manner independent of viral mRNA sensing or IFIT1/3 binding. The mRNA cap structures of rabies and measles viruses were also modified by PCIF1. This identifies a function for PCIF1 and cap-proximal m^6A_m in the host response to VSV infection which likely extends to other viruses.

SARS-CoV-2 requires host proteases for entry, including the serine protease TMPRSS2, and chemical inhibition of these proteases are potential targets for antiviral drug development. We discovered covalent small-molecule ketobenzothiazole (kbt) inhibitors of TMPRSS2 structurally distinct from and with improved activity over existing inhibitors camostat and nafamostat. These molecules potently inhibited TMPRSS2 activity and blocked host cell entry into Calu-3 human lung epithelial cells, both with VSV-SARS-CoV-2 and VSV MERS chimeric viruses and authentic SARS-CoV-2. The leading candidate has excellent metabolic stability, safety, and pharmacokinetics in mice, making it a promising drug candidate for COVID-19 treatment.

Chapter 2:

Viral messenger RNA cap methylation by PCIF1 attenuates the antiviral activity of interferon-β

Michael A. Tartell[1,2], Konstantinos Boulias[3,4], Gabriela Brunsting Hoffmann[4], Louis-Marie Bloyet[1], Eric Lieberman Greer[3,4*] and Sean P. J. Whelan[1*]

[1]Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO, USA, [2]Program in Virology, Harvard Medical School, Boston MA, USA, [3]Department of Pediatrics, Initiative for RNA Medicine, Harvard Medical School, Boston, MA, USA [4]Division of Newborn Medicine, Boston Children's Hospital, Boston, MA, USA

***Corresponding author:**

S. P. J. Whelan email: spjwhelan@wustl.edu

E. L. Greer email: Eric.Greer@childrens.harvard.edu

Data within this chapter is published and a reprint of the article is provided with the archived copy of the dissertation. The published manuscript may also be accessed as follows:

PNAS. Jul 20, 2021. Vol. 118. No. 29, e2025769118. doi: 10.1073/pnas.2025769118.

© 2021 by the National Academy of Sciences of the USA

Author contributions:

M.A.T., K.B., L.-M.B., E.L.G., and S.P.J.W. **designed research**; **M.A.T. performed research**; K.B. and G.B.H. contributed new reagents/analytic tools; **M.A.T.,**L.-M.B., E.L.G., and S.P.J.W. **analyzed data**; and **M.A.T.** and S.P.J.W. **wrote the paper**

ABSTRACT

Interferons induce cell intrinsic responses associated with resistance to viral infection. To overcome the suppressive action of interferons and their effectors, viruses have evolved diverse mechanisms. Using vesicular stomatitis virus (VSV), we report that the host cell N6-adenosine mRNA cap-methylase, phosphorylated C-terminal domain interacting factor 1 (PCIF1), attenuates the antiviral response. We employed cell based and *in vitro* biochemical assays to demonstrate that PCIF1 efficiently modifies VSV mRNA cap structures to $m^7Gpppm^6A_m$, and define the substrate requirements for this modification. Functional assays revealed that the PCIF1 dependent modification of VSV mRNA cap structures is inert with regard to mRNA stability, translation and viral infectivity, but attenuates the antiviral effects of treatment of cells with interferon-β. Cells lacking PCIF1 or expressing a catalytically inactive PCIF1 exhibit an augmented inhibition of viral replication and gene expression following interferon-β treatment. We further demonstrate that the mRNA cap-structures of rabies and measles viruses are also modified by PCIF1 to $m^7Gpppm^6A_m$. This work identifies a function of PCIF1 and cap-proximal m^6A_m in attenuation of the host response to VSV infection that likely extends to other viruses.

Significance:

The cap structure present at the 5’ end of eukaryotic mRNAs regulates RNA stability, translation, and marks mRNA as self, thereby impeding recognition by the innate immune system. Cellular transcripts beginning with adenosine are additionally modified at the N6 position of the 2’-O methylated cap-proximal residue by the methyltransferase PCIF1 to $m^7Gpppm^6A_m$. We define a function for this N6-adenosine methylation in attenuating the interferon-β mediated suppression of viral infection. Cells lacking PCIF1, or defective in its enzymatic activity, augment the cell intrinsic suppressive effect of interferon-β treatment on vesicular stomatitis virus and rabies virus gene expression. VSV and RABV mRNAs are efficiently methylated by PCIF1, suggesting this contributes to viral evasion of innate immune suppression.

INTRODUCTION

Eukaryotic messenger RNAs possess a 5' cap structure that functions in their stability, translation, and helps discriminates host from aberrant RNA by the innate immune system [43, 113, 114, 317]. That mRNA cap structure is formed by the actions of an RNA triphosphatase that converts pppRNA to ppRNA which serves as substrate for an RNA guanylyltransferase to transfer GMP derived from GTP onto the 5' end of the RNA to yield GpppRNA [113, 114, 128]. Methylation of that 5' cap-structure by a guanine-N-7 methylase yields m^7GpppRNA, which is modified by a ribose-2'-O methylase to yield m^7GpppN_m [113, 114]. Known activators of the innate immune system include triphosphate RNA which is recognized by the host pattern recognition receptor, retinoic acid inducible gene-1 (RIG-I) [42, 43], and cap-structures that lack ribose-2'-O methylation which renders translation of those RNAs susceptible to inhibition by interferon-induced protein with tetratricopeptide repeats 1 (IFIT1) [43, 249].

Internal RNA modifications also have important functional consequences for the fate of mRNA, among which is N6-methyladenosine (m^6A). The methyltransferase complex METTL3/METTL14 is responsible for m^6A methylation, which regulates diverse functions in mRNA localization, stability, splicing, and translation [171, 172]. The RNA modification N6, 2'O di-methyladenosine (m^6A_m) present at the cap-proximal position ($m^7Gpppm^6A_m$) is regulated separately from m^6A [157]. Cap proximal m^6A_m is present on approximately 30% of cellular mRNA [67-69, 130, 131, 159], but its function is enigmatic. The host RNA polymerase II associated phosphorylated-CTD interacting factor 1 (PCIF1), catalyzes formation of cap proximal m^6A_m [67, 130, 131, 159] that is

reported to increase [67, 131, 162], decrease [130], or have no consequence [163] for mRNA stability and translation.

Vesicular stomatitis virus (VSV), a non-segmented negative-sense RNA virus, replicates in the host-cell cytoplasm transcribing 5 mRNAs from the viral genome [2]. The viral large polymerase protein (L) contains the enzymatic activities necessary for transcription of the 5 mRNAs, including the co-transcriptional addition of a 5' methylated cap-structure (m^7GpppA_m) and synthesis of the 3' poly-A tail [3]. The polymerase synthesizes the mRNAs by recognizing conserved stop and start sequences within each gene so that each mRNA contains an identical 5' structure m^7GpppA_mACAG [51-53]. VSV mRNAs isolated from cells are additionally N6-methylated at the cap-proximal A_m by a presumed cellular methylase to yield $m^7Gpppm^6A_mACAG$ [60]. The efficiency of VSV transcription is such that at least 65% of total cytoplasmic mRNA corresponds to the 5 VSV mRNAs by 6 hours post infection [32]. The 5 VSV mRNAs and their protein products have been extensively characterized biochemically [4], and as a result provide unique probes into the function(s) of m^6A_m.

Here, we demonstrate that VSV mRNAs are efficiently modified at the cap-proximal nucleotide by host PCIF1. In contrast to the substrate requirements for cellular mRNA modification, the PCIF1-dependent N6-methylation of VSV mRNA is independent of prior guanine-N7-methylation of the mRNA cap. Under basal conditions, VSV mRNA stability and translation are unaffected by the presence of m^6A_m, and viral replication is unaltered. Activation of an antiviral response by treatment of cells with interferon-β uncovers a function for PCIF1. Cells lacking PCIF1 or expressing a catalytically inactive variant of the protein exhibit an augmented suppression of viral

gene expression and infection upon interferon-β treatment. The attenuation of the antiviral activity of interferon-β was dependent upon the catalytic activity of PCIF1, thus defining a functional role of this mRNA cap methylation in evading antiviral suppression of gene expression. This work defines a role of PCIF1 dependent methylation of mRNA cap-structures in the attenuation of the antiviral response in VSV infected cells that likely extends to other viruses.

RESULTS

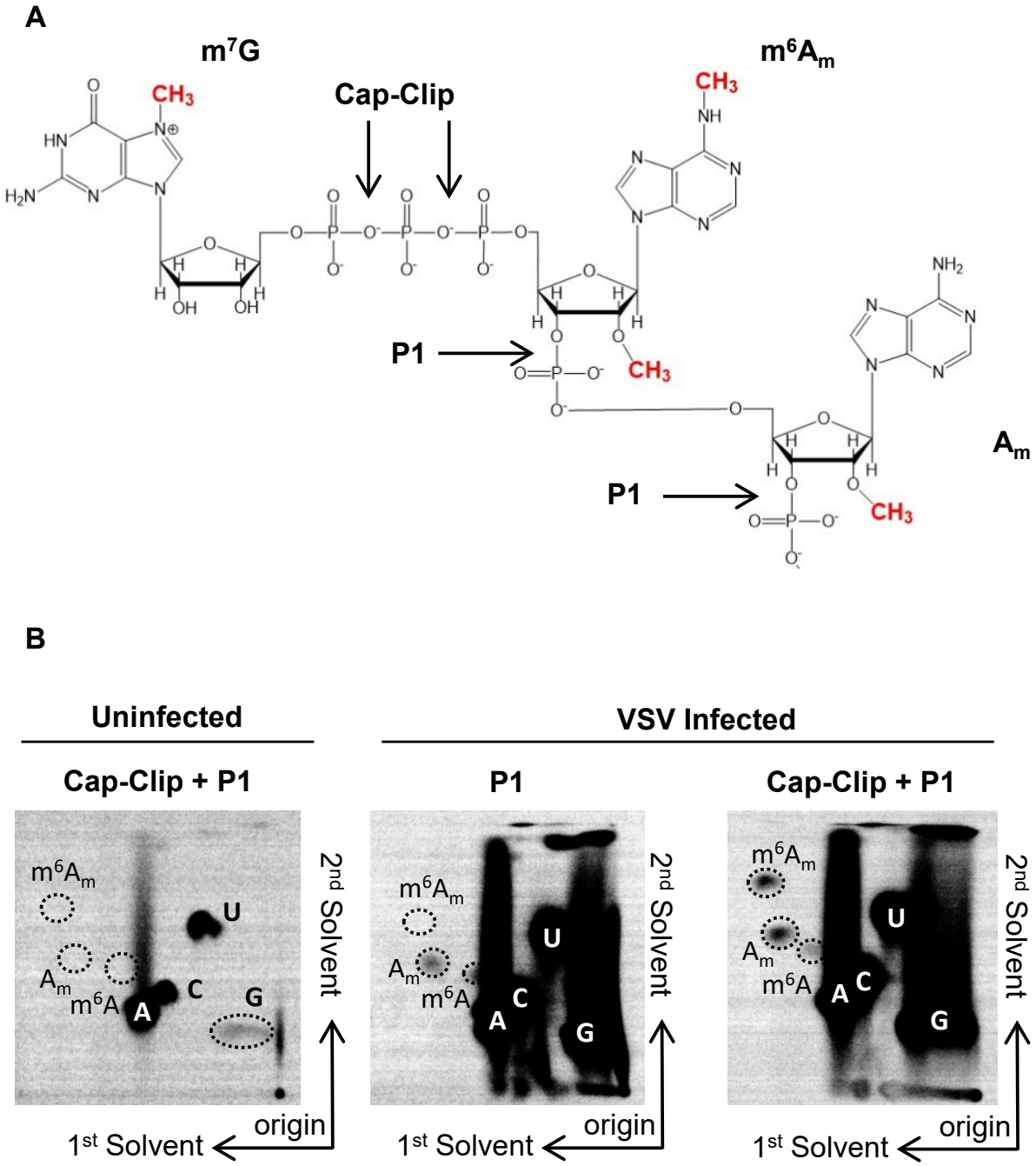
Analysis of VSV mRNA cap-structures isolated from cells.

To examine the methylation status of VSV mRNA cap structures, we infected 293T cells at a MOI of 3, labeled viral RNA by metabolic incorporation of [^{32}P]-phosphoric acid in the presence of actinomycin-D from 3-7 hpi, which selectively inhibits host cell transcription, and isolated total poly(A)+ RNA. Following hydrolysis with nuclease P1 to liberate mononucleotides, and cap-clip acid pyrophosphatase to digest the mRNA cap-structure **(Fig 2.1A)**, the products were resolved by two-dimensional thin-layer chromatography (2D-TLC) [318] **(Fig 2.1B)**. The identity of specific spots was determined by their comigration with co-spotted chemical markers **(Fig A1.S1A)** [162, 318]. Analysis of RNA from uninfected cells yields low levels of products that comigrate with pA, pC, pG, and pU reflecting residual actinomycin-D resistant synthesis of RNA in the cell, but no detectable methylated nucleotides **(Fig 2.1B)**. Nuclease P1 digestion of RNA from infected cells gave rise to abundant pA, pC, pG and pU and two additional spots that comigrate with markers for A_m and m^6A **(Fig 2.1B)**. As nuclease P1 leaves the mRNA cap-structure intact, the presence of A_m and m^6A must reflect internal modifications of the viral mRNA, although our analysis cannot discriminate which positions are modified. Prior studies on VSV mRNAs report the presence of A_m at the second transcribed nucleotide [60], which may account for some of this internal methylation. Further hydrolysis of the RNA purified from infected cells with cap-clip pyrophosphatase leads to the appearance of an additional spot that comigrates with m^6A_m, and an increase in intensity of the A_m spot **(Fig 2.1B)**. This result demonstrates that VSV mRNA cap structures contain $m^7Gpppm^6A_m$, as m^6A_m only appears upon cap-

Figure 2.1: VSV mRNAs contain a 5' $m^7Gpppm^6A_m$ cap-structure.

(A): VSV mRNA cap-structures present in infected cells. VSV mRNAs contain the conserved 5' gene start sequence AACAG, including the m^7GpppA_m cap structures synthesized by the VSV polymerase, N6-methylation at the cap-proximal first nucleotide, and 2'O methylation at the second nucleotide made by the cell. Sites of nuclease P1 and cap-clip pyrophosphatase cleavage are marked. **(B):** VSV contains m^6A_m at the cap-proximal nucleotide. 293T cells were infected with VSV at a MOI of 3, cellular transcription halted by adding 10 μg ml^{-1} actinomycin D at 2.5 hpi, and viral RNA labeled by metabolic incorporation of 100 μCi ml^{-1} [^{32}P] phosphoric acid from 3-7 hpi. Total cellular RNA was extracted and following poly(A) selection hydrolyzed by the indicated nucleases into monophosphates that were resolved by 2D-TLC and detected by phosphorimaging (representative image; n=3). Solvents run in the first and second dimensions are marked.

Figure 2.1 (continued):



structure hydrolysis. We interpret the increase in A_m following cap-clip treatment as reflective of the presence of a minority of transcripts containing only A_m at the cap-proximal nucleotide [60]. Quantitative analysis of each spot reveals over 85% of the cap-proximal nucleotides are m^6A_m, consistent with previous reports [60]. Experiments performed in HeLa cells gave similar results, though no internal m^6A was detectable in these cells **(Fig A1.S1B)**. We conclude that VSV mRNAs synthesized in 293T and Hela cells contain primarily $m^7Gpppm^6A_m$ and $m^7Gpppm^6A_mA_m$ cap-structures.

PCIF1 modifies VSV mRNA

To determine whether VSV mRNA are modified by PCIF1, the cellular N6-methyltransferase for mRNA **(Fig A1.S2)** [67, 130, 131, 159], we infected *PCIF1* knockout (KO) cells and performed RNA analysis as above. Viral mRNA isolated from PCIF1 KO cells (HeLa or 293T) lacked detectable levels of m^6A_m **(Fig 2.2A, Fig A1.S1C)**. Cap-proximal m^6A_m was restored upon add back of PCIF1 but not a catalytically inactive mutant $PCIF1_{SPPG}$ **(Fig 2.2B)**. Methylation reactions performed *in vitro* demonstrate that purified PCIF1 but not $PCIF1_{SPPG}$ are responsible for m^6A_m on VSV mRNA **(Fig 2.2C, A1.S1D).** Collectively these results demonstrate that PCIF1 is necessary and sufficient for formation of cap-proximal m^6A_m on VSV mRNA.

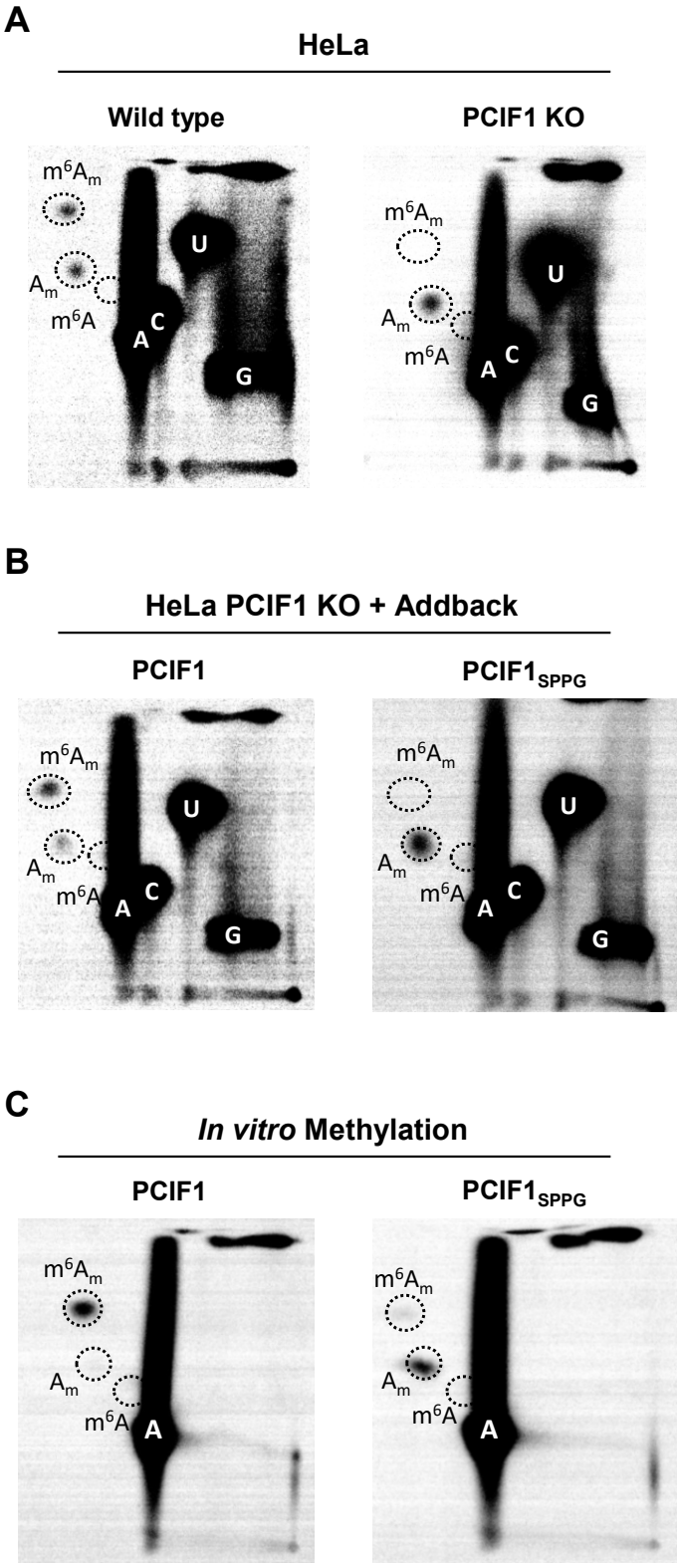
N7-guanosine methylation is dispensable for PCIF1 modification of VSV mRNA

Substrates for modification by PCIF1 require the presence of a methylated m^7G mRNA cap structure [67, 131], and capped RNA lacking m^7G serve as poor substrates for PCIF1 *in vitro* [159]. The obligatory sequential methylation of cellular mRNAs at m^7G and subsequent ribose-2'-O positions [114] precludes tests of the importance of

Figure 2.2: PCIF1 is the cap-proximal N6-methyltransferase.

(A): CRISPR-mediated PCIF1 knockout HeLa cells, or wild type parental cells, were infected with VSV at a MOI of 3. Viral RNA was radiolabeled and analyzed as in Fig 1 (representative images; n=3). **(B):** Add-back of PCIF1, but not a catalytically inactive mutant $PCIF1_{SPPG}$ restores m^6A_m on VSV mRNA. HeLa PCIF1 KO cells stably expressing 3X-FLAG-PCIF1, or 3X-FLAG-$PCIF1_{SPPG}$ were infected with VSV and RNA radiolabeled, digested, and 2D-TLC performed as in A (representative images; n=3). **(C):** PCIF1 N6-methylates VSV mRNA *in vitro*. Purified VSV mRNA, transcribed *in vitro* from viral particles in the presence of [^{32}P]-α-ATP, was used as template for *in vitro* methylation with 50 nM purified PCIF1, or $PCIF1_{SPPG}$. Following hydrolysis the products were visualized by 2D-TLC and phosphorimaging as in Fig 1 (representative images; n=3).

Figure 2.2 (continued):



the 2'-O methylation alone in PCIF1 modification of mRNA. By contrast the mRNA cap methylation reactions of VSV, and by inference other NNS RNA viruses, proceed in the opposite order. Experiments conducted with viral mutants and by reconstitution of cap-methylation *in vitro* demonstrate that cap ribose-2'-O methylation precedes and facilitates the subsequent guanine-N-7 methylation [65], allowing us to explore the impact of 2'-O methylation alone on PCIF1 modification of VSV mRNAs. Messenger RNA synthesized in cells by the viral mutant VSV-L_{G1670A} contains primarily $GpppA_m$ mRNA caps [41], which serve as effective substrates for PCIF1 **(Fig 2.3A)**. In agreement with analysis of viral mRNA made in cells, mRNA transcribed from purified VSV-L_{G1670A} virions are fully methylated by PCIF1 *in vitro* **(Fig 2.3B, A1.S3).** Messenger RNA synthesized by a second viral mutant, VSV-L_{G4A}, that produces unmethylated GpppA cap structures [41], was not modified by PCIF1 **(Fig 2.3C).** A low level of m^6A observed following hydrolysis of RNA produced by VSV-L_{G4A} is consistent with modification at internal positions of the mRNA **(Fig 2.1B).** In agreement with this finding, viral mRNA synthesized by VSV in the presence of the methylation inhibitor *S*-adenosyl-homocysteine (SAH), is poorly methylated by PCIF1 **(Fig 2.3D, A1.S3).** Taken together, this set of results demonstrate in contrast to cellular mRNA modification, VSV mRNA requires ribose-2'-O methylation and not guanine-N-7 methylation.

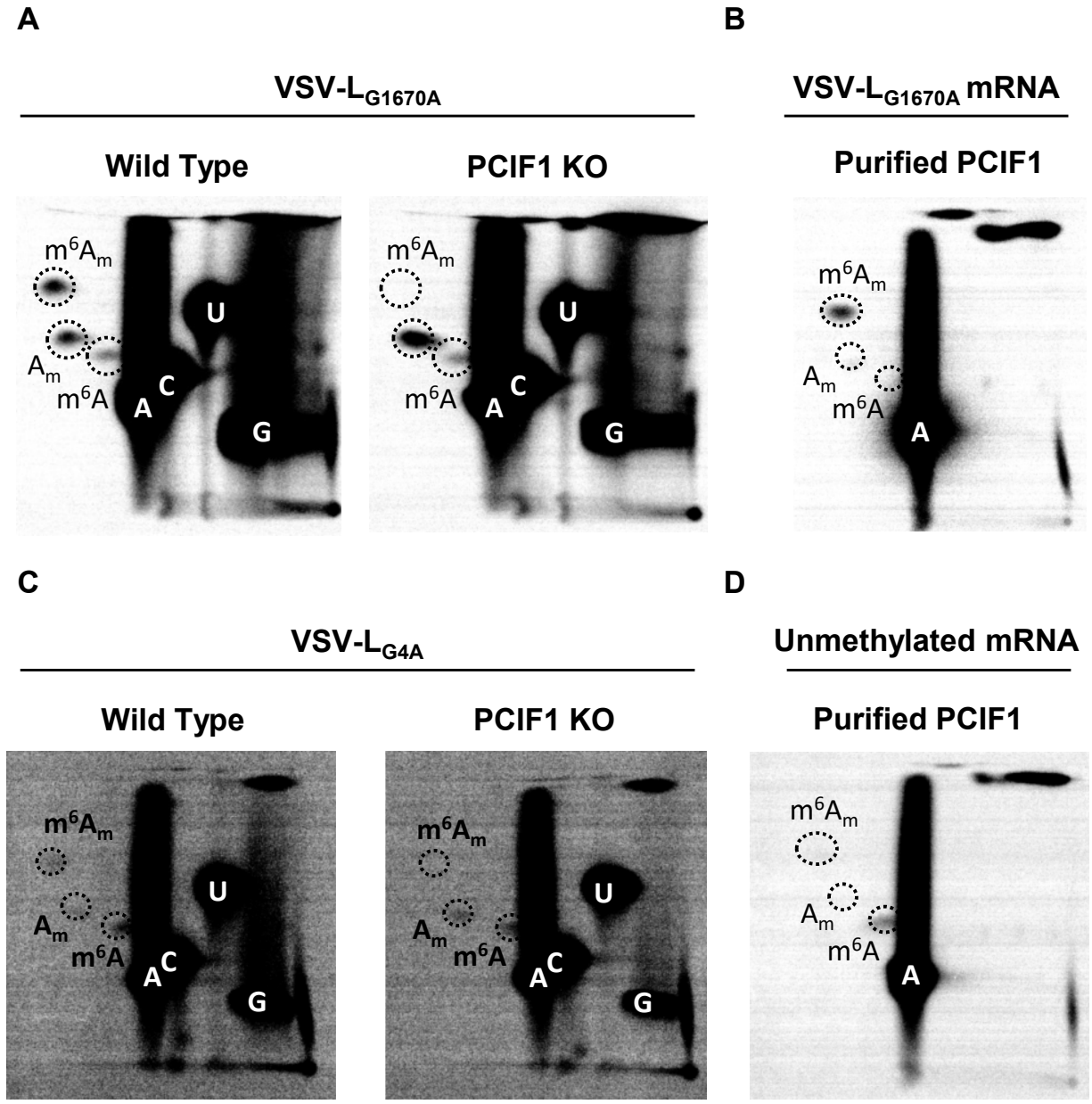
Effect of m^6A_m on viral growth, mRNA stability and translation.

To determine whether PCIF1 modification influences VSV mRNA stability, we compared the decay of modified and unmodified transcripts. Briefly, mRNA was specifically isolated from infected PCIF1 KO cells, *in vitro* methylated with purified PCIF1 where indicated, transfected into uninfected cells, and isolated at the indicated

Figure 2.3: Effect of mRNA cap methylation on PCIF1 modification of VSV mRNA.

(A): The indicated 293T cells were infected with VSV-L_{G1670A} at a MOI of 3 and viral RNA was radiolabeled, extracted and analyzed by 2D TLC as in Fig 2A (representative images, n=3). **(B):** Messenger RNA synthesized *in vitro* by VSV-L_{G1670A} mRNA was incubated with purified PCIF1 and analyzed as in Fig 2C (representative image, n=3). **(C):** The indicated 293T cells were infected with VSV-L_{G4A} as in panel A (representative images, n=3). **(D):** Messenger RNA synthesized by VSV *in vitro* in the presence of 200 μM SAH, was used as substrate for PCIF1 *in vitro* and analyzed as in panel B (representative image, n=3)

Figure 2.3 (continued):



time points. Analysis of the isolated mRNA by electrophoresis on acid-agarose gels demonstrates that viral mRNA stability is unaffected by the presence of m^6A_m (**Fig 2.4A**, two-way ANOVA $p>0.4$). The stability of each individual mRNA appeared unaffected by PCIF1 dependent modification to m^6A_m when transfected into either wild type or PCIF1 knockout cells **(Fig A1.S4)**.

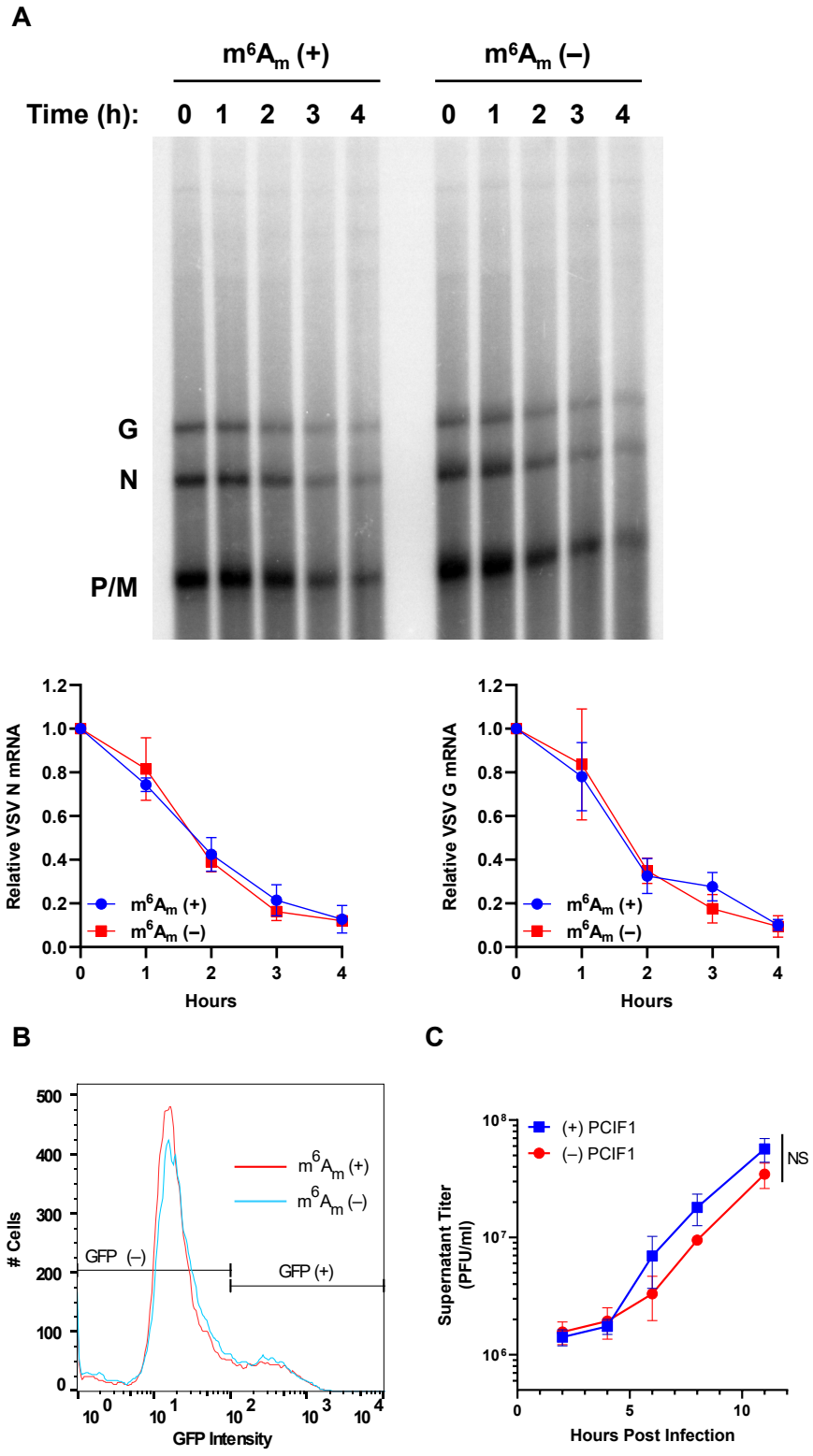
To measure whether mRNA translation was affected by the presence of m^6A_m, we measured the expression of a viral encoded eGFP reporter gene following transfection of methylated or unmethylated mRNA into PCIF1 KO cells. Measurement of eGFP by flow cytometry at 7 h post transfection reveals that neither the fraction of positive cells nor the fluorescence intensity was significantly altered by the presence of m^6A_m **(Fig 2.4B, A1.S5A)**. Similar findings on stability and translation were obtained using a luciferase reporter **(Fig A1.S5B-C)**. To further verify that m^6A_m does not impact the translation of VSV mRNA we co-transfected mRNA from two VSV reporter viruses encoding firefly (Luc) or renilla (Ren) luciferase with opposing methylations. In this competitive translation experiment the ratio of firefly and renilla translated in transfected cells was unaffected by the presence of m^6A_m irrespective of which reporter virus mRNA was modified **(Fig A1.S5D-E)**. Collectively, these results demonstrate that translation of VSV mRNA is unaffected by the presence of m^6A_m.

As neither viral mRNA translation nor stability were altered by the presence of m^6A_m, we next compared the kinetics of viral growth in PCIF1 knockout or addback cells. Cells were infected at a MOI of 3 and viral titers determined by plaque assay at various times post infection. The kinetics of viral growth were also unaffected in cells lacking PCIF1 **(Fig 2.4C, A1.S6)**. Collectively these data demonstrate that despite

Figure 2.4: Effect on m^6A_m on viral mRNA translation and stability

(A-B) VSV mRNA isolated from PCIF1 KO 293T cells and methylated *in vitro* with PCIF1 prior to transfection of 500 ng of RNA into PCIF1 KO HeLa cells and assessed for **(A):** mRNA stability by extraction from cells at the indicated time post-transfection and analyzed by electrophoresis on acid-agarose gels. A representative image is shown along with quantitative analysis of the abundance of the N and G mRNAs (n=3, +/-SD, 2-way ANOVA $p>0.4$). **(B)** mRNA translation by measurement of GFP positive cells and their intensity by flow cytometry (n=3, $0.95>p>0.18$, student's t-test). **(C)** Viral replication assessed in PCIF1 KO HeLa cells expressing 3X-FLAG-PCIF1 or an empty vector following infection with VSV at a MOI of 3. Viral titers were determined by plaque assay on Vero cells at the indicated time post inoculation. (n=3, +/-SD, NS – $p>0.08$, student's t-test, statistics shown are for the 11 h timepoint).

Figure 2.4 (continued):



extensive modification of the viral mRNAs by PCIF1, viral replication and gene expression are unaltered by this modification.

Interferon-β treatment uncovers a role for PCIF1 in the host response to infection.

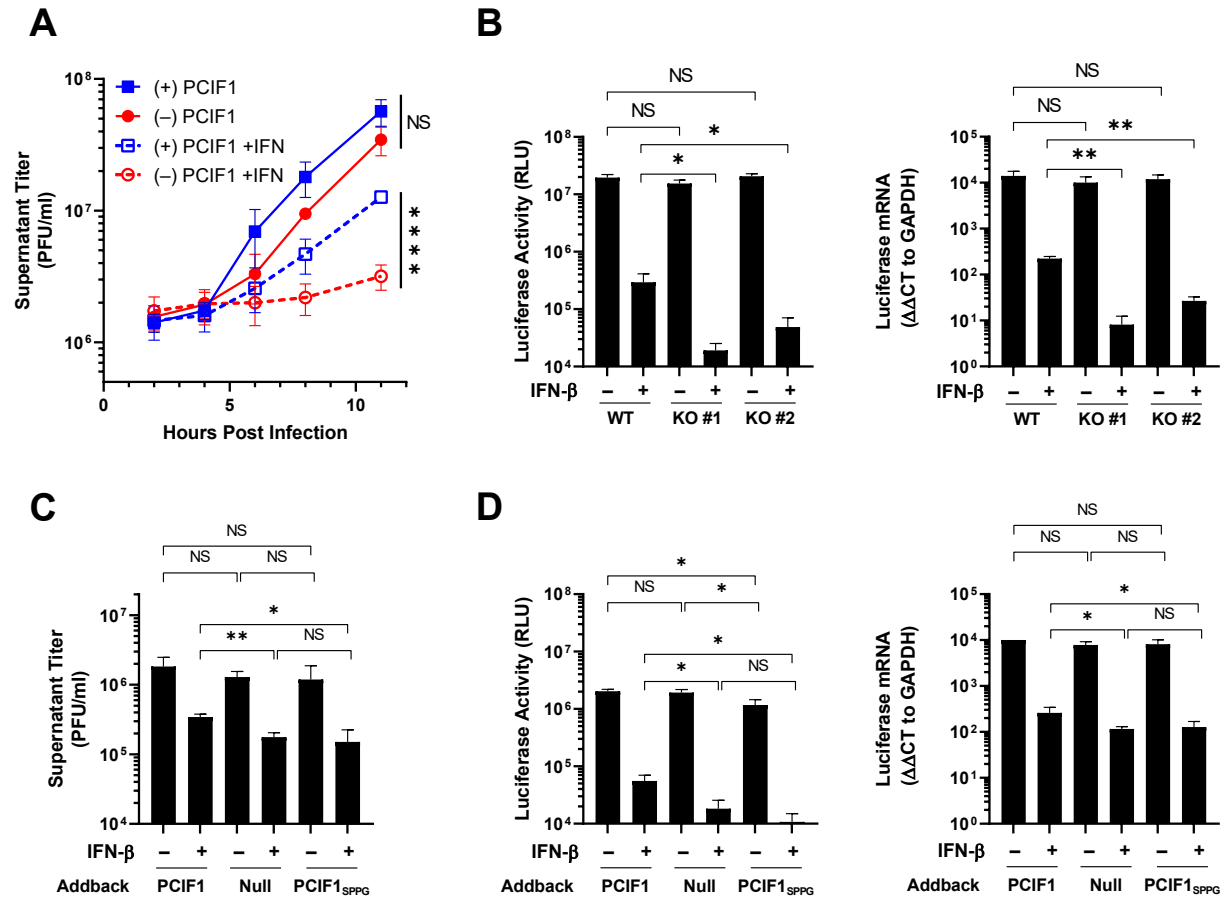
As cap-methylation at the 2'-O position of the first nucleotide helps distinguish self from viral RNA during infection [43] we examined whether PCIF1 modification of mRNA plays a similar role. To examine whether cap-proximal m^6A_m helps counter host cell antiviral responses, we measured how PCIF1 affects the IFN-β mediated inhibition of virus growth. Treatment of cells with IFN-β prior to infection uncovered a PCIF1 dependent attenuation of the antiviral effect **(Fig 2.5A, A1.S6)**. Infection of cells by a VSV-reporter virus that expresses firefly luciferase confirmed that PCIF1 attenuates the suppressive effect of IFN-β on viral gene expression at the RNA and protein levels in a single-round of infection **(Fig A1.S7A)**. We validated this finding in two independent PCIF1 null Hela cell lines, as well as reconstitution with functional PCIF1 **(Fig A1.S7A-B)**. This result suggests that the effects of PCIF1 are restricted to steps of the viral replication cycle up to and including gene expression. To eliminate viral entry as a possible contributor, we transfected ribonucleoprotein cores purified from VSV-Luc into cells, thereby bypassing viral entry. Pretreatment of cells with IFN-β was still accompanied by augmented inhibition of gene expression in cells lacking PCIF1 (Fig A1.S7C). This result demonstrates that the IFN-β mediated suppression of VSV gene-expression is enhanced in cells lacking PCIF1.

The antiviral response in HeLa cells is partly attenuated [319], therefore we examined whether loss of PCIF1 results in a similar IFN-β dependent inhibition of viral

Figure 2.5: Effect of IFN-β pretreatment of cells on viral infection

(A): PCIF1 KO HeLa cells reconstituted with PCIF1 or an empty vector were pretreated with vehicle (0.1% BSA) or 500 U ml^{-1} of interferon-β for 5h and infected with VSV at a MOI of 3. Viral titer was determined at the indicated timepoints by plaque assay on Vero cells (n=3, +/- SD, NS – $p>0.08$, **** $p<0.0001$, student's t-test). **(B):** Wild type (parental) or PCIF1 KO A549 cells were pretreated with vehicle (0.1% BSA) or 500 U ml^{-1} IFN-β for 5 hours prior to infection with VSV expressing a luciferase reporter at a MOI of 5. Cells were lysed 6 hpi and luciferase activity determined as in panel B (n=4, NS – $0.50>p>0.05$, * – $p<0.05$, student's t-test) and mRNA levels verified by qRT-PCR as in panel B (n=3, NS – $0.47>p>0.23$, * – $p<0.05$, ** - $p<0.01$, student's t-test). **(C):** PCIF1 KO HeLa cells were reconstituted with PCIF1, $PCIF1_{SPPG}$ or empty vector, infected with VSV at a MOI of 3, and viral titers measured at 11 hpi by plaque assay (n=3, +/- SD, NS – $0.83>p>0.05$, * – $p<0.05$, ** – $p<0.01$, student's t-test). **(D)**: As in C, except cells were infected with VSV-luc and luciferase protein and RNA measured as in B (n=3, +/- SD, NS – $0.83>p>0.29$, * – $p<0.05$, student's t-test).

Figure 2.5 (continued):



replication in A549 cells. We confirmed that VSV mRNAs were also N6-methylated by PCIF1 in these cells **(Fig A1.S8)**. Pretreatment of A549 cells with IFN-β revealed that, in the absence of PCIF1, viral gene expression was suppressed an additional 10-fold as evident by levels of viral mRNA and protein **(Fig 2.5B)**. Measurements of specific viral proteins following metabolic incorporation of [^{35}S]-met and [^{35}S]-cys followed by analysis of proteins on SDS-PAGE demonstrates that the 3 most abundant viral proteins, N, M and G are further suppressed in cells lacking PCIF1, but cellular translation in uninfected cells is unaffected **(Fig A1.S9)**.

To rule out the possibility of N6-methylation independent activities of PCIF1 mediating this effect, we examined infection in cells expressing a catalytically-inactive mutant of PCIF1, PCIF1$_{SPPG}$ **(Fig 2.2B)**. Both PCIF1$_{SPPG}$ addback and PCIF1 knockout equivalently augmented the effect of IFN-β on viral growth **(Fig 2.5C)** and VSV-luciferase gene expression **(Fig 2.5D)**. Collectively, the above experiments reveal that loss of PCIF1 or its ability to synthesize m^{6}A$_m$ augments the suppressive effect of IFN-β on VSV gene expression, suggesting that m^{6}A$_m$ methylation of viral mRNAs protects against the otherwise antiviral effects of the IFN-mediated innate immune response.

Effect of PCIF1 on innate immune sensing and IFIT proteins

We next examined whether m^{6}A$_m$ functionally impacts known discriminators of the cap-proximal nucleotide in innate immunity, specifically RIG-I and IFIT1/3 [43]. To determine whether sensing of VSV mRNA, for which RIG-I is essential [232], was affected by PCIF1 modification of viral mRNA, we measured induction of IFN-β mRNA during infection. IFN-β mRNA levels were indistinguishable in cells independent of the

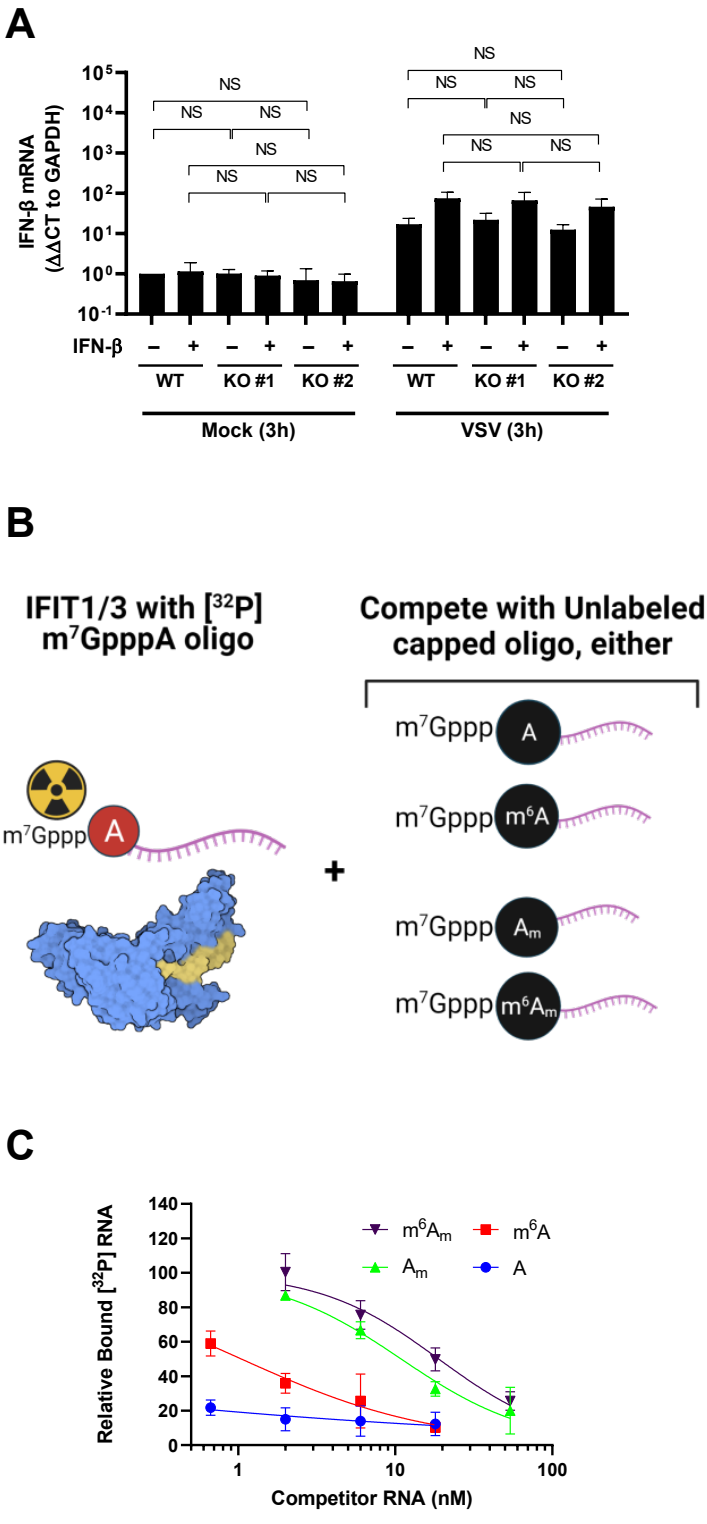
presence of PCIF1 at 3 hours **(Fig 2.6A)** and 6 hours post infection **(Fig A1.S10)**. This result demonstrates that loss of PCIF1 does not alter the RIG-I dependent sensing of VSV RNA in cells.

We next examined whether IFIT1, which represses translation of mRNAs lacking cap-proximal 2'O methylation [249] also discriminates against N6-methylation. To do this, we established a competitive binding assay for capped RNA substrates with a complex of purified IFIT1 and IFIT3 [250]. Briefly, we used as binding substrate a radiolabeled and capped 27-nt synthetic RNA corresponding to the first 27-nucleotides of the VSV genome in competition with unlabeled RNAs containing a cap-proximal, A, m^6A, A_m, or m^6A_m **(Fig 2.6B-C)**. As expected, binding of the radioactive 27-nt capped RNA was efficiently out competed by the identical cold competitor (IC_{50}: < 0.66 nM; **Fig 2.6C**), confirming that IFIT1/3 efficiently binds m^7GpppA capped transcripts [249, 250]. Modification of the competitor RNA to contain a m^7GpppA_m cap structure decreased its effectiveness > 16-fold consistent with published results (IC_{50}: 10.89 nM; **Fig 2.6C**) [249, 250], and further N6 methylation to $m^7Gpppm^6A_m$ resulted in a slight further reduction (IC_{50}: 18.88 nM; **Fig 2.6C**) that failed to rise to the level of statistical significance. At high concentrations of PCIF1 it is possible to N6-methylate cap-structures in vitro that are not 2'-O methylated. Such an m^7Gpppm^6A cap structure competes for IFIT1/3 binding slightly better than m^7GpppA (IC_{50}: 1.05 nM; **Fig 2.6C**). As such RNAs require 2'O methylation to be substrates for PCIF1 in VSV and cellular mRNA, we conclude that the PCIF1 dependent modification of viral RNA does not significantly alter its recognition by IFIT1/3.

Figure 2.6: PCIF1 does not regulate VSV mRNA sensing or IFIT1/3 binding

(A): The indicated A549 cells were pretreated with vehicle (0.1% BSA) or 500 U ml^{-1} IFN-β for 5 hours prior to infection with VSV at a MOI of 5 for 3 hours. Cells were lysed and IFN-β mRNA levels determined by qRT-PCR (n=4, +/- SD, NS – $0.08 < p < 0.80$, student's t-test). **(B):** Schematic for testing IFIT1/3 affinity for differentially methylated cap structures (structure shown from PDB 6C6K [250]). IFIT1-IFIT3 complexes (1 μM) were bound to 1 nM radiolabeled RNA with a m^7GpppA cap in the presence of unlabeled competitor RNAs at the indicated concentration with the indicated cap-proximal nucleotide. High affinity of the competitor for IFIT1/3 will result in displacement of the radiolabeled substrate and loss of signal. Low affinity will allow the radiolabeled substrate to remain bound. **(C):** Radiolabeled RNA bound to IFIT1/3 was measured by filter-binding assay and phosphorimaging, normalized to a control without competitor RNA (n=3, +/- SD, least squares regression line shown).

Figure 2.6 (continued):



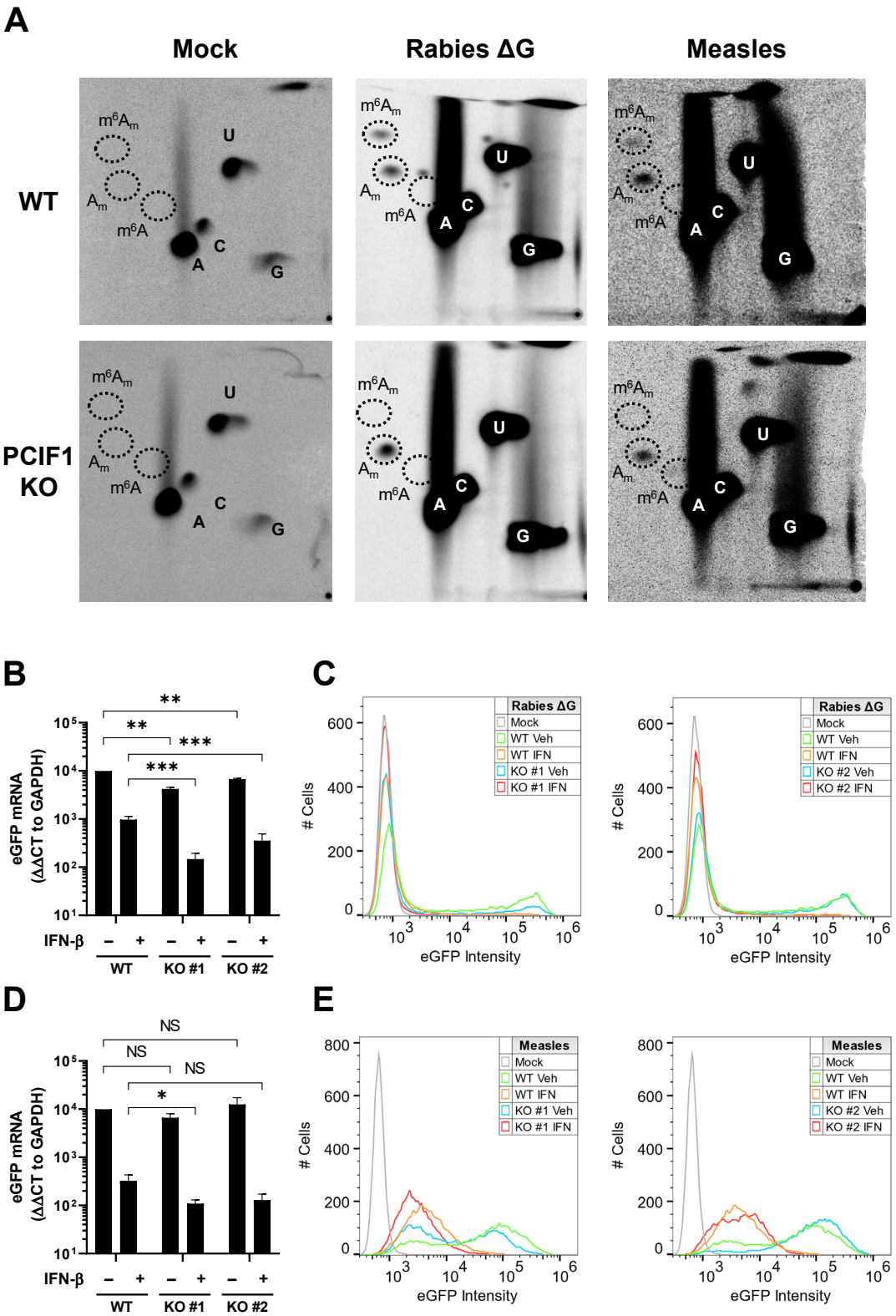
Effect of PCIF1 on messenger RNA from other negative-sense RNA viruses

We next examined whether PCIF1 modifies the mRNA of other negative-sense RNA viruses that replicate in the cytoplasm. For this purpose, we selected rabies virus from the same family as VSV, and whose mRNAs start with the same sequence, and measles virus from the family paramyxoviridae. Both viruses synthesize m^7GpppA_m cap structures [320-322] that were further modified in a PCIF1 dependent manner to $m^7Gpppm^6A_m$ **(Fig 2.7A)**. We note, however, that the PCIF1 dependent methylation of measles virus RNA appears inefficient **(Fig 2.7A)**. We also observed a PCIF1 dependent attenuation of the effect of interferon-β pretreatment of cells on rabies virus gene expression at both the RNA and protein level **(Fig 2.7B-C, A1.S11A)**. Reflecting the inefficient modification of measles virus mRNA we observed an approximately ~2.5 fold PCIF-1 dependent attenuation of the suppressive effect of interferon-β on measles virus mRNA levels (clone #1, p=0.048; clone #2, p=0.072) **(Fig 2.7D-E, A1.S11B)**. Collectively this data demonstrates that PCIF1 N6-methylates other negative-sense RNA virus mRNAs that can serve to attenuate the interferon-β dependent antiviral response.

Figure 2.7: Effect of PCIF1 on other non-segmented negative-stranded RNA viruses

(A): The indicated A549 cells were infected with the indicated viruses at a MOI of 3, cellular transcription halted by adding 10 μg ml^{-1} actinomycin D at 5.5 hpi, and viral RNA labeled by metabolic incorporation of 100 μCi ml^{-1} [^{32}P] phosphoric acid from 6-24 hpi. Viral RNA was extracted and analyzed by 2D-TLC as in Fig 1B. **(B):** The indicated A549 cells were pretreated with vehicle (0.1% BSA) or 500 U ml^{-1} IFN-β concurrent with infection with RABV ΔG at a MOI of 1 for 24 hours. Cells were lysed and eGFP mRNA levels determined by qRT-PCR (n=3, +/- SD, ** – $p < 0.01$, *** – $p < 0.001$, student's t-test). **(C):** Same as B, except cells were fixed in formaldehyde and flow cytometry performed for eGFP expression (n=3, representative histograms shown). **(D):** The indicated A549 cells were pretreated with vehicle (0.1% BSA) or 500 U ml^{-1} IFN-β for 5 hours prior to infection with measles virus at a MOI of 3 for 24 hours. Cells were lysed and eGFP mRNA levels determined by qRT-PCR (n=3, +/- SD, NS – $0.06 < p < 0.47$, * – $p < 0.05$, student's t-test). **(E):** Same as D, except cells were fixed in formaldehyde and flow cytometry performed for eGFP expression (n=3, representative histograms shown). .

Figure 2.7 (continued):



DISCUSSION

The major finding of this study is the identification of a role for PCIF1-mediated m^6A_m methylation in the type I interferon response to VSV infection. We demonstrate that loss of PCIF1 enhances the sensitivity of viral replication to pretreatment of cells with IFN-β by affecting VSV gene expression. PCIF1 is necessary and sufficient for modification of VSV mRNA to yield cap-proximal m^6A_m, and in contrast to cellular mRNA modification, viral mRNAs require prior ribose-2'-O but not guanine-N-7 methylation of the cap-structure. The most parsimonious explanation of our results is that the PCIF1 dependent modification of viral mRNA cap-structures to m^6A_m serves to dampen an IFN-β mediated suppression of gene expression. Mechanistically, how this occurs was not resolved by the present study but we posit that this requires discrimination of modified from unmodified RNA by an interferon stimulated gene (ISG).

Precedent for a role of mRNA cap modifications in the antiviral response already exists. The RIG-I dependent recognition of a 5' triphosphate is suppressed by the presence of an mRNA cap-structure [42, 43], and ribose 2'-O methylation of the cap-structure inhibits the ability of an ISG, IFIT1, to suppress translation of mRNA [43, 249]. The PCIF1 dependent modification of VSV mRNA cap-structures may work by a similar mechanism by helping viral mRNA appear more host like, but this modification has no significant effect on the RIG-I or IFIT1 dependent recognition of viral mRNA (Fig 6). This finding is entirely consistent with the known requirements of a 5' triphosphate for RIG-I dependent recognition. Although we observed that generation of an artificial m^7Gpppm^6A cap structure in vitro reduced binding to IFIT1 it seems unlikely that this ISG plays a major role in this discrimination. This conclusion is based on the known

recognition of m^7GpppA cap-structures by IFIT1 [43, 249, 250] and the requirement for 2'-O modification of mRNAs for their subsequent PCIF1 modification in cells [67, 130, 131]. Additional work will be required to define whether an ISG is required to discriminate between m^6A_m modified and unmodified cap-structures.

Although we establish a role of PCIF1 and m^6A_m in the IFN-β mediated suppression of VSV gene expression, and demonstrate that viral mRNAs are modified by PCIF1, modification of cellular mRNA may also play a role. Cap proximal m^6A_m inhibits the host mRNA decapping enzyme DCP2 [162], which may alter stability of cellular mRNAs including those induced on treatment of cells with IFN-β [323]. This seems unlikely to account for the effects we observe on VSV infection, as the DCP2 dependent decapping and degradation of cellular mRNA would be expected to increase in cells lacking PCIF1, likely dampening rather than augmenting the effect of IFN-β treatment. If the antiviral response is due to m^6A_m modification of cellular mRNA this contrasts with the consequences of 2'-O methylation of the mRNA cap structure of ISG mRNAs which enhances their expression [266]. We are also mindful of the possibility that PCIF1 may have unknown functions in the cell beyond N6-methylation of mRNA. Insects, including *Drosophila*, express an ortholog of PCIF1 that associates with the phosphorylated CTD of PolII, but is catalytically inactive as an RNA N6-methyltransferase [324]. As the catalytic activity of PCIF1 is required for attenuation of the antiviral response we also find this explanation unlikely.

The substrate requirements for PCIF1 modification of VSV mRNA differ to those previously shown in cellular mRNA [67, 159]. Specifically, we found that guanine-N7-methylation was dispensable for N6-methylation, and that ribose 2'-O methylation was

required in its absence. Although we do not understand why this is the case, we recapitulate the substrate specificity *in vitro* making it unlikely that this distinction reflects the cytoplasmic modification of viral mRNA rather than the nuclear modification of cellular mRNA. This altered specificity for modification of the VSV mRNA coupled with the altered recognition specificity of the VSV cap methylation machinery - which requires 2'-O methylation prior to guanine-N-7 methylation - raise the possibility that the structure of the 5' end of VSV mRNAs leads to the altered specificity [41].

The finding that PCIF1 and m^6A_m affect the antiviral response raises the question of why modify cap-proximal A and not other cap-proximal bases. More cellular mRNAs initiate with G than A [68], but guanosine is typically only methylated at the 2'O position in mRNA [231]. This is likely because O6-methylation of guanosine (m^6G) has been shown to have a large fitness cost. Its presence in DNA is highly mutagenic though pairing with thymidine during DNA replication, and when present in RNA, it causes incorrect ribosome decoding [325] and a 1000-fold decrease in the peptide-bond formation rate [326]. N6-methylation in adenosine, by contrast, has minimal effect on these processes [325, 326].

The importance of m^6A_m for most viruses has not been examined. As expected, the mRNAs of DNA viruses which rely on host RNA polymerases for transcription, including adenoviruses [150, 151], simian virus 40 [152], herpes simplex virus 1 [153], and polyomaviruses [154] contain m^6A_m [113, 114]. Vaccinia virus, which replicates in the cytoplasm, also produces m^6A_m containing mRNA [113, 114, 327], likely through modification by PCIF1. Whether such modifications also attenuate the antiviral response remains to be determined. We observed a PCIF-1 dependent attenuation of the

suppressive effect of interferon-β on rabies virus and to a lesser extent measles virus. What impacts the magnitude of PCIF1 dependent attenuation of the suppressive effect of interferon-β between different viruses was not determined. We suggest that this may be related to the different extent of PCIF1 modification of viral RNA, the different kinetics with which VSV, rabies and measles replicate and differences between the viruses in how they regulate antiviral signaling. The evolution by many viruses of their own capping machinery also begs the question of whether viruses have evolved a PCIF1-like cap modifying enzyme, particularly given the potential advantage in the face of an antiviral response. Additional studies with VSV and other viruses will be required to fully define the role of PCIF1 in the host response to infection.

MATERIALS AND METHODS

Cells:

HEK293T, HeLa, A549, Vero CCL81, and BsrT7/5 cells were maintained in humidified incubators at 37 ºC and 5% CO_2 in DMEM supplemented with L-glutamine, sodium pyruvate, glucose, (Corning #10013CV) and 10% fetal bovine serum (Tissue Culture Biologicals #101). Generation of HEK293T and HeLa *PCIF1*-knockout cell lines and addbacks were previously described [131], and A549 *PCIF1*-knockout cell lines were generated and verified using these same methods. Cells were tested regularly using the e-Myco PLUS PCR kit (Bulldog Bio #2523348).

Viruses:

VSV (as rescued from an infectious cDNA clone of VSV, pVSV(1)+), VSV-L_{G1670A}, VSV-L_{G4A}, VSV-luciferase, VSV-RenillaP, VSV-IRES and VSV-eGFP have been described previously [10, 41, 275, 278, 328, 329]. VSV viruses were propagated in BsrT7/5 cells. Rabies ΔG expressing eGFP was kind gift of E. M Calloway [330], and propagated in BsrT7/5 cells expressing SAD B19 RABV G. Measles expressing eGFP was propagated in Vero cells.

Radiolabeling of mRNA:

Cap-proximal nucleotides were radiolabeled as previously described [162]. For specific radiolabeling of VSV mRNA, cells were infected with VSV at a MOI of 3 in serum/phosphate-free DMEM (Gibco #11971-025). 10 μg ml^{-1} Actinomycin D (Sigma #A5156) was added to halt cellular transcription at 2.5 hpi, and at 3 hours post infection, 100 μCi ml^{-1} [^{32}P] phosphoric acid added to label newly synthesized RNA (Perkin Elmer #NEX053H). RNA was harvested in Trizol (Thermofisher #15596018) and poly(A)+

mRNA selected using the NEB Magnetic mRNA Isolation Kit (NEB S1550S). Radiolabeling of measles and rabies mRNA was performed similarly, except ActD was added at 5.5 hours post infection, and radiolabeling continued until 24 hours post infection

Identification of methylated nucleotide levels by two-dimensional thin layer chromatography (2D-TLC):

2 µg mRNA suspended in 6 µl RNAse-free H_2O was digested with 2 units of nuclease P1 (Sigma N8630) for 3 h at 37 ºC. The volume was then increased to 20 µl and RNA further digested with 2 units of Cap-Clip Acid Pyrophosphatase for 3 h (Cell Script #C-CC15011H) in the manufacturer's buffer. 2D-TLC was performed as previously described [318]. Plates were developed in the first dimension with 5 parts isobutyric acid (Sigma #I1754) to 3 parts 0.5 M ammonia (VWR #BDH153312K) for 14 h, and in the second dimension with a solvent of 70 parts isopropanol, 15 parts hydrochloric acid, and 15 parts water for 20 h. RNA species were positively identified by UV-shadowing (254 nm) of co-spotted (non-radioactive) commercially available standards (5' monophosphate forms). The standards A_m and m^6A_m 5' monophosphate were generated by digesting their triphosphate forms (TriLink N-1015 and N-1112) with 1 unit Apyrase (NEB M0398).

In vitro transcription of VSV mRNA and methylation with PCIF1:

VSV or VSV-L_{G1670A} mRNA was synthesized *in vitro* as previously described with 30 µCi [^{32}P]-α-ATP per reaction [41, 331]. RNA was extracted in trizol, poly(A) selected, and extracted in trizol again to concentrate the samples. Purified recombinant GST-

PCIF1 was generated and used to in vitro methylate 225 ng of this mRNA as previously described [131].

Purification, selection, and in vitro methylation of VSV mRNA from cells:

HEK 293T *PCIF1* KO cells were infected at a MOI of 10 for 7 h. RNA was extracted in trizol, and VSV mRNAs selected using a biotinylated oligo against the conserved stop and poly(A) sequence present at the 3' end of VSV mRNAs ("Biotin-VSVstop"). 1.5 nmol oligo was annealed to this mRNA by incubating at 65 ºC for 5 min, followed by cooling on ice for 5 min, and complexes isolated by pulldown with NEB Streptavidin Magnetic Beads (NEB #S1420, manufacturer's protocol). Following cleanup by trizol extraction, mRNA was purified further using the NEB poly(A) magnetic kit as above, and trizol extracted again. This stock of RNA from PCIF1 KO cells was then *in vitro* methylated (or mock methylated) with purified PCIF1 as above.

Determination of VSV mRNA stability:

Biotin-VSVstop selected mRNAs were transfected into HeLa WT or *PCIF1* KO cells. RNA was transfected into separate wells of a 24 well plate ($2x10^5$ cells; 500 ng RNA per well) using Lipofectamine 2000 (Thermofisher #11668019), media was changed at 3 h post transfection, and wells harvested in trizol at 1 h intervals for 4 h. Extracted mRNAs were separated by electrophoresis on acid-agarose gels, which were dried and exposed to a phosphor screen. VSV mRNAs levels were quantified using ImageQuant version 8.2, and normalized to the 0 h timepoint.

Transfection and flow-cytometry analysis of translation of VSV mRNA:

500 ng Biotin-VSVstop selected VSV mRNAs were transfected into $2x10^5$ HeLa WT or *PCIF1* KO cells using Lipofectamine 2000. At 6 hours post transfection, cells

were trypsinized, and washed and resuspended in PBS. Half the cells were analyzed for GFP expression by flow cytometry (BD FACS Calibur); GFP positive cells and mean fluorescence intensity of GFP were calculated in FlowJo (20,000 cells analyzed per replicate). Studies of gene expression with GFP-expressing rabies and measles virus were performed similiarly (24 hours post infection), except flow cytometry was performed on a Beckman Coulter Cytoflex S.

Determination of VSV mRNA gene expression by luciferase luminescence and RT-qPCR:

$2x10^5$ cells (24 well format; transfection experiments) or $4x10^5$ cells (12 well format; IFN experiments) were lysed in 120 µl passive lysis buffer (Promega #E1941). Half the lysate was used to quantify luciferase protein (Promega Luciferase Assay System #E1501) using a Spectramax L luminometer with reagent injectors in technical triplicate. RNA was extracted from the other half of cells in trizol, and 1 µg reverse transcribed using SuperScript III (Invitrogen #18080044), oligo-dT primers (IDT #51011501), and RNase inhibitors (Promega #N2515). Real-time qPCR was performed using Fast SYBR Green (Thermofisher #4385612) in technical duplicate. Relative RNA was calculated as ΔΔCT (normalized to GAPDH) times 10^4.RT-qPCR was performed similarly for rabies and measles virus expressing eGFP.

Detection of Radiolabeled Samples:

Gels were fixed in 30% methanol, 10% acetic acid, washed twice in methanol, and dried using a vacuum pump gel dryer. TLC plates were air dried. Dried gels or plates were then exposed to a phosphor screen and scanned on a Typhoon scanner.

Growth Curve with IFN pretreatment:

HeLa cells were pretreated with 500 U ml^{-1} IFN-β (Tonbo Biosciences 21-8699) or vehicle (0.1% BSA) for 5 h in serum free DMEM. Cells were washed, and infected with VSV at a MOI of 3 for 1 h in serum free DMEM. After 1 hour, the inoculum was removed, cells washed, and supplemented with 2% FBS. At 2, 4, 6, 8, and/or 11 hours post infection, 1% of the supernatant was removed and frozen at -80 ºC. After all samples had been collected, viral titers were determined by plaque assay on vero cells.

Gene expression with IFN pretreatment:

HeLa or A549 cells were pretreated with 500 U ml^{-1} IFN-β as above for 5 h, infected with VSV-Luc at a MOI of 3, and at 6 hours post infection, cells were lysed and processed for luciferase protein and mRNA quantitation. Alternatively, VSV-Luc RNPs were purified as previously described, and 50 ng transfected into HeLa cells instead of infection with virus.

Metabolic Radiolabeling of Protein:

$4x10^5$ A549 cells were pretreated with 500 U ml^{-1} IFN-β as above for 5 h, and infected (or uninfected control) with wild type VSV at MOI 5. At 5 hours post infection, cells were washed, and the media changed to methionine/cysteine free DMEM (Gibco #21013-024). After 40 minutes of starvation, cells were pulse labeled with 30 µCi ml^{-1} [^{35}S] methionine (Perkin Elmer #NEG009T) and 30 µCi ml^{-1} [^{35}S] cysteine (Perkin Elmer #NEG022T) for 20 minutes. Cells were then lysed in SDS sample buffer and run on a low-bis 10% SDS-PAGE gel. Protein translation was determined by phosphorimaging as above. Equal loading was determined by staining with 0.25% Coomassie Brilliant Blue G-250.

IFIT1-IFIT3$_{CTD}$ Complex purification and Filter Binding Assay:

Protein complexes were purified as previously described [250]. MBP and His-tagged IFIT1 and the terminal 87 amino acids of IFIT3 ($IFIT3_{CTD}$) were expressed in Rosetta 2 *E. coli* cells (Novagen, plasmids kind gift of G. Amarasinghe), cultured in Luria Broth at 37 ºC, induced at OD_{600} of 0.6 with 0.5 mM IPTG for 3 hours, resuspended in lysis buffer with 25 mM sodium phosphate (pH 7.5), 250 mM NaCl, and 5 mM 2-mercaptoethol, lysed using a Constant Systems CF1 Cell Disrupter, clarified by centrifugation at 20,000 x g at 4 ºC for 1 hour, and purified by an imidazole-gradient elution from a His-trap purification column (GE) using an AKTA chromatography system (GE). Purified IFIT1 and $IFIT3_{CTD}$ were combined and tags cleaved using TEV protease at 4 ºC for 14 hours. IFIT1-$IFIT3_{CTD}$ complexes were selected by size exclusion using an S200 column (GE), and concentration determined by A280 absorbance.

RNA substrates were prepared as follows. Triphosphate RNA with the 27 nt VSV leader sequence (5'(ppp)ACGAAGACAAACAAACCAUUAUUAUCA 3') were purchased from Trilink Biosciences. Radiolabeled substrate with a m^7GpppA cap was prepared by capping with vaccinia capping system (NEB) and [^{32}P]-GTP (Perkin Elmer) according to manufacturer's instructions. Unlabeled substrate with a m^7GpppA or m^7GpppA_m cap were prepared by capping with vaccinia capping system (NEB) and/or 2'O methyltransferase (NEB). m^7Gpppm^6A or $m^7Gpppm^6A_m$ substrates were prepared by further methylation with purified PCIF1 as described above. To eliminate contaminating uncapped (triphosphate) RNA, samples were further digested by sequential incubation with shrimp alkaline phosphatase (NEB), polynucleotide kinase (NEB), and terminator exonuclease (Lucigen). Samples were trizol extracted into nuclease free water before use, and concentrations measured by Qubit RNA High Sensitivity kit (Invitrogen).

Filter-binding assay was performed as previously described [250, 332]. IFIT1-IFIT3$_{CTD}$ complex (1 μM) was incubated with 1 nM radiolabeled m^{7}GpppA RNA and the indicated concentrations of cold RNA in a total reaction volume of 20 μl in a 96 well plate, including controls with no cold RNA. The binding buffer was 20 mM Tris pH 7.4, 200 mM NaCl, 1 mM DTT, 50 μg/ml BSA, and 10% glycerol. Reactions were incubated for 20 minutes at 37 ºC followed by 40 minutes at 4 ºC, applied to nitrocellulose and nylon membranes using a 96-well dot blot apparatus (Bio-rad), and exposed to a phosphor screen. Relative bound RNA was calculated by dividing the signal from competitor RNA samples from the control reaction without competitor (ImageQuant TL 8.2 software). Least squares fit curves were calculated using Prism 9.

Data analysis and replicates:

All experiments were performed with n=3 or n=4 biological replicates (as indicated). Each qPCR biological replicate is the average of technical duplicates from the same sample. All qPCR biological replicates (except Fig 5F, 6A, 7B-D) were run and analyzed on the same plate, enabling a standard deviation to be calculated for all samples. Each luciferase biological replicate is the average of technical triplicates from the same sample. Statistical tests were performed in Microsoft Excel, graphs were generated in Graphpad Prism 9. Schematics were created using Biorender.com.

Acknowledgments:

Thanks to members of the S.P.J.W. laboratory who provided insightful discussions and advice. Thanks also to S. Hur and G. Amarasinghe for advice on RIG-I and IFIT1/3. This project was supported by NIH Grants F31AI138448 (to M.A.T), T32AI007245 (to S.P.J.W.), AI059371 (to S.P.J.W.), DP2AG055947 (to E.L.G), and R21HG010066 (to E.L.G).

Chapter 2:

A novel class of TMPRSS2 inhibitors potently block SARS-CoV-2 and MERS-CoV viral entry and protect human epithelial lung cells

Matthew Mahoney[1, 2, ±], Vishnu C. Damalanka[1, ±], Michael A. Tartell[3, 4, ±], Dong hee Chung[5], André Luiz Lourenço[5], Dustin Pwee[6], Anne E. Mayer Bridwell[3], Markus Hoffmann[7, 9], Jorine Voss[1], Partha Karmakar[1], Nurit P. Azouz[8], Andrea M. Klingler[8], Paul W. Rothlauf[3, 4], Cassandra E. Thompson[3], Melody Lee[5], Lidija Klampfer[2], Christina Stallings[3], Marc E. Rothenberg[8], Stefan Pöhlmann[7, 9], Sean P. Whelan[3], Anthony J. O'Donoghue[6], Charles S. Craik[5], James W. Janetka[1, 2]*

[±]Equal contributions

[1]*Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, Saint Louis, Missouri, United States*

[2]*ProteXase Therapeutics, Inc., Saint Louis, Missouri, United States*

[3]*Department of Molecular Microbiology, Washington University School of Medicine, Saint Louis, Missouri, United States*

[4]*Program in Virology, Harvard Medical School, Boston, Massachusetts, United States*

[5]*Department of Pharmaceutical Chemistry, University of California, San Francisco, California, United States*

[6]*Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, California, United States*

[7]*Infection Biology Unit, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany*

[8]*Division of Allergy and Immunology, Cincinnati Children's Hospital Medical Center and Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio, United States.*

[9]*Faculty of Biology and Psychology, Georg-August University Göttingen, Göttingen, Germany*

*Correspondence to JWJ; janetkaj@wustl.edu

Data within this chapter is published and a reprint of the article is provided with the archived copy of the dissertation. The published manuscript may also be accessed as follows:

PNAS. Oct 26, 2021. Vol. 118. No. 43, e2108728118. doi: 0.1073/pnas.2108728118.
© 2021 by the National Academy of Sciences of the USA

Author contributions: **M.A.T.**, D.h.C., A.L.L., C.L.S., M.E.R., S.P., S.P.J.W., A.J.O., C.S.C., and J.W.J. **designed research**; M.M., V.C.D., **M.A.T.**, D.h.C., A.L.L., D.P., A.E.M.B., M.H., J.V., P.K., N.P.A., A.M.K., C.E.T., and J.W.J. **performed research**; P.W.R. and M.L. contributed new reagents/analytic tools; M.M., V.C.D., **M.A.T.**, D.h.C., A.L.L., A.E.M.B., M.H., J.V., N.P.A., L.K., C.L.S., M.E.R., S.P., S.P.J.W., A.J.O., C.S.C., and J.W.J. **analyzed data**; and M.M., V.C.D., **M.A.T.**, D.h.C., A.L.L., A.J.O., C.S.C., and J.W.J. **wrote the paper**.

ABSTRACT

The host cell serine protease TMPRSS2 is an attractive therapeutic target for COVID-19 drug discovery. This protease activates the Spike protein of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and of other coronaviruses and is essential for viral spread in the lung. Utilizing rational structure-based drug design (SBDD) coupled to substrate specificity screening of TMPRSS2, we have discovered a novel class of small molecule ketobenzothiazole (kbt) TMPRSS2 inhibitors with significantly improved activity over existing irreversible inhibitors Camostat and Nafamostat. Lead compound MM3122 (**4**) has an IC_{50} of 340 pM against recombinant full-length TMPRSS2 protein, an EC_{50} of 430 pM in blocking host cell entry into Calu-3 human lung epithelial cells of a newly developed VSV SARS-CoV-2 chimeric virus, and an EC_{50} of 74 nM in inhibiting cytopathic effects induced by SARS-CoV-2 virus in Calu-3 cells. Further, MM3122 blocks Middle East Respiratory Syndrome Coronavirus (MERS-CoV) cell entry with an EC_{50} of 870 pM. MM3122 has excellent metabolic stability, safety, and pharmacokinetics in mice with a half-life of 8.6 hours in plasma and 7.5 h in lung tissue, making it suitable for in vivo efficacy evaluation and a promising drug candidate for COVID-19 treatment.

Significance Statement:

MM3122 represents an advanced lead candidate for clinical development as a novel antiviral drug for COVID-19. In addition to being novel drugs, selective these TMRSS2 inhibitors can be used as valuable chemical probes to help elucidate mechanisms of viral pathogenesis. Since TMPRSS2 plays a key role as a viral protein processing protease in the pathogenesis of other coronaviruses (SARS-CoV, MERS-CoV) as well as influenza viruses [47-52], MM3122 and this class of TMPRSS2 inhibitors hold much promise as new drugs to not only treat SARS-CoV-2 infections but also potentially represent broad-spectrum antivirals.

Competing Interest Statement:

The compounds are covered in two patent application filings from Washington University where JWJ and VCD are inventors. JWJ and LK own company stock in ProteXase Therapeutics, who has licensed the two patent filings.

INTRODUCTION

Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) is the newly emerged, highly transmissible coronavirus responsible for the ongoing Coronavirus Disease 2019 (COVID-19) pandemic, which is associated with 136 million cases and almost 3 million deaths worldwide as of April 12, 2021 (https://coronavirus.jhu.edu/map.html). While three vaccines have recently been approved by the FDA, there are still no clinically approved small molecule drugs available for the treatment of this disease except Remdesivir and the effectiveness of the vaccines against immune escape variants might be reduced. Multiple therapeutic strategies have been proposed[333, 334], including both viral and host proteins but none have yet been fully validated for clinical application. One class of protein targets which have shown promising results are proteolytic enzymes including the viral proteases[291, 333, 335, 336], Papain-Like Protease (PLpro) and the 3C-like or 'Main Protease' (3CL or MPro), and several host proteases involved in viral entry, replication, and effects on the immune system creating the life-threatening symptoms of COVID-19 infection.[291, 336, 337]. The latter include various members of the cathepsin family of cysteine proteases including cathepsin L, furin, and the serine proteases factor Xa, plasmin, elastase, tryptase, TMPRSS2 and TMPRSS4.

Coronavirus (SARS CoV-2, SARS-CoV, and MERS) entry is mediated by the viral spike protein, which must be cleaved by host proteases in order to trigger membrane fusion and entry into the host cell after binding to the host cell receptor Angiotensin Converting Enzyme-2 (ACE2) [293, 294, 338, 339]. This is mediated by initial cleavage at the S1/S2 junction of spike, which is thought to occur during

processing in the producer cell, followed by cleavage at the S2' site by either serine proteases at the cell surface or by cathepsin proteases in the late endosome or endolysosome. [294, 339]Whether serine or cathepsin proteases are used for S2' cleavage is cell type dependent. While entry into Calu-3 (human lung epithelial) or HAE (primarily human airway epithelial) cells is cathepsin independent, entry into Vero cells (African green monkey kidney epithelial), which do not express the required serine proteases, depends exclusively on cathepsins [293, 294, 339, 340]

TMPRSS2[341] is a type II transmembrane serine proteases (TTSP)[342] that has been shown to be crucial for host-cell viral entry and spread of SARS-CoV-2[293, 338, 343-345], as well as SARS-CoV[346, 347], Middle East Respiratory Syndrome Coronavirus (MERS-CoV) [348] and influenza A viruses[313, 349-355]. The Spike protein requires proteolytic processing/priming by TMPRSS2 to mediate entry into lung cells, thus small molecule inhibitors of this target offer much promise as new therapeutics for COVID-19 and other coronavirus diseases[293, 338]. TMPRSS2 expression levels dictate the entry route used by SARS-CoV-2 to enter cells as reported recently.[356] In cells that express little or no TMPRSS2, cell entry occurs via the endosomal pathway and cleavage of spike protein is performed by cathepsin L. It has been demonstrated that the TMPRSS2-expressing lung epithelial Calu-3 cells are highly permissive to SARS-CoV-2 infection. The irreversible serine protease inhibitors Camostat[293] and Nafamostat[357] are effective at preventing host cell entry and replication of SARS-CoV-2 in Calu-3 cells through a TMPRSS2-dependent mechanism.[343, 344]

Herein, we report on the discovery of a new class of substrate-based ketobenzothiazole (kbt) inhibitors of TMPRSS2 with potent antiviral activity against SARS-CoV-2 and which are significantly improved over Camostat and Nafamostat. Several compounds were found to be strong inhibitors of viral entry and replication, with EC_{50} values exceeding the potency of Camostat and Nafamostat and without cytotoxicity. Newly developed compound MM3122 (**4**) has excellent pharmacokinetics (PK) and safety in mice and is thus a promising new lead candidate drug for COVID-19 treatment.

RESULTS

1. Hit identification of TMPRSS2 inhibitors

It is well established that several TTSPs including TMPRSS2 not only play a role in infectious diseases[341], but also cancer progression and metastasis[358-360] which is thought to be mainly through its ability to activate hepatocyte growth factor (HGF), the sole ligand for MET receptor tyrosine kinase. This is accomplished via proteolytic processing of the inactive single-chain precursor pro-HGF to a two-chain active form. TMPRSS2 shares pro-HGF as a protein substrate with the other HGF-activating serine proteases, HGFA (HGF-Activator), hepsin and matriptase.[361, 362] Like TMPRSS2, other TTSPs such as matriptase and hepsin have a canonical serine protease domain residing as the C-terminal domain of the protein that is anchored to the cell membrane by an N-terminal type II signal peptide thereby presenting their enzymatic activity outside the cell. We have previously reported on the discovery and anticancer properties of peptidomimetic ketothiazole (kt) and ketobenzothiazole (kbt) inhibitors of HGFA, matriptase and hepsin[363-367] named synthetic HGF Activation Inhibitors or sHAIs. Since TMPRSS2 has an overlapping endogenous substrate specificity profile with HGF-activating proteases, we postulated that our substrate-based sHAIs would also inhibit TMPRSS2.

Antiviral activity of sHAIs (VSV/SARS-CoV-2 Spike Protein Pseudotypes):

Based on our premise that sHAIs would inhibit TMPRSS2, we first selected the two sHAI lead compounds, ZFH7116 (**1**)[365, 368] and VD2173 (**2**) (**Figure 3.1**) for initial antiviral testing and confirmed that they potently inhibit TMPRSS2-dependent host cell viral entry driven by the spike protein of SARS-CoV-2 (**SARS-2 S, Figure 3.2A**) into

Calu-3 lung epithelial cells having an EC_{50} of 307 and 104 nM, respectively. Marked inhibition of SARS-2-S-driven entry was also detected for the irreversible inhibitor Camostat, in line with published data[293, 357]. It is noteworthy that neither compound showed activity in TMPRSS2 negative Vero cells or modulated entry into Calu-3 cells by pseudotypes bearing the VSV glycoprotein (VSVpp VSV-G) (**Figure S6**) suggesting that the inhibition of SARS-CoV-2 spike-driven entry into Calu-3 cells was due to blockade of TMPRSS2.

Antiviral activity of sHAIs (VSV/SARS-CoV-2 Spike Protein Chimeras):

Using a replication-competent, chimeric VSV expressing the SARS-CoV-2 Spike (shown to mimic the Spike-dependent entry of authentic SARS-CoV-2[287] we also demonstrated that **1** and **2** block viral entry into Calu-3 cells in a dose-dependent manner. Both drugs displayed EC_{50} values of 459 and 197 nM) against the chimeric virus VSV-SARS-CoV-2 (**Figure 3.2B**) compared to the pseudotyped virus (**Figure 3.2A**), while showing no activity against VSV-eGFP (VSV G dependent entry), or against either virus in Vero cells (**Figure A2.S7**), which are TMPRSS2 negative. This confirms our initial result and allows for the establishment of a system for screening antiviral activity of these inhibitors using VSV-SARS-CoV-2.

Inhibition of cell bound TMPRS2 enzymatic activity by sHAIs

We tested the inhibitory activity of **1** and **2** on TMPRSS2 proteolytic activity in a cell-based enzyme-based fluorogenic assay, by overexpression of TMPRSS2 in a human cell line, HEK-293T, commonly used for experimentation because of its high transfectability. As previously demonstrated, expression of TMPRSS2 can be accurately measured[369, 370] using a fluorogenic peptide reporter substrate, Boc-QAR-AMC in

Figure 3.1: Structures of the initial TMPRSS2 inhibitors discovered by screening existing HGFA, matriptase and hepsin serine protease inhibitors.

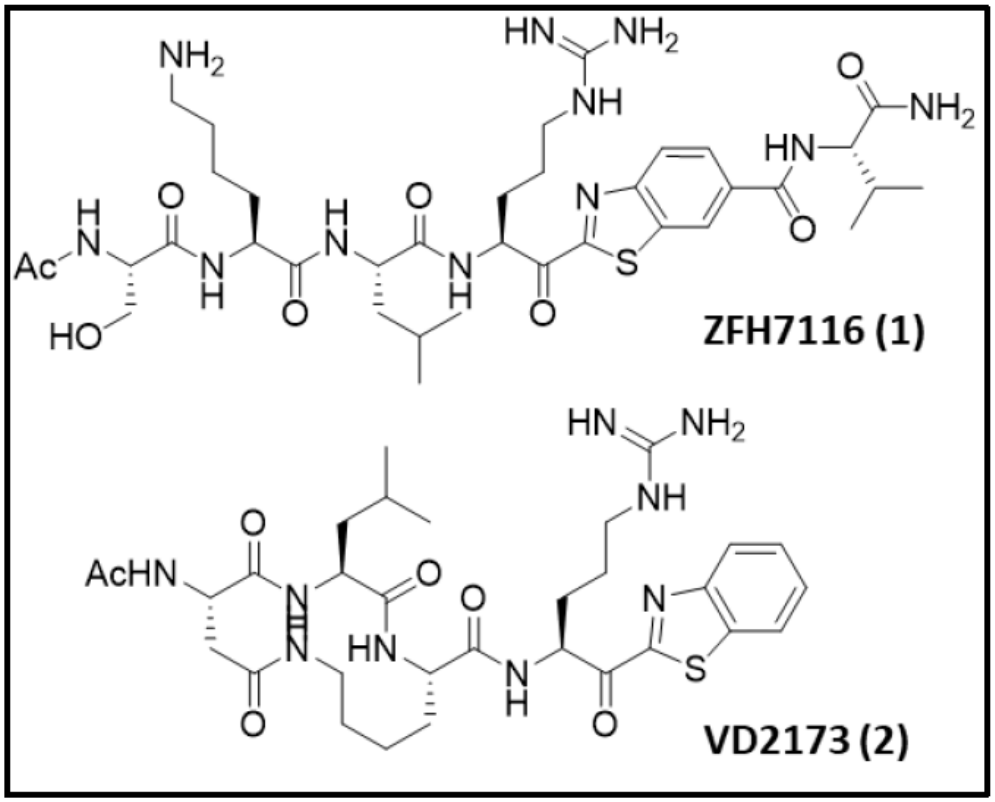
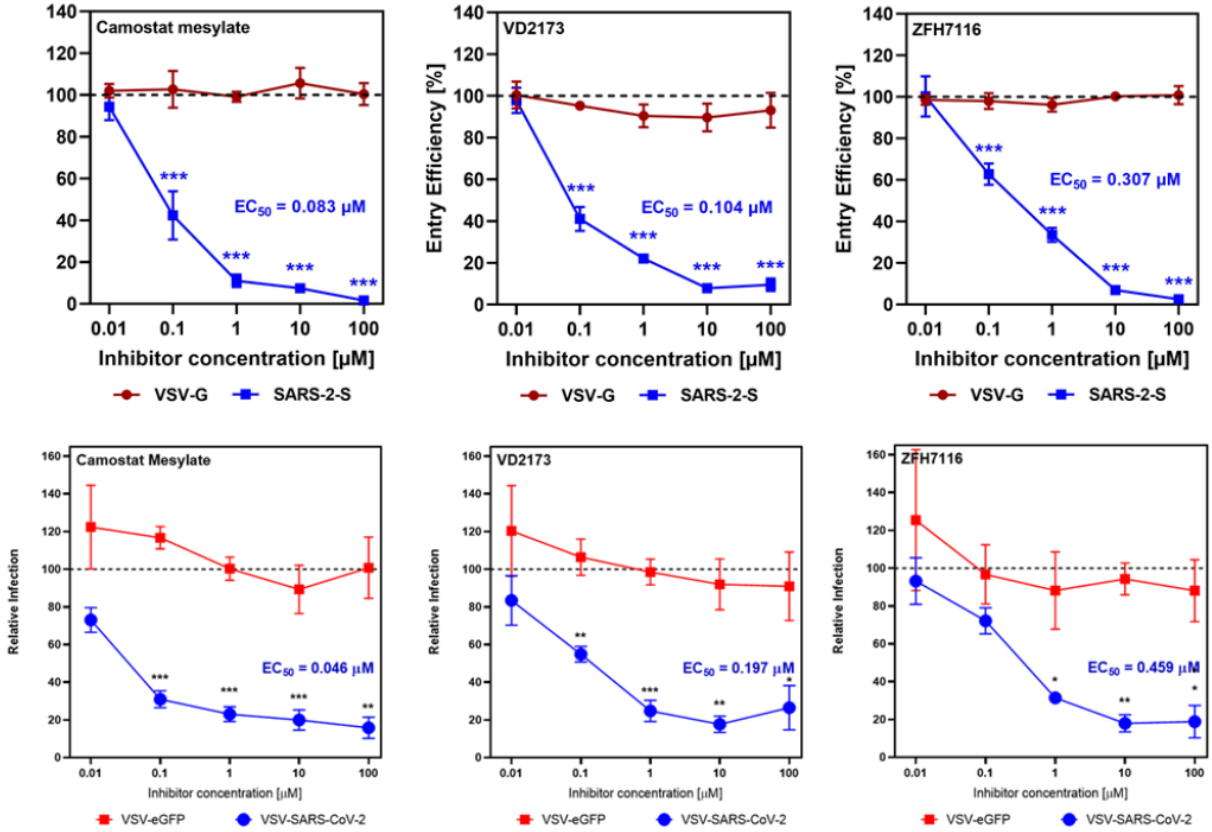


Figure 3.2: Inhibition of SARS-CoV-2 cell entry into Calu-3 lung epithelial cells by ZFH7116 (1) and VD2173 (2) using **(A)** VSV-SARS-CoV2-Spike protein Pseudotype virus and **(B)** VSV-SARS-CoV2-Spike protein Chimeric virus. EC50s are calculated from an average of 3 separate experiments. Camostat was used as a positive control.



cell cultures. **1** inhibited cell based TMPRSS2 enzyme activity in a concentration dependent manner between 10 μM to 10 nM with an IC_{50} of 314 nM (**Figure A2.S4**). **2** was a more potent inhibitor of TMPRSS2 proteolytic activity with IC_{50} of 57 nM. These data confirm that both **1** and **2** mediate their function by potently inhibiting TMPRSS2.

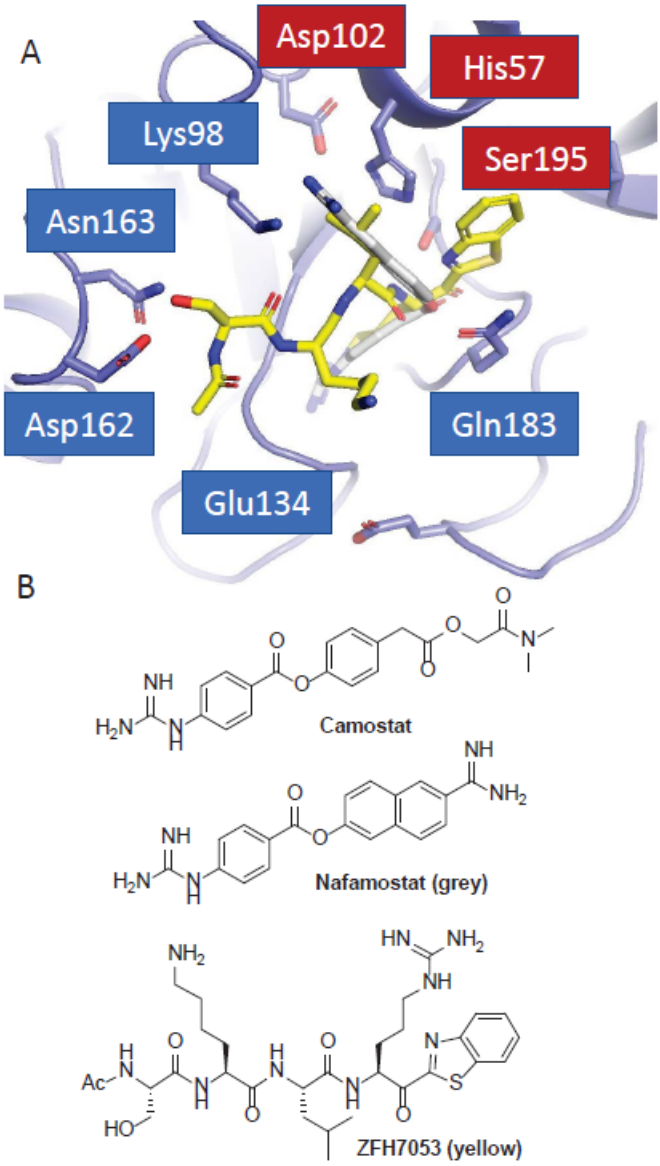
Selectivity data of TMPRSS2 inhibitors in a panel of proteases.

We profiled **1** and **2** for their selectivity against a panel of 43 serine and cysteine proteases (**Tables A2.S1 and A2.S2**) for comparison to that of Camostat and Nafamostat protease selectivity data which has been published[371]. We found **1** was a potent inhibitor of matriptase-2, plasma kallikrein, proteinase K, trypsin, tryptase b2 and G1 but also was a moderate inhibitor of factor Xa, factor XIIa, kallikrein 5 and 14 as well as cysteine protease cathepsin S while showing some activity against cathepsin L as well. There was overlap of **1** with the selectivity profiles of Camostat except they did not show any activity against the cathepsins. Furthermore, Camostat more potently inhibited factor XIIa, plasma kallikrein, matriptase-2, and plasmin while also having increased activity for trypsin, the tryptases and urokinase. Cyclic peptide **2** also more potently inhibited urokinase and factor XIIa relative to **1** but not cathepsin S or L, and instead was a moderate inhibitor of cathepsin B. The ramifications of these different selectivity profiles may explain the varied activity in the VSV-pseudotype and chimeric viral entry results most notably the urokinase potency for which **1** was significantly weaker.

Lead identification of TMPRSS2 inhibitors

We published[366] that Camostat and Nafamostat (**Figure 3.3B**) are inhibitors of matriptase and hepsin, and analog Nafamostat is the most active published inhibitor of

Figure 3.3: (A) Molecular docking model of compound 916 3 (yellow) and Nafamostat (grey) bound to a homology model of TMPRSS2 using Glides/Schrödinger. **(B)** Structures of compound 3, Camostat and Nafamostat.



SARS-CoV-2 cell entry[338]. While we previously pursued this chemical series as HGFA, matriptase and hepsin inhibitors[366], the best series we have developed are small peptide-based molecules like **1, 2** and **3** (Ac-SKLR-kbt; **Figure 3.3B**) which exhibit low nM to picomolar IC_{50}s. These compounds contain a serine trapping ketobenzothiazole (kbt) warhead[363-365, 367] which reacts covalently with the protease, but importantly in a reversible manner, unlike Camostat or Nafamostat in which the inhibition is irreversible. Therefore, we pursued the kbt class of inhibitors for lead identification studies towards more potent and selective TMPRSS2 inhibitors.

Rational design of novel TMPRSS2-selective ketobenzothiazole (kbt) inhibitors

We have employed X-ray co-crystal structure data in the rational design of optimized HGFA, matriptase and hepsin inhibitors with increased potency and selectivity[363-366]. Shown in **Figure 3.3**, here we used a homology model of human TMPRSS2[372] to computationally model our existing inhibitors (using Glide in Schrödinger) where we docked compound **3** (yellow; **Figure 3.3A**) and Nafamostat (grey). For Nafamostat, the naphthyl portion extends deep into the S2 pocket with the benzguanidine in the S1 pocket binding to the conserved S1 Asp189. The P4 Ser of **3** makes two H-bonds to Asp162 and Asn163 outside the S4 pocket, and the Ser backbone carbonyl is also within H-bonding distance to the P2 Lys98. The P3 Lys is near the S4 Glu134 making an electrostatic interaction while the P2 Leu resides in the S2 pocket and the benzothiazole fills the S1' area. The reactive ester of Nafamostat and ketone of the kbt are adjacent to the Ser195-His57-Asp102 catalytic triad. Our model of **3** and Nafamostat bound to TMPRSS2 (**Figure 3.3A**) shows a nice fit to the S1-S4 and S1' pockets of TMPRSS2. For **3**, it appears that the P4 Ser is making a dual H-bond to

the Asp162 and Asn163 residues in TMPRSS2. This area is occupied by a Gln in both hepsin and matriptase but an Asp in HGFA and interestingly the additional Asn163 residue is unique to TMPRSS2. This suggests that a free amine in the P4 could provide selectivity over matriptase and hepsin. The structure also shows that Lys87 resides in the S2 pocket, suggesting a P2 Glu or Asp might be ideal for an inhibitor of TMPRSS2 and lead to potential selectivity over HGFA (P2 Ser), matriptase (P2 Phe) and hepsin (P2 Asn). This Lys87 is also near the P4 Ser sidechain indicating an Asp or Glu in this position could potentially make a third electrostatic interaction to Lys87. Further, the S2 pocket is large and appears that it could accommodate larger sidechains such as Phe, Tyr, or Trp.

PS-SCL Protease Substrate Specificity Profiling of TMPRSS2

To augment our rational design of more potent and selective TMPRSS2 inhibitors, we analyzed existing data on the peptide substrate preferences of TMPRSS2, HGFA, matriptase and hepsin. When comparing the positional scanning of substrate combinatorial libraries (PS-SCL) data of TMPRSS2 (**Figure 3.4**)[358] with that of matriptase [373], hepsin [374] and HGFA [364], it became apparent that there was significant overlap in their preferences for substrates. This data reveals TMPRSS2 is tolerant of many different P2 sidechains but prefers Phe and Ala/Thr like matriptase (**Figure 3.4**) which can also tolerate both large and small groups (prefers Ser/Ala) but hepsin and HGFA both prefer Leu. For P3 TMPRSS2 prefers Gln/Glu and Met whereas it is Lys/Gln for hepsin and matriptase and Lys/Arg for HGFA (**Figure 3.4**). The clearest distinction is in the P4 position where HGFA, matriptase and hepsin all prefer basic

Figure 3.4: Combined PS-SCL for substrate specificity of TMPRSS2, matriptase, hepsin and HGFA. Using protease substrate terminology, the cleavage (hydrolysis) site for protease protein substrates is between the N-terminal P1 and P1' positions of the **P**eptide substrate, while S1 and S1' refers to the **S**ubsite 1 and 1’ of the protease where the P1 and P1’ amino acid sidechain binds.

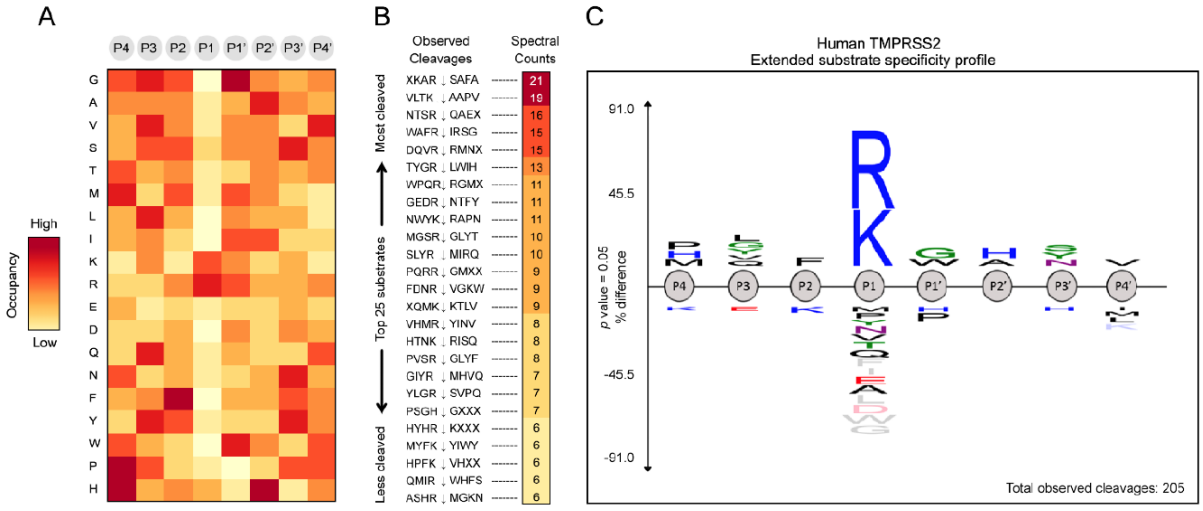
proHGF/MSP	**K/S – Q/K – L/L-R**
TMPRSS2	I(G/P)-Q(E)-F(A/T)-R
Matriptase	K(R) - K(Q) - S(A)-R
Hepsin	K(P) - K(Q) - T(L)-R
HGFA	R(K) - R(K) - L(M)-R

residues like Lys/Arg while TMPRSS2 prefers Ile/Gly and Pro, which is shared attribute with hepsin.

MSP-MS profiling of TMPRSS2 for extended substrate sequence specificity

To strengthen our compound design, we obtained further information on the extended substrate specificity of TMPRSS2 using multiplex substrate profiling by mass spectrometry (MSP-MS).[375] In the MSP-MS assay, a physiochemically diverse library of 228 tetra decapeptides was incubated for several hours with human recombinant TMPRSS2 in activity buffer. At different time points (15, 60 and 240 minutes) aliquots of the reaction mixture were extracted, quenched with 8M guanidium HCl and analyzed by tandem mass spectrometry to allow monitoring of protease-generated cleaved products. As a TTSP with a trypsin fold, TMPRSS2 is known for its high affinity towards substrates containing an arginine residue in the P1 position.[358] This prominent P1 specificity was confirmed by our MSP-MS analysis, which showed a nearly exclusive occupancy of the P1 position by either Arg or Lys on TMPRSS2-generated cleaved products (**Figure 3.5A**). Peptide sequencing by LC-MS/MS enabled the identification of the 25 most preferred substrates for TMPRSS2 in our peptide library (**Figure 3.5B**). An IceLogo frequency plot (**Figure 3.5C**) displays the extended substrate specificity profile of TMPRSS2 at pH 7.5, which reveals the preference for hydrophobic amino acids in both P2 and P1’ positions, which flank the cleavage site. The analysis indicates a preferred cleavage of N-terminal tetrapeptides to be PLFR with other P4 amino acids H and M and P3 amino acids G, Y, V, and Q but only F at P2 which is striking. Largely this data recapitulates what was elucidated in the PS-SCL study (**Figure 3.4**). Based on careful analysis of the computational modeling work with these and the PS-SCL results,

Figure 3.5: (A) Heatmap displaying the overall amino acid frequencies in each of the P4-P4' positions of a positive set of peptides cleaved by human TMPRSS2. Positive enrichment is observed as red and zero was set to yellow. **(B)** Total spectral counts determined by LC-MS/MS highlight peptide sequences as high turnover rates substrates for TMPRSS2. **(C)** IceLogo depicting the extended substrate specificity of human TMPRSS2 based on 205 cleavage events detected through MSP- MS analysis.



we selected additional existing compounds in our library of HGFA/matriptase hepsin inhibitors for testing and synthesized (**Schemes A2.S1-S3**) four new analogs specifically designed for TMPRSS2: Ac-GQFR-kbt (**4**), Ac-PQFR-kbt (**5**), Ac-QFR-kbt (**6**) and Ac-IQFR-kbt (**7**).

Inhibition of VSV-SARS-CoV-2 viral entry

Analogs based on the initial hit compounds **1** and **2** along with the substrate specificity data were Ac-SKLR kbt (**3**), Ac-WFR-kbt (**8**), Ac-SKFR-kt (**9**), Ac-KQFR-kt (**10**), Ac-SQLR-kt (**11**), Ac-FLFR-kbt (**12**), Ac-dWFR-kbt (**13**), dWFR-kbt (**14**), dWFR-kbt-CO_2H (**15**), Ac-WLFR-kbt (**16**), Ac-KQLR-kbt (**17**), Ac-LLR-kt (**18**), Cyclo(DMK)R-kbt (**19**), Cyclo(DQK)R-kbt (**20**), and Cyclo(allylGLY)R-kbt (**21**). Many of these compounds contain a Phe (F) residue in the P2 position as predicted from the PS-SCL study while others contain a Gln (Q) at the P3 position. When tested in the VSV pseudotype assay all showed improved activity relative to **1** (**Table 3.1**) where the best compound was **13**. Compound **13** is a tripeptide containing an unnatural amino acid D-Trp in the P3 position and Phe in the P2 position, predicted from PS-SCL and MSP-MS analysis to be preferred. **17** which contains a P3 Gln also predicted from PS-SCL shows the next best activity with an EC_{50} of 78 nM. Furthermore, **16** also having a P2 Phe showed excellent potency of 150 nM. Corresponding cycloamide analogs of **2**, **19** and **20** retained good activity (EC_{50} of 119 and 138 nM) for inhibition of viral entry of the VSV-pseudotype, but allyl phenyl ether analog **21** showed a significant decrease in activity (EC_{50} 565 nM).

The remaining compounds were tested in the VSV SARS CoV-2 chimeric assay. Excitingly, the most promising results (**Table 3.1**) were derived from our new rationally designed TMPRSS2 inhibitors, **4-7**. These compounds were designed by incorporating

Table 3.1: Structures and inhibition data of compounds for protease enzyme activity and VSV pseudotype and chimera SARS-CoV-2 viral entry into human Calu-3 lung epithelial cells (ND = not determined, NA = not applicable; * Here, n=1).

Name/ Structure	TMPRSS2 IC_{50} (nM)	HGFA IC_{50} (nM)	Matriptase IC_{50} (nM)	Hepsin IC_{50} (nM)	Thrombin IC_{50} (nM)	Factor Xa IC_{50} (nM)	VSV SARS-CoV-2 Calu-3 EC_{50} (nM)	VSV SARS-CoV-2 Chimera Calu-3 EC_{50} (nM)	SARS-CoV-2 Wild-type Calu-3 IC_{50} (nM)
Remdesivir	NA	NA	NA	NA	NA	NA	ND	ND	1271
Camostat	1.5	>20,000	7.0	7.0	>20,000	>20,000	83	21	ND
Nafamostat	0.14	158	0.05	0.9	5,020	4,570	ND	ND	220*
Ac-SKLR-kbt-V (1)	74	23	14	1.0	7530	514	307	489	ND
Cyclo(DLK)R-kbt (2)	2.6	8520	2.6	19	8140	2050	104	197	660
Cyclo(DMK)R-kbt (19)	19	3240	1.0	5.9	18,500	1,390	119	ND	ND
Cyclo(aGLY)R-kbt (21)	197	>20,000	14	20	>20,000	8810	565	ND	ND
Cyclo(DQK)R-kbt (20)	ND	16,100	7.7	22	>20,000	>20,000	138	ND	ND
Ac-WFR-kbt (8)	9.4	329	5.7	7.6	>20,000	22	ND	1377	ND
Ac-SKFR-kt (9)	7.9	114	6.1	17	>20,000	1060	ND	1157	ND
Ac-KQFR-kt (10)	29	116	1.4	1.2	>20,000	158	ND	262	ND
Ac-SQLR-kt (11)	16	364	18	0.68	ND	ND	ND	1320	ND
Ac-LLR-kt (18)	54	506	56	4.6	ND	ND	349	1838	ND
Ac-SKLR-kbt (3)	39	66	6.1	0.32	>20,000	3800	ND	4272	ND
Ac-FLFR-kbt (12)	3.0	228	7.2	2.9	>20,000	26	ND	101	ND
Ac-dWFR-kbt (13)	1.1	27	2.6	1.1	3,700	98	32	ND	ND
dWFR-kbt (14)	39	ND	ND	ND	ND	ND	ND	357	ND
dWFR-kbt-CO_2H (15)	42	ND	ND	ND	ND	ND	ND	105	ND
Ac-WLFR-kbt (16)	6.3	266	12	0.79	>20,000	2.0	150	ND	ND
Ac-KQLR-kbt (17)	ND	60	1.1	0.17	>20,000	258	78	572	ND
Ac-IQFR-kbt (7)	0.25	30	0.92	0.14	>20,000	792	ND	3.6	102
Ac-QFR-kbt (6)	0.31	14	0.13	0.08	>20,000	1.4	ND	0.53	105
Ac-PQFR-kbt (5)	0.28	75	0.32	0.13	>20,000	199	ND	0.86	52
Ac-GQFR-kbt (4)	0.34	32	0.31	0.19	>20,000	700	ND	0.43	74

the preferred sidechains (as determined form PS-SCL data above) at all the P1-P4 positions of the inhibitor. The set of existing compounds all showed good potency with the best **12** showing an EC50 of 101 nM in blocking viral entry of VSV SARS-CoV-2 chimeras into Calu-3 cells, 6-fold more active than the second-best analog in the pseudotype assay, **17**. Excitingly, three of the new rationally designed analogs **4-6** displayed significantly increased potency (**Figure 3.6**) with sub nanomolar EC_{50}s and the fourth new compound **7** is still extremely potent with an EC_{50} of 3.6 nM. This not only constitutes a 230-fold increase in activity for **4** (EC_{50} 0.43 nM) relative to the previous best compound **12** but also a 80-fold improvement over Camostat. Of the other compounds, similar potency compared to **12** was seen from tripeptide **15** (EC_{50} 105 nM) but less so from **14**, analogs of **13** which was the best compound tested in the pseudotype assay. Also showing excellent potency is **10** with an EC_{50} of 262 nM.

Inhibition of VSV-MERS viral entry

To demonstrate broad coronavirus activity, we tested Camostat, **2** and our best new compounds, **4-7** (**Figure 3.6**) based on the VSV chimera results for their activity in blocking host cell entry of a MERS-CoV chimeric virus. Shown in **Table 3.2**, all compounds inhibited MERS-CoV viral entry with nearly equivalent potency as compared to SARS-CoV-2 (**Table 3.1**). Though the TMPRSS2-dependence of MERS-CoV has been demonstrated previously, this provides strong evidence that our compounds have broad-spectrum potential against other coronaviruses in the clinic. To confirm that this activity is indeed TMPRSS2-related, we next set out to produce recombinant TMPRSS2 protein for testing our compounds for their enzyme inhibitory activity.

Figure 3.6: Structures of rationally designed covalent reversible kbt inhibitors (4 through 7) of TMPRSS2

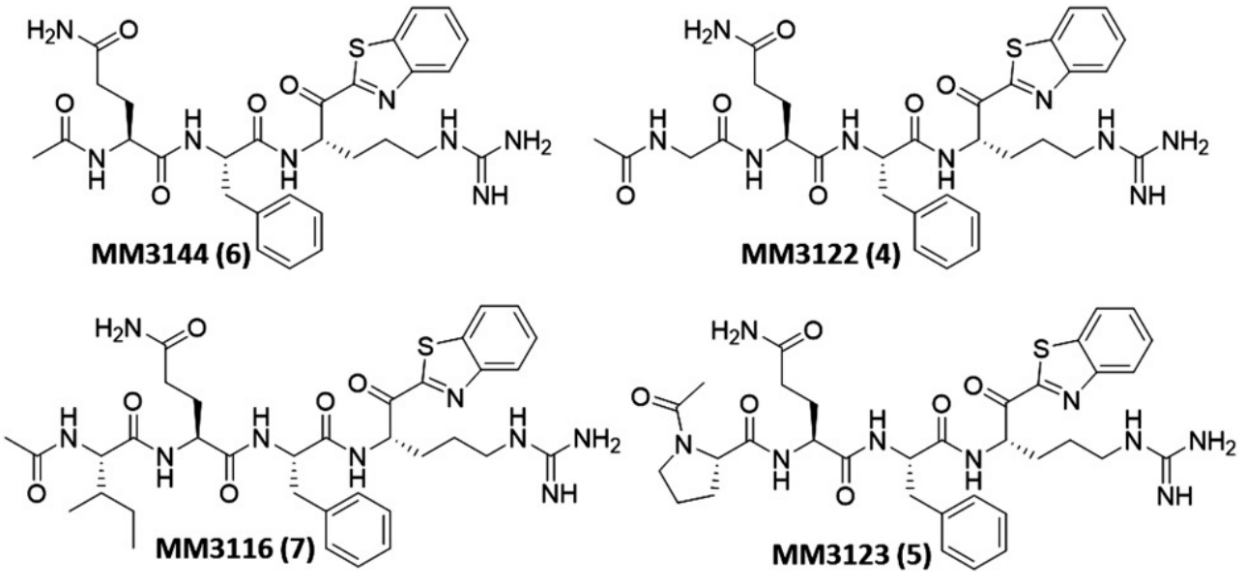


Table 3.2: Activity of lead compounds against VSV pseudotype MERS viral entry into human Calu- 3 lung epithelial cells. Also shown is the plasma stability of selected compounds in both mouse and human plasma (ND = not determined).

Name/ Structure	MERS VSV pseudotype Calu-3 EC50 (nM)	Mouse Plasma Stability t1/2 (min)	Human Plasma Stability t1/2 (min)
Camostat	5.0	ND	ND
2 (VD2173)	273	222	>289
7 (MM3116)	2.3	154	>289
6 (MM3144)	0.55	>289	>289
5 (MM3123)	2.9	>289	>289
4 (MM3122)	0.87	>289	>289

Expression and Purification of Recombinant TMPRSS2 (Protease Domain)

Recombinantly expressed TMPRSS2 protease domain (**Figure A2.S8A**) was purified via Ni-NTA column (**Figure A2.S8C**) then verified by SDS-PAGE and western blot (**Figure A2.S8D** and **A2.S8E)** showing no presence of other proteins. Catalytic activity was tested by combining 150nM of purified TMPRSS2 protease domain with (MCA)-K-KARSAFA-K-(DnP), an 8-mer peptide FRET substrate that was synthesized based on the peptide that was preferred to be most efficiently by TMPRSS2 from our MSP-MS analysis (**Figure 3.5**). Michaelis-Menten kinetics showed a Km of 0.6±0.05 uM (**Figure A2.S8B**). Relative to the full-length TMPRSS2 (see below), the k_{cat} value achieved for the protease domain (0.07/s) was 1000-fold lower than that observed for the construct bearing the LDLR class A domain and SRCR modules N-terminal to the catalytic domain (k_{cat} = 94,54/s). Further investigations are necessary to understand these differences and the interactions among the catalytic domain, the LDLR and the SRCR domains that impact the extended specificity and potentially the catalytic activity of TMPRSS2.

Recombinant TMPRSS2 (Full-length) Enzyme Assay

Using active recombinant full-length TMPRSS2 as described above, along with Boc-QAR-AMC as a fluorogenic substrate[371], we found the K_m was 85.6 μM using an enzyme concentration of 3 nM (**Figure A2.S2**).We used this substrate to test all inhibitors for inhibition of TMPRSS2 proteolytic enzyme activity in a standard kinetic assay using Nafamostat and Camostat as controls, where we determined the compound IC_{50}s over the period of one hour, following 30 min compound incubation with enzyme. The IC_{50} values for Camostat and Nafamostat are 1.5 nM and 0.14 nM

which is similar but slightly improved compared to values previously reported of 6.2 nM and 0.27 nM, respectively.[371] We found the IC_{50} data we generated for TMPRSS2 inhibition closely correlated with the VSV pseudotype and chimera assay data with some exceptions. Multiple compounds were significantly more potent than Camostat and several were equipotent to Nafamostat (**Table 3.1**). The new rationally designed lead compounds **4-7** (**Figure 3.6**) all displayed exquisitely potent sub nanomolar IC_{50} values of which **7** was the best with an IC_{50} of 250 picomolar similar to that of Nafamostat. It should be noted however that Nafamostat is an irreversible inhibitor while the kbt class of inhibitors are reversible, so it is difficult to directly compare IC_{50} values between the two series having different mechanism of inhibition. While the initial hit compound **1** still shows good potency against TMPRSS2 (IC_{50} 74 nM), it was significantly weaker when compared to the rationally designed TMPRSS2 inhibitors which was expected. However, the other initial hit compound **2** is a much more potent inhibitor of TMPRSS2 activity than **1** with an IC_{50} of 2.6 nM, only about 10-fold less active than **4-7**. It is noteworthy that the P3-P1 tripeptide **6** without the P4 sidechain was almost as active as either of the tetrapeptides **4**, **5** or **7** which have a P4 residue. However, this compound **6** and the other tripeptides including **8** and **13** suffer from high inhibition of the plasma protease Factor Xa which is undesirable for further drug discovery due to potential bleeding side effects in patients.

The least active compound was the aryl ether cyclic peptide **21** with an IC_{50} of 197 nM which is however still respectable. It is presumed the larger, more constrained aryl ether ring system is not an ideal fit for binding to the TMPRSS2 active site through bridging the S2-S4 pockets. It is important to note that while TMPRSS2 activity is potent

in all inhibitors, most compounds tested are all still relatively active against the other proteases HGFA, matriptase and hepsin.

The most selective TMPRSS2 analogs are **16** with an IC_{50} of 6.3 nM, being 40-fold more active than HGFA, 2-fold over matriptase and **12** which is 60-fold more selective over HGFA and 2-fold against matriptase. As with other compounds in this series of peptidyl kbt inhibitors, in our experience we have found that deriving selectivity over hepsin is challenging but one of our best examples turns out to be initial hit **2**, which is 4-fold more active against TMPRSS2 relative to hepsin and almost 10-fold more active against matriptase. The consequences of inhibiting these other serine proteases, especially the other HGF-activating proteases, in the COVID-19 scenario is unclear at this time but may be beneficial for treatment rather than detrimental.

There are clear structure-activity relationships (SAR) from the 21 compounds tested. For example, it appears that TMPRSS2 prefers large groups extending beyond the kbt S1' C-terminal portion of the inhibitor since **1** have increased activity relative to unsubstituted kbt analog **3**. Strengthening this hypothesis, the compounds with the smaller ketothiazole (kt) warhead in the P1' position like seen in **9-11** also lose potency relative to analogs with the larger kbt. Also, the P3 position is important for activity where it seems basic groups like Lys (K) are undesired (**9**, **3**, **1**) which seems also true for the P4 position as seen with compound **17**. Finally, as predicted from modeling and the PS-SCL data, Phe is clearly preferred in the P2 position. Therefore, it is likely that diverse new analogs can be pursued with other aromatic sidechains like Trp and Tyr

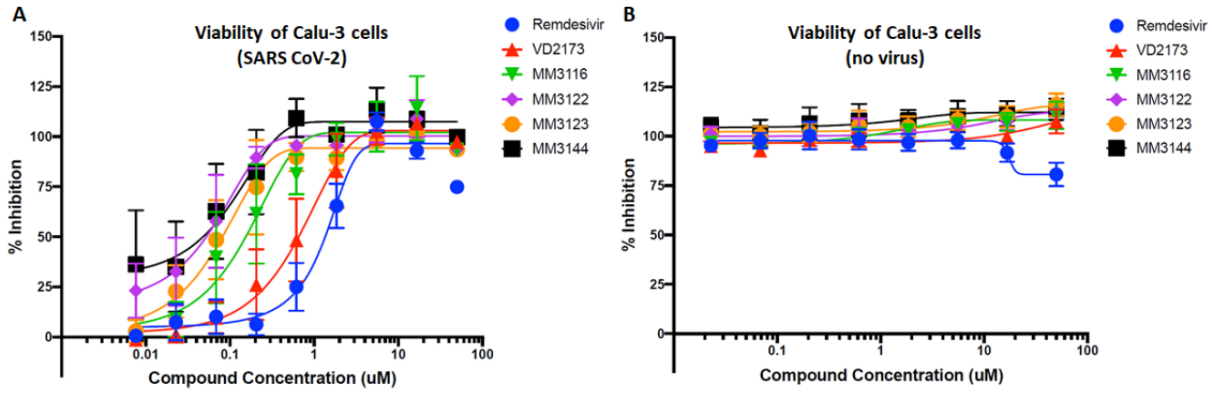
which will also produce inhibitors with exquisite TMPRSS2 activity and potentially more selectivity.

Given the homology model of TMPRSS2 used in this study was exclusive to the catalytic domain of TMPRSS2, both catalytic domain and full length TMPRSS2 were tested against candidate compounds. The inhibition IC_{50} values against the recombinant protease domain (**Figure A2.S1**) versus full length TMPRSS2 shows strikingly different values with estimated IC_{50}s being over 100 mM compared to picomolar to double digit nanomolar IC_{50} values for the commercial full length TMPRSS2 (**Table 3.1** and **Figure A2.S3**). It is not clear from our current studies what the direct cause of this noticeable difference is, but we suspect that the protein domains/modules of TMPRSS2 which are N-terminal to the catalytic domain play an important role in ligand interactions and subsequent catalysis.

Antiviral activity of lead compounds (Wild-type SARS-CoV-2)

After confirming and quantitating TMPRSS2 enzyme inhibition of all compounds against recombinant protein and their selectivity over HGFA, matriptase, hepsin, thrombin, and Factor Xa, we assessed the activity of our most promising lead compounds **2**, **4-7** towards wild-type SARS-CoV-2 by employing a CellTiter-Glo (Promega) cellular viability assay using lung epithelial Calu-3 cells.[376] We used Remdesivir and Nafamostat (n=1) as a positive control. While all compounds showed excellent activity in this assay (**Figure 3.7** and **Table 3.1**), the most potent compounds were the rationally designed TMPRSS2 inhibitors **4** and **5** (**Figure 3.6**) with EC_{50}s of 74 and 52 nM, respectively which is >20-fold improved over Remdesivir and also

Figure 3.7: (A) Inhibition of the cytopathic effects (viral toxicity) of wild-type SARS-CoV-2 Calu-3 by lead TMPRSS2 inhibitors using the CellTiter-Glo (Promega) assay. **(B)** Viability of Calu-3 cells with compounds in the absence of virus (compound toxicity).

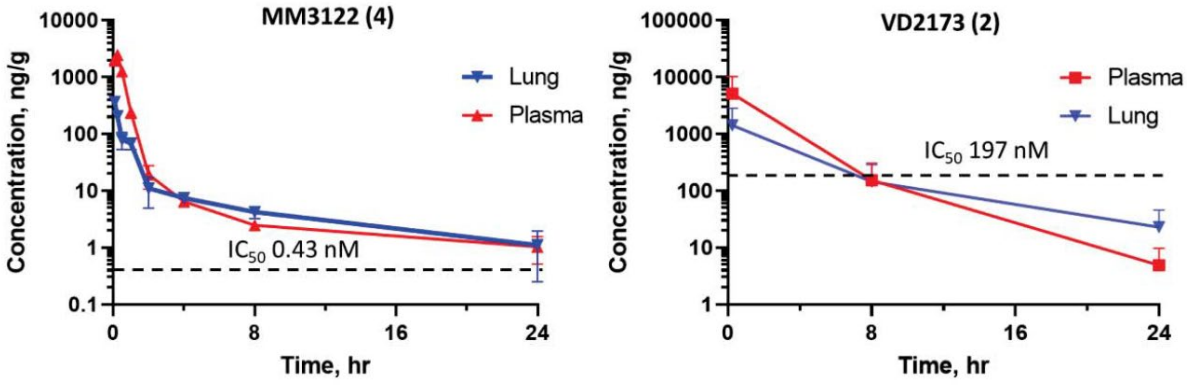


significantly more active than Nafamostat (note: based on n=1). Importantly, all five kbt inhibitors showed no signs of toxicity to Calu-3 cells (up to 50 μM, which is nearly 1000 times the IC_{50} for **5**, while Remdesivir (and Nafamostat) displayed toxicity at the highest concentration tested (50 μM).

Pharmacokinetics and metabolic stability of VD2173 (2) and MM3122 (4)

Compound **2** (VD2173) and lead compounds **4-7** were tested for their in vitro stability in mouse and human blood plasma (**Table 3.2**). All compounds have excellent stability in both mouse and human plasma with >289 min half-life except for **7** which had a half-life of 154 min in mouse plasma and 2 with a 222 min half-life. Based on potency, selectivity and in vitro properties, compound **4** (MM3122) and compound **2** (VD2173) was selected as a lead candidate and was tested for its in vivo pharmacokinetics (PK) in mice. Compounds **5-7** were not tested for PK due their lower antiviral activity and the relatively high Factor Xa activity of **5** and **6**. Shown in **Figure 3.8**, VD2173 (2) has a half-life of 2.5 h in mice with a high AUC and exposure in plasma beyond 24 h. Excitingly, rationally designed TMPRSS2 lead compound MM3122 (4) also had excellent PK with a half-life of 8.6h in plasma. Both VD2173 and MM3122 were subsequently tested for their lung exposure over a period of 24 h after IP dosing. It was found that both compounds attained high levels of compound in the lung and had excellent AUC but with MM3122 being superior with a 7.5 h half-life versus 4.2 h for VD2173. The potency of MM3122 versus VD2173 in both the TMPRSS2 enzyme assay and viral assays is magnitudes higher (500-fold) making MM3122 an ideal lead candidate drug suitable for in vivo efficacy studies of COVID-19 and further optimization to be reported in a future communication.

Figure 3.8: Mouse pharmacokinetics (PK) of MM3122 (4) and VD2173 (2) after IP dosing @16.7 mg/kg. Compound concentrations over time in plasma and lung tissue are shown relative to the IC50 of the compounds in the SARS-CoV-2 VSV hybrid cell entry assay. MM3122 (4) has a half-life (t1/2) of 8.6 h in plasma and 7.5 h in the lung and 2.5 h and 4.2 h, respectively for VD2173 (2).



Acute 7-day toxicity study of MM3122 (4) in mice

We further tested lead compound MM3122 for its safety in mice. MM3122 was administered daily to mice at three different single dose levels of 20, 50, and 100 mg/kg via intraperitoneal (IP) injection over a period of 7 days. No adverse effects were observed in any of the treatment groups and no weight loss or changes to harvested organs (liver, spleen, and kidneys) were noted compared to the control group (**Figure A2.S5**).

DISCUSSION

TMPRSS2 has been shown to be essential for host cell viral entry and replication of SARS-CoV-2 as discussed above. Based on molecular docking studies using a published homology model of TMPRSS2 [372] and substrate specificity data from PS-SCL, we hypothesized that a set of our existing peptidyl ketobenzothiazole (kbt) inhibitors of HGFA, matriptase and hepsin would also inhibit TMPRSS2. Indeed, we demonstrated that these compounds not only potently inhibit TMPRSS2 enzyme activity of recombinant and cell-surface protein, but also host cell entry and toxicity driven by the Spike protein of SARS-CoV-2 into ACE2/TMPRSS2-expressing Calu-3 lung epithelial cells. After further optimization, we identified multiple potent inhibitors of TMPRSS2, with four analogs rationally designed for TMPRSS2 displaying sub nanomolar activity in both the enzyme assay and in blocking the entry of VSV-SARS-CoV-2 chimeras into human Clau-3 epithelial lung cells. In addition, several of these compounds displayed excellent potency against MERS, another prominent coronavirus. We further confirmed potent antiviral activity against the wild-type SARS CoV-2 virus with five lead compounds thus identifying the most promising lead compound MM3122 (**4**). We have clearly established that this novel TMPRSS2 inhibitor MM3122 is more potent than Remdesivir, Camostat and Nafamostat. Importantly, this compound also shows no toxicity to Calu-3 cells in contrast to Remdesivir and Nafamostat which are toxic at higher concentrations. MM3122 has excellent metabolic stability in both mouse and human plasma as well as outstanding pharmacokinetics and safety in mice. Cyclic peptide VD2173 (**2**) and rationally designed TMPRSS2 inhibitor MM3122 both have an impressive >8 h half-life in mice. In addition to its excellent TMPRSS2 inhibition,

pharmacokinetics, and antiviral activity, VD2173 has anticancer efficacy in animal models of lung cancer. Judging from the protease selectivity data and potency of MM3122 and other new compounds, these are predicted to also show corresponding anticancer activity, in particular for HGF-driven prostate cancer for which TMPRSS2 is known to play a key role[358, 377].

Recent studies have identified other TMPRSS-family serine proteases that can mediate SARS-CoV-2 entry, including TMPRSS13 (293T-ACE2 and Calu 3 cells) [312, 339, 378]. and TMPRSS4 (enteroid cells) [345]. Though entry in Calu 3 cells is far less dependent on TMPRSS13 than TMPRSS2 [339, 356], the mechanistic contributions and/or redundancies of these proteases for SARS-CoV-2 entry are not known, nor is their relative inhibition resolved in the scope of this study. Future work in animal models, which can resolve the biological and tissue-specific relevance of protease inhibition, is best poised to answer these questions.

MM3122 represents an advanced lead candidate for clinical development as a novel antiviral drug for COVID-19 and against infections caused by other coronaviruses like MERS in which we showed equipotent activity of MM3122. We are currently optimizing MM3122 and this class of inhibitors for TMPRSS2 selectivity and antiviral potency, and in due course will test MM3122 and other optimized leads in appropriate animal models of COVID-19. In addition to being novel drugs, selective TMRSS2 inhibitors can be used as valuable chemical probes to help elucidate mechanisms of viral pathogenesis, including host cell-virus interactions, Spike protein processing and ACE2 receptor binding. Since TMPRSS2 plays a key role as a viral protein processing protease in the pathogenesis of other coronaviruses (SARS-CoV, MERS-CoV) as well

as influenza viruses [353, 379-383], MM3122 and this new class of inhibitors may be effective not only against COVID-19 but for infections caused by most or all corona and influenza viruses. Thus, these small-molecule inhibitors of TMPRSS2 not only hold much promise as new drugs to treat SARS-CoV-2 infections but also potentially represent broad-spectrum antivirals.

MATERIALS AND METHODS

TMPRSS2-Protease domain expression and purification

Periplasm secreted bacterial expression of serine proteases has been widely reported to yield correctly folded high-quality enzyme due to its reducing environment allowing for disulfide formation and fewer proteases being present compared to the cytoplasm.[384, 385] We therefore employed a similar strategy for TMPRSS2 protease domain. The human TMPRSS2 extracellular domain (residues 106-492) was cloned into a pET28a vector with a *pelB* leader sequence followed by an N-terminal 6x His tag (**Figure 3.7A**) using standard Gibson assembly procedures.[386] *E. Coli* DH5a cells were transformed with the Gibson assembly reaction mixture and resulting plated colonies were selected and verified with Sanger DNA sequencing for correct assembly of the cloned product. Correctly cloned plasmid was used for transforming *E. Coli* BL21(DE3) cells and later used for subsequent expressions. A single colony of BL21(DE3) cells was picked and inoculated into 50 ml of LB 2% glucose, 50 mg/ml Kanamycin (Km) culture and grown overnight at 37 ^{o}C 220 rpm. Liter scale liquid media (LB 0.1% glucose, 50 mg Km) were inoculated with overnight cultures at a starting OD_{600} of 0.05 and expression was induced with a final concentration of IPTG at 1mM at approximately OD_{600} of 0.7. Expressions were carried out at 16 ^{o}C for approximately 72 hours. Cells were collected by centrifugation and the periplasmic extracts were prepared by osmotic shock. Briefly, cell pellets were resuspended in TES buffer (200 mM Tris-HCl pH 8, 0.5 mM EDTA, 500 mM sucrose) and incubation at 4 ^{o}C for 1 hour followed by addition of cold water and incubation at 4 ^{o}C for 45 minutes. The periplasmic extract was isolated with centrifugation (10,000 g for 30 min) at 4 ^{o}C and imidazole (10

mM final concentration) and $MgCl_2$ (100 ml per liter expression culture volume) was added for overnight batch binding with Ni-NTA resin (2 ml slurry per liter of expression culture volume). Purification was carried out with 10 column volumes of Wash buffer 1 (50 mM Tris 250 mM NaCl pH 7.6), subsequently 10 column volumes of wash buffer 2 (50 mM Tris pH 7.6, 250 mM NaCl, 20 mM imidazole) and eluted with elution buffer (50 mM Tris pH 7.6, 250 mM NaCl, 4 mM benzamidine, 1 mM $CaCl_2$, 500 mM imidazole) (**Figure 3.7C**). SDS-PAGE was performed with fractions of eluate, and those showing the correctly sized band for TMPRSS2 protease domain were pooled. Consolidated fractions were buffer exchanged against 25 mM Citrate pH 6, 250 mM NaCl, 4 mM benzamidine, 1 mM $CaCl_2$ to remove excess imidazole and prevent autolysis and concentrated down in the same buffer with 10kDa NMWL Amicon ultra filtration units. Potential aggregates and degradation products were removed by size exclusion chromatography with 25 mM Citrate pH 6, 250 mM NaCl, 4 mM benzamidine, 1 mM $CaCl_2$ and 10% glycerol and Superdex 200 Increase 10/300 GL. TMPRSS2 protease domain containing fractions by SDS-PAGE (**Figure 3.7D**) were pooled. A final western blot against the protease domain (TMPRSS2 monoclonal antibody (M05), clone 2F4) was performed to verify collected sample to be TMPRSS2-protease domain. Pooled sample was flash frozen for subsequent assays (**Figure 3.7E**).

Recombinant enzyme inhibition assays

Recombinant human TMPRSS2 was purchased from Cusabio Technology (CSB-YP023924HU) and assayed in 25 mM Tris-HCl, 150 mM NaCl, 5 mM $CaCl_2$, 0.01% Triton X-100, pH 8.0 at a final concentration of 3 nM. To calculate K_M, Boc-QAR-AMC (Vivitude, MQR-3135-v) was serially diluted in DMSO and then each diluent was further

diluted in assay buffer such the final substrate concentration in the assay ranged from 0.514 μM to 200 μM with a final DMSO concentration of 0.5%. Assays were performed in a total volume of 30 uL in triplicate wells of a black 384-well plates (Nunc 262260) and initial velocity was measured in 35 second intervals for 7 minutes with excitation of 360 nm and emission of 460 nm on a Biotek HTX platereader. K_M and V_{max} were calculated from Michaelis-Menten plots using GraphPad Prism. For inhibitor studies, compounds were serially diluted 3-fold in DMSO and then preincubated with TMPRSS2 in assay buffer for 30 min at room temperature. The reaction was initiated following the addition of Boc-QAR-AMC and fluorescence was measured in 190 second intervals for 90 min. The final concentration of enzyme and substrate were 3 nM and 86.6 μM, respectively and the inhibitor concentrations ranged from 2 μM to 0.15 pM. Assays were performed in quadruplicate plates and IC_{50} was calculated from dose-response curves using Graphpad Prism. IC_{50} values for HGFA, matriptase, hepsin, factor Xa and thrombin were determined using our published assays [365-367].

Multiplex Substrate Profiling by Mass Spectrometry (MSP-MS)

Human recombinant TMPRSS2 (100 nM) expressed in yeast* was incubated at room temperature with a physiochemically diverse library of 228 tetradecapeptides to a final concentration of 500 nM. Aliquots were removed at different time intervals (15, 60 and 240 minutes) and subsequently quenched with 1Eq volume of 8M guanidinium hydrochloride. Samples were desalted using C18 tips (Rainin) and analyzed by LC-MS/MS peptide sequencing using a Quadrupole Orbitrap mass spectrometer (LTQ Orbitrap) coupled to a 10,000 psi nanoACQUITY Ultra Performance Liquid Chromatography (UPLC) System (Waters) for peptide separation by reverse phase

liquid chromatography (RPLC). Peptides were separated over a Thermo ES901 C18 column (75-µm inner diameter, 50-cm length) coupled to an EASY-SprayTM ion source and eluted by applying a flow rate of 300 nL/min in a 65-minute linear gradient from 2–50% in Buffer B (acetonitrile, 0.5% formic acid). Survey scans were recorded over a 325–1500 m/z range and up to the three most intense precursor ions (MS1 features of charge ≥ 2) were selected for collision-induced dissociation (CID). Data was acquired using Xcalibur software and processed as previously described.[387] Briefly, raw mass spectrometry data was processed to generate peak lists using MSConvert. Peak lists were then searched in Protein Prospector v.6.2.2[388] against a database containing the sequences from the 228 tetradecapeptide library. Searches used a mass accuracy tolerance of 20 ppm for precursor ions and 0.8 Da for fragment ions. Variable modifications included N-terminal pyroglutamate conversion from glutamine or glutamate and oxidation of tryptophan, proline, and tyrosine. Searches were subsequently processed using the MSP-xtractor software (http://www.craiklab.ucsf.edu/extractor.html), which extracts the peptide cleavage site and spectral counts of the corresponding cleavage products. Spectral counts were used for the relative quantification of peptide cleavage products.

Expression of Human Recombinant TMPRSS2 in *Pichia pastoris*

Human Recombinant TMPRSS2 was expressed and purified from yeast following protocols previously described[358], with slight modifications. Briefly, the scavenger receptor cysteine-rich (SRCR) and serine protease domain regions of TMPRSS2-N249G were cloned into a pPICZα-B construct (EasySelect Pichia Expression Kit – Invitrogen) according to the manufacturer, transformed into *Pichia pastoris* strain X33,

and selected on plates containing increasing concentrations of Zeocin. A single colony was grown in 10 mL buffered medium with glycerol (BMGY) overnight at 30˚C and 230 rpm. The overnight culture was used to inoculate 1 L BMGY and grown until OD 2-6. Cells were pelleted and resuspended in 100 ml buffered medium with methanol (BMMY). Cells were induced with 5% methanol every 24 hours for 72-96 hours. Secreted TMPRSS2 was precipitated with 70% ammonium sulfate at 4˚C overnight, pelleted at 27,000 x g for 45 minutes and resuspended in 50 mM Tris pH 8, 0.5 M NaCl, 0.01% CHAPS, and solubilized for 2 hours at 4˚C. The solubilized protein was then purified over a gravity column containing soybean trypsin inhibitor immobilized agarose (Pierce). TMPRSS2 was concentrated and then purified with size exclusion chromatography using 50 mM potassium 83 phosphate buffer pH 6, 150 mM NaCl, 1% glycerol to prevent auto-proteolysis."

Cell-based TMPRSS2 fluorogenic enzyme inhibition assay

A PLX304 plasmid–containing human TMPRSS2 open reading frame from the ORFeome Collaboration (Dana-Farber Cancer Institute, Broad Institute of Harvard and Massachusetts Institute of Technology [HsCD00435929]) was obtained from DNASU Plasmid Repository, and a control PLX304 vector was obtained from Addgene (Watertown, MA, USA). HEK-293T cells were grown in DMEM supplemented with 10% FBS and seeded in a black, 96-well plate (75,000 cells/well). The following day, cells were transfected overnight with either a control plasmid (PLX) or TMPRSS2 (PLX-TMPRSS2) via TransIT LT-1 transfection reagent (Mirus Bio) in 100 µL of OptiMEM per well. Twenty-four hours after transfection, the media was replaced with 80 µL of phosphate-buffered saline (PBS). Inhibitors or PBS alone were added to the wells in the

indicated 5 concentrations and incubated at 25°C for 15 minutes. The fluorogenic substrate Boc-QAR-AMC (R&D Biosystems) was then added to each well to a final concentration of 100 μM. Fluorescence (excitation 365 nm, emission 410 nm) was kinetically measured every 15 minutes for a total time of 150 minutes at 37°C using a GloMax plate reader (Promega).

Antiviral activity (VSV SARS-CoV-2 and MERS chimeras):

Viruses used: VSV-eGFP, a recombinant VSV expressing a GFP reporter (depends on the VSV glycoprotein G for entry), has been previously described.[278] VSV-SARS-CoV-2, a replication competent infectious VSV chimera which employs the SARS-CoV-2 Spike (S) protein for viral entry in place of VSV G and expresses eGFP, has also been previously described.[275, 287] VSV-MERS was created in the same manner as VSV-SARS-CoV-2, except the MERS Spike missing the terminal 21 amino acids (HCoV-EMC/2012 strain) was inserted in place of VSV G.

Experimental procedure: Human Calu-3 lung epithelial cells or Vero cells (African green monkey kidney) were seeded in a 96-well black plate in DMEM containing 10% FBS for 24 hours (37ºC and 5% CO2). The next day, cells were pretreated for 2 hours with inhibitor or vehicle control (DMSO) in 50 μl serum-free DMEM, and subsequently infected with VSV-SARS-CoV2, VSV-MERS, or VSV-eGFP at a multiplicity of infection (MOI) of 0.5. At 7 hours post infection (single round of infection), cells were fixed in 2% formaldehyde and nuclei stained with 10 μg/ml Hoechst 33342 (Invitrogen) for 30 minutes at room temperature. Cells were washed once and stored in PBS after fixation, and automated microscopy performed using an InCell 2000 Analyzer (GE Healthcare) in the DAPI and FITC channels (10X objective, 9

fields per well, covering the entire well). Images were analyzed using the Multi Target Analysis Module of the InCell Analyzer 1000 Workstation Software (GE Healthcare) to identify the percentage of GFP-positive cells in each well (top-hat segmentation). The percentage GFP-positive cells for experimental conditions was normalized against the control (DMSO-treated) cells and expressed as relative infectivity. GraphPad Prism (version 8.4.2) was used to calculate the EC_{50} of each drug. Statistics (comparison of VSV-SARS-2 to VSV-eGFP with each drug; Student's t-test) were performed in Microsoft Excel. Three biological replicates were performed.

Antiviral activity (VSV pseudotype SARS-CoV-2)

We analyzed pseudotype entry driven by the spike protein of SARS-CoV-2 (SARS-2-S) or the glycoprotein of vesicular stomatitis virus (VSV-G) into the TMPRSS2-positive human lung cell line Calu-3[293, 357]. VSV-G was used as a control, as it does not depend on TMPRSS2 for host cell entry. Besides Calu-3 cells, we further used Vero cells (African green monkey, kidney) as a control, as these cells do not express TMPRSS2 and therefore any reduction in SARS-2-S-driven entry would be either related to unspecific side effects or cytotoxicity. Each compound was tested in three separate experiments with independent pseudotype batches. We treated cells (96-well format) with different concentration of inhibitor or solvent (DMSO) diluted in medium (50 µl/well) for 2 h at 37 °C and 5 % CO2, then we added 50 µl of pseudotype on top and incubated for 16 h at 37 °C and 5 % CO2. Next, we measured virus-encoded firefly luciferase activity in cell lysates (indicator of pseudotype entry into target cells). The data was normalized against control (DMSO-treated cells = 100 % pseudotype entry) and plotted with GraphPad Prism (version 8.3.0), to calculate effective concentration

EC50 and perform statistics (comparison against the respective control; two-way ANOVA with Dunnett's posttest).

Antiviral activity (Wild-type SARS-CoV-2):

We seeded Calu-3 cells onto a white, flat bottom 96-well plate at 10,000 cells per well in Dulbecco's Modified Eagle Medium supplemented with 10% heat-inactivated fetal bovine serum, 100 U/mL Penicillin-Streptomycin (Life Technologies, Inc., 15140-163), and buffered with 10 mM HEPES and incubated in a humidified 5% CO_2 incubator at 37°C. 16 hours later, the media was removed and replaced with the 50 µl/well of the same media but containing 2% heat-inactivated fetal bovine serum (D-2) instead of 10%. In D-2, compounds were serially diluted 3-fold in a 9-point series and added in a 25 ul volume to the 96-well plate to make final concentrations ranging from 50 µM to 7.6 nM on the assay plate for one hour. We then infected the cells (with compounds still on them) with 25 ul/well of SARS-CoV-2 at an MOI of 0.1 PFU/cell, with a final total volume of 100 µl/well. We incubated the plates for 72 hours as described above and then added 25 µL/well of CellTiter-Glo® Reagent (prepared as directed by manufacturer, Promega, G7573), followed by 5 min shaking and 20 min room temperature incubation before detecting luminescence with a Biotek Synergy H1 plate reader. Compound cytotoxicity was determined using the same assay, but instead of adding virus, we added 25 µl of D-2. These experiments were performed three times, and we performed nonlinear regression on log(inhibitor) vs response with variable slope and four parameters on compiled results from the three experiments. Likewise, IC_{50}s were calculated based on all three biological replicates.

Animal studies

All animal procedures were performed following the guidelines and approval of the Washington University's Institutional Animal Care and Use Committee. Animals were maintained in a controlled temperature (22°C± 0.5°C) and lighting (12L:12D) environment. Standard laboratory chow and water were supplied ad libitum.

Mouse toxicity studies

NSG male and female mice (3 per group) were injected with 0, 20, 50, or 100 mg/kg MM3122 (in 1-5% DMSO/95% saline; IP) daily for 7 days. Animal health was assessed daily for drug toxicity (ruffled fur, discharge from the eyes, weight loss, dehydration, lethargy, hypothermia, abnormal tissue growth) and body weights were measured every 2-3 days. At the end of the experiment animals were sacrificed and organs inspected. Liver, spleen, and both kidneys were collected and weighed.

Mouse pharmacokinetic (PK) studies

MM3122 was tested for its PK in mice (single IP dosing, actual dose 16.7 mg/kg). Eight animal groups contained three animals each (male, CD-1) with plasma drawn and lung tissue harvested at eight different sampling time points (0.08, 0.25, 0.5, 1, 2, 4, 8, 24 h) post dosing. VD2173 was tested for its PK in mice (single IP dosing, actual dose 16.7 mg/kg). Three animal groups contained three animals each (male, CD-1) with plasma drawn and lung tissue harvested from each group of animals at three different sampling time points (0.25, 8, 24 h) post dosing. Bioanalysis of plasma and lung tissue extracts were performed using standard LC/MS/MS techniques.

Acknowledgements:

We thank David Griggs and Scott Campbell at Saint Louis University for conducting the pharmacokinetic studies on MM3122. We also thank Michael Winter at

UCSF for his help with the initial MSP-MS and PS-SCL profiling. This work was funded by grants from the Siteman Cancer Center (#16-FY18-02 and SCC P30CA091842) of Washington University and Barnes Jewish Hospital Foundation (BJHF 4984) in Saint Louis (J.WJ.). This work was also funded by the following awards from the National Institutes of Health (NIH): R43 CA243941 (J.W.J. and L.K.), R43 CA224832 (J.W.J. and L.K.), U19 AI142784 (C.L.S.), P50AI150476 (C.S.C.), U19 AI070235 (M.E.R) and the Campaign Urging Research for Eosinophilic Diseases (CURED) Foundation (M.E.R.). Additional support was provided by a Fast Grant from Emergent Ventures at the Mercatus 9 Center, George Mason University (C.S.C.). S.P. acknowledges funding by BMBF (RAPID Consortium, 01KI1723D and 01KI2006D; RENACO, 01KI20328A; SARS_S1S2 01KI20396; COVIM consortium, 01KX2021), the country of Lower Saxony (14-76103-184) and the German Research Foundation (DFG) (PO 716/11-1, PO 716/14-1). Work with live SARS-CoV-2 was funded by a Burroughs Wellcome Fund Investigators in the Pathogenesis of Infectious Disease Award (C.L.S.).

Chapter 4:

Discussion

Michael Tartell

Summary of Results

We report here the impacts of two key host factors on viral replication. First, we demonstrate a novel role for PCIF1-mediated m^6A_m methylation at the cap-proximal nucleotide in the type I interferon response to VSV infection. PCIF1 is necessary and sufficient for N6-methylation of viral mRNA. In contrast to cellular mRNA, viral mRNA required prior ribose 2'O methylation, but not guanine-N7 methylation, for efficient methylation by PCIF1. While PCIF1 had no impact on viral mRNA stability, mRNA translation, or replication under basal conditions, loss of PCIF1 sensitized VSV infection to the effects of interferon beta (IFN-β) pretreatment. Second, we demonstrate the development of a novel class of inhibitors of the protease TMPRSS2. Treatment with these compounds potently inhibits not only TMPRSS2 enzymatic activity, but entry of chimeric VSV-SARS-CoV-2 and VSV-MERS into Calu-3 epithelial lung cells. They were also potently antiviral against authentic SARS-CoV-2 infection in cell culture, and had outstanding pharmacokinetics and safety profile in mice.

Open Questions in Coronavirus Entry

MM3122 and other inhibitors developed in this work are advanced candidates for pre clinical development as novel antivirals or COVID-19, as well as infection with other viruses which depend on TMPRSS2 for entry (e.g. MERS). These drugs are currently being optimized for testing in animal models of SARS-CoV-2 infection. However, significant questions still remain about the role of proteases in SARS-CoV-2 viral entry, and how the emergence of new variants is altering that landscape.

SARS-CoV-2 spike catalyzes the process of membrane fusion following cleavage by host proteases at the S1/S2 and S2' sites [289]. Current models for these cleavage events include S1/S2 cleavage in the producer cells by furin, followed by S2' cleavage by TMPRSS proteases at the plasma membrane or cathepsins in the endosome [289]. However, requirements for each of these proteases and cleavage events depend heavily on cell type, with TMPRSS2 primarily mediating entry into lung cells, and cathepsins mediating entry into other cell types such as kidney cells [293]. This extends to inhibition of entry using protease inhibitors, as TMPRSS2 inhibitors potently block entry into lung but not kidney cells, and vice versa for cathepsin inhibitors [293]. We recapitulate this in our own data, as our novel TMPRSS2 inhibitors potently inhibited entry into Calu-3 lung epithelial cells, but not entry into TMPRSS2-negative vero cells. Similarly, the furin cleavage site seems to be required for efficient infection of lung cells in an animal model, but is lost upon passage in vero cells [294, 308, 311]. Clear biochemical evidence tracking the precise role of each protease in cleavage of spike across multiple cell types, both in the producer cell and in the newly infected cell, is inadequate.

Understanding the individual contributions of different proteases is also critical for our understanding of viral pathogenesis. It has been hypothesized that a key difference between SARS-CoV and SARS-CoV-2 is the presence of a multi-basic furin cleavage site, enabling S1/S2 cleavage in the producer cell and enhancing its infectivity of lung cells [294, 307, 309]. Competition assays between SARS-CoV-2 and SARS-CoV-2 lacking the furin cleavage site showed a selective advantage for the furin-site-negative virus in TMPRSS2-negative vero cells, but not in vero cells overexpressing TMPRSS2

[308]. Further studies in multiple cell types using VSV chimeras expressing mutant forms of SARS-CoV or SARS-CoV-2 spike proteins could elucidate these mechanisms more precisely. Outside of the laboratory, SARS-CoV-2 has continued to evolve. The spike protein of the Omicron BA.1 variant contains additional mutations reported to reduce its dependency on TMPRSS2, and correspondingly change its replication in different cell types [389]. Replication of omicron relative to the delta variant was reduced in Calu-3 and Caco-2 cells, with no change in nasal airway epithelial cells [389]. Consistent with these results, omicron was less sensitive to the TMPRSS2 inhibitor camostat and more sensitive to the cathepsin inhibitor E64d than delta, and showed reduced cleavage at S1/S2 in producer cells [389]. Whether this has contributed to a change in SARS-CoV-2 pathogenesis and mortality rates is not yet clear, nor whether this shift away from TMPRSS2 would reduce the efficacy of a TMPRSS-targeting antiviral treatment. A combination therapy targeting both TMPRSS- and cathepsin-dependent pathways may be required for optimal efficacy in the face of a constantly evolving virus.

Studies with influenza provide a model of the potential issues associated with this approach. Influenza is capable of being cleaved by many different proteases, including TMPRSS2 [390]. These include TMPRSS4, HAT, TMPRSS13, matriptase, DESC1, KLK5, KLK12, plasmin, furin, and proprotein convertase 5/6 [350, 391]. While inhibition or knockout of TMPRSS2 was effective against some subtypes of flu, they were ineffective against others. For example, TMPRSS2 knockout mice were highly tolerant of infection with H1N1 influenza, but were susceptible to H5H1 or H7N7 [352, 392]. A key contribution to this difference is a multi-basic cleavage site found in high-pathogenic

subtypes of flu which allow cleavage by furin and proprotein convertase 5/6 [351]. Treatment with inhibitors ineffective against these viruses could cause selection for these subtypes over time, resulting in the original treatment becoming ineffective. Additionally, an H3N2 influenza subtype which showed poor pathogenesis in TMPRSS2 knockout mice acquired escape mutations after passage, allowing for processing by an unknown protease and restoring high virulence in infected mice, raising the possibility that protease inhibitor treatments could become ineffective even faster if used on their own [393]. The parallels between influenza and SARS-CoV-2 are notable, including the acquisition of a poly-basic furin cleavage site which changes the infectivity of the virus. Passage studies both in cell culture, and possibly in an animal model, should be performed in order to determine mechanisms by which SARS-CoV-2 could escape protease inhibition. The chimeric VSV system has already been used extensively to isolate escape mutants of antibody neutralization, and could function similar in these assays [394]. A combination of inhibitors effective not only against a wide range of proteases, but also against other parts of the viral life cycle, could potentially reduce the risk of rapid escape.

Open Questions in PCIF1 and m^6A_m

This work describes a novel role for PCIF1-mediated m^6A_m methylation in the IFN-β response to VSV infection. We show that VSV mRNA is N6-methylated by PCIF1, and that PCIF1 protects VSV gene expression against the antiviral response induced by pretreatment with IFN-β. However, significant gaps remain in our understanding of this effect, as well as the role of PCIF1 and m^6A_m in the cell and on an

organismal level. Current findings relating to these issues, future directions of study, and speculation on potential mechanisms based on this data are discussed below.

Phenotype Dependence on Cellular and/or Viral m^6A_m

Although we demonstrate that viral mRNAs are modified with m^6A_m by PCIF1, we have not demonstrated a direct connection between the presence of m^6A_m in viral mRNA and the IFN- and PCIF1-dependent phenotype; knocking out *PCIF1* removes m^6A_m from both viral and cellular mRNA. As such, there are two potential models to explain our observations. In the viral mRNA dependent model, an unknown cellular factor discriminates binding to viral mRNA based on the presence or absence of m^6A_m. In this way, m^6A_m would serve as a self-marker for self/non-self recognition by the cell. IFIT1/3 is an example of this mechanism, which binds to and suppresses translation of mRNAs lacking cap-proximal 2'O methylation [250]. Though our work has ruled out IFIT1/3 and RNA sensors such as RIG-I, there may be unknown RNA binding proteins which specifically discriminate m^6A_m binding.

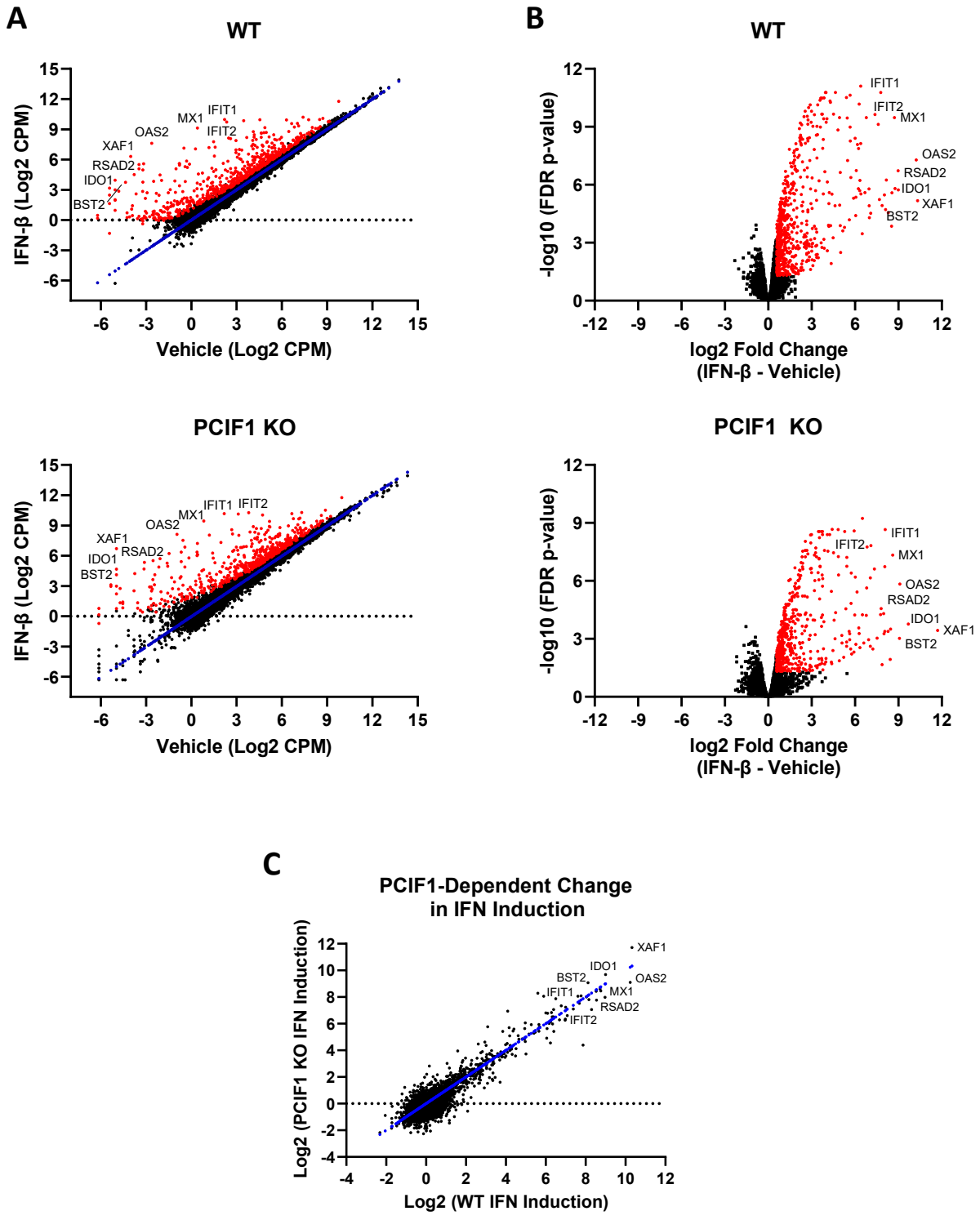
In contrast, a cellular mRNA dependent model posits that the presence of m^6A_m in cellular mRNA changes the interferon response in the cell. Under this model, PCIF1 and m^6A_m would suppress some aspect of the antiviral response, as we see an augmented antiviral effect in *PCIF1* knockout cells. Though we cannot rule out this model, it stands in contrast to known effects of cap-proximal 2'O methylation, which enhances the expression of certain ISGs [266]. It also seems unlikely that the cell would dampen its innate immune response by default, as PCIF1 is ubiquitously expressed and m^6A_m is highly abundant in cells [67]. To test whether the transcriptional

response to interferon is altered by PCIF1, we pretreated wild type parental or *PCIF1* knockout A549 cells with IFN-β for 6 hours, and performed differential expression analysis by RNA-seq **(Fig 4.1)**. We saw robust upregulation of genes upon IFN-β stimulation **(Fig 4.1A)**, with known ISGs constituting the most significant hits **(Fig 4.1B)**. However, the level of IFN-induction for each gene was largely unaffected by *PCIF1*-knockout **(Fig 4.1C)**. Though we cannot rule out that a small subset of genes could be causing a change in the antiviral response to VSV, there was no broad change in the expression of ISGs in the absence of PCIF1.

To further test these two models, we performed a mass spectrometry based screen for RNA-binding proteins bound to the 5' UTR of VSV. Briefly, we purchased a 30 nt long RNA oligo containing a 5' cap1 structure, the first 30 nt of VSV M mRNA, and a 3' biotin ($m^7GpppA_mACAGAUAUCACAAUCUAAGUGUUAUCCCA$-biotin.BB). We then further N6-methylated or mock methylated this oligo with purified PCIF1, generating an "m^6A_m oligo" and "A_m oligo" differing only in their cap-proximal N6-methylation. We bound these oligos to streptavidin beads and used them as bait to select RNA-binding proteins in IFN-β treated lysates from wild type or *PCIF1*-knockout cells, and identified bound proteins by liquid chromatography followed by mass spectrometry. We verified this approach by performing a competition experiment for known cap-binding protein eIF4E. An oligonucleotide containing a 5' triphosphate corresponding to the VSV leader sequence (5' pppACGAAGACAAACAAACCAUUAUUAUCA) was capped and 2'O methylated using vaccinia capping enzymes and 2'O methyltransferase. The triphosphate or capped oligo was then added to the RNA-protein binding reaction to compete for eIF4E binding

Figure 4.1: No change in bulk IFN-β induced transcripts. The indicated A549 cells were treated with vehicle (0.1% BSA) or 500 U ml-1 IFN-β for 5 hours. RNA was extracted in Trizol and differential expression analysis performed by RNA-seq (n=2). Red indicates significant differences between conditions for a gene (Benjamini-Hochberg false discovery rate $p<0.05$), black no significant difference ($p>0.05$), and blue the line of identity. (A) Log2 counts per million mapped reads (CPM) are plotted for the indicated condition within each cell type. (B) Volcano plot indicating the log2 fold change in CPM with IFN-β treatment (Log2 CPM IFN-β – Log2 CPM Vehicle) versus significance. (C) Changes in the IFN-β induction of each gene with PCIF1 knockout. IFN-β induction in wild type cells (Log2 CPM IFN-β – Log2 CPM Vehicle) versus the same in knockout cells is shown.

Fig 4.1 (continued):



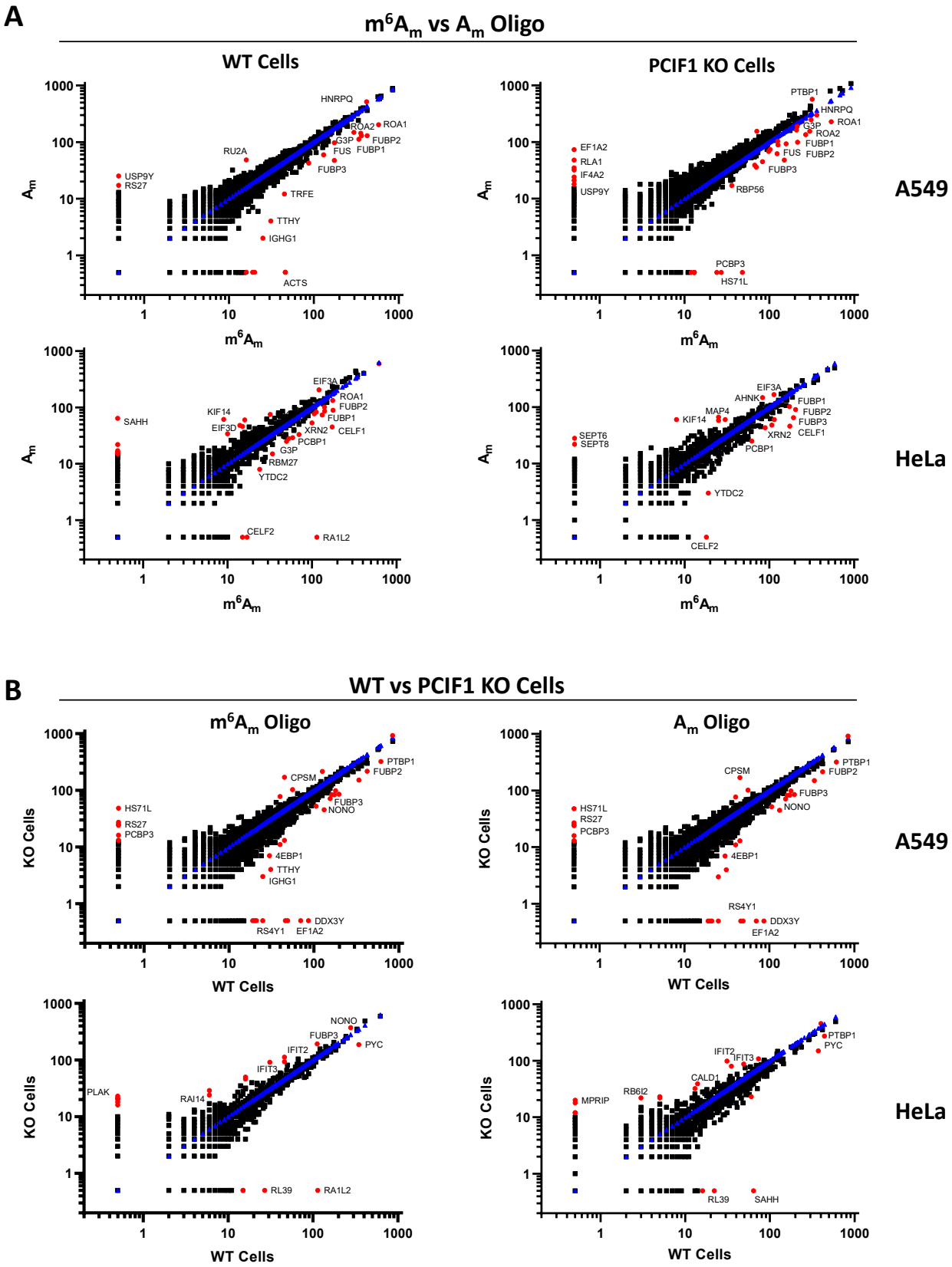
with the biotinylated capped oligo. Up to 200 times excess of the triphosphate oligo had no impact on eIF4E binding, while as little 2 times excess of the capped oligo proved an effective competitor **(Fig A3.1)**, demonstrating that this approach can resolve cap-specific differences in protein binding.

Using this approach, we screened for m^6A_m-dependent differences with the same cellular background, and cellular *PCIF1*-dependent differences with the same methylated oligo background **(Table A3.T1)**. Outlying "hits" were determined by Fisher's exact test with Benjamini-Hochberg false discovery rate correction. The m^6A_m-oligo was enriched for ~1.5 to 3-fold more of several related proteins **(Fig 4.2A)**, including far-upstream binding proteins 1, 2, and 3 (FUBP 1-3), and heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1, identified in this dataset as ROA1). *PCIF1*-knockout produced no consistent set of differentially enriched proteins across HeLa and A549 cells, with the exception of polypyrimidisne tract binding protein 1 (PTBP1), which was ~1.5-fold more highly enriched with wild type cells **(Fig 4.2B)**. Taken with the lack of transcriptional differences in the A549 cells this data favors a viral mRNA dependent model, in which certain RNA binding proteins specifically discriminate viral mRNA based on the presence of m^6A_m, as opposed to the cellular dependent model.

No reader proteins have yet been reported for m^6A_m. Though these candidates have not yet been validated, nor their role in VSV replication examined, they have reported roles in viral translation. hnRNPA1, FUBP1, and FUBP3 bind to different sites in the internal ribosomal entry site (IRES) in the 5' UTR of enterovirus 71 and hepatitis C virus (HCV) and enhance its translation, while FUBP2 represses translation [395-400]. hnRNPA1 also positively regulated replication of Sindbis virus and hepatitis C

Figure 4.2: Identification of N6-methylation specific RNA binding proteins. Wild type (parental) or PCIF1 knockout cells from the indicated cell lines were treated with IFN-β for 5 hours (A549) or 3 hours (HeLa). Lysates from these cells were bound to an RNA oligo with a 3' biotin and 5' cap, containing either a m^6A_m or A_m methylated cap-proximal nucleotide, bound to streptavidin beads. RNA-binding proteins specific for these oligos were isolated and identified by mass spectrometry. Protein spectrum counts for each condition are displayed. Red indicates statistical significance by Fisher's exact test with Benjamini-Hochberg false discovery rate correction ($p < 0.0008$). (A) Binding to differentially methylated oligos in the same cell lysate background. (B) Differences in binding to identically methylated oligos in a wild type or PCIF1 knockout cell background.

Fig 4.2 (continued):



virus [395, 400, 401]. hnRNPA1 and other hnRNP proteins relocalize to the cytoplasm during VSV infection, but were implicated in regulation of apoptotic signaling, not viral gene expression (though this was not tested under IFN-β induced conditions) [102]. hnRNPA2B1, which has ~60% identity to hnRNPA1 and was a hit in the A549 cells, also regulates RNA export of ISGs in an m^6A-dependent manner [180].

Whether these proteins represent true m^6A_m-reader proteins, and what effect they have on viral replication during the interferon response, remains to be explored definitively. However, the combination of RNA-seq and proteomics data support a viral mRNA-dependent model in which a post-transcriptional regulator binds with higher affinity to m^6A_m viral mRNAs and positively regulates their expression. While this could be due to protection against RNA degradation, as m^6A_m protects against decapping by DCP2 [162], the aforementioned effects of FUBP1-3 and hnRNPA1 on EV71, Sindbis, and HCV favor a mechanism involving viral translation. Each of these viruses carries out eIF4E-independent translation of their mRNA using an IRES, and this process is enhanced by FUBP1, FUBP3, and hnRNPA1. VSV induces eIF4E-dependent translational shutdown without interferon pretreatment [91], so it is unlikely this acts solely as a binding site for eIF4E-independent translation. More likely, this mechanism helps overcome interferon-induced translational repression of viral mRNA. This may also require a second signal from viral infection (possibly nucleic acid sensing), as interferon pretreatment absent infection does not result in bulk translational repression (**Fig A1.S9**, and previously reported [402]). VSV mRNAs have been reported to have equal sensitivity to eIF2α phosphorylation (induced by protein kinase R (PKR) and thapsigargin) as cellular mRNAs [403], though this was not tested in the presence of

interferon or PCIF1 knockout. It is possible that translation of VSV mRNAs could further decrease in the absence of PCIF1, such that eIF2α phosphorylation would preferentially repress VSV mRNA translation over cellular translation without m^6A_m. Alternatively, this translational repression could be specific for viral RNA through an unknown mechanism, as not all ISGs affecting translation have been well characterized (e.g. PARPs, ISG20) [246, 404]

Cap-Proximal Nucleotide Methylation During the IFN Response

Though cap methylations are known to impact self/non-self recognition during the innate immune response, it is unclear whether cap methylations are preferentially installed on particular transcripts or under particular conditions. CMTR1, which is transcriptionally upregulated by interferon and was first characterized as interferon stimulated gene 95 [405], has been shown to regulate the expression of certain ISGs [266]. CMTR1 knockdown resulted in an impaired antiviral response due to IFIT1-mediated translational repression of these ISGs, though an RNA element present in a gene not regulated by CMTR1 (IFIT3) alleviated this effect [266]. Interestingly, though CMTR1 is upregulated by interferon, pretreatment with interferon was reported to decrease 2'O methylation at the first nucleotide in vaccinia virus infected chick cells (both for viral and cellular mRNA) and at the second nucleotide in reovirus infected mouse cells (reovirus synthesizes its own 2'O methylation at the first nucleotide) [406, 407]. Similar results were reported for VSV, due an interferon-induced increase in the ratio of S-adenosyl-homocysteine (SAH – an inhibitor of methyltransferases) to S-adenosyl-methionine (SAM – the methyl donor for methyltransferases) [408]. Clear data is lacking on the regulation of cap nucleotide methylation during interferon

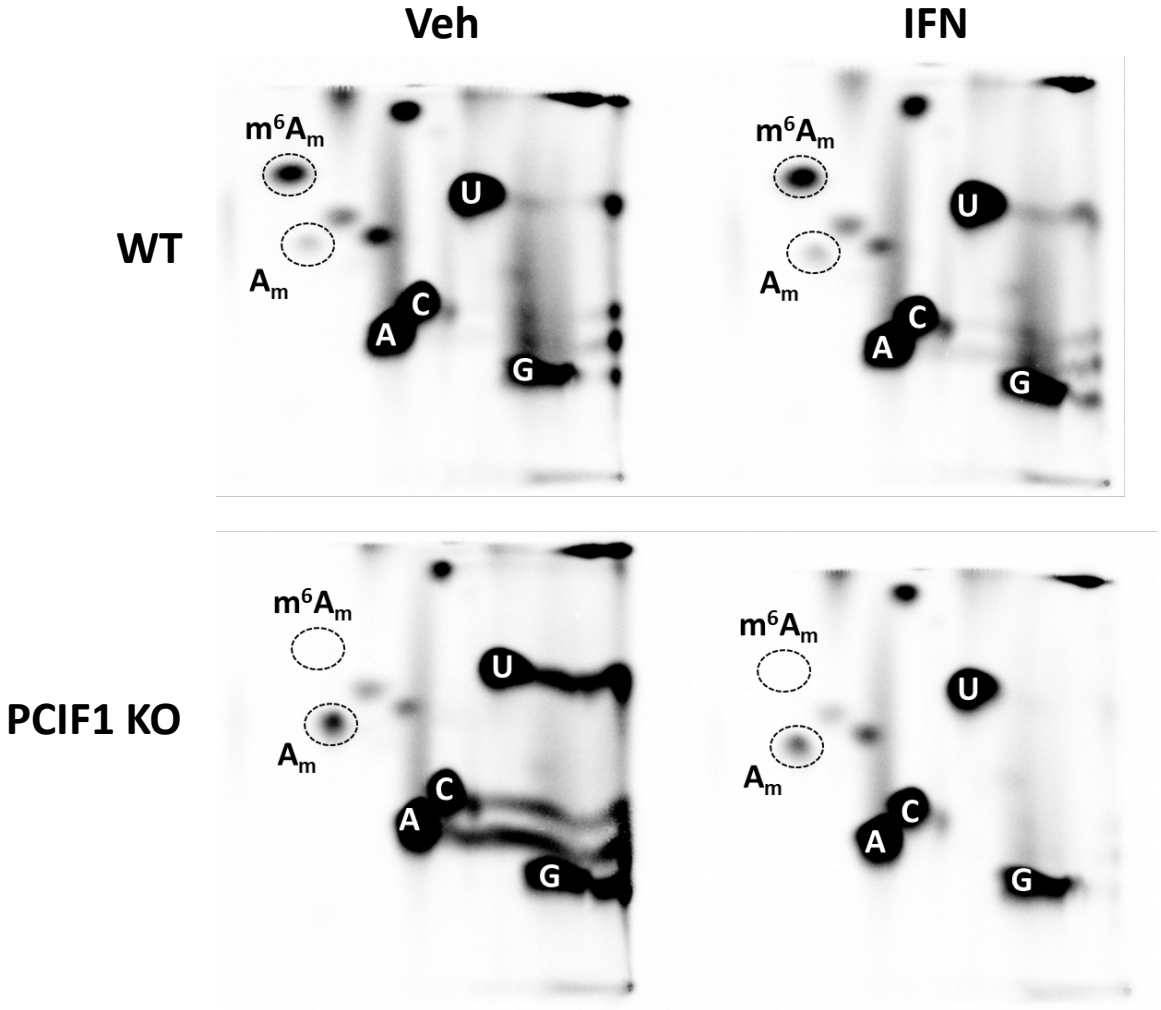
stimulation, and what role that plays in self/non-self recognition mechanisms that rely on 2'O- or N6-methylation.

To elucidate this, we attempted to identify the cap-proximal nucleotide composition of nascent transcripts synthesized during the interferon response. Wild type (parental) or *PCIF1* KO A549 cells were treated with vehicle (0.1% BSA) or 500 units ml^{-1} of IFN-β, and nascent transcripts labeled with 5-ethynyl uridine (EU). Poly(A)+ was isolated, EU(+) RNA biotinylated, and isolated by streptavidin pulldown. Cap-proximal nucleotides were specifically radiolabeled and analyzed by 2D-TLC. No change in the m^6A_m to A_m ratio was observed with interferon treatment **(Fig 4.3)**. *PCIF1* knockout cells also showed a consistent reduction in A_m with IFN-pretreatment across 3 replicates **(Fig 4.3)**, consistent with the aforementioned reports. This data indicates that ISGs are not preferentially methylated with m^6A_m, though there still may be regulation at the individual transcript level that cannot be resolved here.

Recruitment of Host Factors to VSV mRNA

VSV mRNA synthesis occurs exclusively in the cytoplasm, while PCIF1 is an RNA Pol II-associated enzyme which has been shown to localize to the nucleus by IF microscopy [160, 161]. Though we demonstrated that VSV mRNAs are N6-methylated by PCIF1, we did not determine whether it re-localizes to the cytoplasm or is recruited to viral mRNA. Certain nuclear proteins re-localize to the cytoplasm during VSV infection in an M-dependent manner [101, 102, 108], and this effect is abrogated by a single point mutation in M (M_{M51R}) [102]. Though how this occurs is poorly understood, PCIF1 could be re-localizing from the nucleus by the same mechanism. Since over 60% of

Figure 4.3: Cap-proximal nucleotides synthesized during IFN-β treatment. Wild type (parental) or PCIF1 KO A549 cells were treated with vehicle (0.1% BSA) or 500 U ml-1 IFN-β. One hour post treatment, nascent RNAs were labeled with 2.5 mM 5-ethynyl uridine (EU) for an additional 7 hours. RNA was extracted, poly(A) selected, and incorporated EU biotinylated. Labeled RNA was isolated by streptavidin pulldown, cap-proximal nucleotides radiolabeled by successive decapping, dephosphorylation, phosphorylation with [32]P gamma-ATP, digested to mononucleotides, and analyzed by 2D-TLC and phosphorimaging(representative image, n=3).

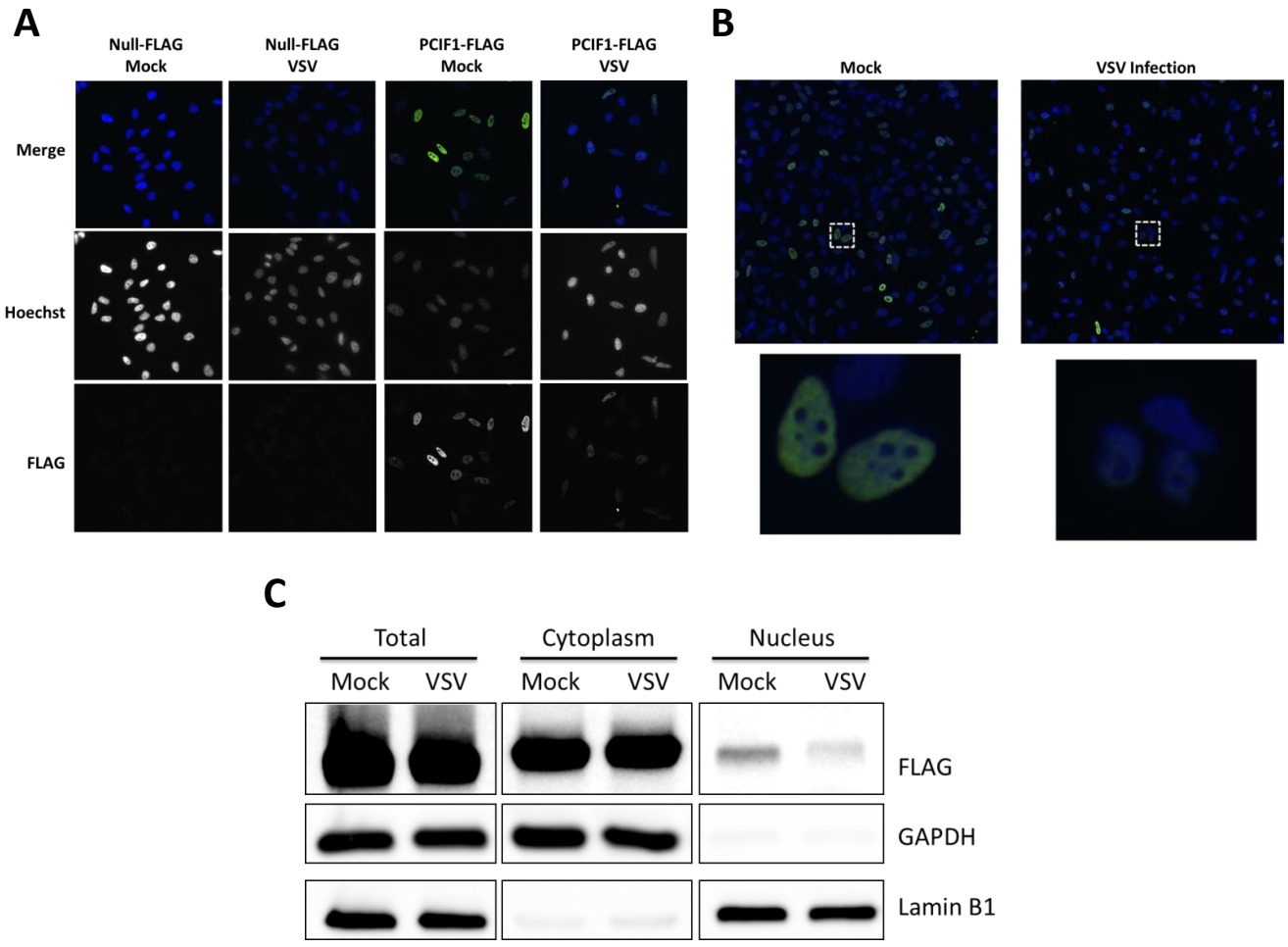


transcripts in an infected cell are viral transcripts at 6 hours post infection [32], even non-specific methylation by PCIF1 following re-localization to the cytoplasm could plausibly result in substantial methylation of viral transcripts. As PCIF1 contains a WW-domain, and the PPxY late domain of M is known to bind to the WW-domains of cellular proteins, it is also possible that PCIF1 could be actively recruited by viral proteins [160] [409].

To determine the localization of PCIF1 during infection, we used *PCIF1*-knockout HeLa cells stably expressing FLAG-tagged PCIF1 [131]. Cells were infected or mock infected with VSV at a MOI 3 for 6 hours, and IF microscopy performed for FLAG-PCIF1. In mock infected cells, PCIF1 was exclusively seen in the nucleus **(Fig 4.4A)**. VSV infection resulted in a decrease in nuclear PCIF1, but no proportional increase in the cytoplasm **(Fig 4.4A-B)**. Whether this occurred due to degradation of total PCIF1 levels during infection or a technical issue in detecting cytoplasmic PCIF1 was unclear. We then performed nuclear cytoplasmic fractionation under the same infection conditions. Total FLAG-PCIF1 was unchanged during infection, and nuclear PCIF1 was decreased during infection consistent with our microscopy result **(Fig 4.4C)**. Surprisingly, the majority of PCIF1 was present in the cytoplasmic fraction even without infection, a result not consistent with our microscopy **(Fig 4.4C)**. GAPDH and Lamin were heavily enriched in the cytoplasmic and nuclear fractions respectively, indicating a successful fractionation **(Fig 4.4C)**.

As PCIF1 did re-localize to the cytoplasm upon VSV infection, we sought to determine whether PCIF1-dependent methylation was dependent on M. We metabolically radiolabeled mRNA during infection with VSV, or VSV containing the

Figure 4.4: Localization of PCIF1 during VSV infection. PCIF1 KO HeLa cells reconstituted with FLAG-PCIF1 or an empty vector were infected or mock infected with VSV at a MOI of 3 for 6 hours. (A) Cells were fixed in 2% paraformaldehyde and stained with anti-FLAG antibody or Hoeschst. Confocal microscopy was performed using a 60X objective (representative image, n=3). (B) Multiple images from the indicated condition in A were combined together to create a larger field of vision. (C) Cells were lysed by passage through a 27G needle, and cytoplasmic and nuclear fractions isolated. Fractions were run on a 12% SDS-PAGE gel, transferred to membrane, and analyzed by immunoblot for the indicated proteins (representative image, n=3)

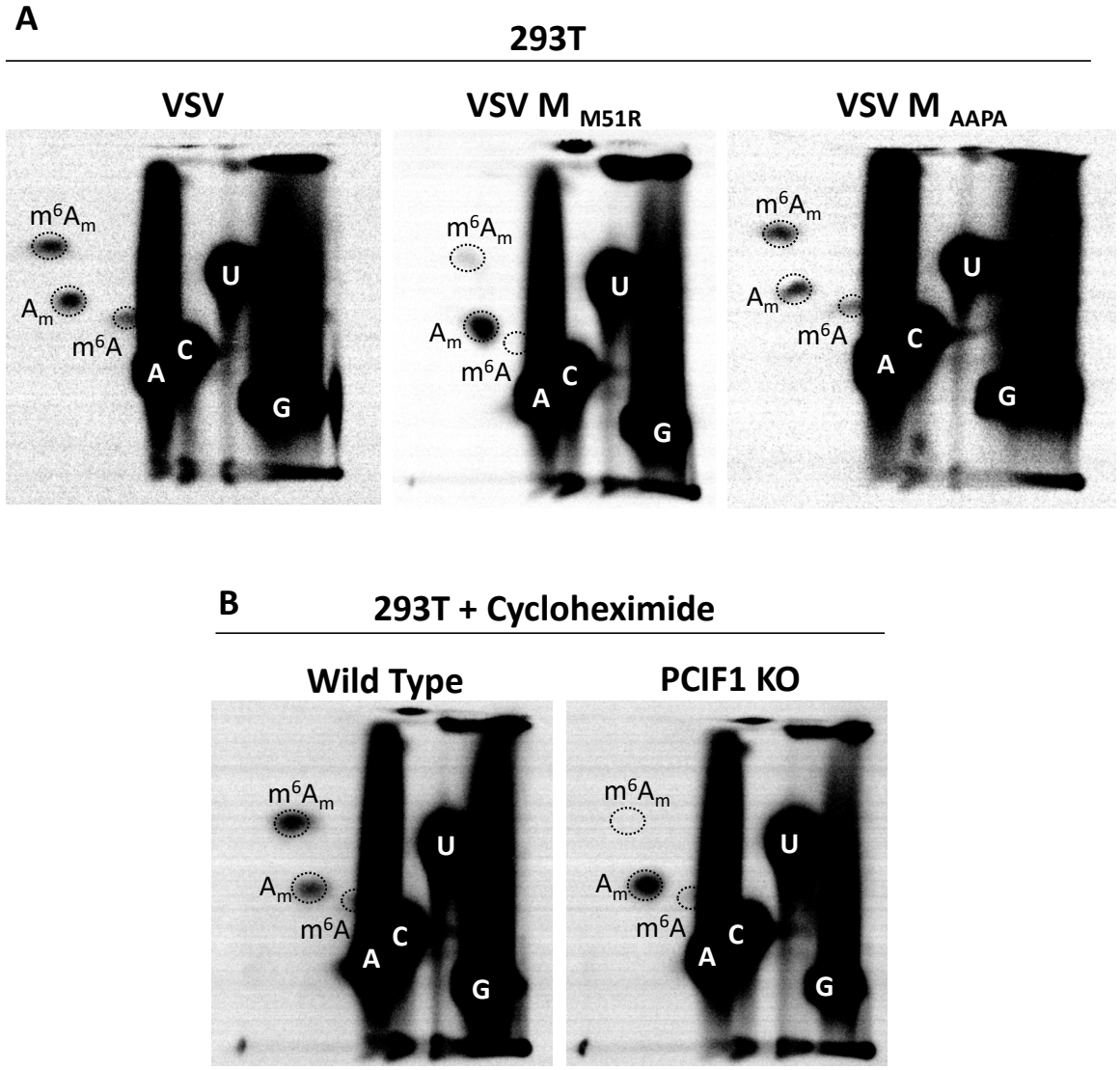


following mutations in M: M51R, known to be deficient in re-localization of nuclear proteins [102], and M with the late domain PPPY (amino acids 24-27) mutated to APPA (M_{APPA}) [35, 36]. VSV M_{M51R}, but not VSV M_{APPA}, was deficient in N6-methylation **(Fig 4.5A)**. Whether VSV M_{M51R} is also deficient in PCIF1 re-localization remains to be tested. PCIF1-dependent methylation was unaffected by the presence of cycloheximide during infection **(Fig 4.5B)**, indicating that new protein synthesis was not necessary for methylation by PCIF1. VSV RNA synthesis takes place in cytoplasmic inclusions, the formation of which depends on viral protein synthesis [16, 410]. As PCIF1-dependent methylation does not require active protein synthesis, it likely takes place independently of viral inclusion bodies.

Together, these data indicate that PCIF1 re-localizes from the nucleus to the cytoplasm during infection, and that PCIF1-dependent methylation requires VSV M. However, significant mechanistic questions remain unresolved. The purpose of such a large cytoplasmic population of PCIF1, why it does not appear by microscopy, or whether it localizes to some other cellular compartment not resolved by our approach is unknown. However, it is consistent with previous fractionation studies. A cap-specific mRNA N6-methyltransferase was originally isolated and purified from cytoplasmic fractions of HeLa cells [157]. In a similar procedure, though 2'O methyltransferase activity at the cap-proximal nucleotide was predominantly nuclear, activity at the second nucleotide was predominantly cytoplasmic [142]. By IF microscopy, CMTR1 was exclusively nuclear, while CMTR2 was both nuclear and cytoplasmic [141]. Given the critical role for methylation in self/non-self recognition, and that many RNA viruses replicate in the cytoplasm, the purpose of cytoplasmic methyltransferases is unclear.

Figure 4.5: Viral determinants of N6-methylation. 293T cells were infected with the indicated viruses at a MOI of 5, cellular transcription halted by adding 10 μg ml-1 actinomycin D at 2.5 hpi, and viral RNA labeled by metabolic incorporation of 100 μCi ml-1 [32P] phosphoric acid from 3-7 hpi. Total cellular RNA was extracted, poly(A) selected, and digested with Cap-Clip pyrophosphatase and nuclease p1. The resulting nucleotide monophosphates were resolved by 2D-TLC and detected by phosphorimaging (representative image shown). (A) Infection with VSV (n=3), VSV containing a point mutation in MM51R (n=3), or VSV with M24-PPPY-27 mutated to M24-APPA-27 (VSV MAppA) (n=2). (B) 293T wild type or PCIF1 KO cells were infected with VSV at a MOI of 100 in the presence of 100 μg ml-1 cycloheximide. Radiolabeling and 2D-TLC were performed as in A (n=3).

Fig 4.5 (continued):



One possible function for this location is cytoplasmic recapping, a process in which mRNAs which have been decapped as the first step in XRN1-dependent degradation can be recapped in the cytoplasm by cellular capping enzyme (RNGTT, normally found in the nucleus) and reenter the pool of translating mRNAs [411]. Though this process is only partially understood, cytoplasmic methyltransferases could play a role in this process.

Cap-Binding Determinants in VSV mRNAs

Although we often use VSV as a tool to study cellular processes, there are significant differences between VSV and cellular mRNAs. Notably, the average length of human 5' UTRs is 210 nucleotides, while the longest 5' UTR of VSV is M at 42 nt, and N, P, and L each have a 5' UTR of 13 nt or fewer [50, 87]. Previous work has indicated that a sequence-independent element in the 5' UTR of VSV drives preferential translation of viral RNA [412]. A viral mutant deficient in N7-methylation of the 5' cap is efficiently translated and sensitive to eIF4E depletion [32], despite eIF4E having a ~750-fold affinity decrease for caps lacking N7-methylation [413]. Similarly, we found that PCIF1 efficiently methylates VSV mRNAs lacking N7-methylation, in contrast to other biochemical and structural studies [67, 131, 159]. These multiple lines of evidence suggest that VSV mRNA binds to certain cap-binding proteins even in the absence of N7-methylation. The RNA structure of the VSV 5' UTR has not yet been determined. It is possible that even such a short sequence could contain some secondary structure element capable of recruiting key cellular factors. It is unlikely that N6-methylation contributes to that specific process, as that would not explain why PCIF1 binds its non-N6-methylated substrate, nor why the PCIF1-dependent phenotype only occurs under

conditions of interferon pre-stimulation (which was not examined in previous studies). Detailed biochemical studies of RNA binding proteins for the VSV 5' UTR, as well as RNA structural studies, will be needed to fully elucidate these mechanisms.

Infection of *PCIF1*-knockout Mice

In order to determine whether PCIF1 played a role during VSV infection of mice, we generated a *PCIF1*-knockout mouse and verified it to be null for both PCIF1 and m^6A_m **(Fig 4.6)**. We then infected mice, both male and female, with VSV at 5 different doses (10^6, 10^5, 10^4, 10^3, and 10^2 PFU, as measured by plaque assay on vero cells). We monitored these mice daily for survival and weight loss for 14 days. Mice were sacrificed at the end of the monitoring period, if they lost more than 80% of their starting body weight, or if they showed severe symptoms of viral infection (e.g. paralysis). Both male and female *PCIF1*-knockout mice had a lower 50% lethal dose (LD_{50}) compared to infection with wild type mice **(Table 4.1, Fig 4.7A)**. Males had a ~3-fold lower LD_{50}, while females had a ~5-fold lower LD_{50} **(Fig 4.7A)**. However, this difference in survival only reached statistical significance in the 10^5 PFU dose in male mice (**Fig 4.7B**, Mantel-Cox test $p<0.05$). There were differences in the body weight accumulation of mock infected mice, though in contrasting directions in males and females (**Fig 4.7D**). *PCIF1*-knockout males lost more weight during infection with the 10^4 PFU dose (**Fig 4.7D**). There was no statistical difference in the titers of infectious virus isolated from the brains of infected mice, as titered on vero cells **(Fig. 4.7C)**.

Intranasal infection of mice with VSV causes IFN-β expression in the olfactory bulb, leading to a protective antiviral state in the brain which arrests virus spread and

Figure 4.6: PCIF1 knockout mice have no detectable m^6A_m. (A) Western blots of liver, kidney and brain samples from WT and PCIF1 KO mice demonstrate the absence of PCIF1 protein. (B) PCIF1 KO mice reveal complete absence of m6Am as assessed by (B) UHPLCms/ms and (C) 2D-TLC. Left panels display 2D TLC images and right graph represents quantification of nucleotide intensity.

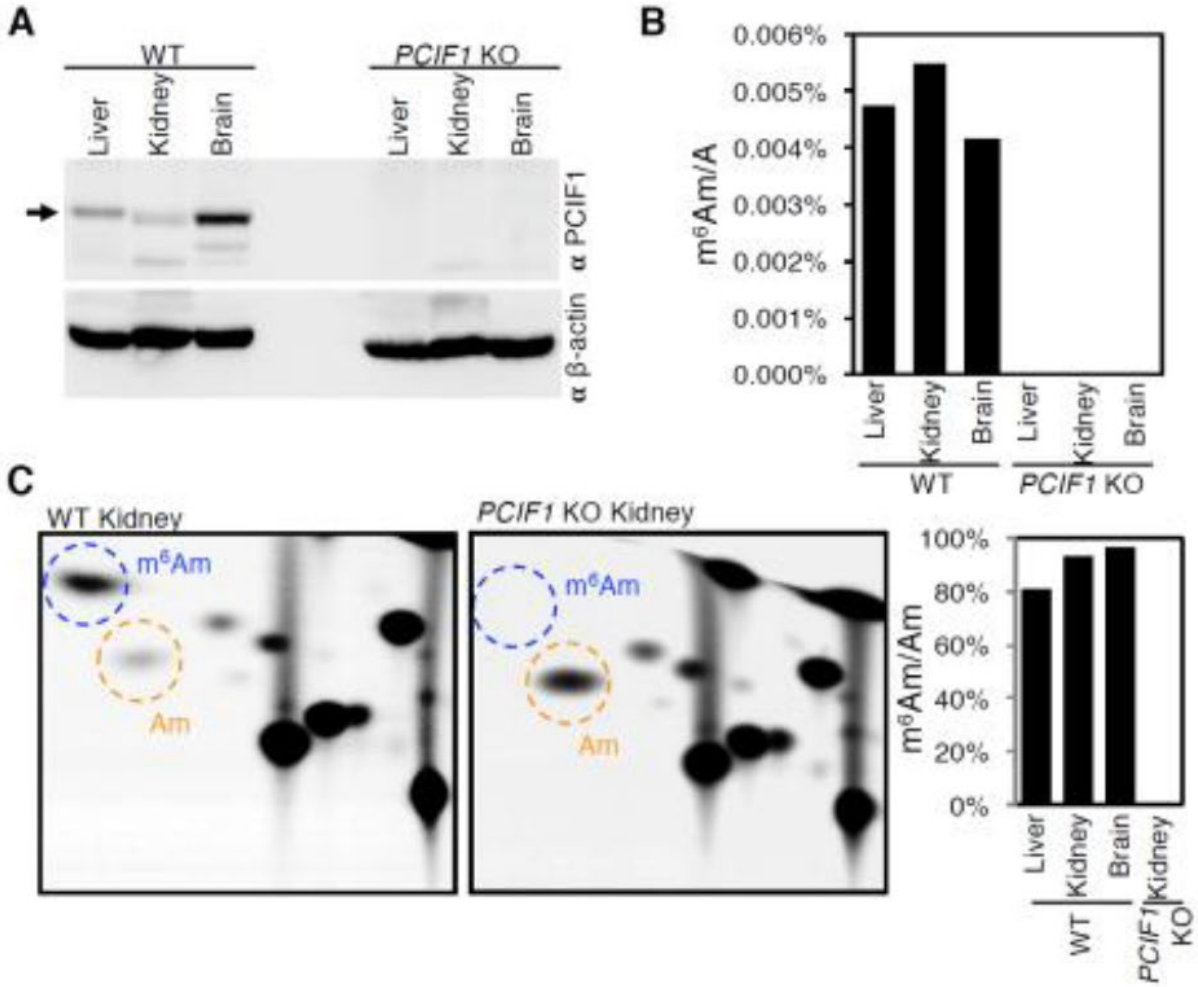


Table 4.1: Determining the 50% lethal dose in male and female wild type and PCIF1 knockout mice challenged with VSV

Deaths / # Mice

	Male	Male	Female	Female	Total	Total
Challenge Dose (PFU)	**Wild Type**	**PCIF1 Knockout**	**Wild Type**	**PCIF1 Knockout**	**Wild Type**	**PCIF1 Knockout**
10^6	6/6	6/6	0/0	0/0	6/6	6/6
10^5	5/7	6/6	6/7	6/6	11/14	12/12
10^4	5/6	6/6	4/7	5/6	9/15	11/12
10^3	3/6	3/6	4/7	4/6	7/13	7/12
10^2	1/6	1/6	0/6	0/6	1/12	1/12

Figure 4.7: Infection of mice with VSV. Wild type (parental) or PCIF1-knockout ("KO") 3-4 week old C57/BL6 mice were infected intranasally with the indicated dose of VSV for 14 days. Mice were sacrificed prior to day 14 if they shows greater than 80% body weight loss or severe symptoms of viral infection. The brains of sacrificed mice were dissected and VSV titers measured by plaque assay. (A) 50% lethal dose of VSV for each sex and mouse background and was calculated according to the method of Reed and Muench. (B) Survival curves for the indicated doses of VSV and sex of the mouse. Significant differences between survival curves were calculated by Mantel-Cox test (* - $p < 0.05$, NS – $p > 0.06$). (C) VSV titers present in brains from symptomatic mice. No statistically significant difference between any condition (2-way ANOVA, $p > 0.2$). (D) Weight of mice at the indicated day, dose, and sex. Significant differences in weight curves were calculated by 2-way ANOVA (**** - $p < 0.0001$, * - $p < 0.05$, NS - $p > 0.06$

Fig 4.7 (continued):

A

LD50 (pfu)

	Wild Type	PCIF1 KO
Male	1930	710
Female	3930	750

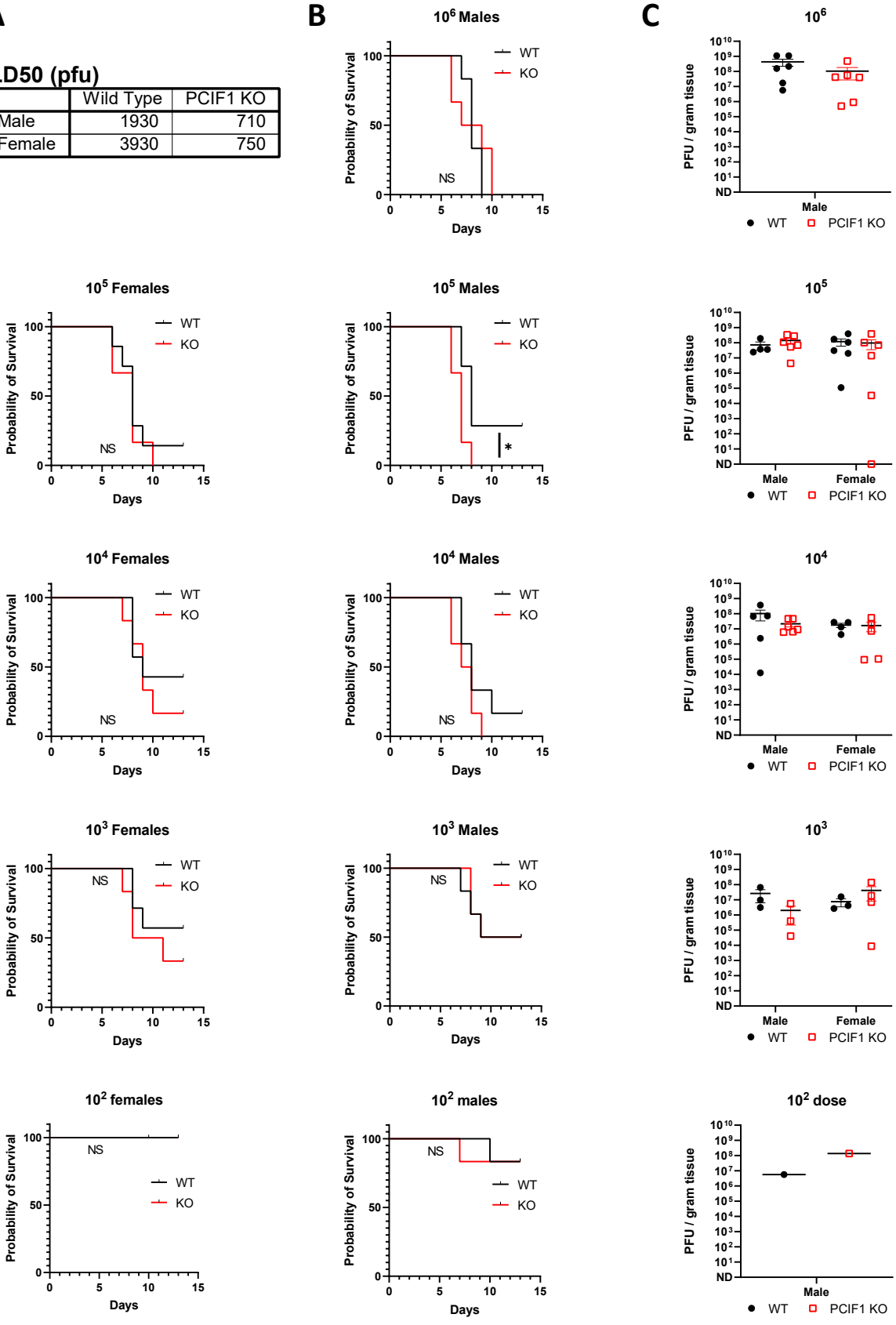
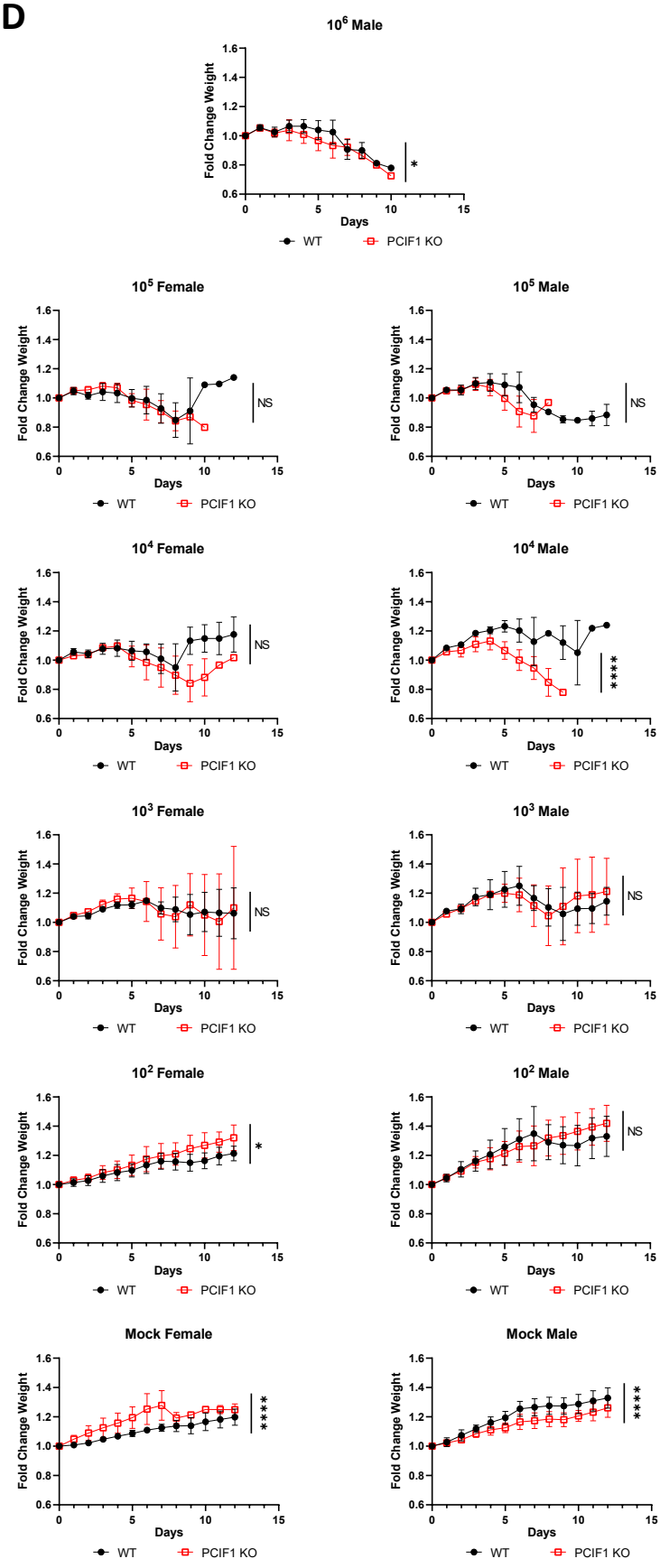


Fig 4.7 (continued): **D**



promotes survival [414, 415]. Based on the interferon-dependent phenotype we observed in human cell culture, we hypothesized *PCIF1*-knockout cells would have a higher rate of survival and LD_{50} value following VSV infection. Instead, we observed a lower LD_{50} value in *PCIF1*-knockout mice **(Fig 4.7A)**. There also appeared to be some sex-dependent effect. *PCIF1*-knockout male mice without infection gained less weight than wild type mice, while the opposite was true of their female counterparts **(Fig 4.7D)**. Similarly, male *PCIF1*-knockout mice had greater weight loss and lesser survival at the 10^4 and 10^5 PFU doses respectively, while the same in female mice did not rise to the level of statistical significance (**Fig 4.7D**, $p>0.06$, 2-way ANOVA for weight loss; $p>0.37$, Mantel-Cox test for survival). At most viral doses, and in all infected brain samples, *PCIF1*-knockout had no effect **(Fig 4.7B-D)**. Without a clear mechanistic understanding of how PCIF1 functions in VSV infection in cell culture, an explanation for the discrepancy between the cell culture and animal studies is highly speculative. One possibility is that PCIF1 is required for proper self/non-self recognition and functioning of the innate immune system, and that its deletion could cause dysregulation during viral infection. Alternatively, the kinetics of interferon pre-stimulation performed in cell culture may not be relevant to an organism. Viral infection induces multiple interferons, which upregulate different sets of ISGs with distinct kinetics [416]. The influence of PCIF1 on the antiviral state induced by 5 hours IFN-β pretreatment in cell culture may not be relevant to a multi-day infection of a mouse. Finally, there could be critical variations between human and mouse factors depending on cap-proximal N6-methylation. There are significant differences between the ability of human and mouse IFIT proteins to bind and sense 2'O methylation at the cap-proximal nucleotide [257, 264]. Though we

showed that human IFIT1/3 complexes do not discriminate A_m from m^6A_m, we did not test murine IFIT proteins, nor have we tested the PCIF1- and IFN-dependent phenotype in mouse cells. Much remains to be understood about the physiological role of PCIF1 during viral infection.

Why Is PCIF1 Nonessential?

Given the ubiquity of m^6A_m caps in mammalian mRNA, the generation of a *PCIF1* knockout mouse with only small differences in weight gain is somewhat surprising. In contrast, *CMTR1* knockout is embryonic lethal in mice [417]. Similar to our findings, another group reported the generation of a *PCIF1*-knockout mouse with differences in weight gain and stability of a subset of RNA transcripts, but no other negative effects [324]. This implies that PCIF1 is not an essential gene and instead plays a more specialized role in the cell. Consistent with this, there is still a substantial minority of A_m transcripts despite m^6A_m being a majority of cap-proximal adenosines [67, 130, 131, 418]. If N6-methylation played an essential role in the cell, we might expect a higher proportion of m^6A_m in mRNA. Our finding that PCIF1 is not involved in sensing of viral nucleic acids is consistent with this, as the minority of A_m transcripts in the cell would constantly be sensed as non-self. We would speculate that there is redundancy in methylation sites as self markers. This is supported by our IFIT-binding data. While N6-methylation or 2'O methylation alone produced substantial reductions in binding of IFIT1/3 to an mRNA oligo, the presence of both methylations had little effect over 2'O methylation alone. There may also be redundancy in antiviral mechanisms upregulated by interferon. Many ISGs are known to target similar steps in viral replication [246], and

may be responsive to different forms of regulation than self/non-self recognition of viral mRNA.

The role cap-proximal nucleotide identity plays in the cell is still unclear. G_m and m^6A_m make up the vast majority of methylated cap-proximal nucleotides in the cell [68]. However, no difference has been shown in RNA stability or translation of A_m, C_m, G_m, and U_m cap-proximal nucleotides [131]. Perplexingly, though m^6A_m mRNAs had longer half-lives and higher translational efficiency than their other cap-proximal nucleotide counterparts, this was not dependent on N6-methylation or PCIF1 [131]. Cap-proximal nucleotide identity impacts eIF4E-affinity *in vitro* (A or G preferred over C), and may enhance translation in cells under conditions of glucose starvation [419]. Of the cap-proximal nucleotides, only A is N6-methylated. This may be because O6-methylation of guanosine has a significant fitness cost [325, 326]. The full landscape behind 5' UTR sequence, length, and secondary structure is not yet known, and the specific contributions of cap-proximal m^6A_m over A_m, G_m, C_m, and U_m remain a topic for future study.

PCIF1 may also have other roles in the cell which are not yet understood. The *Drosophila* genome encodes for a homolog of *PCIF1* which is catalytically inactive as an N6-methyltransferase but associates with RNA Pol II [324]. What role this protein plays in *Drosophila*, which does not contain m^6A_m, is unknown. The evolution of the catalytic activity of PCIF1 parallels the evolution of the innate immune and interferon response. *Drosophila* and mosquitoes lack m^6A_m while it is present in zebrafish and higher mammals [114, 130], which parallels the emergence of interferon systems in fish [420].

The evolution of the IFN-dependent effect of PCIF1, and whether this has any impact on the host range of a particular virus, remains to be studied.

Potential Role for m^6A_m in Other Viruses

m^6A_m could potentially be incorporated into any viral RNA with a cap-proximal adenosine. We demonstrate that the mRNA of both VSV and rabies are N6-methylated by PCIF1, as well as measles to a lesser extent. The mechanism behind the difference in efficiency of methylation is unknown and could be due to differences in 5' UTR sequence or viral recruitment of PCIF1. Many other negative-sense RNA viruses, including sendai, mumps, and Newcastle disease viruses synthesize a cap1 mRNA that initiates on A, and could plausibly be methylated by PCIF1 [321]. As a cap-snatching virus, it is likely that influenza also contains m^6A_m, though this has not yet been reported [421]. m^6A_m has not yet been studied or reported in positive-sense RNA viruses. DNA viruses transcribed by host machinery do contain m^6A_m, such as adenoviruses, simian virus 40, polyomavirus, and herpes simplex virus 1 [150, 152-154]. Vaccinia virus, which encodes its own replication machinery expressed in the cytoplasm, is also N6-methylated [327], likely by PCIF1.

We would speculate that the functional impact of m^6A_m in the mRNA of these viruses would dependent on their viral life cycles. Since the effect we observe is interferon-dependent, viruses which robustly target these pathways may be minimally affected. For example, vaccinia virus is insensitive to a 20 hour interferon pretreatment of HeLa cells [422]. As discussed earlier, we have not definitively shown whether the presence of m^6A_m in viral or cellular mRNA mediates this effect. If it is dependent on

the presence of m^6A_m in viral mRNA, it will likely differentially affect viruses based on their ability to efficiently incorporate m^6A_m. PCIF1 may be antiviral against viruses which cannot hijack its function and incorporate m^6A_m for themselves, for example, negative-sense RNA viruses which initiate transcription on guanosine. PCIF1 was recently reported to restrict HIV infection, and is targeted for degradation by the HIV viral protein R (vpr) [423]. This would parallel the functionality of 2'O methylation in the innate immune response. Cap-proximal 2'O methylation protects gene expression against restriction by IFIT1 [249, 266]. When installed on viral RNA by a viral methyltransferase, 2'O methylation enhances viral gene expression [249, 250]. When installed on cellular ISGs by the cellular methyltransferase CMTR1, 2'O methylation enhances ISG expression, and as a result attenuates viral gene expression [266]. Because viral and cellular 2'O methylation of many RNA viruses are carried out by different enzymes, the specific function of cellular CMTR1 can be categorized as antiviral. However, because PCIF1 methylates both viral and cellular mRNAs, its positive or negative effect on a virus likely changes depending on how the virus can hijack PCIF1-dependent mechanisms to its own advantage. The role of PCIF1 with viruses which acquire caps synthesized by host machinery (e.g. cap-snatching RNA viruses or DNA viruses transcribed by RNA Pol II) may similarly depend on what proportion of those cellular caps initiated on A, and how these viruses bias cellular translation in their favor. Finally, while our data show the effect of PCIF1 specifically on gene expression from mRNA, the role of PCIF1 in genomic RNA remains unknown. Positive-sense RNA viruses with a capped genome could possibly be further N6-methylated by PCIF1. Though we have shown that m^6A_m does not impact viral RNA

sensing by the cell, there may be advantages in incorporating m^6A_m so early in the viral life cycle, and could possibly impact the host range of infection.

A Model for PCIF1 and m^6A_m in Viral Infection

We establish here a novel role for m^6A_m and PCIF1 in the type I interferon response against RNA viruses. Based on our above findings and speculation, we propose the following model. Early during VSV infection of the cell, the VSV M causes re-localization of nuclear factors, including PCIF1, to the cytoplasm. There, PCIF1 N6-methylates VSV mRNA with cap-proximal m^6A_m. Under conditions of interferon pre-stimulation, m^6A_m enables recruitment of specific RNA-binding proteins to viral mRNA, which in turn drives positive regulation of viral gene expression. In the absence of viral infection, this same mechanism may positively regulate expression of a subset of cellular genes under conditions of stress. Though consistent with our current data, much of this model remains speculative. Whether m^6A_m and PCIF1 regulate a subset of RNA viruses or a wide spectrum of viral families is also unclear. We look forward to future studies to validate and extend these ideas toward novel mechanisms of the interaction between virus and host.

REFERENCE

1. Pyle, J.D., S.P.J. Whelan, and L.M. Bloyet, *Structure and function of negative-strand RNA virus polymerase complexes.* Enzymes, 2021. **50**: p. 21-78.

2. Baltimore, D., A.S. Huang, and M. Stampfer, *Ribonucleic acid synthesis of vesicular stomatitis virus, II. An RNA polymerase in the virion.* Proc Natl Acad Sci U S A, 1970. **66**(2): p. 572-6.

3. Liang, B., et al., *Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy.* Cell, 2015. **162**(2): p. 314-327.

4. Knipe, D.M. and P. Howley, *Fields Virology*. 6 ed. 2013, Philadelphia: Wolters Kluwer, Lippincott Williams & Wilkins.

5. Ball, L.A. and C.N. White, *Order of transcription of genes of vesicular stomatitis virus.* Proc Natl Acad Sci U S A, 1976. **73**(2): p. 442-6.

6. Ball, L.A., *Transcriptional mapping of vesicular stomatitis virus in vivo.* J Virol, 1977. **21**(1): p. 411-4.

7. Abraham, G. and A.K. Banerjee, *Sequential transcription of the genes of vesicular stomatitis virus.* Proc Natl Acad Sci U S A, 1976. **73**(5): p. 1504-8.

8. Matlin, K.S., et al., *Pathway of vesicular stomatitis virus entry leading to infection.* J Mol Biol, 1982. **156**(3): p. 609-31.

9. Superti, F., et al., *Entry pathway of vesicular stomatitis virus into different host cells.* J Gen Virol, 1987. **68 (Pt 2)**: p. 387-99.

10. Cureton, D.K., et al., *Vesicular stomatitis virus enters cells through vesicles incompletely coated with clathrin that depend upon actin for internalization.* PLoS Pathog, 2009. **5**(4): p. e1000394.

11. Johannsdottir, H.K., et al., *Host cell factors and functions involved in vesicular stomatitis virus entry.* J Virol, 2009. **83**(1): p. 440-53.

12. Kim, I.S., et al., *Mechanism of membrane fusion induced by vesicular stomatitis virus G protein.* Proc Natl Acad Sci U S A, 2017. **114**(1): p. E28-E36.

13. Emerson, S.U. and Y. Yu, *Both NS and L proteins are required for in vitro RNA synthesis by vesicular stomatitis virus.* J Virol, 1975. **15**(6): p. 1348-56.

14. Emerson, S.U. and R.R. Wagner, *Dissociation and reconstitution of the transcriptase and template activities of vesicular stomatitis B and T virions.* J Virol, 1972. **10**(2): p. 297-309.

15. Emerson, S.U. and R.R. Wagner, *L protein requirement for in vitro RNA synthesis by vesicular stomatitis virus.* J Virol, 1973. **12**(6): p. 1325-35.

16. Heinrich, B.S., et al., *Protein expression redirects vesicular stomatitis virus RNA synthesis to cytoplasmic inclusions.* PLoS Pathog, 2010. **6**(6): p. e1000958.

17. Das, S.C., et al., *Visualization of intracellular transport of vesicular stomatitis virus nucleocapsids in living cells.* J Virol, 2006. **80**(13): p. 6368-77.

18. Villarreal, L.P., M. Breindl, and J.J. Holland, *Determination of molar ratios of vesicular stomatitis virus induced RNA species in BHK21 cells.* Biochemistry, 1976. **15**(8): p. 1663-7.

19. Iverson, L.E. and J.K. Rose, *Localized attenuation and discontinuous synthesis during vesicular stomatitis virus transcription.* Cell, 1981. **23**(2): p. 477-84.

20. Herman, R.C., et al., *Polycistronic vesicular stomatitis virus RNA transcripts.* Proc Natl Acad Sci U S A, 1980. **77**(8): p. 4662-5.

21. Follett, E.A., et al., *Virus replication in enucleate cells: vesicular stomatitis virus and influenza virus.* J Virol, 1974. **13**(2): p. 394-9.

22. Grubman, M.J., J.A. Weinstein, and D.A. Shafritz, *Studies on the mechanism for entry of vesicular stomatitis virus glycoprotein G mRNA into membrane-bound polyribosome complexes.* J Cell Biol, 1977. **74**(1): p. 43-57.

23. Grubman, M.J., E. Ehrenfeld, and D.F. Summers, *In vitro synthesis of proteins by membrane-bound polyribosomes from vesicular stomatitis virus-infected HeLa cells.* J Virol, 1974. **14**(3): p. 560-71.

24. Knipe, D.M., D. Baltimore, and H.F. Lodish, *Separate pathways of maturation of the major structural proteins of vesicular stomatitis virus.* J Virol, 1977. **21**(3): p. 1128-39.

25. Knipe, D., J.K. Rose, and H.F. Lodish, *Translation of individual species of vesicular stomatitis viral mRNA.* J Virol, 1975. **15**(4): p. 1004-11.

26. Soria, M., S.P. Little, and A.S. Huang, *Characterization of vesicular stomatitis virus nucleocapsids. I. Complementary 40 S RNA molecules in nucleocapsids.* Virology, 1974. **61**(1): p. 270-80.

27. Wertz, G.W., *Isolation of possible replicative intermediate structures from vesicular stomatitis virus-infected cells.* Virology, 1978. **85**(1): p. 271-85.

28. Testa, D., P.K. Chanda, and A.K. Banerjee, *In vitro synthesis of the full-length complement of the negative-strand genome RNA of vesicular stomatitis virus.* Proc Natl Acad Sci U S A, 1980. **77**(1): p. 294-8.

29. Masters, P.S. and A.K. Banerjee, *Complex formation with vesicular stomatitis virus phosphoprotein NS prevents binding of nucleocapsid protein N to nonspecific RNA.* J Virol, 1988. **62**(8): p. 2658-64.

30. Patton, J.T., N.L. Davis, and G.W. Wertz, *N protein alone satisfies the requirement for protein synthesis during RNA replication of vesicular stomatitis virus.* J Virol, 1984. **49**(2): p. 303-9.

31. Peluso, R.W. and S.A. Moyer, *Initiation and replication of vesicular stomatitis virus genome RNA in a cell-free system.* Proc Natl Acad Sci U S A, 1983. **80**(11): p. 3198-202.

32. Neidermyer, W.J., Jr. and S.P.J. Whelan, *Global analysis of polysome-associated mRNA in vesicular stomatitis virus infected cells.* PLoS Pathog, 2019. **15**(6): p. e1007875.

33. Newcomb, W.W. and J.C. Brown, *Role of the vesicular stomatitis virus matrix protein in maintaining the viral nucleocapsid in the condensed form found in native virions.* J Virol, 1981. **39**(1): p. 295-9.

34. Obiang, L., et al., *Phenotypes of vesicular stomatitis virus mutants with mutations in the PSAP motif of the matrix protein.* J Gen Virol, 2012. **93**(Pt 4): p. 857-865.

35. Jayakar, H.R., K.G. Murti, and M.A. Whitt, *Mutations in the PPPY motif of vesicular stomatitis virus matrix protein reduce virus budding by inhibiting a late step in virion release.* J Virol, 2000. **74**(21): p. 9818-27.

36. Soh, T.K. and S.P. Whelan, *Tracking the Fate of Genetically Distinct Vesicular Stomatitis Virus Matrix Proteins Highlights the Role for Late Domains in Assembly.* J Virol, 2015. **89**(23): p. 11750-60.

37. Ogino, T. and T.J. Green, *RNA Synthesis and Capping by Non-segmented Negative Strand RNA Viral Polymerases: Lessons From a Prototypic Virus.* Front Microbiol, 2019. **10**: p. 1490.

38. Jenni, S., et al., *Structure of the Vesicular Stomatitis Virus L Protein in Complex with Its Phosphoprotein Cofactor.* Cell Rep, 2020. **30**(1): p. 53-60 e5.

39. Ogino, T. and A.K. Banerjee, *Unconventional mechanism of mRNA capping by the RNA-dependent RNA polymerase of vesicular stomatitis virus.* Mol Cell, 2007. **25**(1): p. 85-97.

40. Li, J., et al., *A conserved motif in region v of the large polymerase proteins of nonsegmented negative-sense RNA viruses that is essential for mRNA capping.* J Virol, 2008. **82**(2): p. 775-84.

41. Li, J., J.T. Wang, and S.P. Whelan, *A unique strategy for mRNA cap methylation used by vesicular stomatitis virus.* Proc Natl Acad Sci U S A, 2006. **103**(22): p. 8493-8.

42. Hornung, V., et al., *5'-Triphosphate RNA is the ligand for RIG-I.* Science, 2006. **314**(5801): p. 994-7.

43. Leung, D.W. and G.K. Amarasinghe, *When your cap matters: structural insights into self vs non-self recognition of 5' RNA by immunomodulatory host proteins.* Curr Opin Struct Biol, 2016. **36**: p. 133-41.

44. Ma, Y., et al., *mRNA cap methylation influences pathogenesis of vesicular stomatitis virus in vivo.* J Virol, 2014. **88**(5): p. 2913-26.

45. Colonno, R.J. and A.K. Banerjee, *Complete nucleotide sequence of the leader RNA synthesized in vitro by vesicular stomatitis virus.* Cell, 1978. **15**(1): p. 93-101.

46. Colonno, R.J. and A.K. Banerjee, *In vitro RNA transcription by the New Jersey serotype of vesicular stomatitis virus. II. Characterization of the leader RNA.* J Virol, 1978. **26**(1): p. 188-94.

47. Colonno, R.J. and A.K. Banerjee, *Mapping and initiation studies on the leader RNA of vesicular stomatitis virus.* Virology, 1977. **77**(1): p. 260-8.

48. Colonno, R.J. and A.K. Banerjee, *A unique RNA species involved in initiation of vesicular stomatitis virus RNA transcription in vitro.* Cell, 1976. **8**(2): p. 197-204.

49. Emerson, S.U., *Reconstitution studies detect a single polymerase entry site on the vesicular stomatitis virus genome.* Cell, 1982. **31**(3 Pt 2): p. 635-42.

50. Rose, J.K., *Complete intergenic and flanking gene sequences from the genome of vesicular stomatitis virus.* Cell, 1980. **19**(2): p. 415-21.

51. Barr, J.N., S.P. Whelan, and G.W. Wertz, *cis-Acting signals involved in termination of vesicular stomatitis virus mRNA synthesis include the conserved AUAC and the U7 signal for polyadenylation.* J Virol, 1997. **71**(11): p. 8718-25.

52. Stillman, E.A. and M.A. Whitt, *Mutational analyses of the intergenic dinucleotide and the transcriptional start sequence of vesicular stomatitis virus (VSV) define sequences required for efficient termination and initiation of VSV transcripts.* J Virol, 1997. **71**(3): p. 2127-37.

53. Wang, J.T., L.E. McElvain, and S.P. Whelan, *Vesicular stomatitis virus mRNA capping machinery requires specific cis-acting signals in the RNA.* J Virol, 2007. **81**(20): p. 11499-506.

54. Ehrenfeld, E. and D.F. Summers, *Adenylate-rich sequences in vesicular stomatitis virus messenger ribonucleic acid.* J Virol, 1972. **10**(4): p. 683-8.

55. Ehrenfeld, E., *Polyadenylation of vesicular stomatitis virus mRNA.* J Virol, 1974. **13**(5): p. 1055-60.

56. Banerjee, A.K. and D.P. Rhodes, *In vitro synthesis of RNA that contains polyadenylate by virion-associated RNA polymerase of vesicular stomatitis virus.* Proc Natl Acad Sci U S A, 1973. **70**(12): p. 3566-70.

57. Villarreal, L.P. and J.J. Holland, *Synthesis of poly(A) in vitro by purified virions of vesicular stomatitis virus.* Nat New Biol, 1973. **246**(149): p. 17-9.

58. Schubert, M., et al., *Site on the vesicular stomatitis virus genome specifying polyadenylation and the end of the L gene mRNA.* J Virol, 1980. **34**(2): p. 550-9.

59. Hunt, D.M., E.F. Smith, and D.W. Buckley, *Aberrant polyadenylation by a vesicular stomatitis virus mutant is due to an altered L protein.* J Virol, 1984. **52**(2): p. 515-21.

60. Moyer, S.A. and A.K. Banerjee, *In vivo methylation of vesicular stomatitis virus and its host-cell messenger RNA species.* Virology, 1976. **70**(2): p. 339-51.

61. Moyer, S.A., et al., *Methylated and blocked 5' termini in vesicular stomatitis virus in vivo mRNAs.* Cell, 1975. **5**(1): p. 59-67.

62. Abraham, G., D.P. Rhodes, and A.K. Banerjee, *The 5' terminal structure of the methylated mRNA synthesized in vitro by vesicular stomatitis virus.* Cell, 1975. **5**(1): p. 51-8.

63. Tekes, G., A.A. Rahmeh, and S.P. Whelan, *A freeze frame view of vesicular stomatitis virus transcription defines a minimal length of RNA for 5' processing.* PLoS Pathog, 2011. **7**(6): p. e1002073.

64. Ogino, M. and T. Ogino, *5'-Phospho-RNA Acceptor Specificity of GDP Polyribonucleotidyltransferase of Vesicular Stomatitis Virus in mRNA Capping.* J Virol, 2017. **91**(6).

65. Rahmeh, A.A., et al., *Ribose 2'-O methylation of the vesicular stomatitis virus mRNA cap precedes and facilitates subsequent guanine-N-7 methylation by the large polymerase protein.* J Virol, 2009. **83**(21): p. 11043-50.

66. Li, J., E.C. Fontaine-Rodriguez, and S.P. Whelan, *Amino acid residues within conserved domain VI of the vesicular stomatitis virus large polymerase protein essential for mRNA cap methyltransferase activity.* J Virol, 2005. **79**(21): p. 13373-84.

67. Akichika, S., et al., *Cap-specific terminal N (6)-methylation of RNA by an RNA polymerase II-associated methyltransferase.* Science, 2019. **363**(6423).

68. Wei, C.M., A. Gershowitz, and B. Moss, *5'-Terminal and internal methylated nucleotide sequences in HeLa cell mRNA.* Biochemistry, 1976. **15**(2): p. 397-401.

69. Wei, C., A. Gershowitz, and B. Moss, *N6, O2'-dimethyladenosine a novel methylated ribonucleoside next to the 5' terminal of animal cell and virus mRNAs.* Nature, 1975. **257**(5523): p. 251-3.

70. Lu, M., et al., *N(6)-methyladenosine modification enables viral RNA to escape recognition by RNA sensor RIG-I.* Nat Microbiol, 2020. **5**(4): p. 584-598.

71. Lu, M., et al., *Nonsegmented Negative-Sense RNA Viruses Utilize N (6)-Methyladenosine (m(6)A) as a Common Strategy To Evade Host Innate Immunity.* J Virol, 2021. **95**(9).

72. Lodish, H.F. and J.K. Rose, *Relative importance of 7-methylguanosine in ribosome binding and translation of vesicular stomatitis virus mRNA in wheat germ and reticulocyte cell-free systems.* J Biol Chem, 1977. **252**(4): p. 1181-8.

73. Morrison, T.G., *Site of synthesis of membrane and nonmembrane proteins of vesicular stomatitis virus.* J Biol Chem, 1975. **250**(17): p. 6955-62.

74. Lee, A.S., R. Burdeinick-Kerr, and S.P. Whelan, *A ribosome-specialized translation initiation pathway is required for cap-dependent translation of vesicular stomatitis virus mRNAs.* Proc Natl Acad Sci U S A, 2013. **110**(1): p. 324-9.

75. Jackson, R.J., C.U. Hellen, and T.V. Pestova, *The mechanism of eukaryotic translation initiation and principles of its regulation.* Nat Rev Mol Cell Biol, 2010. **11**(2): p. 113-27.

76. Hinnebusch, A.G. and J.R. Lorsch, *The mechanism of eukaryotic translation initiation: new insights and challenges.* Cold Spring Harb Perspect Biol, 2012. **4**(10).

77. Jan, E., I. Mohr, and D. Walsh, *A Cap-to-Tail Guide to mRNA Translation Strategies in Virus-Infected Cells.* Annu Rev Virol, 2016. **3**(1): p. 283-307.

78. Kumar, P., C.U. Hellen, and T.V. Pestova, *Toward the mechanism of eIF4F-mediated ribosomal attachment to mammalian capped mRNAs.* Genes Dev, 2016. **30**(13): p. 1573-88.

79. Thompson, M.K. and W.V. Gilbert, *mRNA length-sensing in eukaryotic translation: reconsidering the "closed loop" and its implications for translational control.* Curr Genet, 2017. **63**(4): p. 613-620.

80. Hinnebusch, A.G., *The scanning mechanism of eukaryotic translation initiation.* Annu Rev Biochem, 2014. **83**: p. 779-812.

81. Valasek, L.S., et al., *Embraced by eIF3: structural and functional insights into the roles of eIF3 across the translation cycle.* Nucleic Acids Res, 2017. **45**(19): p. 10948-10968.

82. Pestova, T.V., S.I. Borukhov, and C.U. Hellen, *Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons.* Nature, 1998. **394**(6696): p. 854-9.

83. Pestova, T.V., et al., *The joining of ribosomal subunits in eukaryotes requires eIF5B.* Nature, 2000. **403**(6767): p. 332-5.

84. Both, G.W., A.K. Banerjee, and A.J. Shatkin, *Methylation-dependent translation of viral messenger RNAs in vitro.* Proc Natl Acad Sci U S A, 1975. **72**(3): p. 1189-93.

85. Muthukrishnan, S., et al., *Influence of 5'-terminal m7G and 2'--O-methylated residues on messenger ribonucleic acid binding to ribosomes.* Biochemistry, 1976. **15**(26): p. 5761-8.

86. Rose, J.K. and H.F. Lodish, *Translation in vitro of vesicular stomatitis virus mRNA lacking 5'-terminal 7-methylguanosine.* Nature, 1976. **262**(5563): p. 32-7.

87. Pesole, G., et al., *Structural and functional features of eukaryotic mRNA untranslated regions.* Gene, 2001. **276**(1-2): p. 73-81.

88. Duncan, R., S.C. Milburn, and J.W. Hershey, *Regulated phosphorylation and low abundance of HeLa cell initiation factor eIF-4F suggest a role in translational control. Heat shock effects on eIF-4F.* J Biol Chem, 1987. **262**(1): p. 380-8.

89. Uttam, S., et al., *eIF4E-Dependent Translational Control: A Central Mechanism for Regulation of Pain Plasticity.* Front Genet, 2018. **9**: p. 470.

90. Mudd, J.A. and D.F. Summers, *Protein synthesis in vesicular stomatitis virus-infected HeLa cells.* Virology, 1970. **42**(2): p. 328-40.

91. Connor, J.H. and D.S. Lyles, *Vesicular stomatitis virus infection alters the eIF4F translation initiation complex and causes dephosphorylation of the eIF4E binding protein 4E-BP1.* J Virol, 2002. **76**(20): p. 10177-87.

92. Pelletier, J., et al., *Targeting the eIF4F translation initiation complex: a critical nexus for cancer development.* Cancer Res, 2015. **75**(2): p. 250-63.

93. Siddiqui, N. and N. Sonenberg, *Signalling to eIF4E in cancer.* Biochem Soc Trans, 2015. **43**(5): p. 763-72.

94. Thoreen, C.C., et al., *A unifying model for mTORC1-mediated regulation of mRNA translation.* Nature, 2012. **485**(7396): p. 109-13.

95. Beretta, L., et al., *Rapamycin blocks the phosphorylation of 4E-BP1 and inhibits cap-dependent initiation of translation.* EMBO J, 1996. **15**(3): p. 658-64.

96. Hara, K., et al., *Regulation of eIF-4E BP1 phosphorylation by mTOR.* J Biol Chem, 1997. **272**(42): p. 26457-63.

97. Hopkins, K.C., et al., *Virus-induced translational arrest through 4EBP1/2-dependent decay of 5'-TOP mRNAs restricts viral infection.* Proc Natl Acad Sci U S A, 2015. **112**(22): p. E2920-9.

98. Le Sage, V., et al., *Adapting the Stress Response: Viral Subversion of the mTOR Signaling Pathway.* Viruses, 2016. **8**(6).

99. Welnowska, E., et al., *Translation of mRNAs from vesicular stomatitis virus and vaccinia virus is differentially blocked in cells with depletion of eIF4GI and/or eIF4GII.* J Mol Biol, 2009. **394**(3): p. 506-21.

100. Connor, J.H., et al., *Replication and cytopathic effect of oncolytic vesicular stomatitis virus in hypoxic tumor cells in vitro and in vivo.* J Virol, 2004. **78**(17): p. 8960-70.

101. Yuan, H., B.K. Yoza, and D.S. Lyles, *Inhibition of host RNA polymerase II-dependent transcription by vesicular stomatitis virus results from inactivation of TFIID.* Virology, 1998. **251**(2): p. 383-92.

102. Pettit Kneller, E.L., J.H. Connor, and D.S. Lyles, *hnRNPs Relocalize to the cytoplasm following infection with vesicular stomatitis virus.* J Virol, 2009. **83**(2): p. 770-80.

103. Ferran, M.C. and J.M. Lucas-Lenard, *The vesicular stomatitis virus matrix protein inhibits transcription from the human beta interferon promoter.* J Virol, 1997. **71**(1): p. 371-7.

104. Ahmed, M., et al., *Ability of the matrix protein of vesicular stomatitis virus to suppress beta interferon gene expression is genetically correlated with the inhibition of host RNA and protein synthesis.* J Virol, 2003. **77**(8): p. 4646-57.

105. Quan, B., et al., *Vesiculoviral matrix (M) protein occupies nucleic acid binding site at nucleoporin pair (Rae1 * Nup98).* Proc Natl Acad Sci U S A, 2014. **111**(25): p. 9127-32.

106. Faria, P.A., et al., *VSV disrupts the Rae1/mrnp41 mRNA nuclear export pathway.* Mol Cell, 2005. **17**(1): p. 93-102.

107. Stojdl, D.F., et al., *VSV strains with defects in their ability to shutdown innate immunity are potent systemic anti-cancer agents.* Cancer Cell, 2003. **4**(4): p. 263-75.

108. Redondo, N., et al., *Impact of Vesicular Stomatitis Virus M Proteins on Different Cellular Functions.* PLoS One, 2015. **10**(6): p. e0131137.

109. Lee, A.S., et al., *eIF3d is an mRNA cap-binding protein that is required for specialized translation initiation.* Nature, 2016. **536**(7614): p. 96-9.

110. Lee, A.S., P.J. Kranzusch, and J.H. Cate, *eIF3 targets cell-proliferation messenger RNAs for translational activation or repression.* Nature, 2015. **522**(7554): p. 111-4.

111. Lamper, A.M., et al., *A phosphorylation-regulated eIF3d translation switch mediates cellular adaptation to metabolic stress.* Science, 2020. **370**(6518): p. 853-856.

112. Meyer, K.D., et al., *5' UTR m(6)A Promotes Cap-Independent Translation.* Cell, 2015. **163**(4): p. 999-1010.

113. Furuichi, Y. and A.J. Shatkin, *Viral and cellular mRNA capping: past and prospects.* Adv Virus Res, 2000. **55**: p. 135-84.

114. Furuichi, Y., *Discovery of m(7)G-cap in eukaryotic mRNAs.* Proc Jpn Acad Ser B Phys Biol Sci, 2015. **91**(8): p. 394-409.

115. Miura, K., et al., *The 5'-terminal nucleotide sequences of the double-stranded RNA of human reovirus.* Proc Natl Acad Sci U S A, 1974. **71**(10): p. 3979-83.

116. Furuichi, Y. and K.I. Miura, *Identity of the 3'-terminal sequences in ten genome segments of silkworm cytoplasmic polyhedrosis virus.* Virology, 1973. **55**(2): p. 418-25.

117. Furuichi, Y. and K. Miura, *A blocked structure at the 5' terminus of mRNA from cytoplasmic polyhedrosis virus.* Nature, 1975. **253**(5490): p. 374-5.

118. Furuichi, Y. and K. Miura, *The 3'-termini of the genome RNA segments of silkworm cytoplasmic polyhedrosis virus.* J Mol Biol, 1972. **64**(3): p. 619-32.

119. Furuichi, Y., et al., *Reovirus messenger RNA contains a methylated, blocked 5'-terminal structure: m-7G(5')ppp(5')G-MpCp.* Proc Natl Acad Sci U S A, 1975. **72**(1): p. 362-6.

120. Miura, K., K. Watanabe, and M. Sugiura, *5'-Terminal nucleotide sequences of the double-stranded RNA of silkworm cytoplasmic polyhedrosis virus.* J Mol Biol, 1974. **86**(1): p. 31-48.

121. Furuichi, Y., *"Methylation-coupled" transcription by virus-associated transcriptase of cytoplasmic polyhedrosis virus containing double-stranded RNA.* Nucleic Acids Res, 1974. **1**(6): p. 809-22.

122. Desrosiers, R., K. Friderici, and F. Rottman, *Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells.* Proc Natl Acad Sci U S A, 1974. **71**(10): p. 3971-5.

123. Rottman, F., A.J. Shatkin, and R.P. Perry, *Sequences containing methylated nucleotides at the 5' termini of messenger RNAs: possible implications for processing.* Cell, 1974. **3**(3): p. 197-9.

124. Wei, C.M. and B. Moss, *Methylated nucleotides block 5'-terminus of vaccinia virus messenger RNA.* Proc Natl Acad Sci U S A, 1975. **72**(1): p. 318-22.

125. Abraham, G., D.P. Rhodes, and A.K. Banerjee, *Novel initiation of RNA synthesis in vitro by vesicular stomatitis virus.* Nature, 1975. **255**(5503): p. 37-40.

126. Keith, J.M., S. Muthukrishnan, and B. Moss, *Effect of methylation of the N6 position of the penultimate adenosine of capped mRNA on ribosome binding.* J Biol Chem, 1978. **253**(14): p. 5039-41.

127. Topisirovic, I., et al., *Cap and cap-binding proteins in the control of gene expression.* Wiley Interdiscip Rev RNA, 2011. **2**(2): p. 277-98.

128. Shuman, S., *Capping enzyme in eukaryotic mRNA synthesis.* Prog Nucleic Acid Res Mol Biol, 1995. **50**: p. 101-29.

129. Shuman, S., *What messenger RNA capping tells us about eukaryotic evolution.* Nat Rev Mol Cell Biol, 2002. **3**(8): p. 619-25.

130. Sendinc, E., et al., *PCIF1 Catalyzes m6Am mRNA Methylation to Regulate Gene Expression.* Mol Cell, 2019. **75**(3): p. 620-630 e9.

131. Boulias, K., et al., *Identification of the m(6)Am Methyltransferase PCIF1 Reveals the Location and Functions of m(6)Am in the Transcriptome.* Mol Cell, 2019. **75**(3): p. 631-643 e8.

132. Shatkin, A.J. and J.L. Manley, *The ends of the affair: capping and polyadenylation.* Nat Struct Biol, 2000. **7**(10): p. 838-42.

133. Shatkin, A.J., *Capping of eucaryotic mRNAs.* Cell, 1976. **9**(4 PT 2): p. 645-53.

134. McCracken, S., et al., *5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II.* Genes Dev, 1997. **11**(24): p. 3306-18.

135. Mao, X., B. Schwer, and S. Shuman, *Yeast mRNA cap methyltransferase is a 50-kilodalton protein encoded by an essential gene.* Mol Cell Biol, 1995. **15**(8): p. 4167-74.

136. Ho, C.K., et al., *The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II.* J Biol Chem, 1998. **273**(16): p. 9577-85.

137. Tsukamoto, T., et al., *Cloning and characterization of three human cDNAs encoding mRNA (guanine-7-)-methyltransferase, an mRNA cap methylase.* Biochem Biophys Res Commun, 1998. **251**(1): p. 27-34.

138. Tsukamoto, T., et al., *Isolation and characterization of the yeast mRNA capping enzyme beta subunit gene encoding RNA 5'-triphosphatase, which is essential for cell viability.* Biochem Biophys Res Commun, 1997. **239**(1): p. 116-22.

139. Shibagaki, Y., et al., *mRNA capping enzyme. Isolation and characterization of the gene encoding mRNA guanylytransferase subunit from Saccharomyces cerevisiae.* J Biol Chem, 1992. **267**(14): p. 9521-8.

140. Belanger, F., et al., *Characterization of hMTr1, a human Cap1 2'-O-ribose methyltransferase.* J Biol Chem, 2010. **285**(43): p. 33037-33044.

141. Werner, M., et al., *2'-O-ribose methylation of cap2 in human: function and evolution in a horizontally mobile family.* Nucleic Acids Res, 2011. **39**(11): p. 4756-68.

142. Langberg, S.R. and B. Moss, *Post-transcriptional modifications of mRNA. Purification and characterization of cap I and cap II RNA (nucleoside-2'-)-methyltransferases from HeLa cells.* J Biol Chem, 1981. **256**(19): p. 10054-60.

143. Decroly, E., et al., *Conventional and unconventional mechanisms for capping viral mRNA.* Nat Rev Microbiol, 2011. **10**(1): p. 51-65.

144. Yuan, P., et al., *Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site.* Nature, 2009. **458**(7240): p. 909-13.

145. Guilligay, D., et al., *The structural basis for cap binding by influenza virus polymerase subunit PB2.* Nat Struct Mol Biol, 2008. **15**(5): p. 500-6.

146. Dias, A., et al., *The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit.* Nature, 2009. **458**(7240): p. 914-8.

147. Morin, B., et al., *The N-terminal domain of the arenavirus L protein is an RNA endonuclease essential in mRNA transcription.* PLoS Pathog, 2010. **6**(9): p. e1001038.

148. Reguera, J., F. Weber, and S. Cusack, *Bunyaviridae RNA polymerases (L-protein) have an N-terminal, influenza-like endonuclease domain, essential for viral cap-dependent transcription.* PLoS Pathog, 2010. **6**(9): p. e1001101.

149. Mir, M.A., et al., *Storage of cellular 5' mRNA caps in P bodies for viral cap-snatching.* Proc Natl Acad Sci U S A, 2008. **105**(49): p. 19294-9.

150. Moss, B. and F. Koczot, *Sequence of methylated nucleotides at the 5'-terminus of adenovirus-specific RNA.* J Virol, 1976. **17**(2): p. 385-92.

151. Sommer, S., et al., *The methylation of adenovirus-specific nuclear and cytoplasmic RNA.* Nucleic Acids Res, 1976. **3**(3): p. 749-65.

152. Haegeman, G. and W. Fiers, *Characterization of the 5'-terminal capped structures of late simian virus 40-specific mRNA.* J Virol, 1978. **25**(3): p. 824-30.

153. Moss, B., et al., *5'-Terminal and internal methylated nucleosides in herpes simplex virus type 1 mRNA.* J Virol, 1977. **23**(2): p. 234-9.

154. Flavell, A.J., et al., *Multiple 5' terminal cap structures in late polyoma virus RNA.* Cell, 1979. **16**(2): p. 357-71.

155. Goodfellow, I., et al., *Calicivirus translation initiation requires an interaction between VPg and eIF 4 E.* EMBO Rep, 2005. **6**(10): p. 968-72.

156. Ensinger, M.J. and B. Moss, *Modification of the 5' terminus of mRNA by an RNA (guanine-7-)-methyltransferase from HeLa cells.* J Biol Chem, 1976. **251**(17): p. 5283-91.

157. Keith, J.M., M.J. Ensinger, and B. Mose, *HeLa cell RNA (2'-O-methyladenosine-N6-)-methyltransferase specific for the capped 5'-end of messenger RNA.* J Biol Chem, 1978. **253**(14): p. 5033-9.

158. Pillutla, R.C., et al., *Recombinant human mRNA cap methyltransferase binds capping enzyme/RNA polymerase IIo complexes.* J Biol Chem, 1998. **273**(34): p. 21443-6.

159. Sun, H., et al., *Cap-specific, terminal N(6)-methylation by a mammalian m(6)Am methyltransferase.* Cell Res, 2019. **29**(1): p. 80-82.

160. Fan, H., et al., *PCIF1, a novel human WW domain-containing protein, interacts with the phosphorylated RNA polymerase II.* Biochem Biophys Res Commun, 2003. **301**(2): p. 378-85.

161. Hirose, Y., et al., *Human phosphorylated CTD-interacting protein, PCIF1, negatively modulates gene expression by RNA polymerase II.* Biochem Biophys Res Commun, 2008. **369**(2): p. 449-55.

162. Mauer, J., et al., *Reversible methylation of m(6)Am in the 5' cap controls mRNA stability.* Nature, 2017. **541**(7637): p. 371-375.

163. Wei, J., et al., *Differential m(6)A, m(6)Am, and m(1)A Demethylation Mediated by FTO in the Cell Nucleus and Cytoplasm.* Mol Cell, 2018. **71**(6): p. 973-985 e5.

164. Tong, J., et al., *The Emerging Role of RNA Modifications in the Regulation of Antiviral Innate Immunity.* Front Microbiol, 2022. **13**: p. 845625.

165. Li, N. and T.M. Rana, *Regulation of antiviral innate immunity by chemical modification of viral RNA.* Wiley Interdiscip Rev RNA, 2022: p. e1720.

166. Barbieri, I. and T. Kouzarides, *Role of RNA modifications in cancer.* Nat Rev Cancer, 2020. **20**(6): p. 303-322.

167. Kumar, S. and T. Mohapatra, *Deciphering Epitranscriptome: Modification of mRNA Bases Provides a New Perspective for Post-transcriptional Regulation of Gene Expression.* Front Cell Dev Biol, 2021. **9**: p. 628415.

168. Thompson, M.G., M.T. Sacco, and S.M. Horner, *How RNA modifications regulate the antiviral response.* Immunol Rev, 2021. **304**(1): p. 169-180.

169. Netzband, R. and C.T. Pager, *Epitranscriptomic marks: Emerging modulators of RNA virus gene expression.* Wiley Interdiscip Rev RNA, 2020. **11**(3): p. e1576.

170. Mauer, J. and S.R. Jaffrey, *FTO, m(6) Am , and the hypothesis of reversible epitranscriptomic mRNA modifications.* FEBS Lett, 2018. **592**(12): p. 2012-2022.

171. Meyer, K.D. and S.R. Jaffrey, *Rethinking m(6)A Readers, Writers, and Erasers.* Annu Rev Cell Dev Biol, 2017. **33**: p. 319-342.

172. Meyer, K.D. and S.R. Jaffrey, *The dynamic epitranscriptome: N6-methyladenosine and gene expression control.* Nat Rev Mol Cell Biol, 2014. **15**(5): p. 313-26.

173. Zhou, J., et al., *Dynamic m(6)A mRNA methylation directs translational control of heat shock response.* Nature, 2015. **526**(7574): p. 591-4.

174. Shi, H., et al., *YTHDF3 facilitates translation and decay of N(6)-methyladenosine-modified RNA.* Cell Res, 2017. **27**(3): p. 315-328.

175. Wang, X., et al., *N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency.* Cell, 2015. **161**(6): p. 1388-99.

176. Kretschmer, J., et al., *The m(6)A reader protein YTHDC2 interacts with the small ribosomal subunit and the 5'-3' exoribonuclease XRN1.* RNA, 2018. **24**(10): p. 1339-1350.

177. Mao, Y., et al., *m(6)A in mRNA coding regions promotes translation via the RNA helicase-containing YTHDC2.* Nat Commun, 2019. **10**(1): p. 5332.

178. Xiao, W., et al., *Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing.* Mol Cell, 2016. **61**(4): p. 507-519.

179. Alarcon, C.R., et al., *HNRNPA2B1 Is a Mediator of m(6)A-Dependent Nuclear RNA Processing Events.* Cell, 2015. **162**(6): p. 1299-308.

180. Wang, L., M. Wen, and X. Cao, *Nuclear hnRNPA2B1 initiates and amplifies the innate immune response to DNA viruses.* Science, 2019. **365**(6454).

181. Gokhale, N.S. and S.M. Horner, *RNA modifications go viral.* PLoS Pathog, 2017. **13**(3): p. e1006188.

182. Gokhale, N.S., et al., *Altered m(6)A Modification of Specific Cellular Transcripts Affects Flaviviridae Infection.* Mol Cell, 2020. **77**(3): p. 542-555 e8.

183. Kennedy, E.M., et al., *Posttranscriptional m(6)A Editing of HIV-1 mRNAs Enhances Viral Gene Expression.* Cell Host Microbe, 2017. **22**(6): p. 830.

184. Lichinchi, G., et al., *Dynamics of the human and viral m(6)A RNA methylomes during HIV-1 infection of T cells.* Nat Microbiol, 2016. **1**: p. 16011.

185. Lichinchi, G., et al., *Dynamics of Human and Viral RNA Methylation during Zika Virus Infection.* Cell Host Microbe, 2016. **20**(5): p. 666-673.

186. Price, A.M., et al., *Direct RNA sequencing reveals m(6)A modifications on adenovirus RNA are necessary for efficient splicing.* Nat Commun, 2020. **11**(1): p. 6016.

187. Dominissini, D., et al., *The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA.* Nature, 2016. **530**(7591): p. 441-6.

188. Safra, M., et al., *The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution.* Nature, 2017. **551**(7679): p. 251-255.

189. Schwartz, S., *m(1)A within cytoplasmic mRNAs at single nucleotide resolution: a reconciled transcriptome-wide map.* RNA, 2018. **24**(11): p. 1427-1436.

190. Zhou, H., et al., *Evolution of a reverse transcriptase to map N(1)-methyladenosine in human messenger RNA.* Nat Methods, 2019. **16**(12): p. 1281-1288.

191. Woo, H.H. and S.K. Chambers, *Human ALKBH3-induced m(1)A demethylation increases the CSF-1 mRNA stability in breast and ovarian cancer cells.* Biochim Biophys Acta Gene Regul Mech, 2019. **1862**(1): p. 35-46.

192. Furuse, Y., *RNA Modifications in Genomic RNA of Influenza A Virus and the Relationship between RNA Modifications and Viral Infection.* Int J Mol Sci, 2021. **22**(17).

193. Yang, X., et al., *5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader.* Cell Res, 2017. **27**(5): p. 606-625.

194. Squires, J.E., et al., *Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA.* Nucleic Acids Res, 2012. **40**(11): p. 5023-33.

195. Legrand, C., et al., *Statistically robust methylation calling for whole-transcriptome bisulfite sequencing reveals distinct methylation patterns for mouse RNAs.* Genome Res, 2017. **27**(9): p. 1589-1596.

196. Huang, T., et al., *Genome-wide identification of mRNA 5-methylcytosine in mammals.* Nat Struct Mol Biol, 2019. **26**(5): p. 380-388.

197. Xing, J., et al., *NSun2 Promotes Cell Growth via Elevating Cyclin-Dependent Kinase 1 Translation.* Mol Cell Biol, 2015. **35**(23): p. 4043-52.

198. Courtney, D.G., et al., *Extensive Epitranscriptomic Methylation of A and C Residues on Murine Leukemia Virus Transcripts Enhances Viral Gene Expression.* mBio, 2019. **10**(3).

199. Courtney, D.G., et al., *Epitranscriptomic Addition of m(5)C to HIV-1 Transcripts Regulates Viral Gene Expression.* Cell Host Microbe, 2019. **26**(2): p. 217-227 e6.

200. Dubin, D.T. and V. Stollar, *Methylation of Sindbis virus "26S" messenger RNA.* Biochem Biophys Res Commun, 1975. **66**(4): p. 1373-9.

201. Dubin, D.T., et al., *Sindbis virus messenger RNA: the 5'-termini and methylated residues of 26 and 42 S RNA.* Virology, 1977. **77**(2): p. 457-70.

202. Arango, D., et al., *Acetylation of Cytidine in mRNA Promotes Translation Efficiency.* Cell, 2018. **175**(7): p. 1872-1886 e24.

203. Tsai, K., et al., *Acetylation of Cytidine Residues Boosts HIV-1 Gene Expression by Increasing Viral RNA Stability.* Cell Host Microbe, 2020. **28**(2): p. 306-312 e6.

204. Gelinas, J.F., et al., *Enhancement of replication of RNA viruses by ADAR1 via RNA editing and inhibition of RNA-activated protein kinase.* J Virol, 2011. **85**(17): p. 8460-6.

205. Borchardt, E.K., N.M. Martinez, and W.V. Gilbert, *Regulation and Function of RNA Pseudouridylation in Human Cells.* Annu Rev Genet, 2020. **54**: p. 309-336.

206. Carlile, T.M., et al., *mRNA structure determines modification by pseudouridine synthase 1.* Nat Chem Biol, 2019. **15**(10): p. 966-974.

207. Martinez, N.M., C. Schaening-Burgos, and W.V. Gilbert, *Pseudouridine site assignment by high-throughput in vitro RNA pseudouridylation and sequencing.* Methods Enzymol, 2021. **658**: p. 277-310.

208. Li, X., et al., *Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome.* Nat Chem Biol, 2015. **11**(8): p. 592-7.

209. Eyler, D.E., et al., *Pseudouridinylation of mRNA coding sequences alters translation.* Proc Natl Acad Sci U S A, 2019. **116**(46): p. 23068-23074.

210. Martinez, N.M., et al., *Pseudouridine synthases modify human pre-mRNA co-transcriptionally and affect pre-mRNA processing.* Mol Cell, 2022. **82**(3): p. 645-659 e9.

211. McIntyre, W., et al., *Positive-sense RNA viruses reveal the complexity and dynamics of the cellular and viral epitranscriptomes during infection.* Nucleic Acids Res, 2018. **46**(11): p. 5776-5791.

212. Schlee, M. and G. Hartmann, *Discriminating self from non-self in nucleic acid sensing.* Nat Rev Immunol, 2016. **16**(9): p. 566-80.

213. Li, D. and M. Wu, *Pattern recognition receptors in health and diseases.* Signal Transduct Target Ther, 2021. **6**(1): p. 291.

214. Jensen, S. and A.R. Thomsen, *Sensing of RNA viruses: a review of innate immune receptors involved in recognizing RNA virus invasion.* J Virol, 2012. **86**(6): p. 2900-10.

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

216. Akira, S., K. Takeda, and T. Kaisho, *Toll-like receptors: critical proteins linking innate and acquired immunity.* Nat Immunol, 2001. **2**(8): p. 675-80.

217. Takeda, K. and S. Akira, *Toll-like receptors in innate immunity.* Int Immunol, 2005. **17**(1): p. 1-14.

218. Daffis, S., et al., *Toll-like receptor 3 has a protective role against West Nile virus infection.* J Virol, 2008. **82**(21): p. 10349-58.

219. Gowen, B.B., et al., *TLR3 deletion limits mortality and disease severity due to Phlebovirus infection.* J Immunol, 2006. **177**(9): p. 6301-7.

220. Guillot, L., et al., *Involvement of toll-like receptor 3 in the immune response of lung epithelial cells to double-stranded RNA and influenza A virus.* J Biol Chem, 2005. **280**(7): p. 5571-80.

221. Negishi, H., et al., *A critical link between Toll-like receptor 3 and type II interferon signaling pathways in antiviral innate immunity.* Proc Natl Acad Sci U S A, 2008. **105**(51): p. 20446-51.

222. Oshiumi, H., et al., *The TLR3/TICAM-1 pathway is mandatory for innate immune responses to poliovirus infection.* J Immunol, 2011. **187**(10): p. 5320-7.

223. Rudd, B.D., et al., *Differential role for TLR3 in respiratory syncytial virus-induced chemokine expression.* J Virol, 2005. **79**(6): p. 3350-7.

224. Diebold, S.S., et al., *Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA.* Science, 2004. **303**(5663): p. 1529-31.

225. Heil, F., et al., *Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8.* Science, 2004. **303**(5663): p. 1526-9.

226. Jurk, M., et al., *Human TLR7 or TLR8 independently confer responsiveness to the antiviral compound R-848.* Nat Immunol, 2002. **3**(6): p. 499.

227. Kane, M., et al., *Innate immune sensing of retroviral infection via Toll-like receptor 7 occurs upon viral entry.* Immunity, 2011. **35**(1): p. 135-45.

228. Wang, J.P., et al., *Flavivirus activation of plasmacytoid dendritic cells delineates key elements of TLR7 signaling beyond endosomal recognition.* J Immunol, 2006. **177**(10): p. 7114-21.

229. Kariko, K., et al., *Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA.* Immunity, 2005. **23**(2): p. 165-75.

230. Kariko, K., et al., *Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability.* Mol Ther, 2008. **16**(11): p. 1833-40.

231. Limbach, P.A., P.F. Crain, and J.A. McCloskey, *Summary: the modified nucleosides of RNA.* Nucleic Acids Res, 1994. **22**(12): p. 2183-96.

232. Kato, H., et al., *Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses.* Nature, 2006. **441**(7089): p. 101-5.

233. Devarkar, S.C., et al., *Structural basis for m7G recognition and 2'-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I.* Proc Natl Acad Sci U S A, 2016. **113**(3): p. 596-601.

234. Kim, G.W., et al., *N (6)-Methyladenosine modification of hepatitis B and C viral RNAs attenuates host innate immunity via RIG-I signaling.* J Biol Chem, 2020. **295**(37): p. 13123-13133.

235. Chen, S., et al., *N6-methyladenosine modification of HIV-1 RNA suppresses type-I interferon induction in differentiated monocytic cells and primary macrophages.* PLoS Pathog, 2021. **17**(3): p. e1009421.

236. Durbin, A.F., et al., *RNAs Containing Modified Nucleotides Fail To Trigger RIG-I Conformational Changes for Innate Immune Signaling.* mBio, 2016. **7**(5).

237. Zust, R., et al., *Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5.* Nat Immunol, 2011. **12**(2): p. 137-43.

238. Lamers, M.M., B.G. van den Hoogen, and B.L. Haagmans, *ADAR1: "Editor-in-Chief" of Cytoplasmic Innate Immunity.* Front Immunol, 2019. **10**: p. 1763.

239. Liddicoat, B.J., et al., *RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself.* Science, 2015. **349**(6252): p. 1115-20.

240. Yang, S., et al., *Adenosine deaminase acting on RNA 1 limits RIG-I RNA detection and suppresses IFN production responding to viral and endogenous RNAs.* J Immunol, 2014. **193**(7): p. 3436-45.

241. Balachandran, S., et al., *Activation of the dsRNA-dependent protein kinase, PKR, induces apoptosis through FADD-mediated death signaling.* EMBO J, 1998. **17**(23): p. 6888-902.

242. Balachandran, S., et al., *Essential role for the dsRNA-dependent protein kinase PKR in innate immunity to viral infection.* Immunity, 2000. **13**(1): p. 129-41.

243. Zilberstein, A., et al., *Isolation of two interferon-induced translational inhibitors: a protein kinase and an oligo-isoadenylate synthetase.* Proc Natl Acad Sci U S A, 1978. **75**(10): p. 4734-8.

244. Hur, S., *Double-Stranded RNA Sensors and Modulators in Innate Immunity.* Annu Rev Immunol, 2019. **37**: p. 349-375.

245. Chung, H., et al., *Human ADAR1 Prevents Endogenous RNA from Triggering Translational Shutdown.* Cell, 2018. **172**(4): p. 811-824 e14.

246. Schoggins, J.W., *Interferon-Stimulated Genes: What Do They All Do?* Annu Rev Virol, 2019. **6**(1): p. 567-584.

247. Schoggins, J.W. and C.M. Rice, *Interferon-stimulated genes and their antiviral effector functions.* Curr Opin Virol, 2011. **1**(6): p. 519-25.

248. Schoggins, J.W., et al., *A diverse range of gene products are effectors of the type I interferon antiviral response.* Nature, 2011. **472**(7344): p. 481-5.

249. Hyde, J.L. and M.S. Diamond, *Innate immune restriction and antagonism of viral RNA lacking 2-O methylation.* Virology, 2015. **479-480**: p. 66-74.

250. Johnson, B., et al., *Human IFIT3 Modulates IFIT1 RNA Binding Specificity and Protein Stability.* Immunity, 2018. **48**(3): p. 487-499 e5.

251. Abbas, Y.M., et al., *Structure of human IFIT1 with capped RNA reveals adaptable mRNA binding and mechanisms for sensing N1 and N2 ribose 2'-O methylations.* Proc Natl Acad Sci U S A, 2017. **114**(11): p. E2106-E2115.

252. Deymier, S., et al., *ISG20: an enigmatic antiviral RNase targeting multiple viruses.* FEBS Open Bio, 2022.

253. Imam, H., et al., *Interferon-stimulated gene 20 (ISG20) selectively degrades N6-methyladenosine modified Hepatitis B Virus transcripts.* PLoS Pathog, 2020. **16**(2): p. e1008338.

254. Liu, Y., et al., *Interferon-inducible ribonuclease ISG20 inhibits hepatitis B virus replication through directly binding to the epsilon stem-loop structure of viral RNA.* PLoS Pathog, 2017. **13**(4): p. e1006296.

255. Stadler, D., et al., *Interferon-induced degradation of the persistent hepatitis B virus cccDNA form depends on ISG20.* EMBO Rep, 2021. **22**(6): p. e49568.

256. Zhou, Z., et al., *Antiviral activities of ISG20 in positive-strand RNA virus infections.* Virology, 2011. **409**(2): p. 175-88.

257. Daugherty, M.D., et al., *Evolution-guided functional analyses reveal diverse antiviral specificities encoded by IFIT1 genes in mammals.* Elife, 2016. **5**.

258. Fleith, R.C., et al., *IFIT3 and IFIT2/3 promote IFIT1-mediated translation inhibition by enhancing binding to non-self RNA.* Nucleic Acids Res, 2018. **46**(10): p. 5269-5285.

259. Diamond, M.S. and M. Farzan, *The broad-spectrum antiviral functions of IFIT and IFITM proteins.* Nat Rev Immunol, 2013. **13**(1): p. 46-57.

260. Diamond, M.S., *IFIT1: A dual sensor and effector molecule that detects non-2'-O methylated viral RNA and inhibits its translation.* Cytokine Growth Factor Rev, 2014. **25**(5): p. 543-50.

261. Miedziak, B., et al., *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): p. 58-68.

262. Yang, Z., et al., *Crystal structure of ISG54 reveals a novel RNA binding structure and potential functional mechanisms.* Cell Res, 2012. **22**(9): p. 1328-38.

263. Tran, V., et al., *Influenza virus repurposes the antiviral protein IFIT2 to promote translation of viral mRNAs.* Nat Microbiol, 2020. **5**(12): p. 1490-1503.

264. Mears, H.V. and T.R. Sweeney, *Mouse Ifit1b is a cap1-RNA-binding protein that inhibits mouse coronavirus translation and is regulated by complexing with Ifit1c.* J Biol Chem, 2020. **295**(51): p. 17781-17801.

265. Takata, M.A., et al., *CG dinucleotide suppression enables antiviral defence targeting non-self RNA.* Nature, 2017. **550**(7674): p. 124-127.

266. Williams, G.D., et al., *The mRNA Cap 2'-O-Methyltransferase CMTR1 Regulates the Expression of Certain Interferon-Stimulated Genes.* mSphere, 2020. **5**(3).

267. Zheng, Q., et al., *The RNA helicase DDX46 inhibits innate immunity by entrapping m(6)A-demethylated antiviral transcripts in the nucleus.* Nat Immunol, 2017. **18**(10): p. 1094-1103.

268. Xu, J., et al., *The RNA helicase DDX5 promotes viral infection via regulating N6-methyladenosine levels on the DHX58 and NFkappaB transcripts to dampen antiviral innate immunity.* PLoS Pathog, 2021. **17**(4): p. e1009530.

269. Winkler, R., et al., *m(6)A modification controls the innate immune response to infection by targeting type I interferons.* Nat Immunol, 2019. **20**(2): p. 173-182.

270. Rubio, R.M., et al., *RNA m(6) A modification enzymes shape innate responses to DNA by regulating interferon beta.* Genes Dev, 2018. **32**(23-24): p. 1472-1484.

271. McFadden, M.J. and S.M. Horner, *N(6)-Methyladenosine Regulates Host Responses to Viral Infection.* Trends Biochem Sci, 2021. **46**(5): p. 366-377.

272. McFadden, M.J., et al., *Post-transcriptional regulation of antiviral gene expression by N6-methyladenosine.* Cell Rep, 2021. **34**(9): p. 108798.

273. McFadden, M.J., et al., *FTO Suppresses STAT3 Activation and Modulates Proinflammatory Interferon-Stimulated Gene Expression.* J Mol Biol, 2021: p. 167247.

274. Wang, H., et al., *NSUN2-mediated m^5C methylation of IRF3 mRNA negatively regulates type I interferon responses.* bioRxiv, 2021: p. 2021.07.09.451748.

275. Whelan, S.P., et al., *Efficient recovery of infectious vesicular stomatitis virus entirely from cDNA clones.* Proc Natl Acad Sci U S A, 1995. **92**(18): p. 8388-92.

276. Kahn, J.S., et al., *Recombinant vesicular stomatitis virus expressing respiratory syncytial virus (RSV) glycoproteins: RSV fusion protein can mediate infection and cell fusion.* Virology, 1999. **254**(1): p. 81-91.

277. Carette, J.E., et al., *Ebola virus entry requires the cholesterol transporter Niemann-Pick C1.* Nature, 2011. **477**(7364): p. 340-3.

278. Chandran, K., et al., *Endosomal proteolysis of the Ebola virus glycoprotein is necessary for infection.* Science, 2005. **308**(5728): p. 1643-5.

279. Robinson, L.R. and S.P. Whelan, *Infectious Entry Pathway Mediated by the Human Endogenous Retrovirus K Envelope Protein.* J Virol, 2016. **90**(7): p. 3640-9.

280. Robinson-McCarthy, L.R., et al., *Reconstruction of the cell entry pathway of an extinct virus.* PLoS Pathog, 2018. **14**(8): p. e1007123.

281. Jae, L.T., et al., *Deciphering the glycosylome of dystroglycanopathies using haploid screens for lassa virus entry.* Science, 2013. **340**(6131): p. 479-83.

282. Piccinotti, S. and S.P. Whelan, *Rabies Internalizes into Primary Peripheral Neurons via Clathrin Coated Pits and Requires Fusion at the Cell Body.* PLoS Pathog, 2016. **12**(7): p. e1005753.

283. Raaben, M., et al., *NRP2 and CD63 Are Host Factors for Lujo Virus Cell Entry.* Cell Host Microbe, 2017. **22**(5): p. 688-696 e5.

284. Stubbs, S.H., et al., *Vesicular Stomatitis Virus Chimeras Expressing the Oropouche Virus Glycoproteins Elicit Protective Immune Responses in Mice.* mBio, 2021. **12**(4): p. e0046321.

285. Ganaie, S.S., et al., *Lrp1 is a host entry factor for Rift Valley fever virus.* Cell, 2021. **184**(20): p. 5163-5178 e24.

286. Bakkers, M.J.G., et al., *CD164 is a host factor for lymphocytic choriomeningitis virus entry.* Proc Natl Acad Sci U S A, 2022. **119**(10): p. e2119676119.

287. Case, J.B., et al., *Neutralizing Antibody and Soluble ACE2 Inhibition of a Replication-Competent VSV-SARS-CoV-2 and a Clinical Isolate of SARS-CoV-2.* Cell Host Microbe, 2020. **28**(3): p. 475-485 e5.

288. Zhou, Z., Y. Qiu, and X. Ge, *The taxonomy, host range and pathogenicity of coronaviruses and other viruses in the Nidovirales order.* Anim Dis, 2021. **1**(1): p. 5.

289. Jackson, C.B., et al., *Mechanisms of SARS-CoV-2 entry into cells.* Nat Rev Mol Cell Biol, 2022. **23**(1): p. 3-20.

290. Li, F., *Structure, Function, and Evolution of Coronavirus Spike Proteins.* Annu Rev Virol, 2016. **3**(1): p. 237-261.

291. Seth, S., J. Batra, and S. Srinivasan, *COVID-19: Targeting Proteases in Viral Invasion and Host Immune Response.* Front Mol Biosci, 2020. **7**: p. 215.

292. Rahbar Saadat, Y., et al., *Host Serine Proteases: A Potential Targeted Therapy for COVID-19 and Influenza.* Front Mol Biosci, 2021. **8**: p. 725528.

293. Hoffmann, M., et al., *SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor.* Cell, 2020. **181**(2): p. 271-280 e8.

294. Peacock, T.P., et al., *The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets.* Nat Microbiol, 2021. **6**(7): p. 899-909.

295. Shang, J., et al., *Structural basis of receptor recognition by SARS-CoV-2.* Nature, 2020. **581**(7807): p. 221-224.

296. Lan, J., et al., *Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor.* Nature, 2020. **581**(7807): p. 215-220.

297. Wang, Q., et al., *Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2.* Cell, 2020. **181**(4): p. 894-904 e9.

298. Wrapp, D., et al., *Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation.* Science, 2020. **367**(6483): p. 1260-1263.

299. Cai, Y., et al., *Distinct conformational states of SARS-CoV-2 spike protein.* Science, 2020. **369**(6511): p. 1586-1592.

300. Ke, Z., et al., *Structures and distributions of SARS-CoV-2 spike proteins on intact virions.* Nature, 2020. **588**(7838): p. 498-502.

301. Yuan, Y., et al., *Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains.* Nat Commun, 2017. **8**: p. 15092.

302. Li, F., et al., *Structure of SARS coronavirus spike receptor-binding domain complexed with receptor.* Science, 2005. **309**(5742): p. 1864-8.

303. Wang, N., et al., *Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4.* Cell Res, 2013. **23**(8): p. 986-93.

304. Tong, P., et al., *Memory B cell repertoire for recognition of evolving SARS-CoV-2 spike.* Cell, 2021. **184**(19): p. 4969-4980 e15.

305. Hansen, J., et al., *Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail.* Science, 2020. **369**(6506): p. 1010-1014.

306. Robbiani, D.F., et al., *Convergent antibody responses to SARS-CoV-2 in convalescent individuals.* Nature, 2020. **584**(7821): p. 437-442.

307. Hoffmann, M., H. Kleine-Weber, and S. Pohlmann, *A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells.* Mol Cell, 2020. **78**(4): p. 779-784 e5.

308. Johnson, B.A., et al., *Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis.* Nature, 2021. **591**(7849): p. 293-299.

309. Shang, J., et al., *Cell entry mechanisms of SARS-CoV-2.* Proc Natl Acad Sci U S A, 2020. **117**(21): p. 11727-11734.

310. Sasaki, M., et al., *SARS-CoV-2 variants with mutations at the S1/S2 cleavage site are generated in vitro during propagation in TMPRSS2-deficient cells.* PLoS Pathog, 2021. **17**(1): p. e1009233.

311. Liu, Z., et al., *Identification of Common Deletions in the Spike Protein of Severe Acute Respiratory Syndrome Coronavirus 2.* J Virol, 2020. **94**(17).

312. Hoffmann, M., et al., *Camostat mesylate inhibits SARS-CoV-2 activation by TMPRSS2-related proteases and its metabolite GBPA exerts antiviral activity.* EBioMedicine, 2021. **65**: p. 103255.

313. Shen, L.W., et al., *TMPRSS2: A potential target for treatment of influenza virus and coronavirus infections.* Biochimie, 2017. **142**: p. 1-10.

314. Ou, T., et al., *Hydroxychloroquine-mediated inhibition of SARS-CoV-2 entry is attenuated by TMPRSS2.* PLoS Pathog, 2021. **17**(1): p. e1009212.

315. Bosch, B.J., W. Bartelink, and P.J. Rottier, *Cathepsin L functionally cleaves the severe acute respiratory syndrome coronavirus class I fusion protein upstream of rather than adjacent to the fusion peptide.* J Virol, 2008. **82**(17): p. 8887-90.

316. Ebert, D.H., et al., *Cathepsin L and cathepsin B mediate reovirus disassembly in murine fibroblast cells.* J Biol Chem, 2002. **277**(27): p. 24609-17.

317. Furuichi, Y., A. LaFiandra, and A.J. Shatkin, *5'-Terminal structure and mRNA stability.* Nature, 1977. **266**(5599): p. 235-9.

318. Kruse, S., et al., *A novel synthesis and detection method for cap-associated adenosine modifications in mouse mRNA.* Sci Rep, 2011. **1**: p. 126.

319. Wang, G., et al., *Attenuation of the Innate Immune Response against Viral Infection Due to ZNF598-Promoted Binding of FAT10 to RIG-I.* Cell Rep, 2019. **28**(8): p. 1961-1970 e4.

320. Horwitz, J.A., et al., *Structure of a rabies virus polymerase complex from electron cryo-microscopy.* Proc Natl Acad Sci U S A, 2020. **117**(4): p. 2099-2107.

321. Noton, S.L. and R. Fearns, *Initiation and regulation of paramyxovirus transcription and replication.* Virology, 2015. **479-480**: p. 545-54.

322. Ogino, M., et al., *The Rabies Virus L Protein Catalyzes mRNA Capping with GDP Polyribonucleotidyltransferase Activity.* Viruses, 2016. **8**(5).

323. Li, Y., et al., *Dcp2 decapping protein modulates mRNA stability of the critical interferon regulatory factor (IRF) IRF-7.* Mol Cell Biol, 2012. **32**(6): p. 1164-72.

324. Pandey, R.R., et al., *The Mammalian Cap-Specific m(6)Am RNA Methyltransferase PCIF1 Regulates Transcript Levels in Mouse Tissues.* Cell Rep, 2020. **32**(7): p. 108038.

325. You, C., X. Dai, and Y. Wang, *Position-dependent effects of regioisomeric methylated adenine and guanine ribonucleosides on translation.* Nucleic Acids Res, 2017. **45**(15): p. 9059-9067.

326. Hudson, B.H. and H.S. Zaher, *O6-Methylguanosine leads to position-dependent effects on ribosome speed and fidelity.* RNA, 2015. **21**(9): p. 1648-59.

327. Boone, R.F. and B. Moss, *Methylated 5'-terminal sequences of vaccinia virus mRNA species made in vivo at early and late times after infection.* Virology, 1977. **79**(1): p. 67-80.

328. Cureton, D.K., R. Burdeinick-Kerr, and S.P. Whelan, *Genetic inactivation of COPI coatomer separately inhibits vesicular stomatitis virus entry and gene expression.* J Virol, 2012. **86**(2): p. 655-66.

329. Franz, K.M., et al., *STING-dependent translation inhibition restricts RNA virus replication.* Proc Natl Acad Sci U S A, 2018. **115**(9): p. E2058-E2067.

330. Wickersham, I.R., et al., *Retrograde neuronal tracing with a deletion-mutant rabies virus.* Nat Methods, 2007. **4**(1): p. 47-9.

331. Whelan, S.P. and G.W. Wertz, *Transcription and replication initiate at separate sites on the vesicular stomatitis virus genome.* Proc Natl Acad Sci U S A, 2002. **99**(14): p. 9178-83.

332. Rio, D.C., *Filter-binding assay for analysis of RNA-protein interactions.* Cold Spring Harb Protoc, 2012. **2012**(10): p. 1078-81.

333. Mirtaleb, M.S., et al., *Potential therapeutic agents to COVID-19: An update review on antiviral therapy, immunotherapy, and cell therapy.* Biomed Pharmacother, 2021. **138**: p. 111518.

334. McKee, D.L., et al., *Candidate drugs against SARS-CoV-2 and COVID-19.* Pharmacol Res, 2020. **157**: p. 104859.

335. Cui, W., K. Yang, and H. Yang, *Recent Progress in the Drug Development Targeting SARS-CoV-2 Main Protease as Treatment for COVID-19.* Frontiers in Molecular Biosciences, 2020. **7**(398).

336. Luan, B., et al., *Targeting Proteases for Treating COVID-19.* J Proteome Res, 2020. **19**(11): p. 4316-4326.

337. Mellott, D.M., et al., *A Clinical-Stage Cysteine Protease Inhibitor blocks SARS-CoV-2 Infection of Human and Monkey Cells.* ACS Chem Biol, 2021.

338. Bojkova, D., et al., *SARS-CoV-2 and SARS-CoV differ in their cell tropism and drug sensitivity profiles.* bioRxiv, 2020: p. 2020.04.03.024257.

339. Laporte, M., et al., *The SARS-CoV-2 and other human coronavirus spike proteins are fine-tuned towards temperature and proteases of the human airways.* PLoS Pathog, 2021. **17**(4): p. e1009500.

340. Park, J.E., et al., *Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism.* Proc Natl Acad Sci U S A, 2016. **113**(43): p. 12262-12267.

341. Choi, S.Y., et al., *Type II transmembrane serine proteases in cancer and viral infections.* Trends Mol Med, 2009. **15**(7): p. 303-12.

342. Bugge, T.H., T.M. Antalis, and Q. Wu, *Type II transmembrane serine proteases.* The Journal of biological chemistry, 2009. **284**(35): p. 23177-81.

343. Bestle, D., et al., *TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells.* Life Science Alliance, 2020. **3**(9): p. e202000786.

344. Li, F., et al., *Distinct mechanisms for TMPRSS2 expression explain organ-specific inhibition of SARS-CoV-2 infection by enzalutamide.* Nat Commun, 2021. **12**(1): p. 866.

345. Zang, R., et al., *TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes.* Sci Immunol, 2020. **5**(47).

346. Bertram, S., et al., *Influenza and SARS-coronavirus activating proteases TMPRSS2 and HAT are expressed at multiple sites in human respiratory and gastrointestinal tracts.* PLoS One, 2012. **7**(4): p. e35876.

347. Simmons, G., et al., *Proteolytic activation of the SARS-coronavirus spike protein: cutting enzymes at the cutting edge of antiviral research.* Antiviral Research, 2013. **100**(3): p. 605-14.

348. Leow, M.K., *Correlating cell line studies with tissue distribution of DPP4/TMPRSS2 and human biological samples may better define the viral tropism of MERS-CoV.* J Infect Dis, 2013. **208**(8): p. 1350-1.

349. Limburg, H., et al., *TMPRSS2 Is the Major Activating Protease of Influenza A Virus in Primary Human Airway Cells and Influenza B Virus in Human Type II Pneumocytes.* J Virol, 2019. **93**(21).

350. Garten, W., et al., *Influenza virus activating host proteases: Identification, localization and inhibitors as potential therapeutics.* European Journal of Cell Biology, 2015. **94**(7-9): p. 375-83.

351. Sakai, K., et al., *The host protease TMPRSS2 plays a major role in in vivo replication of emerging H7N9 and seasonal influenza viruses.* J Virol, 2014. **88**(10): p. 5608-16.

352. Hatesuer, B., et al., *Tmprss2 is essential for influenza H1N1 virus pathogenesis in mice.* PLoS Pathog, 2013. **9**(12): p. e1003774.

353. Baron, J., et al., *Matriptase, HAT, and TMPRSS2 activate the hemagglutinin of H9N2 influenza A viruses.* J Virol, 2013. **87**(3): p. 1811-20.

354. Abe, M., et al., *TMPRSS2 is an activating protease for respiratory parainfluenza viruses.* J Virol, 2013. **87**(21): p. 11930-5.

355. Bottcher, E., et al., *Proteolytic activation of influenza viruses by serine proteases TMPRSS2 and HAT from human airway epithelium.* J Virol, 2006. **80**(19): p. 9896-8.

356. Koch, J., et al., *TMPRSS2 expression dictates the entry route used by SARS-CoV-2 to infect host cells.* EMBO J, 2021: p. e107821.

357. Hoffmann, M., et al., *Nafamostat Mesylate Blocks Activation of SARS-CoV-2: New Treatment Option for COVID-19.* Antimicrob Agents Chemother, 2020. **64**(6).

358. Lucas, J.M., et al., *The androgen-regulated protease TMPRSS2 activates a proteolytic cascade involving components of the tumor microenvironment and promotes prostate cancer metastasis.* Cancer Discov, 2014. **4**(11): p. 1310-25.

359. Tanabe, L.M. and K. List, *The role of type II transmembrane serine protease-mediated signaling in cancer.* FEBS J, 2017. **284**(10): p. 1421-1436.

360. Martin, C.E. and K. List, *Cell surface-anchored serine proteases in cancer progression and metastasis.* Cancer Metastasis Rev, 2019. **38**(3): p. 357-387.

361. Damalanka, V.C. and J.W. Janetka, *Recent progress on inhibitors of the type II transmembrane serine proteases, hepsin, matriptase and matriptase-2.* Future Med Chem, 2019. **11**(7): p. 743-769.

362. Owusu, B.Y., et al., *Hepatocyte Growth Factor, a Key Tumor-Promoting Factor in the Tumor Microenvironment.* Cancers (Basel), 2017. **9**(4).

363. Damalanka, V.C., S.A. Wildman, and J.W. Janetka, *Piperidine carbamate peptidomimetic inhibitors of the serine proteases HGFA, matriptase and hepsin.* Medchemcomm, 2019. **10**(9): p. 1646-1655.

364. Damalanka, V.C., et al., *Discovery of Selective Matriptase and Hepsin Serine Protease Inhibitors: Useful Chemical Tools for Cancer Cell Biology.* J Med Chem, 2019. **62**(2): p. 480-490.

365. Han, Z., et al., *alpha-Ketobenzothiazole Serine Protease Inhibitors of Aberrant HGF/c-MET and MSP/RON Kinase Pathway Signaling in Cancer.* ChemMedChem, 2016. **11**(6): p. 585-99.

366. Franco, F.M., et al., *Structure-based discovery of small molecule hepsin and HGFA protease inhibitors: Evaluation of potency and selectivity derived from distinct binding pockets.* Bioorg Med Chem, 2015. **23**(10): p. 2328-43.

367. Han, Z., et al., *Inhibitors of HGFA, Matriptase, and Hepsin Serine Proteases: A Nonkinase Strategy to Block Cell Signaling in Cancer.* ACS Med Chem Lett, 2014. **5**(11): p. 1219-24.

368. Owusu, B.Y., et al., *Targeting the tumor-promoting microenvironment in MET-amplified NSCLC cells with a novel inhibitor of pro-HGF activation.* Oncotarget, 2017. **8**(38): p. 63014-63025.

369. Hempel, T., et al., *Molecular mechanism of inhibiting the SARS-CoV-2 cell entry facilitator TMPRSS2 with camostat and nafamostat.* Chemical Science, 2021. **12**(3): p. 983-992.

370. Azouz, N.P., et al., *Alpha 1 Antitrypsin is an Inhibitor of the SARS-CoV-2-Priming Protease TMPRSS2.* bioRxiv, 2020.

371. Shrimp, J.H., et al., *An Enzymatic TMPRSS2 Assay for Assessment of Clinical Candidates and Discovery of Inhibitors as Potential Treatment of COVID-19.* ACS Pharmacol Transl Sci, 2020. **3**(5): p. 997-1007.

372. Stefano, R., et al., *Homology Modeling of TMPRSS2 Yields Candidate Drugs That May Inhibit Entry of SARS-CoV-2 into Human Cells*. 2020.

373. Takeuchi, T., et al., *Cellular localization of membrane-type serine protease 1 and identification of protease-activated receptor-2 and single-chain urokinase-type plasminogen activator as substrates.* J Biol Chem, 2000. **275**(34): p. 26333-42.

374. Herter, S., et al., *Hepatocyte growth factor is a preferred in vitro substrate for human hepsin, a membrane-anchored serine protease implicated in prostate and ovarian cancers.* Biochemical Journal, 2005. **390**: p. 125-136.

375. O'Donoghue, A.J., et al., *Global identification of peptidase specificity by multiplex substrate profiling.* Nat Methods, 2012. **9**(11): p. 1095-100.

376. Chung, D.-H., et al., *A cell based high-throughput screening approach for the discovery of new inhibitors of respiratory syncytial virus.* Virology Journal, 2013. **10**(1): p. 19.

377. Lucas, J.M., et al., *The androgen-regulated type II serine protease TMPRSS2 is differentially expressed and mislocalized in prostate adenocarcinoma.* The Journal of pathology, 2008. **215**(2): p. 118-25.

378. Kishimoto, M., et al., *TMPRSS11D and TMPRSS13 Activate the SARS-CoV-2 Spike Protein.* Viruses, 2021. **13**(3).

379. Whittaker, G.R. and M.R. Straus, *Human matriptase/ST 14 proteolytically cleaves H7N9 hemagglutinin and facilitates the activation of influenza A/Shanghai/2/2013 virus in cell culture.* Influenza Other Respir Viruses, 2019.

380. Marsault, E., et al., *New Peptidomimetic Inhibitors of Matriptase to Fight Influenza.* Journal of Peptide Science, 2014. **20**: p. S224-S224.

381. Colombo, E., et al., *New matriptase inhibitors as a potential treatment against influenza.* Abstracts of Papers of the American Chemical Society, 2013. **245**.

382. Beaulieu, A., et al., *Matriptase Proteolytically Activates Influenza Virus and Promotes Multicycle Replication in the Human Airway Epithelium.* Journal of Virology, 2013. **87**(8): p. 4237-4251.

383. Hamilton, B.S., D.W.J. Gludish, and G.R. Whittaker, *Cleavage Activation of the Human-Adapted Influenza Virus Subtypes by Matriptase Reveals both Subtype and Strain Specificities.* Journal of Virology, 2012. **86**(19): p. 10579-10586.

384. Singh, P., et al., *Effect of signal peptide on stability and folding of Escherichia coli thioredoxin.* PLoS One, 2013. **8**(5): p. e63442.

385. Vasquez, J.R., et al., *An expression system for trypsin.* Journal of Cellular Biochemistry, 1989. **39**(3): p. 265-276.

386. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nature Methods, 2009. **6**(5): p. 343-345.

387. Ivry, S.L., et al., *Global Protease Activity Profiling Provides Differential Diagnosis of Pancreatic Cysts.* Clin Cancer Res, 2017. **23**(16): p. 4865-4874.

388. Salcedo, E.C., et al., *Global Protease Activity Profiling Identifies HER2-Driven Proteolysis in Breast Cancer.* ACS Chem Biol, 2021.

389. Meng, B., et al., *Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity.* Nature, 2022. **603**(7902): p. 706-714.

390. Laporte, M. and L. Naesens, *Airway proteases: an emerging drug target for influenza and other respiratory virus infections.* Curr Opin Virol, 2017. **24**: p. 16-24.

391. Bottcher-Friebertshauser, E., H.D. Klenk, and W. Garten, *Activation of influenza viruses by proteases from host cells and bacteria in the human airway epithelium.* Pathog Dis, 2013. **69**(2): p. 87-100.

392. Kuhn, N., et al., *The Proteolytic Activation of (H3N2) Influenza A Virus Hemagglutinin Is Facilitated by Different Type II Transmembrane Serine Proteases.* J Virol, 2016. **90**(9): p. 4298-4307.

393. Sakai, K., et al., *A mutant H3N2 influenza virus uses an alternative activation mechanism in TMPRSS2 knockout mice by loss of an oligosaccharide in the hemagglutinin stalk region.* J Virol, 2015. **89**(9): p. 5154-8.

394. Liu, Z., et al., *Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization.* Cell Host Microbe, 2021. **29**(3): p. 477-488 e4.

395. Lin, J.Y., et al., *hnRNP A1 interacts with the 5' untranslated regions of enterovirus 71 and Sindbis virus RNA and is required for viral replication.* J Virol, 2009. **83**(12): p. 6106-14.

396. Huang, P.N., et al., *Far upstream element binding protein 1 binds the internal ribosomal entry site of enterovirus 71 and enhances viral translation and viral growth.* Nucleic Acids Res, 2011. **39**(22): p. 9633-48.

397. Shih, S.R., V. Stollar, and M.L. Li, *Host factors in enterovirus 71 replication.* J Virol, 2011. **85**(19): p. 9658-66.

398. Tolbert, M., et al., *HnRNP A1 Alters the Structure of a Conserved Enterovirus IRES Domain to Stimulate Viral Translation.* J Mol Biol, 2017. **429**(19): p. 2841-2858.

399. Huang, H.I., et al., *Interactome analysis of the EV71 5' untranslated region in differentiated neuronal cells SH-SY5Y and regulatory role of FBP3 in viral replication.* Proteomics, 2016. **16**(17): p. 2351-62.

400. Gui, H., et al., *hnRNP A1 interacts with the genomic and subgenomic RNA promoters of Sindbis virus and is required for the synthesis of G and SG RNA.* J Biomed Sci, 2010. **17**: p. 59.

401. Kim, C.S., et al., *An RNA-binding protein, hnRNP A1, and a scaffold protein, septin 6, facilitate hepatitis C virus replication.* J Virol, 2007. **81**(8): p. 3852-65.

402. Garaigorta, U. and F.V. Chisari, *Hepatitis C virus blocks interferon effector function by inducing protein kinase R phosphorylation.* Cell Host Microbe, 2009. **6**(6): p. 513-22.

403. Connor, J.H. and D.S. Lyles, *Inhibition of host and viral translation during vesicular stomatitis virus infection. eIF2 is responsible for the inhibition of viral but not host translation.* J Biol Chem, 2005. **280**(14): p. 13512-9.

404. Wu, N., et al., *The interferon stimulated gene 20 protein (ISG20) is an innate defense antiviral factor that discriminates self versus non-self translation.* PLoS Pathog, 2019. **15**(10): p. e1008093.

405. Haline-Vaz, T., T.C. Silva, and N.I. Zanchin, *The human interferon-regulated ISG95 protein interacts with RNA polymerase II and shows methyltransferase activity.* Biochem Biophys Res Commun, 2008. **372**(4): p. 719-24.

406. Kroath, H., et al., *RNA methylation in vaccinia virus-infected chick embryo fibroblasts treated with homologous interferon.* Nucleic Acids Res, 1978. **5**(7): p. 2441-54.

407. Desrosiers, R.C. and P. Lengyel, *Impairment of reovirus mRNA 'cap' methylation in interferon-treated mouse L929 cells.* Biochim Biophys Acta, 1979. **562**(3): p. 471-80.

408. de Ferra, F. and C. Baglioni, *Increase in S-adenosylhomocysteine concentration in interferon-treated HeLa cells and inhibition of methylation of vesicular stomatitis virus mRNA.* J Biol Chem, 1983. **258**(4): p. 2118-21.

409. Harty, R.N., et al., *A proline-rich motif within the matrix protein of vesicular stomatitis virus and rabies virus interacts with WW domains of cellular proteins: implications for viral budding.* J Virol, 1999. **73**(4): p. 2921-9.

410. Heinrich, B.S., et al., *Phase Transitions Drive the Formation of Vesicular Stomatitis Virus Replication Compartments.* mBio, 2018. **9**(5).

411. Trotman, J.B. and D.R. Schoenberg, *A recap of RNA recapping.* Wiley Interdiscip Rev RNA, 2019. **10**(1): p. e1504.

412. Whitlow, Z.W., J.H. Connor, and D.S. Lyles, *Preferential translation of vesicular stomatitis virus mRNAs is conferred by transcription from the viral genome.* J Virol, 2006. **80**(23): p. 11733-42.

413. Liu, W., et al., *Structural basis for nematode eIF4E binding an m(2,2,7)G-Cap and its implications for translation initiation.* Nucleic Acids Res, 2011. **39**(20): p. 8820-32.

414. van den Pol, A.N., S. Ding, and M.D. Robek, *Long-distance interferon signaling within the brain blocks virus spread.* J Virol, 2014. **88**(7): p. 3695-704.

415. Detje, C.N., et al., *Upon intranasal vesicular stomatitis virus infection, astrocytes in the olfactory bulb are important interferon Beta producers that protect from lethal encephalitis.* J Virol, 2015. **89**(5): p. 2731-8.

416. Bolen, C.R., et al., *Dynamic expression profiling of type I and type III interferon-stimulated hepatocytes reveals a stable hierarchy of gene expression.* Hepatology, 2014. **59**(4): p. 1262-72.

417. Bult, C.J., et al., *Mouse Genome Database (MGD) 2019.* Nucleic Acids Res, 2019. **47**(D1): p. D801-D806.

418. Tartell, M.A., et al., *Methylation of viral mRNA cap structures by PCIF1 attenuates the antiviral activity of interferon-beta.* Proc Natl Acad Sci U S A, 2021. **118**(29).

419. Tamarkin-Ben-Harush, A., et al., *Cap-proximal nucleotides via differential eIF4E binding and alternative promoter usage mediate translational response to energy stress.* Elife, 2017. **6**.

420. Levraud, J.P., et al., *IFN-Stimulated Genes in Zebrafish and Humans Define an Ancient Arsenal of Antiviral Immunity.* J Immunol, 2019. **203**(12): p. 3361-3373.

421. Courtney, D.G., et al., *Epitranscriptomic Enhancement of Influenza A Virus Gene Expression and Replication.* Cell Host Microbe, 2017. **22**(3): p. 377-386 e5.

422. Paez, E. and M. Esteban, *Resistance of vaccinia virus to interferon is related to an interference phenomenon between the virus and the interferon system.* Virology, 1984. **134**(1): p. 12-28.

423. Zhang, Q., et al., *HIV reprograms host m(6)Am RNA methylome by viral Vpr protein-mediated degradation of PCIF1.* Nat Commun, 2021. **12**(1): p. 5543.

Appendix 1:

Supplemental Material to Chapter 2

Data within this chapter is published and a reprint of the article is provided with the archived copy of the dissertation. The published manuscript may also be accessed as follows:

PNAS. Jul 20, 2021. Vol. 118. No. 29, e2025769118. doi: 10.1073/pnas.2025769118.

© 2021 by the National Academy of Sciences of the USA

PNAS
www.pnas.org

Supplementary Information for

Messenger RNA cap methylation by PCIF1 attenuates the interferon-β induced antiviral state.

Michael A. Tartell[1,2], Konstantinos Boulias[3,4], Gabriela Brunsting Hoffmann[4], Louis-Marie Bloyet[1], Eric Lieberman Greer[3,4*] and Sean P. J. Whelan[1*]

[1]Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO, USA, [2]Program in Virology, Harvard Medical School, Boston MA, USA, [3]Department of Pediatrics, Initiative for RNA Medicine, Harvard Medical School, Boston, MA, USA [4]Division of Newborn Medicine, Boston Children's Hospital, Boston, MA, USA

***Corresponding author:**

S. P. J. Whelan email: spjwhelan@wustl.edu

E. L. Greer email: Eric.Greer@childrens.harvard.edu

This PDF file includes:

Figures A1.S1 to A1.S11

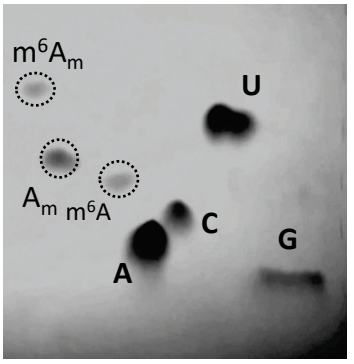
Table A1.S1

Figure A1.S1: VSV mRNAs in HeLa cells contain a 5' $m^7Gpppm^6A_m$ cap-structure. (A): Chemical standards used to identify nucleotide species visualized by UV shadowing (254 nm). **(B):** HeLa cells were infected with VSV at a MOI of 3, cellular transcription halted by adding 10 µg ml^{-1} actinomycin D at 2.5 hpi, and viral RNA metabolically labeled with 100 µCi ml^{-1} [^{32}P] phosphoric acid from 3-7 hpi. RNA was extracted, poly(A) selected and incubated with the indicated nucleases and the products resolved by 2D-TLC and detected by phosphorimaging. Wild type HeLa cells (representative image; n=3). **(C):** 293T PCIF1 KO cells **(D):** In vitro transcribed VSV mRNA (input to Fig 2C).

Figure A1.S1 (continued):

A

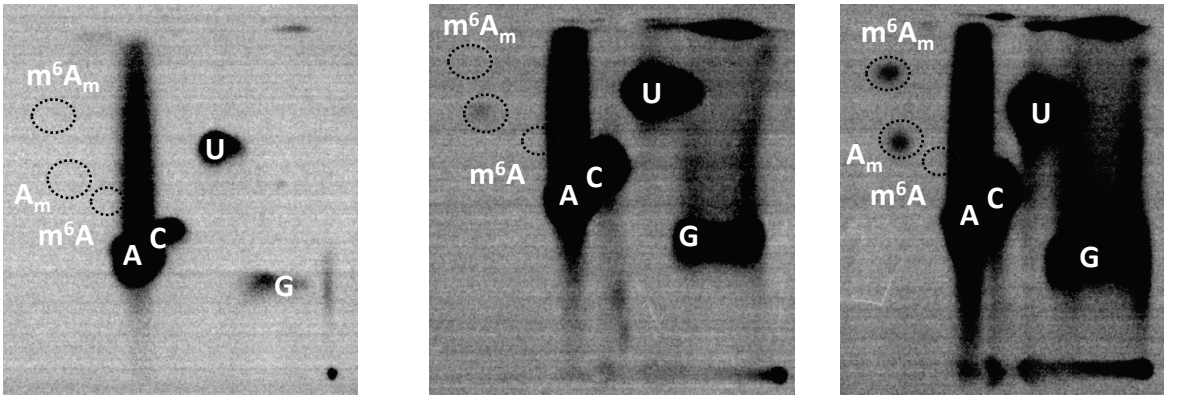
Standards



B

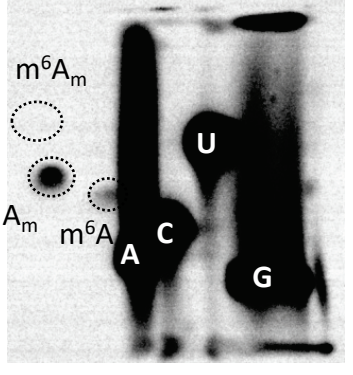
Uninfected | **VSV Infected**

Cap-Clip + P1 | **P1** | **Cap-Clip + P1**



C

PCIF1 KO: VSV Infected



D

***In Vitro* Methylation Input**

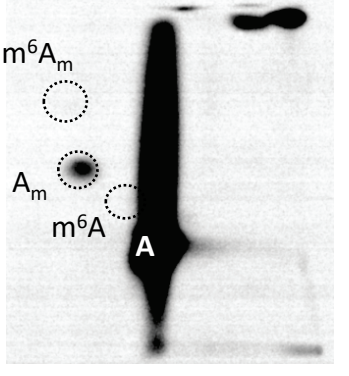


Figure A1.S2: PCIF1 methylates viral mRNA.

293T cells were infected with VSV at a MOI of 3, poly(A)+ RNA purified at 6 hpi, and cap-proximal nucleotide identity determined by selective radiolabeling. Poly(A)+ RNA was decapped with Cap-Clip, and the exposed 5' phosphate of the cap-proximal nucleotide was radiolabeled with [^{32}P] γ-ATP by sequential treatment with shrimp alkaline phosphatase and polynucleotide kinase. Hydrolyzed nucleotide monophosphates were resolved by 2D-TLC and detected by phosphorimaging (representative images; n=3). **(A):** Parental wild type 293T cells. **(B):** PCIF1 knockout 293T cells.

Figure A1.S2 (continued):

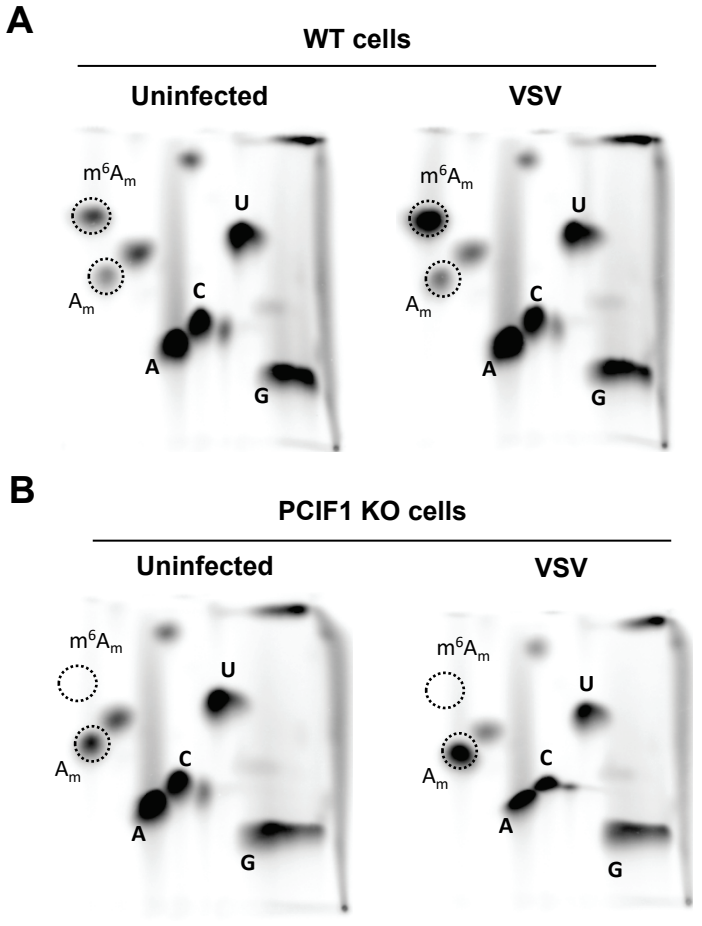


Figure A1.S3: Cap-methylation requirements for i*n vitro* methylation of VSV mRNA. VSV mRNA was transcribed *in vitro* using purified virions from the indicated viruses used in Fig 3, and in the presence or absence of 200 μM SAH (a methylation inhibitor) as noted, followed by *in vitro* methylation with no enzyme ("input"), purified PCIF1, or purified $PCIF1_{SPPG}$. 2D-TLC was performed on the products to determine the relative amounts of m^6A_m and A_m present (representative images; n=3).

Figure A1.S3 (continued):

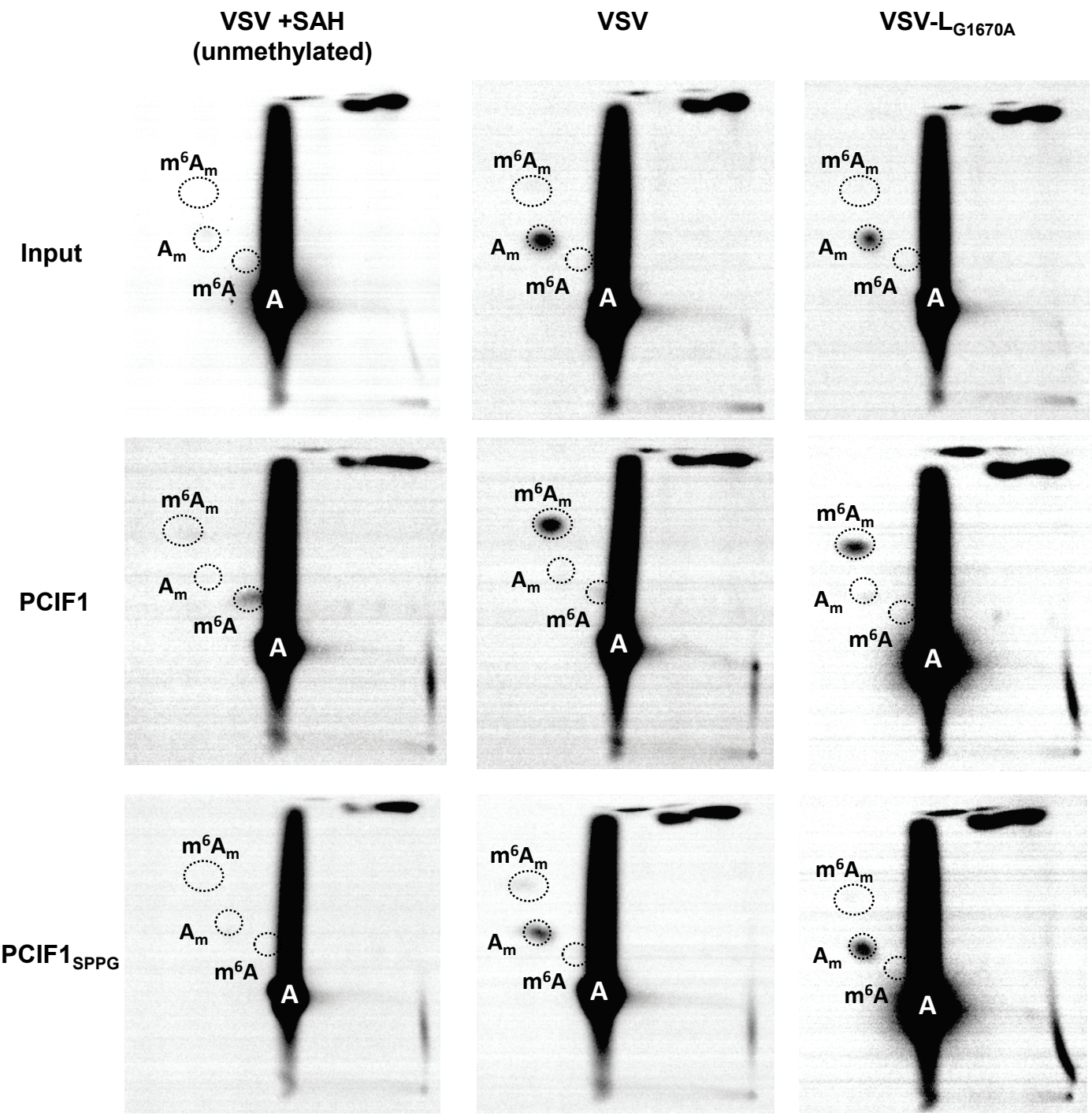


Figure A1.S4: Effect of m^6A_m on VSV mRNA stability.

Purified stocks of VSV radiolabeled mRNA were generated by infecting 293T PCIF1 KO cells at a MOI of 10 with VSV as in Fig 1, followed by purification. Purified RNA (500 ng) was in vitro methylated with PCIF1, transfected into HeLa cells and RNA amounts assessed by re-extraction from cells at the indicated times followed by **(A)** electrophoresis on acid-agarose gels and phosphorimaging (see Fig 4) (n=3, +/- SD). **(B):** Data from A was used to calculate half lives in (n=3, +/- SD). There is no significant difference between any decay curve (2 way ANOVA, 0.75>p>0.4), or calculated half life (student's t-test, 0.98 >p>0.06).

Figure A1.S4 (continued):

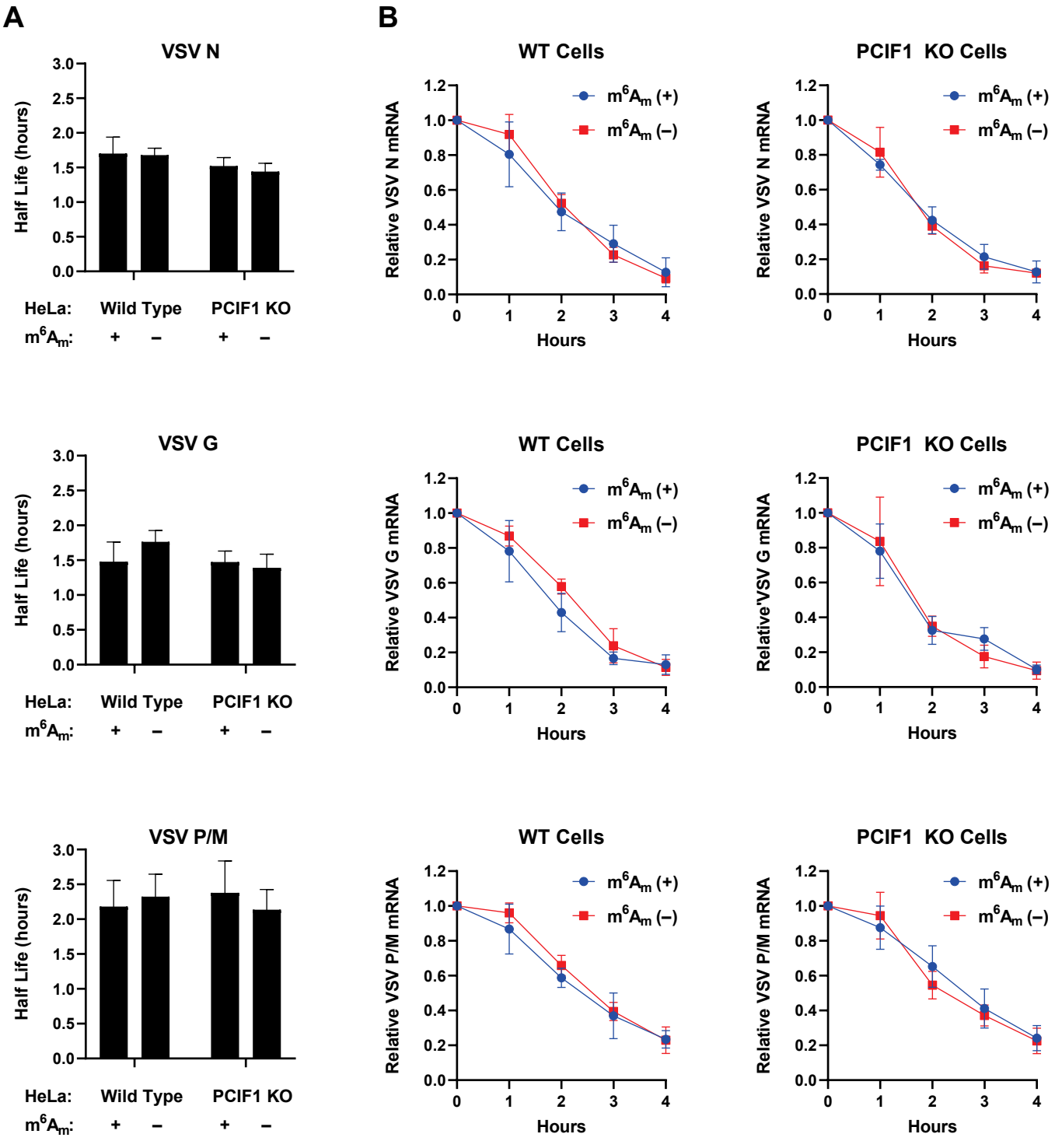


Figure A1.S5: Effect of m^6A_m on translation of VSV mRNA reporters.

Purified stocks of VSV mRNA from the indicated virus were generated by infecting 293T PCIF1 KO cells at a MOI of 10 with VSV as in Fig 4, followed by purification of extracted RNA by poly(A) selection and a biotinylated oligonucleotide against the conserved VSV-stop sequence. 500 ng RNA was then mock or in vitro methylated with purified PCIF1, and transfected into HeLa wild type or PCIF1 knockout cells. **(A):** N6-methylation does not impact translation of a GFP reporter. VSV-eGFP mRNA was transfected into cells, and flow cytometry performed at 6 hpi. Mean fluorescence intensity (n=3, +/- SD, NS – $p>0.18$, student's t-test,) and percent GFP positive cells (n=3, +/- SD, NS – $p>0.89$, student's t-test) are shown, with a representative florescence microscopy image of transfected cells. **(B):** N6-methylation does not impact translation of a luciferase reporter. 500 ng VSV-luciferase mRNA with the indicated methylation was transfected into the indicated HeLa cells. Cells were lysed at 6 hpi, and luciferase levels were measured using a Promega Luciferase Assay kit (n=3, +/- SD, NS – $p>0.70$, student's t-test). **(C):** RNA was extracted from lysate from (B), RT-PCR performed with oligo-dT primers, and qPCR performed for luciferase RNA (n=3, normalized to GAPDH, +/- SD, NS – $p>0.08$, student's t-test). **(D):** Translation of an m^6A_m (+) reporter does not outcompete a co-transfected m^6A_m (–) reporter. 300 ng purified VSV-Luc (firefly) and VSV-RenP (renilla) mRNAs with opposing methylation status (m^6A_m (+) firefly with m^6A_m (–) renilla, and vice versa) were transfected into the indicated HeLa cells for 8 hours. Cells were lysed and luciferase levels of both reporters using a Promega Dual-Luciferase kit. Relative luminescence units (RLU) are shown (n=3, +/- SD, NS – $0.98 > p > 0.11$, student's t-test). Ratios of Firefly to Renilla are shown above each condition.

Figure A1.S5 (continued):

(E): No change in luciferase RNA levels from (D). RNA from (D) was extracted and qPCR performed as in C for Firefly and Renilla luciferase transcripts (n=3, normalized to GAPDH, +/- SD, 0.49>p >0.05, student's t-test).

Figure A1.S5 (continued):

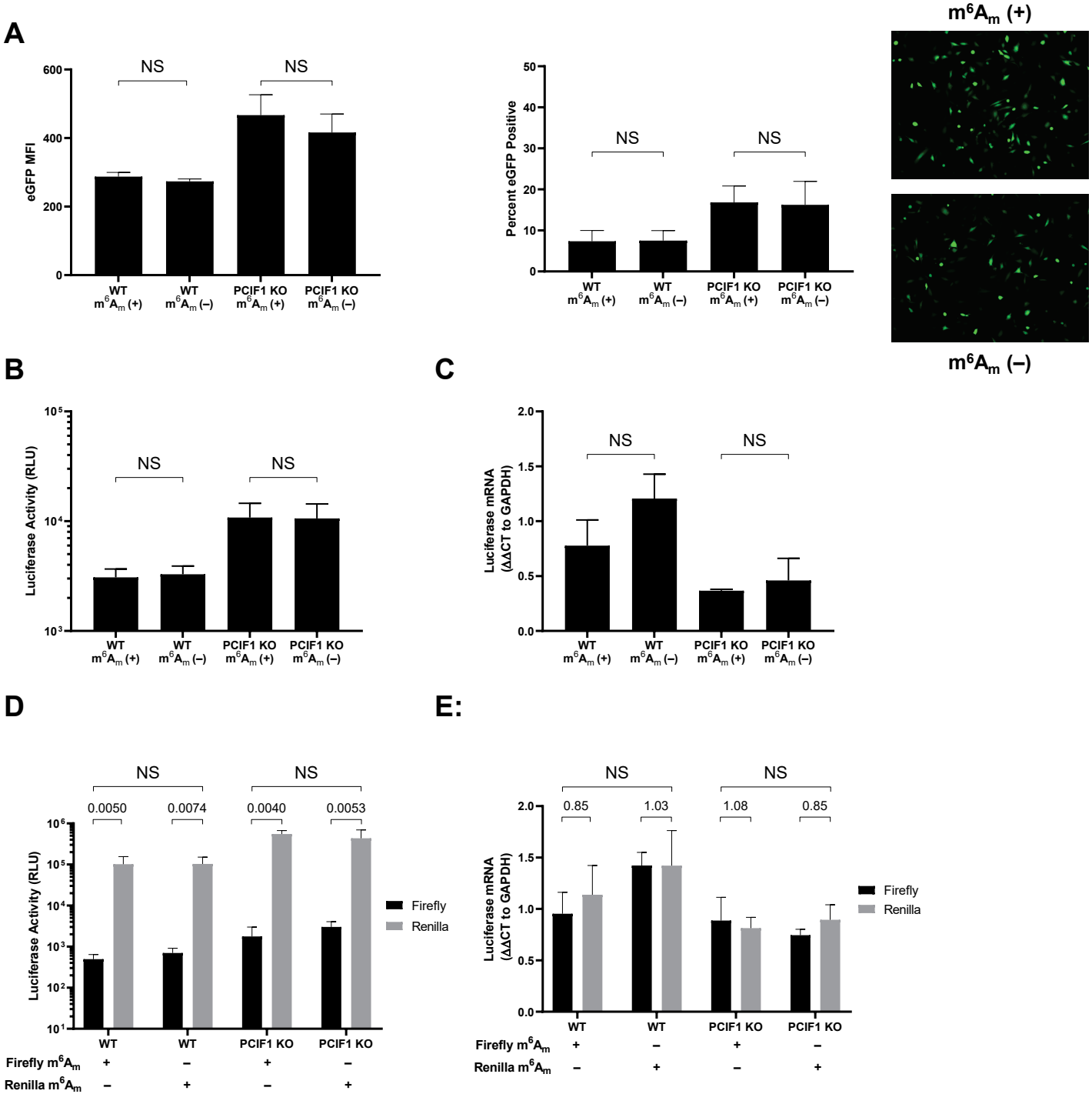


Figure A1.S6: Effect of IFN-β pretreatment of cells on viral infection in a second *PCIF1*-addback clone

PCIF1 KO HeLa cells (different single cell clone from Fig 4C, 5A) reconstituted with PCIF1 or an empty vector were pretreated with vehicle (0.1% BSA) or 500 U ml^{-1} interferon-β for 5h. Treatment media was removed from the cells, followed by infection with VSV WT at a MOI of 3. After 1 hour, the inoculum was removed, cells washed, and initial treatment media added back to cells. At 2, 4, 6, 8, and 11 hpi, 1% of the supernatant was removed, and plaque assays performed on Vero cells to determine the titer of VSV in each sample. Growth curve of supernatant virus (n=3, +/-SD. NS – $p>0.51$, ** - $p<0.01$, student's t-test. Statistics shown are for the 11h timepoint).

Figure A1.S6 (continued):

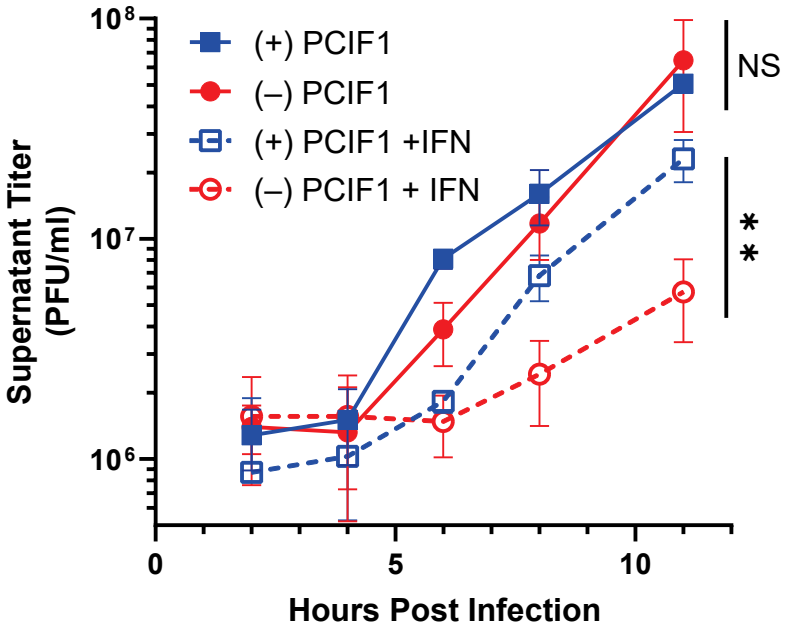


Figure A1.S7: Effect of IFN-β pretreatment of cells on viral infection in multiple single-cell clones of PCIF1 KO cells.

(A): PCIF1 KO HeLa cells reconstituted with PCIF1 or an empty vector were pretreated with IFN-β for 5h, then infected with VSV expressing a luciferase reporter (VSV-Luc) at a MOI of 3. Cells were lysed at 6 hpi. Half the lysate was used to measure luciferase using a Promega Luciferase Assay kit (n=4, +/-SD, NS – $p>0.80$, ** – $p<0.01$, student's t-test). RNA was extracted from the other half in Trizol and RT-PCR performed using oligo-dT, followed by qPCR for luciferase mRNA (normalized to GAPDH, n=4, +/-SD, NS – $p>0.05$, * – $p<0.05$, student's t-test). **(B):** As in A, except cells were transfected with 500 ng ribonucleoprotein cores of VSV-Luc. (n=4, +/-SD, NS – $p>0.19$, * – $p<0.05$, student's t-test). **(C):** Wild type (parental) HeLa or PCIF1 KO cells (two independent clones) were pretreated with IFN-β for 5h, then infected with VSV expressing a luciferase reporter (VSV-Luc) at a MOI of 3. Luciferase protein was measured as in A (n=4, NS – $p>0.20$, * - $p<0.05$, ** - $p<0.01$, student's t-test). Luciferase mRNA was also measured as in A (normalized to GAPDH, n=4, NS – $p>0.18$, * - $p<0.05$, ** - $p<0.01$, student's t-test).

Figure A1.S7 (continued):

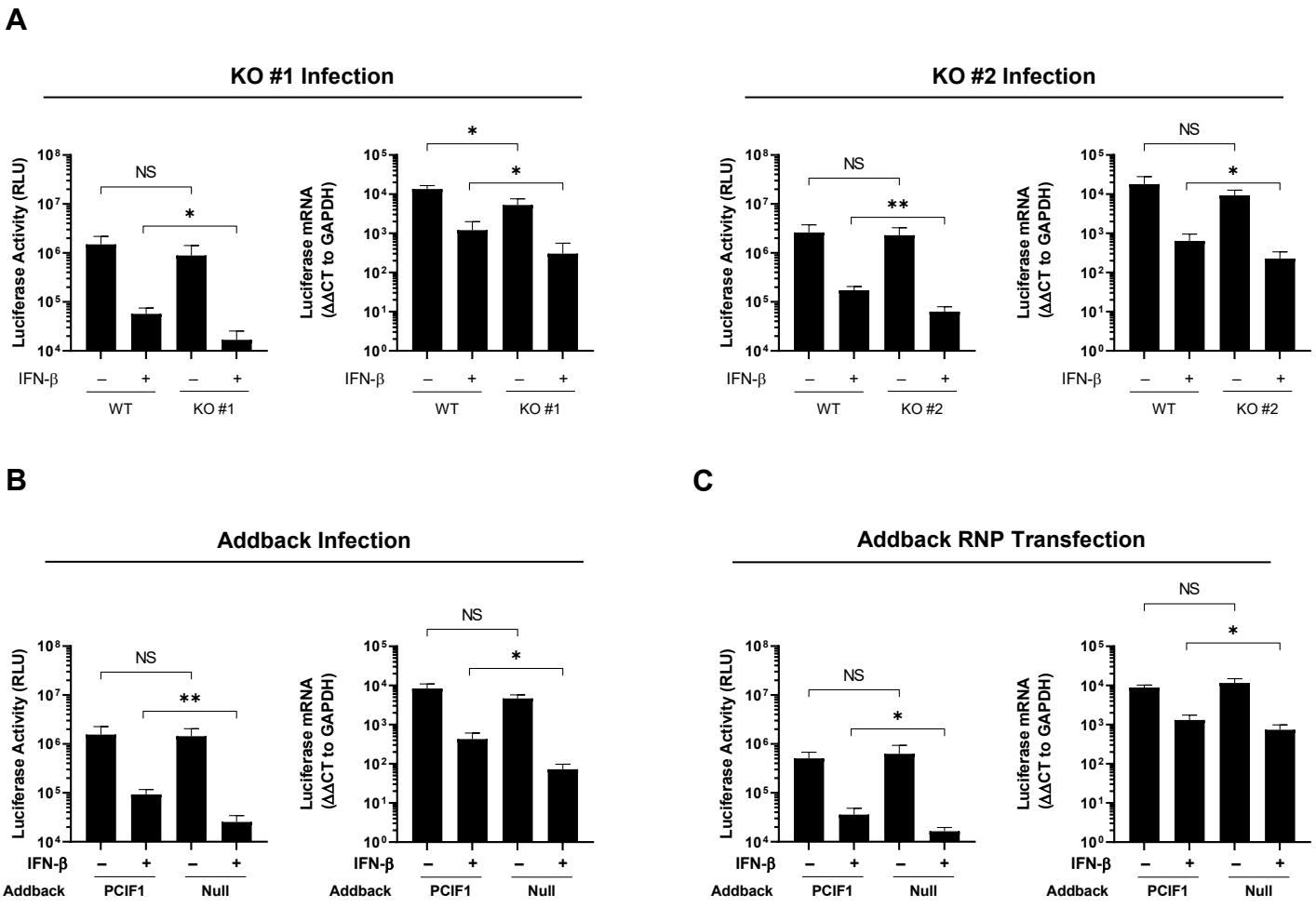


Figure A1.S8: VSV mRNA in A549 cells contain m^6A_m

A549 WT or PCIF1 KO cells were infected with VSV at a MOI of 5, and viral mRNA specifically radiolabeled, extracted, and digested as in Fig 2A. Released nucleotide monophosphates were resolved by 2D-TLC and detected by phosphorimaging (representative images, n=3) **(A)** VSV mRNA from wild type cells digested with the indicated enzymes. **(B)** VSV mRNA from PCIF1 KO cells.

Figure A1.S8 (continued):

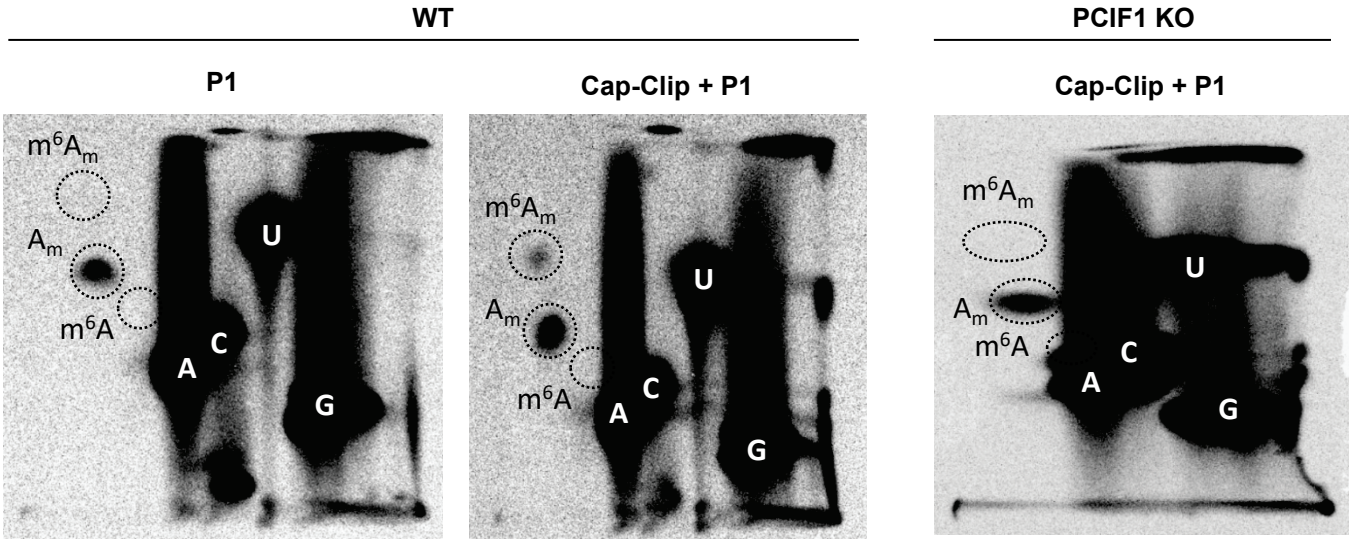


Figure A1.S9: Effect of IFN-β pretreatment and PCIF1 knockout on rates of viral translation

(A): The indicated A549 cells were pretreated with vehicle (0.1% BSA) or 500 U ml^{-1} IFN-β for 5 hours prior to infection with VSV at a MOI of 5 for 6 hours. Rates of viral translation were assessed by metabolic incorporation of [^{35}S]-met and [^{35}S]-cys into viral proteins for 20 minutes as described in methods. Proteins were analyzed by SDS-PAGE and visualized by phosphorimager (representative image, n=3). **(B):** Quantification of VSV M protein in A (n=3, +/- SD, * – $p < 0.05$, NS – $p > 0.58$). **(C):** Gels used for PAGE-analysis of radiolabeled products in A were stained with 0.25% Coomassie Brilliant Blue R-250 in 10% acetic acid, followed by destaining in 10% acetic acid. Gels were visualized by OD laser scanning on a GE Typhoon 5. Staining indicated even protein loading (representative image shown corresponding to autoradiogram in Fig 5D, n=3).

Figure A1.S9 (continued):

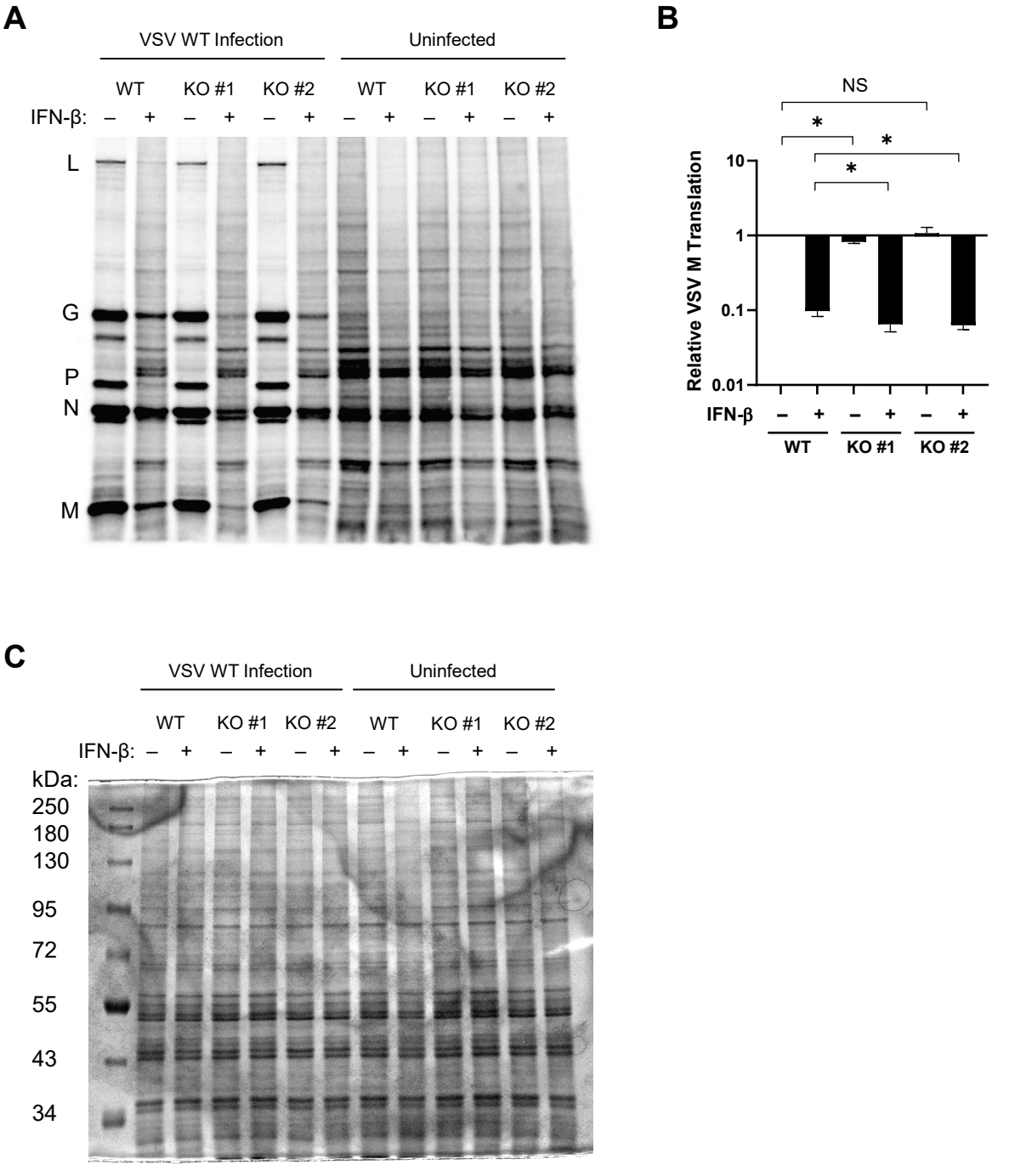


Figure A1.S10: PCIF1 does not regulate VSV mRNA sensing late in infection

The indicated A549 cells were pretreated with vehicle (0.1% BSA) or 500 U ml^{-1} IFN-β for 5 hours prior to infection with VSV at a MOI of 5 for 6 hours. Cells were lysed and IFN-β mRNA levels determined by qRT-PCR (n=4, +/- SD, NS – $0.08 < p < 0.80$, student's t-test).

Figure A1.S10 (continued):

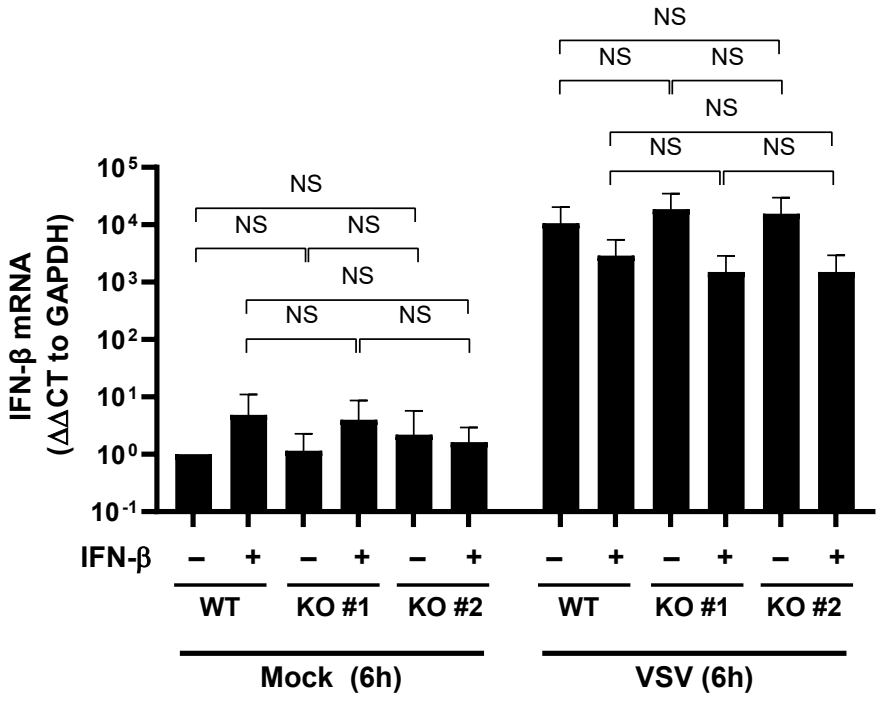
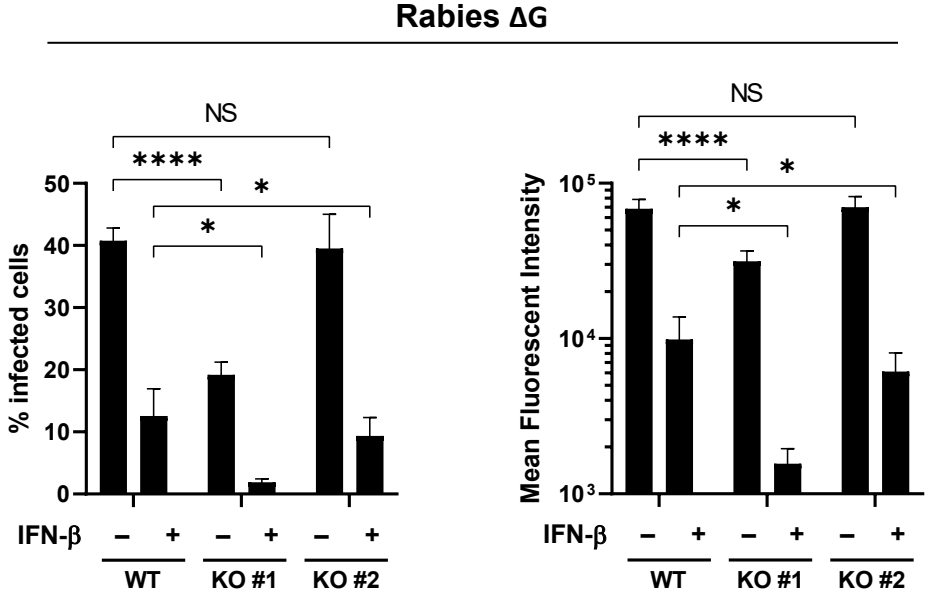


Figure A1.S11: Effect of PCIF1 on viral-encoded eGFP expression

(A): The indicated A549 cells were pretreated with vehicle (0.1% BSA) or 500 U ml^{-1} IFN-β concurrent with infection with RABV ΔG at a MOI of 1 for 24 hours. Cells were fixed in formaldehyde and flow cytometry performed for eGFP expression. Percent infected cells were defined as GFP positive, measured by gating on uninfected cells. Mean fluorescent intensity was measured based on the total population (n=3, +/- SD, NS – $0.56 < p < 0.58$, * – $p < 0.05$, ** – $p < 0.01$, **** – $p < 0.0001$, student's t-test). **(B):** The indicated A549 cells were pretreated with vehicle (0.1% BSA) or 500 U ml^{-1} IFN-β for 5 hours prior to infection with measles virus at a MOI of 3 for 24 hours. Cells were fixed in formaldehyde and flow cytometry performed for eGFP expression. Percent infected cells were defined as GFP positive, measured by gating on uninfected cells. Mean fluorescent intensity was measured based on the total population (n=3, +/- SD, NS – $0.09 < p < 0.46$, student's t-test).

Figure A1.S11 (continued):

A



B

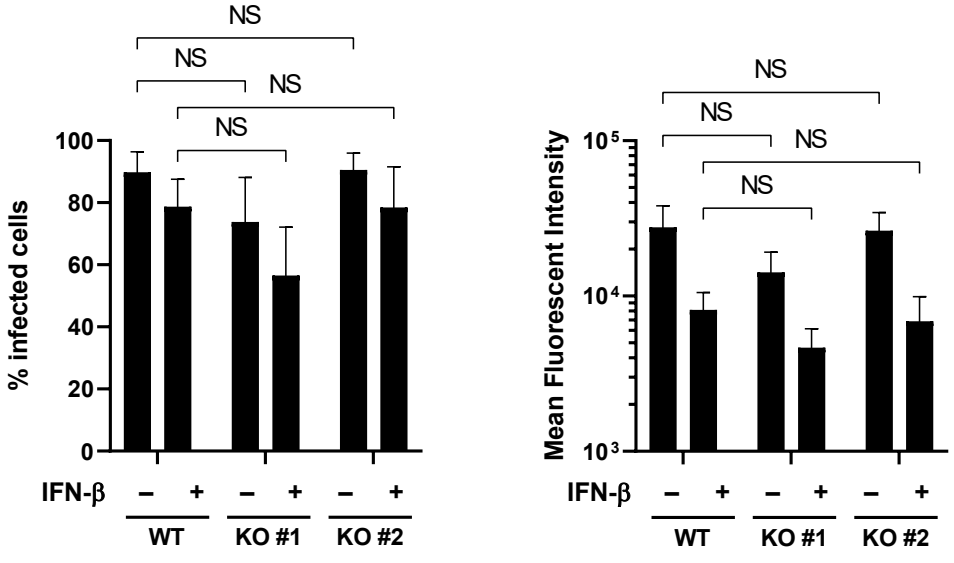


Table A1.S1: Oligonucleotides used for qPCR or mRNA selection

Primer	Direction	Sequence
Firefly Luciferase	Forward	CAACTGCATAAGGCTATGAAGAGA
Firefly Luciferase	Reverse	ATTTGTATTCAGCCCATATCGTTT
Renilla Luciferase	Forward	GAGCATCAAGATAAGATCAAAGCA
Renilla Luciferase	Reverse	CTTCACCTTTCTCTTTGAATGGTT
eGFP	Forward	GAACCGCATCGAGCTGAA
eGFP	Reverse	TGCTTGTGCGCCATGATATAG
IFIT1	Forward	AACTTAATGCAGGAACATGACAA
IFIT1	Reverse	CTGCCAGTCTGCCCATGTG
GAPDH	Forward	AGCCTCAAGATCATCAGCAAT
GAPDH	Reverse	ATGGACTGTGGTCATGAGTCCTT
Biotin-T(16)-VSVstop		5'biotin-TTTTTTTTTTTTTTTTCATA

Appendix 2:

Supplemental Material to Chapter 3

Data within this chapter is published and a reprint of the article is provided with the archived copy of the dissertation. The published manuscript may also be accessed as follows:

PNAS. Oct 26, 2021. Vol. 118. No. 43, e2108728118. doi: 0.1073/pnas.2108728118.

© 2021 by the National Academy of Sciences of the USA

Supplementary Material for

A novel class of TMPRSS2 inhibitors potently block SARS-CoV-2 and MERS-CoV viral entry and protect human epithelial lung cells

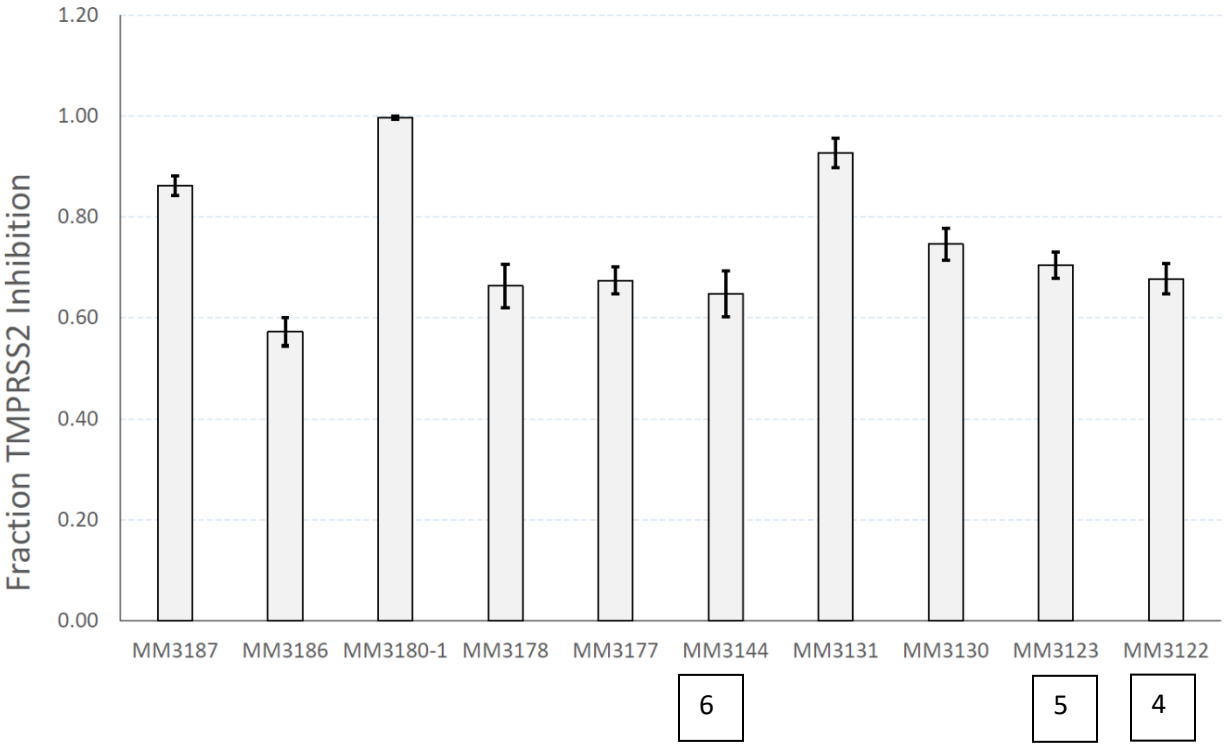
Matthew Mahoney[1, 2], Vishnu C. Damalanka[1], Michael A. Tartell[3, 4], Dong hee Chung[5], André Luiz Lourenço[5], Dustin Pwee[6], Anne E. Mayer Bridwell[3], Markus Hoffmann[7, 9], Jorine Voss[1], Partha Karmakar[1], Nurit P. Azouz[8], Andrea M. Klingler[8], Paul W. Rothlauf[3, 4], Cassandra E. Thompson[3], Melody Lee[5], Lidija Klampfer[2], Christina Stallings[3], Marc E. Rothenberg[8], Stefan Pöhlmann[7, 9], Sean P. Whelan[3], Anthony J. O'Donoghue[6], Charles S. Craik[5], James W. Janetka[1, 2]

Table of Contents

1. Single point inhibition data using recombinant protease domain/FRET substrate for **4-6**. [Figure S1]
2. Protease selectivity data for **1** (ZFH7116) and **2** (VD2173). [Tables S1 and S2]
3. Synthesis and NMR and HPLC-MS spectra of new compounds **2**, **4-7**, **19-21**. [Schemes S1-S3]
4. K_m curve for Boc-QAR-AMC using full-length TMPRSS2 (3 nM). [Figure S2]
5. IC_{50} inhibition curves of full-length TMPRSS2/Boc-QAR-AMC. [Figure S3]
6. Cell-based enzyme activity of compounds **1** and **2** in HEK-293 cells. [Figure S4]
7. Acute toxicity of MM3122 (**4**) data. [Figure S5]
8. Activity of **1** and **2** and Camostat using Vero cells in pseudotype [Figure S6] and chimeric VSV-SARS-CoV-2 viruses [Figure S7].
9. Expression and purification of TMPRSS2 protease domain. [Figure S8]

1. Single point inhibition data using recombinant protease domain and FRET peptide.

Figure A2.S1: Compounds (300 µM) were tested against the protease domain of Human Recombinant TMPRSS2 (3 nM) having (MCA)K-KARSAFA-K(Dnp) as substrate (1nM)



2. Enzyme selectivity data of ZFH7116 (1), VD2173 (2), Camostat and Nafamostat [3] from a panel 43 serine and cysteine proteases.

Table A2.S1. Protease selectivity of ZFH7116 (1) and VD2173 (2). NA = not active; all data provided in supplementary

Target	VD2173 IC50 (M)	ZFH7116 IC50 (M)	Camostat IC50 (M)	Nafamostat IC50 (M)
Cathepsin B	2.24E-07	NA	NA	NA
Cathepsin H	1.04E-05	NA	NA	NA
Cathepsin L	NA	7.14E-06	NA	NA
Cathepsin S	1.99E-06	**8.95E-08**	>1.00E-05	NA
Cathepsin V	NA	1.00E-05	NA	NA
Chymotrypsin	NA	>2.00E-05	NA	NA
FVIIa	6.15E-06	2.54E-06	3.55E-06	2.75E-07
FXa	4.89E-07	**8.66E-08**	9.91E-06	1.11E-06
FXIa	**7.89E-09**	**1.60E-08**	3.46E-09	8.56E-10
Kallikrein 1	1.98E-05	2.28E-07	>1.00E-05	2.39E-06
Kallikrein 5	3.27E-07	**8.07E-08**	1.01E-06	6.37E-07
Kallikrein 12	5.85E-06	5.60E-07	1.31E-06	3.59E-07
Kallikrein 13	1.47E-06	5.99E-07	8.45E-07	3.02E-07
Kallikrein 14	**4.86E-08**	**1.95E-08**	9.99E-07	2.15E-09
Matriptase 2	**<1.02E-09**	**<1.02E-09**	7.80E-09	<5.08E-10
Papain	**2.67E-08**	1.97E-05	NA	NA
Plasma Kallikrein	**1.39E-08**	**2.34E-09**	8.36E-10	<5.08E-10
Plasmin	1.73E-06	1.34E-07	5.62E-09	1.04E-09
Proteinase A	7.59E-06	1.88E-06	NA	NA
Proteinase K	**2.21E-08**	**5.19E-09**	NA	NA
Thrombin a	2.14E-07	1.33E-06	3.62E-06	3.03E-07
Trypsin	**<1.02E-09**	**<1.02E-09**	5.24E-10	<5.08E-10
Tryptase b2	**1.10E-09**	**<1.02E-09**	<5.08E-10	<5.08E-10
Tryptase g1	**2.34E-09**	**<1.02E-09**	<5.08E-10	<5.08E-10
Urokinase	**5.01E-09**	7.82E-06	1.64E-08	<5.08E-10

Table A2.S2. Full protease selectivity data of ZFH7116 (1) and VD2173 (2).

Report of Protease Profiling for:	Washington Univ. St. Louis	Quotation #	20200430-WUSM-JJ-Pro-RV02

Two compounds were received as powder stock and resuspended to 10 mM in DMSO.

Compound ID	Concentration (mM)	Original Stock Volume (uL)	Exact weight (mg)	MW
VD2173	10	382	3.0	785.84
ZFH7116	10	349	3.2	918.00

The compound was tested in a 10-dose IC50 with a 3-fold serial dilution starting at 20 uM against 43 proteases.

Control compounds were tested in a 10-dose IC50 with 3-fold serial dilution starting at 10 uM*.
(*Start at different concentrations for some enzymes)

Compound fluorescence : Compounds exhibit no fluorescent background that could interfere with the assay.

The protease activities were monitored as a time-course measurement of the increase in fluorescence signal from fluorescently-labeled peptide substrate, and initial linear portion of slope (signal/min) was analyzed.

Data pages include slope, % Enzyme Activity (No inhibitor control as 100% Activity), curve fit, and IC50.

The obtained IC50 values are summarized in the table below.
(Curve fits were performed when the activities at the highest concentration of compounds were less than 65%.)

Summary Table:

		Compound IC50 (M)			
	Target:	VD2173	ZFH7116	Control Compound IC50 (M)	Control compound ID
1	Calpain 1			4.71E-08	E64
2	Caspase 1			2.92E-08	IETD-CHO
3	Caspase 2			5.47E-07	IETD-CHO
4	Caspase 3			1.07E-09	DEVD-CHO
5	Caspase 4	>2.00E-05		2.13E-06	IETD-CHO
6	Caspase 5			1.15E-08	IETD-CHO
7	Caspase 6		>2.00E-05	1.24E-08	DEVD-CHO
8	Caspase 7			2.31E-09	DEVD-CHO
9	Caspase 8			2.83E-09	IETD-CHO
10	Caspase 9			2.64E-08	IETD-CHO
11	Caspase 10			9.17E-09	IETD-CHO
12	Caspase 11			5.67E-07	IETD-CHO
13	Caspase 14			6.82E-08	WEHD-CHO
14	Cathepsin B	2.24E-07		7.33E-09	E64
15	Cathepsin C			1.18E-06	E64
16	Cathepsin G			4.04E-06	Chymostatin
17	Cathepsin H	1.04E-05		3.56E-08	E64
18	Cathepsin L		7.14E-06	1.24E-08	E64
19	Cathepsin S	1.99E-06	8.95E-08	1.73E-09	E64
20	Cathepsin V		1.00E-05	3.89E-09	E64
21	Chymase			1.01E-08	Chymostatin
22	Chymotrypsin		>2.00E-05	1.41E-09	Chymostatin
23	Elastase			4.83E-09	Sivelestat
24	FVIIa	6.15E-06	2.54E-06	5.75E-08	PCI 27483
25	FXa	4.89E-07	8.66E-08	1.86E-06	Gabexate mesylate (GM)
26	FXIa	7.89E-09	1.60E-08	2.39E-07	Gabexate mesylate (GM)
27	Kallikrein 1	1.98E-05	2.28E-07	4.61E-06	Leupeptin
28	Kallikrein 5	3.27E-07	8.07E-08	4.94E-06	Gabexate mesylate (GM)
29	Kallikrein 7			4.31E-05	Gabexate mesylate (GM)
30	Kallikrein 12	5.85E-06	5.60E-07	9.43E-08	Gabexate mesylate (GM)
31	Kallikrein 13	1.47E-06	5.99E-07	1.22E-05	Gabexate mesylate (GM)
32	Kallikrein 14	4.86E-08	1.95E-08	6.51E-07	Gabexate mesylate (GM)
33	Matriptase 2	<1.02E-09	<1.02E-09	3.91E-07	Gabexate mesylate (GM)
34	Papain	2.67E-08	1.97E-05	2.40E-10	E64
35	Plasma Kallikrein	1.39E-08	2.34E-09	1.48E-07	Gabexate mesylate (GM)
36	Plasmin	1.73E-06	1.34E-07	2.65E-07	Gabexate mesylate (GM)
37	Proteinase A	7.59E-06	1.88E-06	2.44E-04	Leupeptin
38	Proteinase K	2.21E-08	5.19E-09	4.51E-08	Proteinase K inhibitor
39	Thrombin a	2.14E-07	1.33E-06	1.57E-06	Gabexate mesylate (GM)
40	Trypsin	<1.02E-09	<1.02E-09	3.07E-08	Gabexate mesylate (GM)
41	Tryptase b2	1.10E-09	<1.02E-09	9.70E-09	Gabexate mesylate (GM)
42	Tryptase g1	2.34E-09	<1.02E-09	1.33E-08	Gabexate mesylate (GM)
43	Urokinase	5.01E-09	7.82E-06	2.44E-08	Gabexate mesylate (GM)

* Empty cells indicate no inhibition or compound activity that could not be fit to an IC50 curve	* IC50 value higher than 2.00E-05 M is estimated based on the best curve	* IC50 value lower than 1.02E-09 M is estimated based on the best curve

Report of Protease Profiling for:	Washington Univ. St. Louis	Quotation #	20200604-WUSM-JJ-PRO

2 compounds were located at RBC as 10 mM DMSO stock.

Compound ID	Concentration (mM)	Original Stock Volume (uL)	Exact weight (mg)	MW
VD2173	10	382	3.0	785.84
ZFH7116	10	349	3.2	918.00

The compounds were tested in a 10-dose IC50 with a 3-fold serial dilution starting at 20 uM against 1 protease.

Control compounds were tested in a 10-dose IC50 with 3-fold serial dilution starting at 1 uM.
(*Start at different concentrations for some enzymes)

Compound fluorescence : Compounds exhibit no fluorescent background that could interfere with the assay.

The protease activities were monitored as a time-course measurement of the increase in fluorescence signal
from fluorescently-labeled peptide substrate, and initial linear portion of slope (signal/min) was analyzed.

Data pages include slope, % Enzyme Activity (No inhibitor control as 100% Activity), curve fit, and IC50.

The obtained IC50 values are summarized in the table below.
(Curve fits were performed when the activities at the highest concentration of compounds were less than 65%)

Summary Table:

		Compound IC50 (M)			
	Target:	VD2173	ZFH7116	Control Compound IC50 (M)	Control compound ID
1	Furin			1.10E-09	Furin Inhibitor 1

* Empty cells indicate no inhibition or compound activity that could not be fit to an IC50 curve

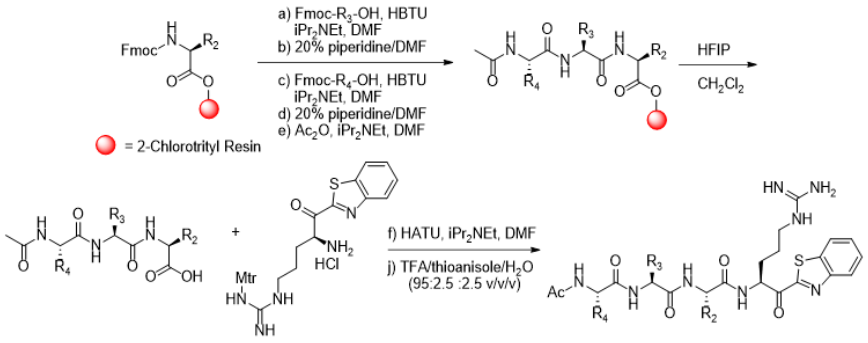
3. Synthesis and NMR and HPLC-MS spectra of new compounds 2, 4-7, 19-21.

Synthesis of peptidyl ketobenzothiazoles (kbts), 4-7

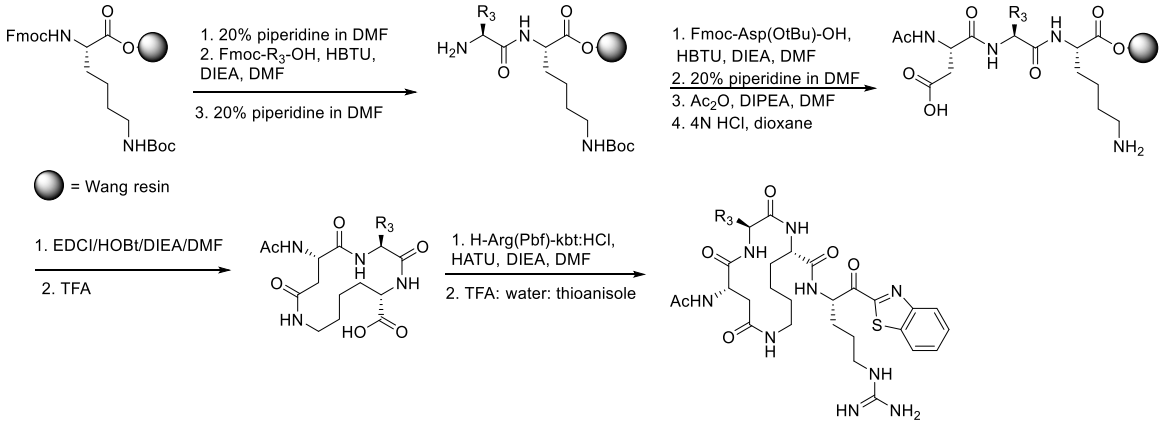
We have previously published methodology for the synthesis of the kbt peptide inhibitors [1]. As shown in **Scheme 1**, we first construct the tripeptide (e.g. **6**) or tetrapeptide (e.g. **4**) on the 2-chlorotrityl resin using standard Fmoc-solid phase peptide synthesis protocols including HBTU for amide bond coupling and 20% piperidine for Fmoc deprotection steps. The peptide is then capped with an acetyl group using acetic anhydride. Cleavage of the 2-chlorotrityl resin without removal of the protecting groups is accomplished using HFIP, then H-Arg(Mtr)-kbt is installed with HATU or EDC/HOBt in DMF. Final deprotection of the amino acid sidechains using TFA:water:thioanisole (95:2.5:2.5 %v/v) generates the target compounds which are purified by reverse phase prep HPLC.

A2.

Scheme S1. Synthesis of acyclic P4-P1 tetrapeptide ketobenzothiazoles (kbts), 4-7.



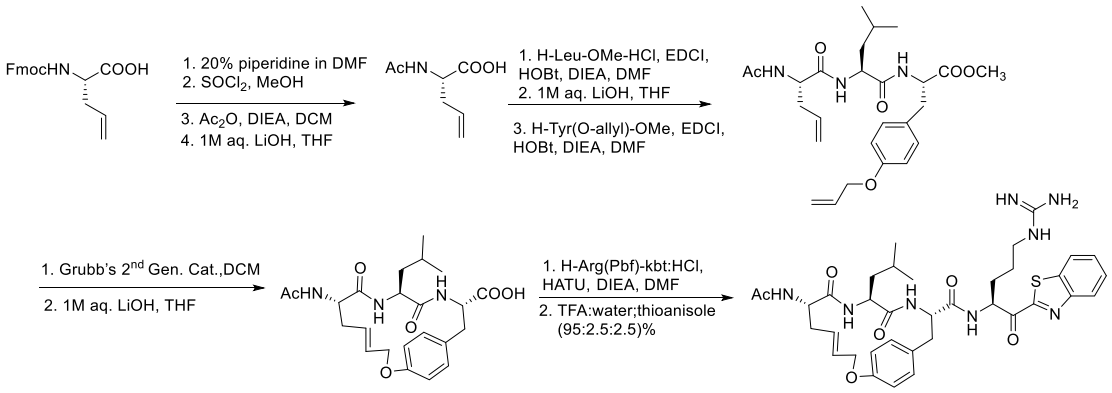
Scheme A2.S2. Cycloamide synthesis (2, 19 and 20)



Cyclic peptide VD2173 (**2**) was synthesized following the generic procedure outlined in **Scheme 2** in which is also used for the construction of **19** and **20**. Synthesis of the tripeptide cyclization precursors are synthesized on Wang resin using standard Fmoc solid phase peptide synthesis (SPPS) using 2HBTU for coupling steps and 20% piperidine for Fmoc removal steps. After the final Fmoc deprotection, the tripeptide is acetylated and then the Asp and Lys protecting groups are removed with 4N HCl. Cyclization is performed using EDCI and HOBt on the resin followed by TFA cleavage, then a final HATU coupling

with H-Arg-(Pbf)-kbt followed by Arg deprotection to give the cyclic peptides which are purified by reverse phase prep HPLC.

Scheme A2.S3. Synthesis of cyclic peptide aryl ether 21



Compound **21** was prepared from Fmoc-allylglycine as shown in **Scheme 3**. N-Deprotection followed by esterification and acetylation gives Ac-allylglycine which is then coupled to H-Leu-OMe using EDC/HOBt. The resulting dipeptide ester is hydrolyzed with LiOH and then coupled to H-Tyr(Oallyl)-OMe once again with EDC/HOBt to yield the cyclization precursor. Olefin metathesis cyclization is accomplished with Grubbs 2nd generation catalyst to give key aryl allyl ether cyclic peptide intermediate. Ester hydrolysis followed by amide coupling to H-Arg(Pbf)-kbt and final deprotection with TFA as before gives **21** which is purified by prep HPLC.

General synthesis, purification, and analytical chemistry procedures. Starting materials, reagents, and solvents were purchased from commercial vendors unless otherwise noted. ^{1}H NMR spectra were measured on a Varian 400 MHz NMR instrument. The chemical shifts were reported as δ ppm relative to TMS using residual solvent peak as the reference unless otherwise noted. The following abbreviations were used to express the multiplicities: s = singlet; d = doublet; t = triplet; q = quartet; m = multiplet; br = broad. High-performed liquid chromatography (HPLC) was carried out on GILSON GX-281 using Waters C18 5μM, 4.6*50mm and Waters Prep C18 5μM, 19*150mm reverse phase columns, eluted with a gradient system of 5:95 to 95:5 acetonitrile:water with a buffer consisting of 0.05% TFA. Mass spectra (MS) were performed on HPLC/MSD using electrospray ionization (ESI) for detection. All reactions were monitored by thin layer chromatography (TLC) carried out on Merck silica gel plates (0.25 mm thick, 60F254), visualized by using UV (254 nm) or dyes such as $KMnO_4$, *p*-Anisaldehyde and CAMA (Cerium Ammonium Molybdate or Hanessian's Stain). Silica gel chromatography was carried out on a Teledyne ISCO CombiFlash purification system using pre-packed silica gel columns (12g to 330g sizes). All compounds used for biological assays are greater than 95% purity based on NMR and HPLC by absorbance at 220 nm and 254 nm wavelengths.

General procedure for synthesis of acyclic peptides[1]

Solid phase peptide coupling and deprotection: Into a reaction vial (with a fritted glass filter) under nitrogen containing H-Leu-2-Cl trityl or H-Phe-2-Cl trityl resin (0.714 g, 0.5 mmol) was added DMF/CH_2Cl_2 (15/15 mL). The mixture was shaken at RT for 30 min and then filtered. The resin was washed with DMF (10 mL) 2 times. A mixture of Fmoc-AA-OH (2.5 mmol) in DMF (20 mL), HBTU (0.853 g, 2.25 mmol) and $^{i}Pr_2NEt$ (0.87 mL, 5 mmol) was stirred at RT for 10 min and then added to the resin. The resultant heterogeneous mixture was shaken at RT overnight and then filtered. The resin was washed with DMF (20 mL x 4), dried and then piperidine/DMF (20% v/v, 30 mL) was added. The mixture was shaken for 1-4 h at RT, then filtered and was washed with DMF (10 mL x 4). Following the Fmoc deprotection of the dipeptide, the dipeptide is carried on to the next step or coupling of another Fmoc-AA-OH is performed in an identical fashion as described above and then subsequently a final Fmoc deprotection to the tripeptide.

Acetyl capping and cleavage from resin: The peptide-containing resin was suspended in 30 mL of 0.5 M Ac_2O and 1 M $^{i}Pr_2NEt$ in DMF and shaken at RT for 1 h. The reaction was filtered, and resin washed with DMF (10 mL x 4) followed by CH_2Cl_2 (10 mL x 4). The resin was then suspended in 30 ml of 25% v/v HFIP/ CH_2Cl_2 and shaken for 1 h. The reaction was filtered, and the filtrate was concentrated and dried *in vacuo*.

Coupling of Arg (Pbf)-kbt:HCl and final deprotection. To crude peptide acid (400 mg, 1.0 mmol) dissolved in dry DMF (10 mL) under a nitrogen atmosphere at 0 °C was added HATU (456 mg, 1.20 mmol) followed by stirring for 15 min. Next, Arg(Pbf)-kbt:HCl (638mg; 1.10 mmol) and iPr_2NEt (0.87 mL, 5.0 mmol) were added to the reaction at 0 °C. The reaction was allowed to reach room temperature and then stirred for an additional 2-3 h. DMF was removed under vacuum and water (250 mL) was added to the residue. The precipitate formed was filtered and washed with water (2 x 50 mL) then dried under vacuum. The precipitate was suspended in 10 mL TFA/thioanisole/water (95:2.5:2.5 *v/v/v*) and stirred for2 h at RT. The solvent was removed, and cold ether (100 mL) was added. The resulting precipitate was collected by centrifugation and the crude product was purified by HPLC (C_{18}, 15 x 150 mm column; eluent: acetonitrile/water (0.05% TFA) to give the final compound.

Ac-WFR-kbt, 8. ^{1}H NMR (399 MHz, DMSO-d_6) d ppm 1.48 (br. s., 1 H), 1.54 - 1.69 (m, 1 H), 1.74 (m, 4 H), 2.70 - 2.89 (m, 2 H), 2.90 - 3.22 (m, 4 H), 4.37 - 4.74 (m, 1 H), 5.41 - 5.58 (m, 1 H), 6.86 - 7.10 (m, 4 H), 7.11 - 7.36 (m, 8 H), 7.40 - 7.57 (m, 2 H), 7.62 - 7.76 (m, 2 H), 7.93 - 8.12 (m, 2 H), 8.22 - 8.34 (m, 2 H), 8.60 (t, *J*=7.59 Hz, 1 H), 10.76 (s, 1 H). ESI-MS [M+H]+ calcd for for $C_{35}H_{38}N_8O_4S$+ 667.28, found 667.5.

Ac-dWFR-kbt, 13. ^{1}H NMR (399 MHz, DMSO-d_6) d ppm 1.56 - 2.07 (m, 6 H), 1.93 - 2.04 (m, 1 H), 2.52 - 2.81 (m, 4 H), 2.96 - 3.08 (m, 1 H), 4.39 - 4.50 (m, 1 H), 4.57 - 4.70 (m, 1 H), 5.46 - 5.58 (m, 1 H), 6.90 - 7.09 (m, 4 H), 7.09 - 7.36 (m, 8 H), 7.44 - 7.59 (m, 2 H), 7.62 - 7.76 (m, 2 H), 7.99 (d, *J*=7.79 Hz, 1 H), 8.28 (t, *J*=8.37 Hz, 2 H), 8.47 (d, *J*=8.95 Hz, 1 H), 8.63 (d, *J*=6.62 Hz, 1 H), 10.74 (s, 1 H). ESI-MS [M+H]+ calcd for $C_{35}H_{38}N_8O_4S$+ 667.28, found 667.5.

Ac-QFR-kbt, 6. ^{1}H NMR (400 MHz, DMSO-d_6) d ppm 0.85 - 1.23 (m, 10 H), 1.23 - 1.49 (m, 4 H), 2.15 (dd, *J*=13.89, 8.41 Hz, 1 H), 2.30 (dd, *J*=13.69, 6.26 Hz, 1 H), 2.37 - 2.48 (m, 2 H), 3.41 - 3.52 (m, 1 H), 3.79 - 3.89 (m, 1 H), 4.86 - 4.96 (m, 1 H), 6.24 - 6.42 (m, 7 H), 6.84 (quin, *J*=7.53 Hz, 3 H), 7.33 (d, *J*=7.43 Hz, 1 H), 7.37 - 7.45 (m, 3 H), 7.73 (d, *J*=7.04 Hz, 1 H). ESI-MS [M+H]+ calcd for $C_{29}H_{36}N_8O_5S$+ 609.26, found 609.5.

Ac-IQFR-kbt, 7. H NMR (400 MHz, DMSO-d_6) d ppm 0.70 - 0.84 (m, 7 H) 1.00 - 1.14 (m, 1 H) 1.31 - 1.48 (m, 1 H) 1.54 - 1.84 (m, 7 H) 1.84 - 2.13 (m, 7 H) 2.72 - 2.84 (m, 1 H) 2.95 - 3.07 (m, 1 H) 3.10 - 3.21 (m, 2 H) 4.08 (t, *J*=6.85 Hz, 1 H) 4.12 - 4.21 (m, 1 H) 4.51 - 4.64 (m, 1 H) 5.45 - 5.55 (m, 1 H) 6.81 (br. s., 1 H)7.10 - 7.31 (m, 6 H) 7.49 - 7.57 (m, 1 H) 7.64 - 7.74 (m, 2 H) 7.95 (t, *J*=9.00 Hz, 2 H) 8.15 (d, *J*=7.83 Hz, 1 H) 8.28 (t, *J*=8.61 Hz, 2 H) 8.56 (d, *J*=5.87 Hz, 1 H). ESI-MS [M+H]+ calcd for $C_{35}H_{47}N_9O_6S$+ 722.34, found 722.6.

Ac-GQFR-kbt, 4 (MM3122). ^{1}H NMR (400 MHz, DMSO-*d*6) δ ppm 1.51 - 1.69 (m, 3 H), 1.70 - 1.83(m, 2 H), 1.87 (s, 3 H), 1.91 - 2.11 (m, 3 H), 2.71 - 2.86 (m, 1 H), 2.98 - 3.08 (m, 1 H), 3.11 - 3.22 (m, 2 H), 3.68 (d, *J*=5.48 Hz, 2 H), 4.10 - 4.23 (m, 1 H), 4.55 (br. s., 1 H), 5.50 (br. s., 1 H), 6.80 (br. s., 1 H), 7.09 - 7.32 (m, 7 H), 7.53 (br. s., 1 H), 7.63 - 7.75 (m, 2 H), 8.02 (d, *J*=8.22 Hz, 1 H), 8.09 - 8.21 (m, 2 H), 8.23 - 8.34 (m, 2 H), 8.52 (d, *J*=6.26 Hz, 1 H).). ESI-MS [M+H]+ calcd for $C_{31}H_{39}N_9O_6S$+ , 666.28, found 666.50.

Ac-PQFR-kbt, 5. ^{1}H NMR (400 MHz, DMSO-d_6) δ ppm 0.87 - 1.54 (m, 19 H), 2.09 -2.24 (m, 1 H), 2.37 - 2.49 (m, 3 H), 2.76 - 2.89 (m, 1 H), 2.96 - 3.09(m, 1 H), 3.35 - 3.57 (m, 2 H), 3.78 - 3.94 (m, 1 H), 4.90 (br. s., 1 H), 6.25 - 6.53 (m, 6 H), 6.76 - 6.90 (m, 2 H), 7.23 (d, J=7.43 Hz, 1 H), 7.33 (d, J=7.83 Hz, 1 H), 7.37 - 7.48 (m, 2 H), 8.44 (d, J=5.48 Hz, 1H).). ESI-MS [M+H]+ calcd for $C_{34}H_{43}N_9O_6S$+ 706.31, found 706.50.

H-dWFR-kbt, 14:

Boc-dWF-OMe: Boc-D-Trp-OH (1 g, 3.28 mmol) and HCl.Phe-OMe (0.708g, 3.28 mmol) was taken in dry dichloromethane (10 mL) under nitrogen atmosphere and the reaction mixture was cooled to 0 ^{o}C and N,N-diisopropylethylamine (1.7 mL, 9.84 mmol) and propylphosphonic anhydride (1.9 mL, 3.28 mmol, 50% solution in EtOAc) were added to the solution drop wise respectively. The reaction mixture was then stirred at 25 ^{o}C under nitrogen atmosphere for 1 hour and the completion of the reaction was confirmed by LC-MS monitoring. On completion the reaction mixture was diluted with 10 mL dichloromethane and washed with 10% citric acid solution, saturated sodium bicarbonate solution and brine respectively. The organic layer was dried over sodium

sulfate and concentrated under reduced pressure. The crude product was triturated with hexane to obtain the title product in pure form as a white solid. Yield: 1.3 g (92.8%). Chemical formula: $C_{23}H_{33}N_3O_5$, Exact Mass: 465.55, MS(ESI): found: $[M + Na]^+$, 488.58.

Boc-dWF-OH: Boc-WF-OMe (0.130 g, 0.279 mmol) was taken in a 1:1 mixture of THF and water and LiOH.H_2O (0.035g, 0.873 mmol) was added to it. The reaction mixture was stirred for 30 minutes at 25 ^oC and the completion of the reaction was confirmed by LCMS monitoring. On completion, the THF was evaporated under reduced pressure and the remaining water layer was cooled to 0 ^oC. The water layer was then brought to pH 6.5 by slow addition of 0.5 M HCl solution in water. The crude product precipitates out on addition of HCl and it was isolated by filtration. The crude product was dried under reduced pressure and triturated with diethyl ether to obtain the pure title product in pure form as white solid. Yield: 0.102g (80%). Exact Mass: 451.21, MS(ESI): found: $[M + H]^+$, 452.26.

Boc-dWFR(Mtr)-kbt: Boc-dWF-OH (35 mg, .077 mmol) and HATU (43.9 mg, 0.115 mmol) was taken in dry DMF under nitrogen atmosphere and the reaction mixture was cooled to 0 ^oC. N,N-diisopropylethylamine (0.04mL, 0.231 mmol) was then added drop wise to the reaction mixture and the reaction mixture was allowed to stir for 15 minutes followed by addition of HCl.Arg(Mtr)-kbt (41.75 mg, .077 mmol). The reaction mixture was stirred for 12 hours at 25 ^oC under nitrogen atmosphere and the completion of the reaction was confirmed by LC-MS monitoring. On completion, the reaction mixture was diluted with EtOAc and washed with 10% citric acid solution, saturated sodium bicarbonate solution and brine respectively. The organic layer was dried over sodium sulfate and concentrated under reduced pressure. The crude product was directly taken to the next step without further purification. Chemical formula: $C_{48}H_{56}N_8O_8S_2$, Exact Mass: 936.37, MS(ESI): found: $[M + H]^+$, 937.17.

H-dWFR-kbt (14): Boc-dWFR(Mtr)-kbt (85 mg, crude product from previous step) was taken in 5 mL TFA:thioanisole:H_2O (95:2.5:2.5) and the reaction mixture was stirred for 6 hours at 25 ^oC. The completion of the reaction was confirmed by LC-MS monitoring. On completion, the reaction mixture was concentrated under reduced pressure and triturated with diethyl ether to obtain the crude product as brown solid. The crude product was then subjected to reverse phase semi-preparative HPLC (Stationary phase: C18 column, mobile phase: H_2O-Acetonitrile with 0.1% TFA in each, 15-65% Acetonitrile in H_2O gradient for 20 minutes) to obtain the pure title product Yield: 25 mg (42% over two steps). Chemical formula: $C_{33}H_{36}N_8O_3S$, ^1H NMR (400 MHz, METHANOL-d4) δ ppm 8.21 (d, *J* = 6.65 Hz, 1H), 8.11 (d, *J* = 7.43 Hz, 1H), 7.57 - 7.67 (m, 4H), 7.12 - 7.36 (m, 10H), 7.01 - 7.08(m, 2H), 6.84 (s, 1H), 3.48 (d, *J* = 1.96 Hz, 1H), 3.13 (d, *J* = 6.65 Hz, 4H), 3.01 - 3.07 (m, 1H), 2.75 - 2.84 (m, 1H), 2.65 (s, 2H), 2.03 (s, 5H), 1.41(d, *J* = 8.61 Hz, 1H)Exact Mass: 624.26, MS (ESI): found: $[M + H]^+$, 625.5.

H-dWFR-kbt-COOH, 15: The title compound was synthesized using the same procedure as **14** using H-Arg(Mtr)-kbt-COOH:HCl[1]. Yield: 24 mg (43% over two steps). Chemical formula: $C_{34}H_{36}N_8O_5S$, ^1H NMR (400 MHz,

METHANOL-d4) δ ppm 10.58 (br. s., 0H), 8.82 (s, 1H), 8.22 - 8.32 (m, 2H), 7.70 (dd, *J* = 8.02, 16.24 Hz, 0H), 7.57 - 7.64 (m, 1H),7.37 (d, *J* = 6.26 Hz, 1H), 7.22 (d, *J* = 13.30 Hz, 0H), 7.11 - 7.18 (m, 2H), 7.02 - 7.10 (m, 1H), 5.55 - 5.67 (m, 1H), 4.47 - 4.58 (m, 0H), 4.35 (dd, *J*= 5.09, 10.17 Hz, 0H), 4.05 - 4.31 (m, 2H), 3.37 - 3.55 (m, 1H), 3.07 - 3.27 (m, 3H), 2.65 (s, 1H), 2.18 (dd, *J* = 6.06, 13.11 Hz,1H), 1.57 - 1.95 (m,3H), 1.16 - 1.52 (m, 2H), 1.06 (td, *J* = 7.14, 13.89 Hz, 1H), 0.91 - 1.00 (m, 2H), 0.65 - 0.81 (m, 6H) Exact Mass: 668.253, MS (ESI): found: $[M + H]^+$, 669.5.

Ac-Cyclo(DLK)-R- ketobenzothiazole (2, VD2173). Into a reaction vessel (with fritted glass for resin support) containing Fmoc-L-Lys(Boc) Wang resin (5 g, 1.7 mmol), DCM (40 mL) was added. The mixture was shaken at RT for 15 min and then filtered. To the dry resin was added piperidine/DMF (20% *v/v*, 40 mL) and the mixture was shaken for 30 min at RT, then filtered. The resin was washed with DMF (2 x 30mL) and DCM (2 x 30 mL). Fmoc-Leu-OH (1.8 g, 5.1 mmol), HBTU (2.25 g, 5.95 mmol), *i*Pr_2NEt (1.31 g, 10.2 mmol), and DMF (50 mL) were added to the vessel and shaken for 12 h, then filtered. The resin was washed with DCM (2 x 40 mL) and DMF (2 x 40 mL), then piperidine/DMF (20% *v/v*, 40 mL) was added and the reaction was shaken for 30 min at RT, then filtered. The resin washed with DCM (2 x 30 mL) and DMF (2 x 30mL). Fmoc-Asp(OtBu)-OH (2.10 g, 5.1 mmol), HBTU (2.25 g, 5.95 mmol), *i*Pr_2NEt (1.31 g, 10.2 mmol), and DMF (50 mL) were added to the vessel and shaken for 12 h, then filtered. The resin was washed with DCM (2 x 40 mL) and DMF (2 x 40 mL). The peptide resin was then suspended in a solution of Ac_2O (1.04 g, 10.2 mmol), and *i*Pr_2NEt (3.07 g, 23.8 mmol) in 40 mL of DMF. The mixture was shaken at RT for 1-2 h, filtered and resin washed with DCM (2 x 40 mL) followed by DMF (2 x 40 mL). To the resin was added 40 mL of dry 4M HCl in 1, 4-dioxane followed by shaking for 30-40 min. at RT. The reaction was filtered, and the resin washed with DCM (2 x 40 mL) followed by DMF (2 x 40 mL). EDCI (0.98 g, 5.1 mmol), HOBt (0.78 g, 5.1 mmol), *i*Pr_2NEt (1.1 g, 8.5 mmol), and DMF (80 mL) were added to the resin and the resulting mixture was shaken for overnight at RT. The mixture was filtered and the resin and washed with DCM (2 x 40 mL) followed by DMF (2 x 40 mL). To the acetyl capped peptide resin was added TFA (2 x 35 mL) and shaken for 30 min. The mixture was filtered, and the resin washed with DCM (2 x 40 mL). The filtrate was concentrated, and cold ether was added to the residue yielding the crude product as a precipitate which was purified by flash chromatography to give an off-white sold (400 mg).

The macrocyclic tripeptide acid (400 mg, 1.0 mmol) was dissolved in dry DMF (10 mL) under a nitrogen atmosphere at 0 °C and HATU (456 mg, 1.20 mmol) was added followed by stirring for 15 min, and then the addition of Arg(Pbf)-kbt:HCl (638mg; 1.10 mmol) and *i*Pr_2NEt (0.87 mL, 5.0 mmol) at 0 °C. The reaction is allowed to reach RT and then stirred for 2-3 h. The DMF was removed in vacuo and water (250 mL) was added to the resulting residue. The precipitate formed was filtered and washed with water (2 x 50 mL) and dried. To this precipitate was added 10 mL of TFA/thioanisole/water (95:2.5:2.5 *v/v/v*) and the mixture was stirred for 2

h at RT. The solvent was removed, and then cold ether (100 mL) was added. The crude product was collected by centrifugation. The crude product was purified by HPLC (C_{18}, 15 x 150 mm column; eluent: acetonitrile/water (0.05% TFA) to give the title compound as a white solid. Overall yield (20%).1H NMR (400MHz ,DMSO-d_6) δ ppm = 8.51 (d, *J* = 6.7 Hz, 1 H), 8.26 (dd, *J* = 8.0, 15.1 Hz, 1 H), 7.98 (d, *J* = 7.4 Hz, 1 H), 7.93 - 7.83 (m, 2 H), 7.73 - 7.63 (m, 2 H), 7.53 (br. s., 1 H), 5.44 - 5.33 (m, 1 H), 4.60 - 4.48 (m, 1 H), 4.29 - 4.18 (m, 1 H), 3.42 (br. s., 4 H), 3.19 - 3.06 (m, 3 H), 2.96 (br. s., 1 H), 1.84 (s, 3 H), 1.78 - 1.69 (m, 1 H), 1.65 - 1.33 (m, 8 H), 1.23 - 1.07 (m, 2 H), 0.89 - 0.74 (m, 7 H). ESI-MS [M+H]+ calcd for $C_{31}H_{46}N_9O_6S$+ 672.33, found 672.5.

Ac-Cyclo(DQK)-R- ketobenzothiazole, 20. Synthesized like VD2173. Overall yield (30%).^{1}H NMR (400MHz, DMSO-d_6) δ ppm = 9.19 (s, 1 H), 8.98 - 8.81 (m, 2 H), 8.78 - 8.67 (m, 1 H), 8.60 - 8.50 (m, 1 H), 8.38 - 8.24 (m, 1 H), 8.24 - 8.13 (m, 1 H), 7.89 - 7.81 (m, 1 H), 7.41 - 7.32 (m, 1 H), 7.29 - 7.11 (m, 1 H), 6.21 - 5.98 (m, 2 H), 5.25 - 5.08 (m, 1 H), 4.94 - 4.82 (m, 1 H), 4.05 (br. s., 2 H), 3.79 (t, *J* = 6.1 Hz, 6 H), 3.15 – 3.04 (m, 6 H), 2.69 - 2.56 (m, 3 H), 2.44 - 2.32 (m, 3 H), 2.04 (s, 3 H), 2.25 (br. s., 5 H). ESI-MS [M+H]+ calcd for $C_{30}H_{43}N_{10}O_7S$+ 687.30, found 687.50.

Ac-Cyclo(DMK)-R- ketobenzothiazole, 19. Synthesized like VD2173. Overall yield (27%). ^{1}H NMR (400MHz, DMSO-d6) δ ppm = 8.54 (d, *J* = 6.3 Hz, 1 H), 8.31 - 8.07 (m, 2 H), 7.93 (d, *J* = 8.2 Hz, 1 H), 7.68 (br. s., 1 H), 7.49 (br. s., 1 H), 5.40 (br. s., 1 H), 4.51 (br. s., 1 H), 4.22 (br. s., 1 H), 3.15 (br. s., 6 H), 2.00 (d, *J* = 1.2 Hz, 4 H), 1.88 - 1.77 (s, 3 H), 1.58 (br. s., 3 H), 1.39 - 1.03 (m, 16 H). ESI-MS [M+H]+ calcd for $C_{30}H_{43}N_9O_6S_2$+ 690.28, found 690.40.

Ac-Cyclo(Allyl-Y)-R- ketobenzothiazole, 21. Fmoc-(L)-glycine (3.5 g, 10 mmol) stirred in 20% piperidine in DMF (20 mL) for 1 hr. Solvent was removed under reduced pressure, product triturated with DCM and hexanes (1:3), filtered the product and washed with hexanes, dried and used in the next reaction. Above material was dissolved in methanol (10 mL) and cooled the reaction to 0 °C followed by added thionyl chloride (2 mL) dropwise and stirred for 10 min and ice bath was replaced by a water bath, and the reaction mixture heated to ~50 °C for 3 hr while stirring. Removal of the solvent left a white residue which was washed with diethyl ether (100 mL) and collected by vacuum filtration to yield the amino acid methyl ester hydrochloride as a solid (1.7 g). Above ester (500 mg; 3.02 mmol) was taken in DCM (10 mL) and added DIEA (1.58 mL; 9.06 mmol) and Ac_2O (0.86 mL; 9.06 mmol) at RT and stirred for 3 hrs. Solvent was removed under reduced pressure and crude was purified by flash chromatography using EtOAc and Hexanes (1:9). A solution of ester (395 mg; 2.5 mmol) in THF (3 mL) was treated with 1M aqueous LiOH (3 mL) and the reaction mixture was stirred for 3 h at RT, and the absence of starting material was monitored by TLC. After the solvent was evaporated off, the residue was diluted with water and the pH was adjusted to ~3.0 using 5% aq. HCl. The product was extracted with ethyl acetate (2 x 50 mL) and the combined organic layer washe with brine (20 mL), dried over anhydrous Na_2SO_4, filtered off and concentrated, which is used in the next step without further purification. N-acetyl allyl glycine acid (167 mg;

1.06 mmol) in DMF (5 mL) was stirred with peptide coupling reagent EDCI/HOBt or HATU (1.3 eq) for 30 min. The reaction was cooled to 0-5 °C and charged with amino acid methyl ester hydrochloride (1.1 eq.) followed by diisopropylethylamine (3.0 eq.). After 15 min, allowed the reaction was brought to RT and stirred overnight. Solvent was removed under reduced pressure and the residue partitioned between EtOAc and 5% aq. HCl. The separated organic layer was washed with aq. 5% HCl, saturated $NaHCO_3$ solution (2x) and brine (1x) then dried over anhydrous Na_2SO_4. The crude product was purified by silica gel column chromatography using EtOAc and Hexanes (2:8). A solution of the ester (343 mg; 1.2 mmol) in THF (4 mL) was treated with 1M aqueous LiOH (4 mL). The reaction mixture was stirred for 3 h at RT, and the absence of starting material was monitored by TLC. After the solvent was evaporated off, the residue was diluted with water and the pH was adjusted to ~3.0 using 5% aq. HCl. The crude product was extracted into ethyl acetate (3 x 100 mL). The combined organic layers were washed with brine (25 mL) and dried over anhydrous Na_2SO_4. N-acetyl dipeptide acid (135 mg; 0.5 mmol) was stirred with EDCI (1.3 eq) and HOBt (1.3 eq) in DMF (3 mL) for 30 min. The reaction was cooled to 0–5 °C and H-L- O-allyl Tyr-OMe. HCl (130 mg: 0.55 mmol) followed by DIEA (3.0 eq). After 15 min, allowed the reaction to RT and stirred overnight. Solvent was removed under reduced pressure and the residue partitioned between EtOAc and 5% aq. HCl. The separated organic layer was washed with aq. 5% HCl, saturated $NaHCO_3$ solution (2x) and brine (1x) then dried over anhydrous Na_2SO_4. The crude product was purified by silica gel column chromatography using EtOAc and Hexanes (3:7).

A solution of acyclic diene precursor (150 mg, 0.3076 mmol) in DCM (280 mL, 0.2 Mol.) degassed for 30 min by purging nitrogen gas and then Grubbs 2nd generation catalyst (26 mg, 10 mol %) was added. The reaction was refluxed for 30 min and then additional Grubbs 2nd generation (13 mg, 5 mol %) was added. The reaction was refluxed for 18 h under nitrogen atmosphere. After depletion of the starting material as monitored by TLC and LCMS, the reaction was cooled to RT and quenched by adding activated charcoal (100 mg) followed by stirring for 1 h. The mixture was filtered through celite bed and washed generously with DCM. The filtrate was concentrated in vacuo and the crude product was purified by silica gel chromatography to yield an off-white solid. A solution of the macrocyclic ester (50 mg: 0.10 mmol) in MeOH (2 mL) was treated with 1M aq. LiOH (2 mL) at RT for 3 hrs, and the absence of starting material was monitored by TLC. After the solvent was evaporated off, the residue was diluted with water and the pH was adjusted to ~3.0 using 5% aq. HCl. The product was extracted with ethyl acetate (3 x 50 mL) and the combined organic layer washe with brine (20 mL), dried over anhydrous Na_2SO_4, filtered off and concentrated, which is used in the next step without further purification.

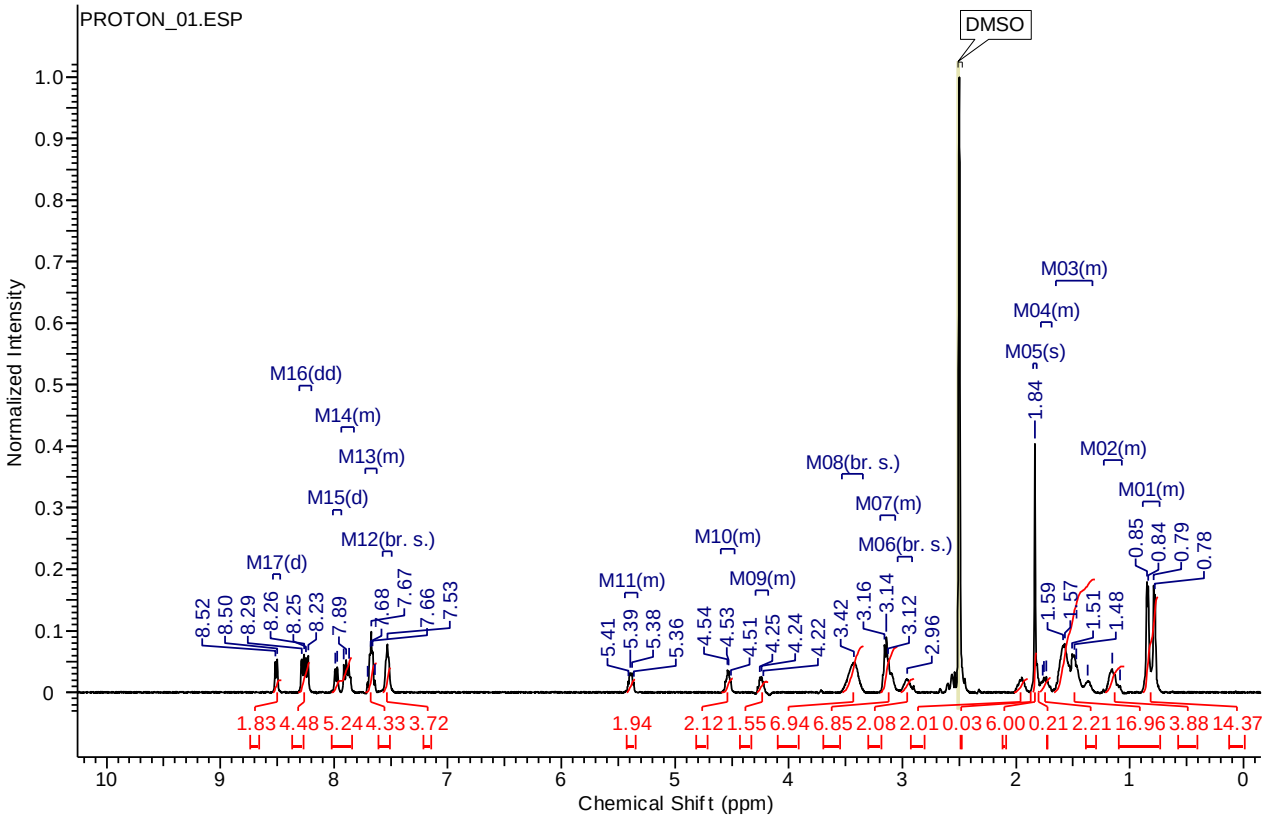
The macrocyclic acid (45 mg, 0.101 mmol) was dissolved in dry DMF (3 mL) under a nitrogen atmosphere at 0 °C and HATU (50 mg, 1.30 mmol) was added followed by stirring for 15 min, and then the addition of Arg(Pbf)-kbt:HCl (65 mg; 0.111 mmol) and *i*Pr_2NEt (70 uL, 0.404 mmol) at 0 °C. The reaction is allowed to reach RT and

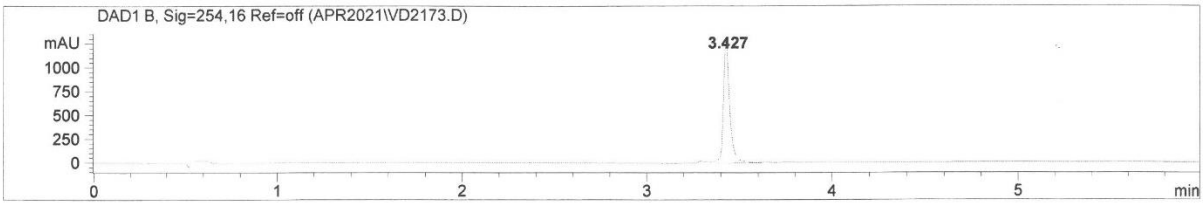
then stirred for 2-3 h. The DMF was removed in vacuo and water (250 mL) was added to the resulting residue. The precipitate formed was filtered and washed with water (2 x 50 mL) and dried. To this precipitate was added 2.5 mL of TFA/thioanisole/water (95:2.5:2.5 *v/v/v*) and the mixture was stirred for 2 h at RT. The solvent was removed, and then cold ether (35 mL) was added. The crude product was collected by centrifugation. The crude product was purified by HPLC (C_{18}, 15 x 150 mm column; eluent: acetonitrile/water (0.05% TFA) to give the title compound as a white solid. Overall yield (32%). ^{1}H NMR (400MHz, DMSO-d6) δ ppm = 8.79 (d, *J* = 7.4 Hz, 1 H), 8.33 - 8.25 (m, 2 H), 8.20 (d, *J* = 9.0 Hz, 1 H), 8.13 (d, *J* = 7.4 Hz, 1 H), 7.75 - 7.64 (m, 2 H), 7.53 (br. s., 1 H), 7.07 (d, *J* = 7.8 Hz, 2 H), 6.67 (d, *J* = 7.8 Hz, 1 H), 5.60 - 5.36 (m, 2 H), 4.72 - 4.54 (m, 2 H), 4.37 - 4.25 (m, 1 H), 4.17 (br. s., 1 H), 3.16 (d, *J* = 6.3 Hz, 2 H), 2.98 (d, *J* = 11.7 Hz, 3 H), 1.79 (s, 3 H), 1.64 (br. s., 4 H), 1.41 - 1.32 (m, 3 H), 1.24 - 1.14 (m, 4 H), 0.83 - 0.72 (m, 10 H).

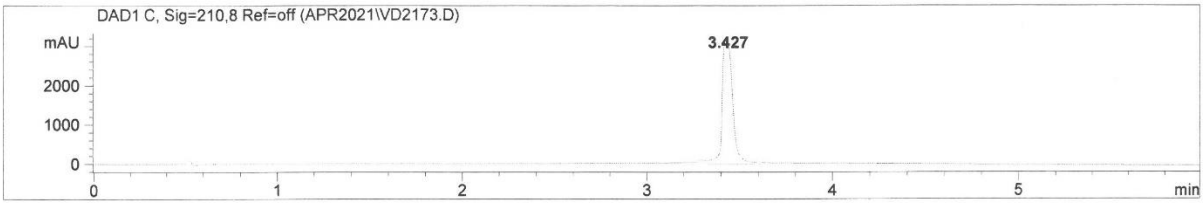
The synthesis of **9**, **10**, **11**, and **18** have been previously reported.[2]

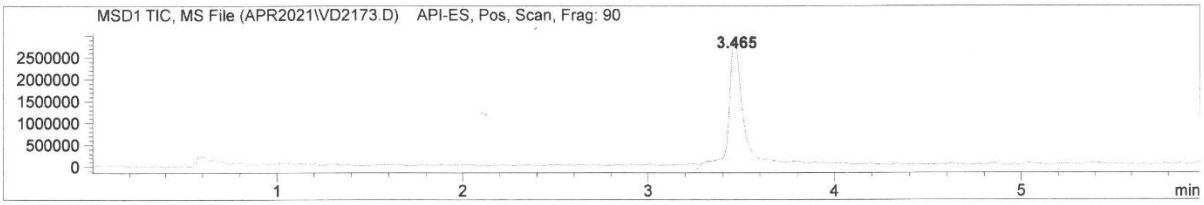
The synthesis of **1**, **3**, **12**, **16**, and **17** are as previously described.[1]

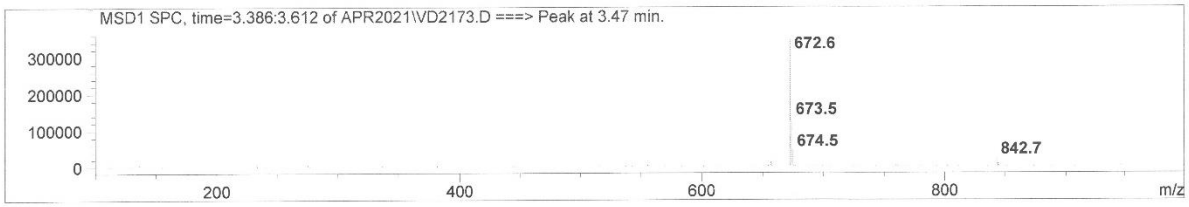
VD2173 (Compound 2)



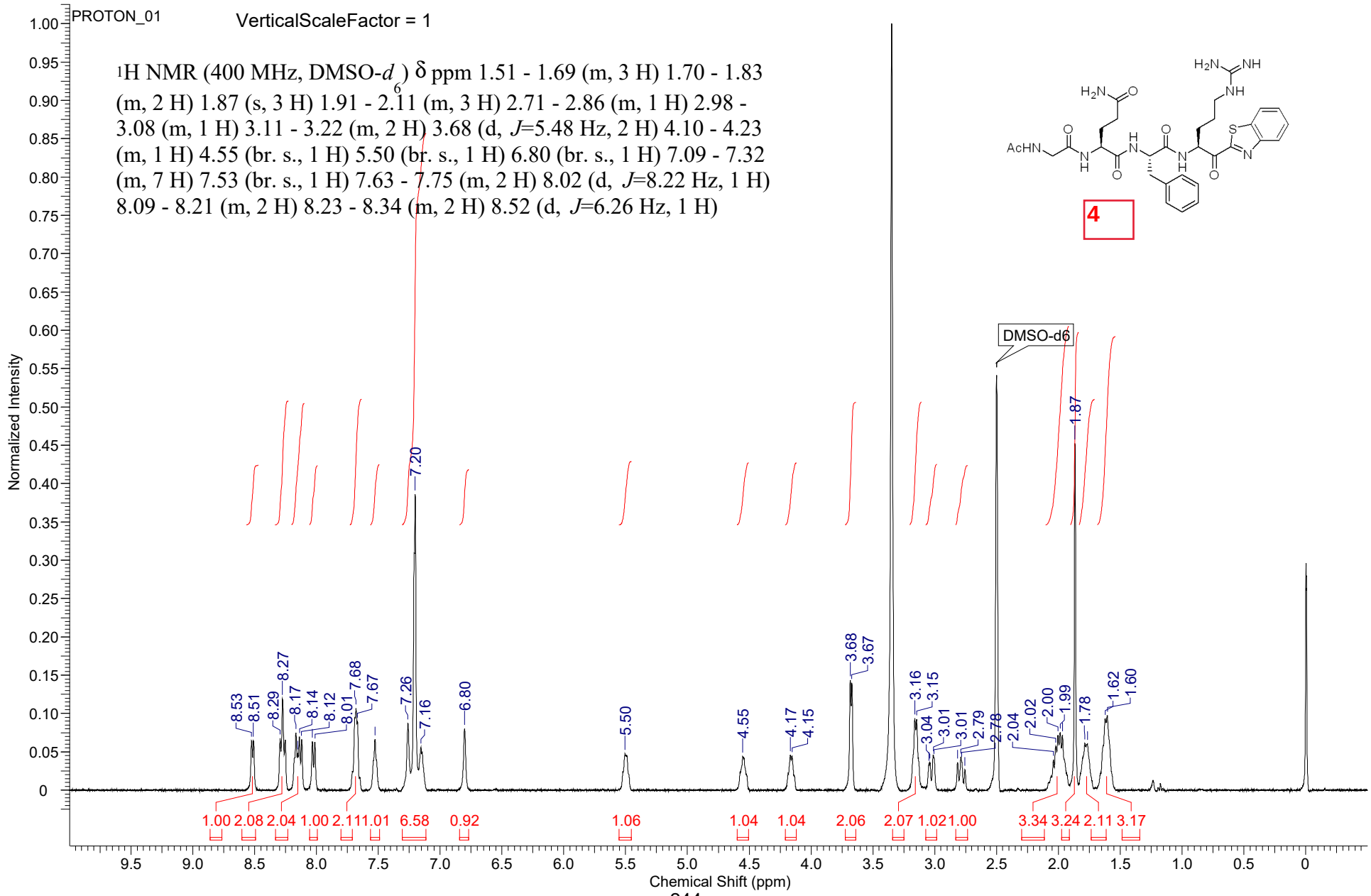




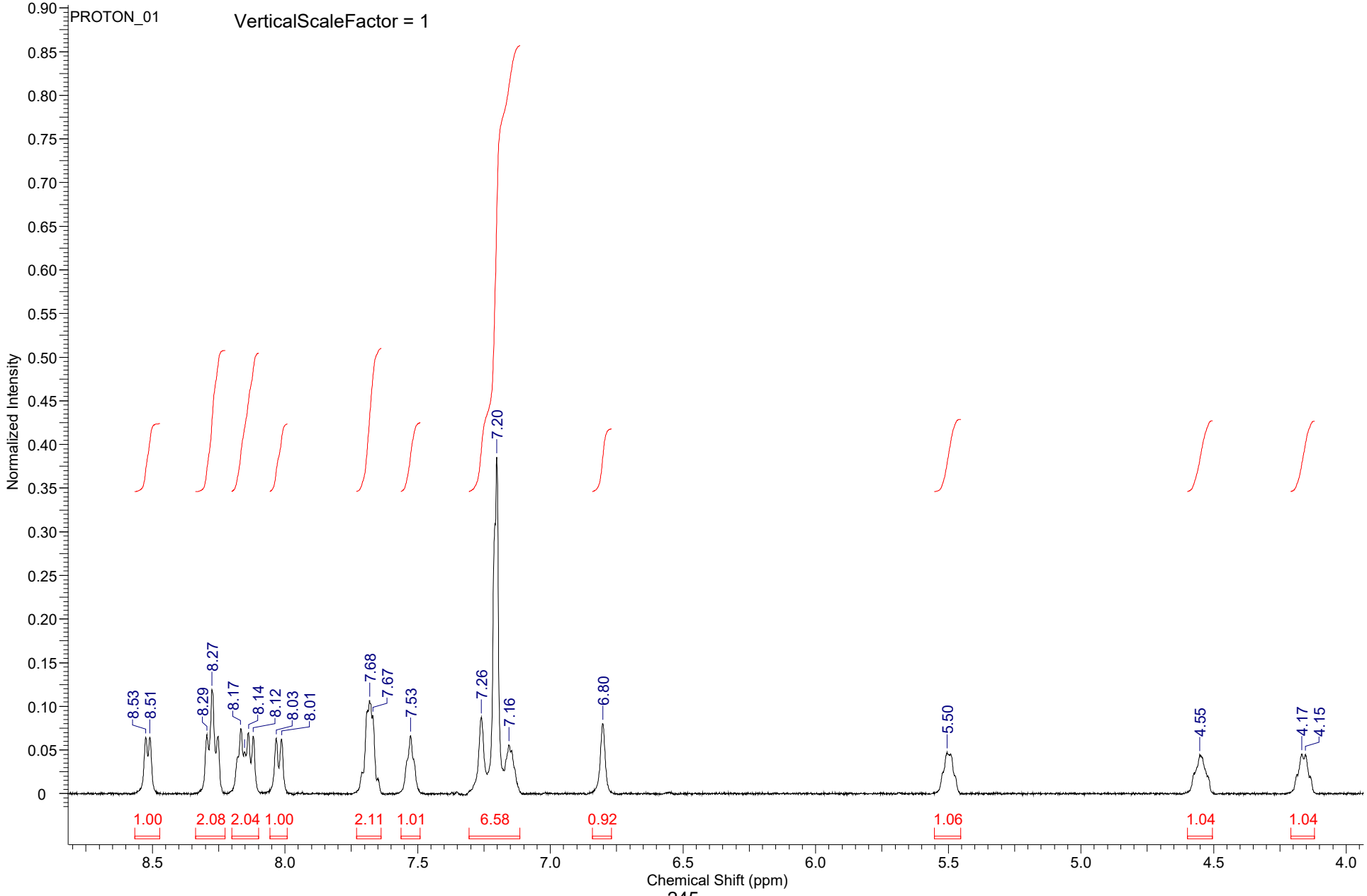




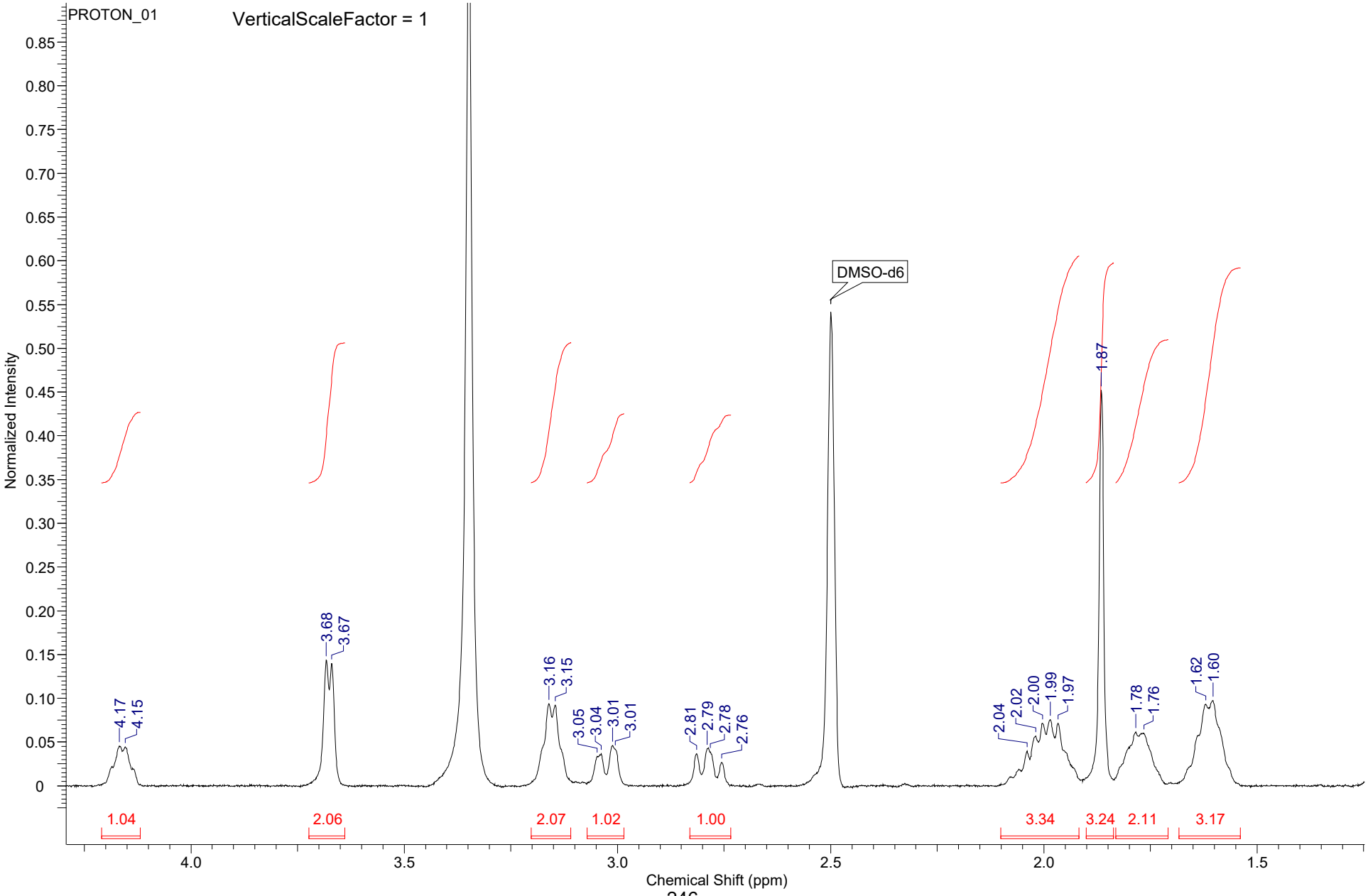
Acquisition Time (sec)	2.5559	***Comment***	MM3122 Ac-GQFR-kbt			***Date***	Jun 16 2020
Date Stamp	Jun 16 2020	***File Name***	C:\Mahoney_Raw_NMR\MM3122_20200616_01\PROTON_01.fid\fid				
Frequency (MHz)	399.75	***Nucleus***	1H	***Number of Transients***	8	***Original Points Count***	16384
Points Count	16384	***Pulse Sequence***	s2pul	***Receiver Gain***	46.00	***Solvent***	DMSO-d6
Spectrum Offset (Hz)	2411.2898	***Spectrum Type***	STANDARD	***Sweep Width (Hz)***	6410.26	***Temperature (degree C)***	25.000



Acquisition Time (sec)	2.5559	***Comment***	MM3122 Ac-GQFR-kbt			***Date***	Jun 16 2020
Date Stamp	Jun 16 2020	***File Name***	C:\Mahoney_Raw_NMR\MM3122_20200616_01\PROTON_01.fid\fid				
Frequency (MHz)	399.75	***Nucleus***	1H	***Number of Transients***	8	***Original Points Count***	16384
Points Count	16384	***Pulse Sequence***	s2pul	***Receiver Gain***	46.00	***Solvent***	DMSO-d6
Spectrum Offset (Hz)	2411.2898	***Spectrum Type***	STANDARD	***Sweep Width (Hz)***	6410.26	***Temperature (degree C)***	25.000



Acquisition Time (sec)	2.5559	***Comment***	MM3122 Ac-GQFR-kbt			***Date***	Jun 16 2020
Date Stamp	Jun 16 2020	***File Name***	C:\Mahoney_Raw_NMR\MM3122_20200616_01\PROTON_01.fid\fid				
Frequency (MHz)	399.75	***Nucleus***	1H	***Number of Transients***	8	***Original Points Count***	16384
Points Count	16384	***Pulse Sequence***	s2pul	***Receiver Gain***	46.00	***Solvent***	DMSO-d6
Spectrum Offset (Hz)	2411.2898	***Spectrum Type***	STANDARD	***Sweep Width (Hz)***	6410.26	***Temperature (degree C)***	25.000

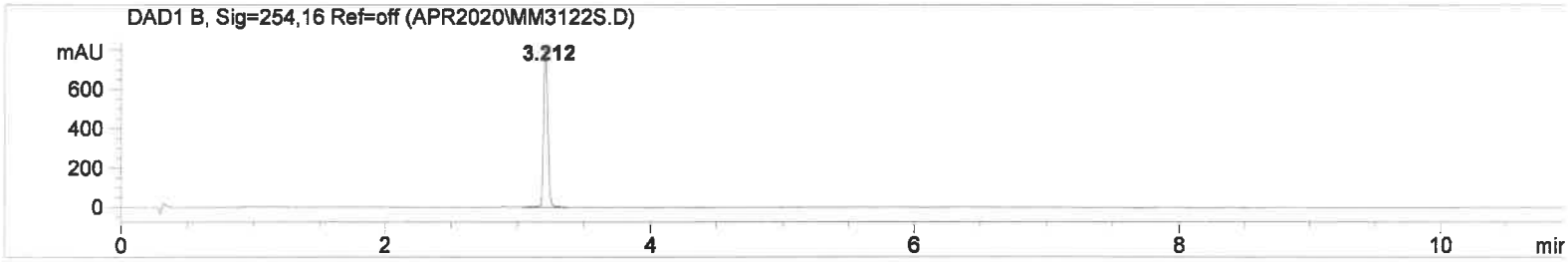


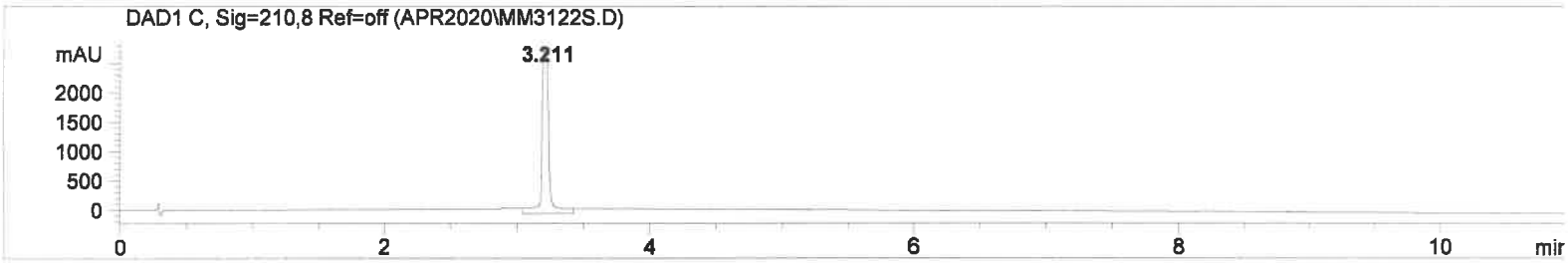
Ac-Gly-Gln-Phe-Arg-Ket ·TFA

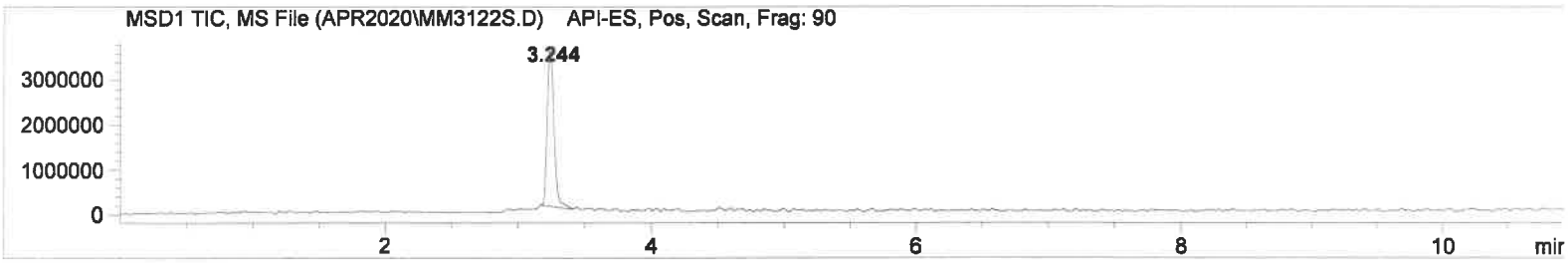
5.18.20

MM3122S

Injection Date : 5/18/2020 10:22:39 AM
Sample Name : MM3122S Location : Vial 4
Acq. Operator : Janetka
Acq. Instrument : Instrument 1 Inj Volume : 5 ul
Acq. Method : C:\HPCHEM\1\METHODS\MM10H1MM.M
Last changed : 5/15/2020 12:07:07 PM







DAD1 B, Sig=254,16 Ref=off

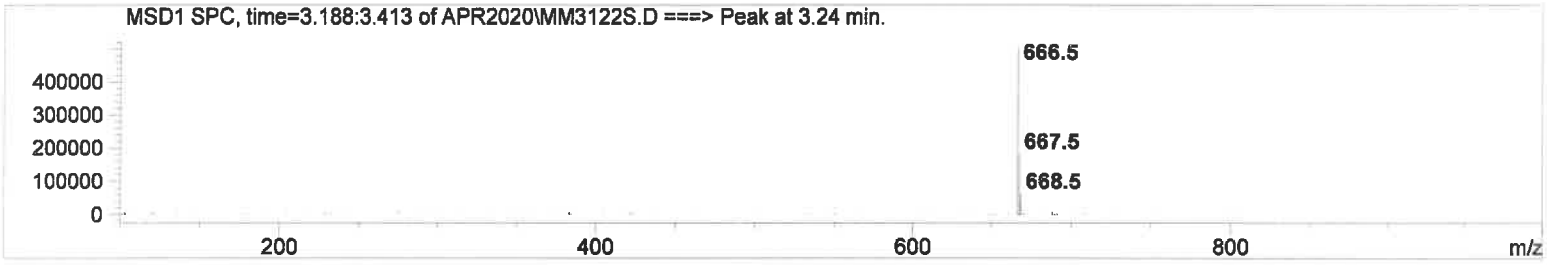
#	Meas. Ret	Area	Area %
1	3.21	1.813e3	100.00

DAD1 C, Sig=210,8 Ref=off

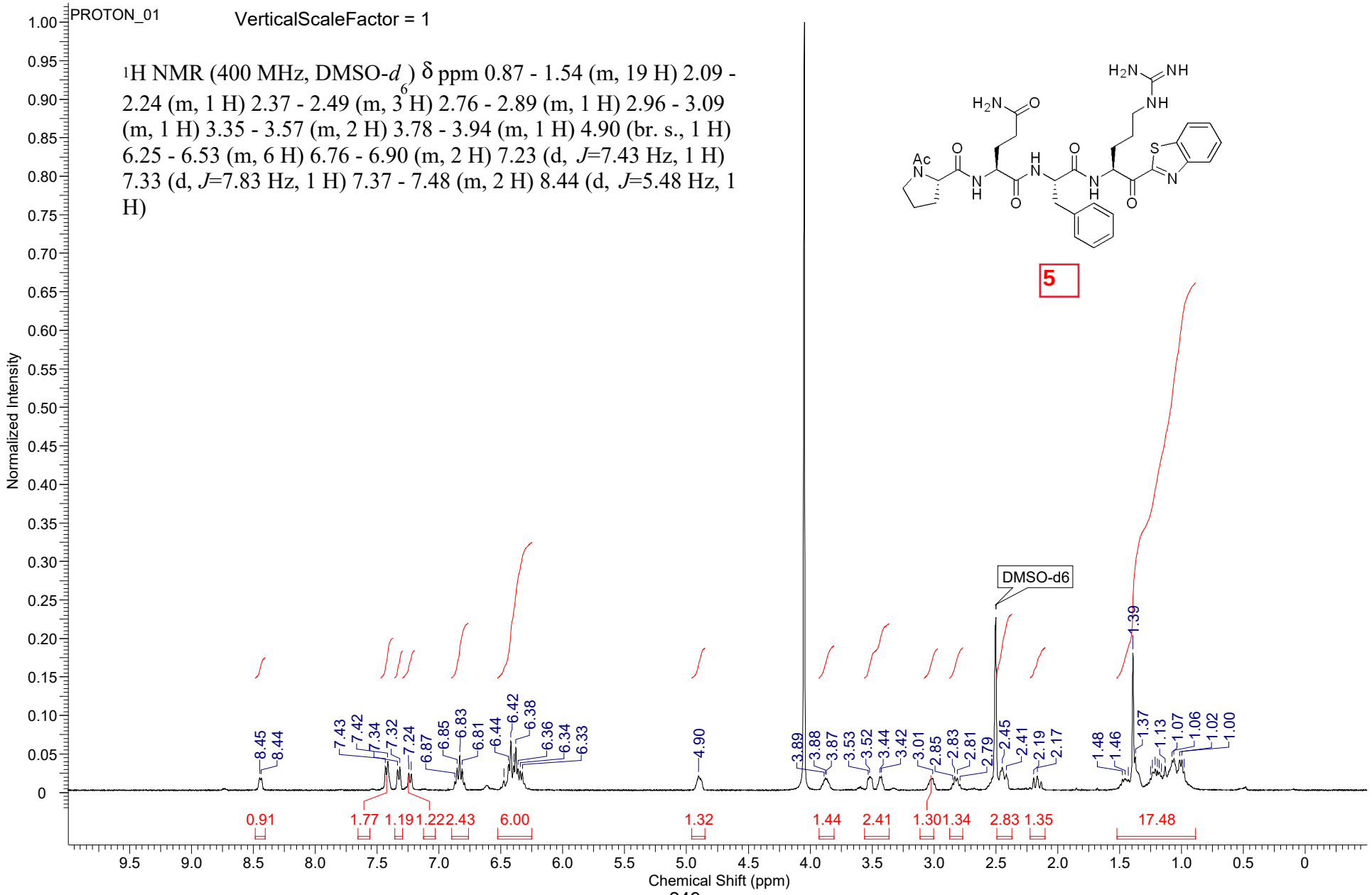
#	Meas. Ret	Area	Area %
1	3.21	1.050e4	100.00

MSD1 TIC, MS File

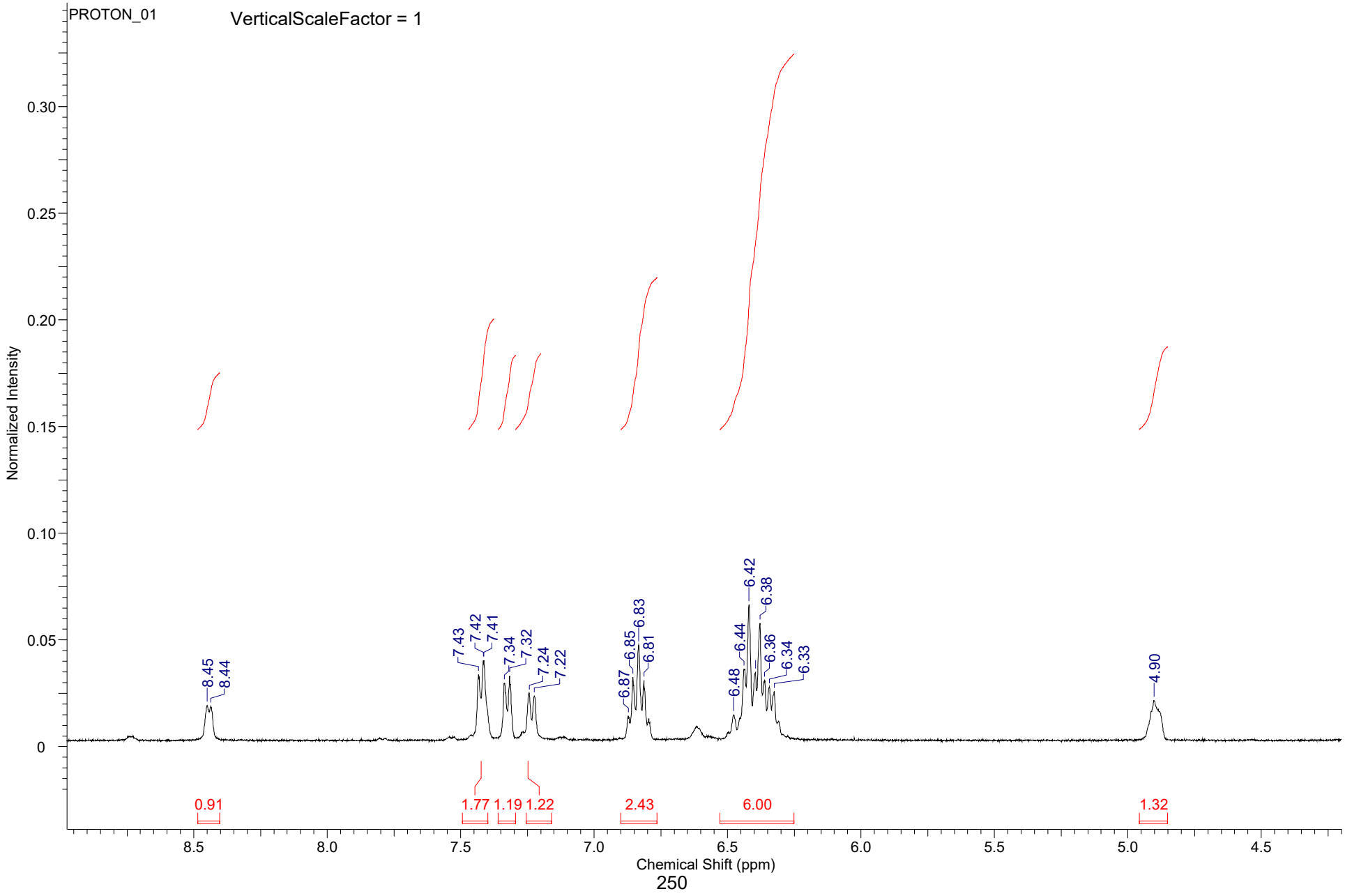
#	Meas. Ret	Area	Area %
1	3.24	1.212e7	100.00



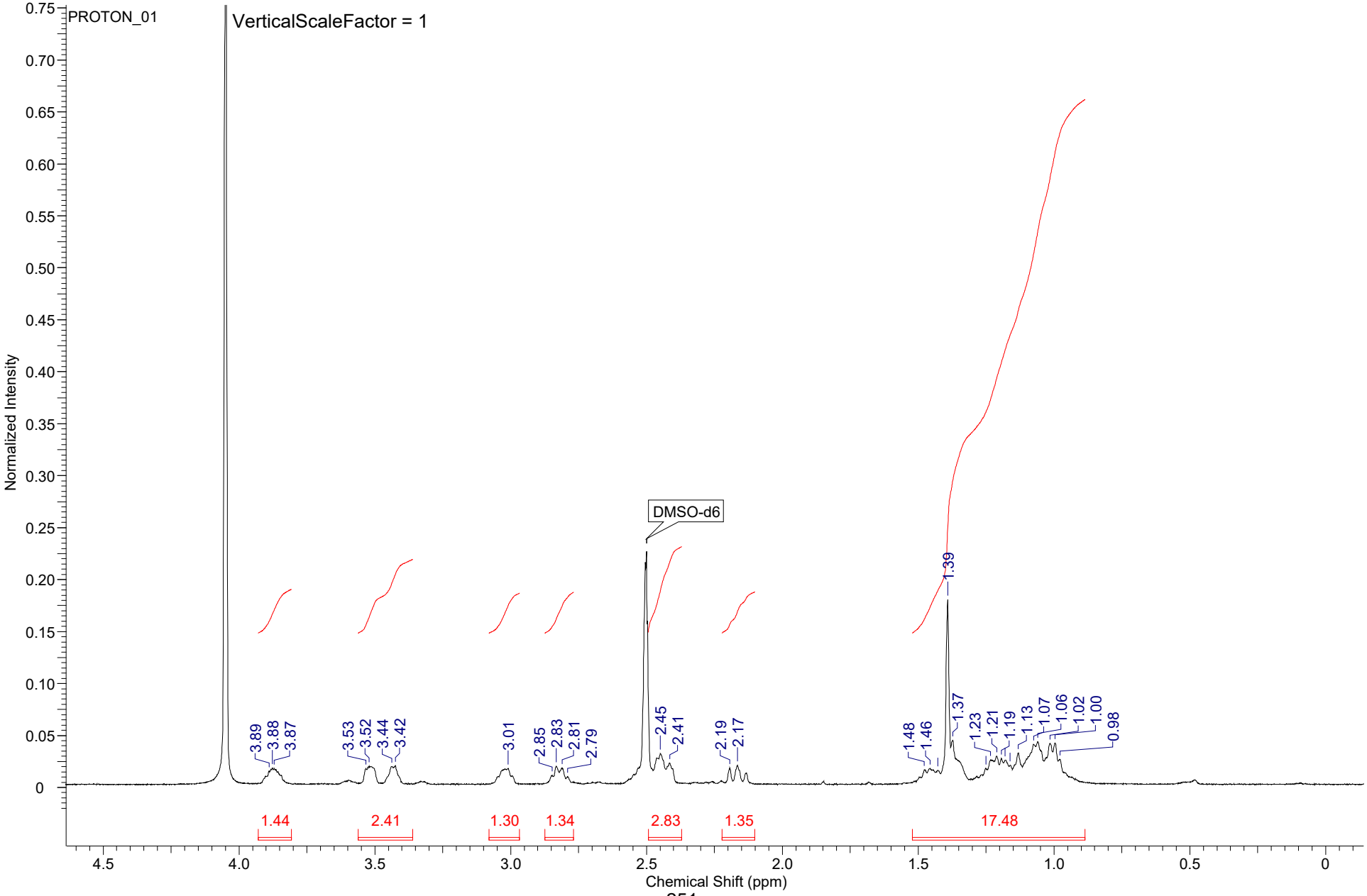
Acquisition Time (sec)	2.5559	***Comment***	MM3123 Ac-PQFR-kbt			***Date***	Jun 17 2020
Date Stamp	Jun 17 2020	***File Name***	C:\Mahoney_Raw_NMR\MM3131_20200617_01\PROTON_01.fid\fid				
Frequency (MHz)	399.75	***Nucleus***	1H	***Number of Transients***	8	***Original Points Count***	16384
Points Count	16384	***Pulse Sequence***	s2pul	***Receiver Gain***	46.00	***Solvent***	DMSO-d6
Spectrum Offset (Hz)	2408.1597	***Spectrum Type***	STANDARD	***Sweep Width (Hz)***	6410.26	***Temperature (degree C)***	25.000



Acquisition Time (sec)	2.5559	***Comment***	MM3123 Ac-PQFR-kbt			***Date***	Jun 17 2020
Date Stamp	Jun 17 2020	***File Name***	C:\Mahoney_Raw_NMR\MM3131_20200617_01\PROTON_01.fid\fid				
Frequency (MHz)	399.75	***Nucleus***	1H	***Number of Transients***	8	***Original Points Count***	16384
Points Count	16384	***Pulse Sequence***	s2pul	***Receiver Gain***	46.00	***Solvent***	DMSO-d6
Spectrum Offset (Hz)	2408.1597	***Spectrum Type***	STANDARD	***Sweep Width (Hz)***	6410.26	***Temperature (degree C)***	25.000

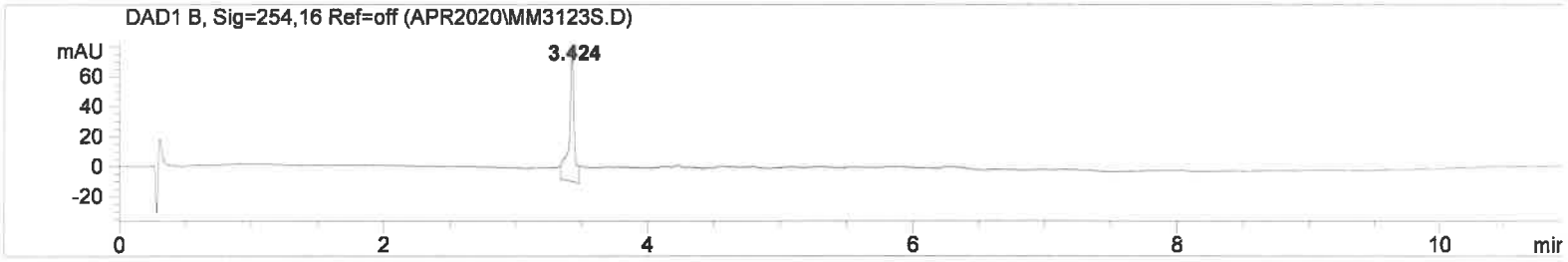


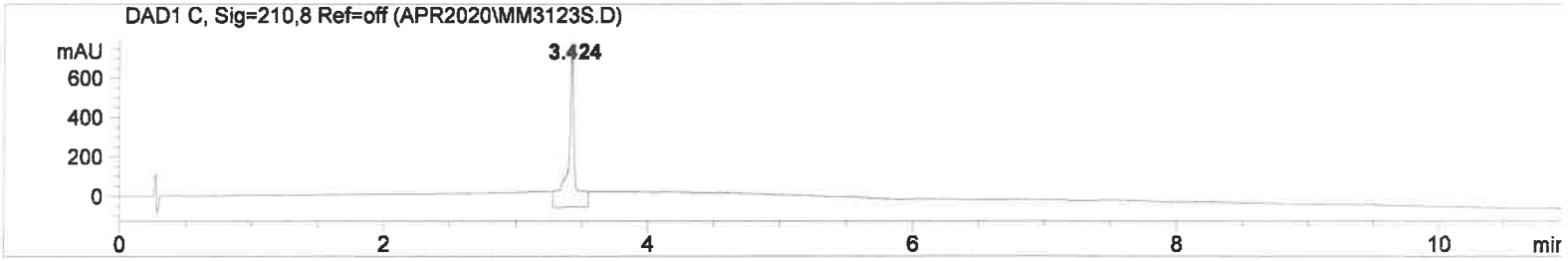
Acquisition Time (sec)	2.5559	***Comment***	MM3123 Ac-PQFR-kbt			***Date***	Jun 17 2020
Date Stamp	Jun 17 2020	***File Name***	C:\Mahoney_Raw_NMR\MM3131_20200617_01\PROTON_01.fid\fid				
Frequency (MHz)	399.75	***Nucleus***	1H	***Number of Transients***	8	***Original Points Count***	16384
Points Count	16384	***Pulse Sequence***	s2pul	***Receiver Gain***	46.00	***Solvent***	DMSO-d6
Spectrum Offset (Hz)	2408.1597	***Spectrum Type***	STANDARD	***Sweep Width (Hz)***	6410.26	***Temperature (degree C)***	25.000

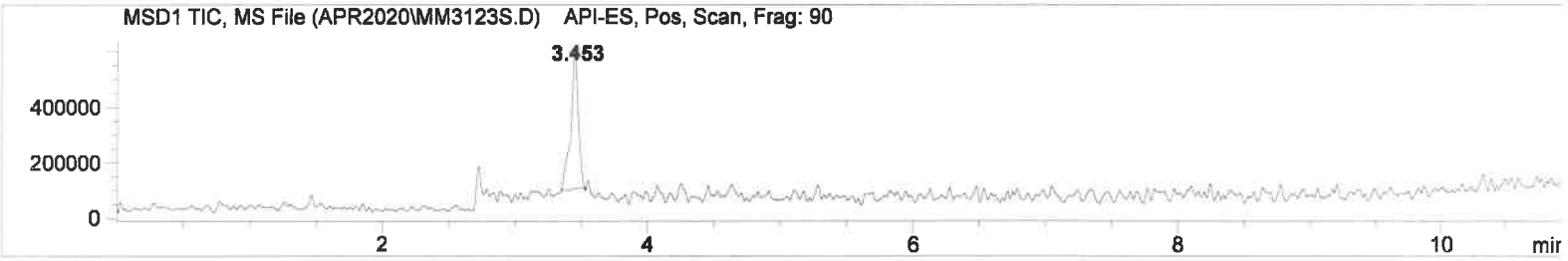


Ac-Pro-Gln-Phe-Arg-Ket-TFA
MW: 705.84
5.18.20
MM3123S

Injection Date : 5/18/2020 10:36:21 AM
Sample Name : MM3123S Location : Vial 5
Acq. Operator : Janetka
Acq. Instrument : Instrument 1 Inj Volume : 5 ul
Acq. Method : C:\HPCHEM\1\METHODS\MM10H1MM.M
Last changed : 5/15/2020 12:07:07 PM







DAD1 B, Sig=254,16 Ref=off

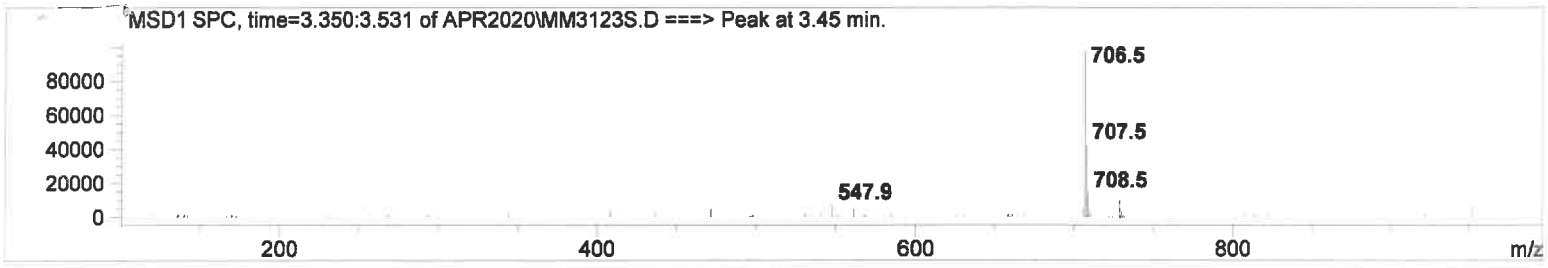
#	Meas. Ret	Area	Area %
1	3.42	2.332e2	100.00

DAD1 C, Sig=210,8 Ref=off

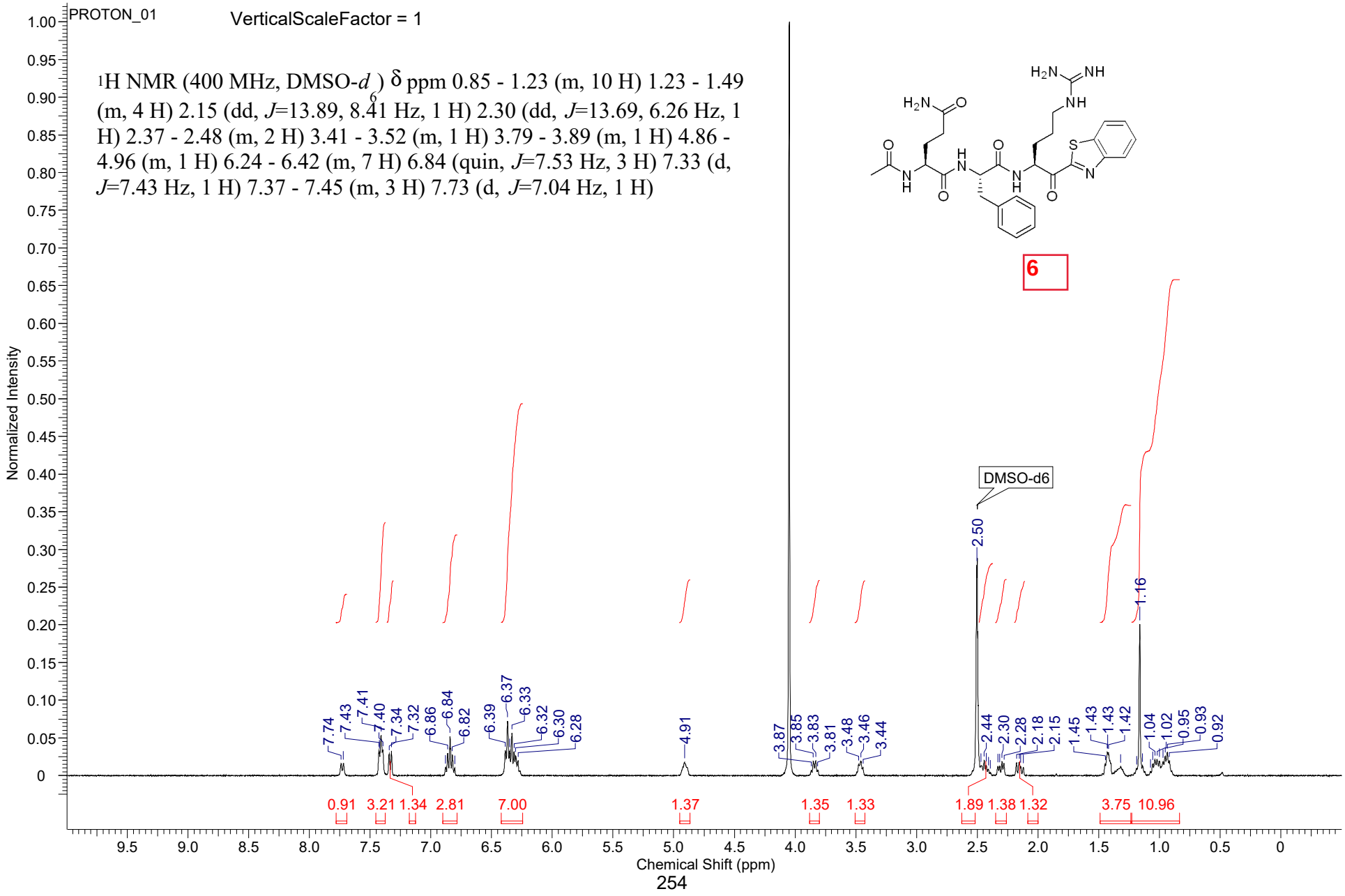
#	Meas. Ret	Area	Area %
1	3.42	2.760e3	100.00

MSD1 TIC, MS File

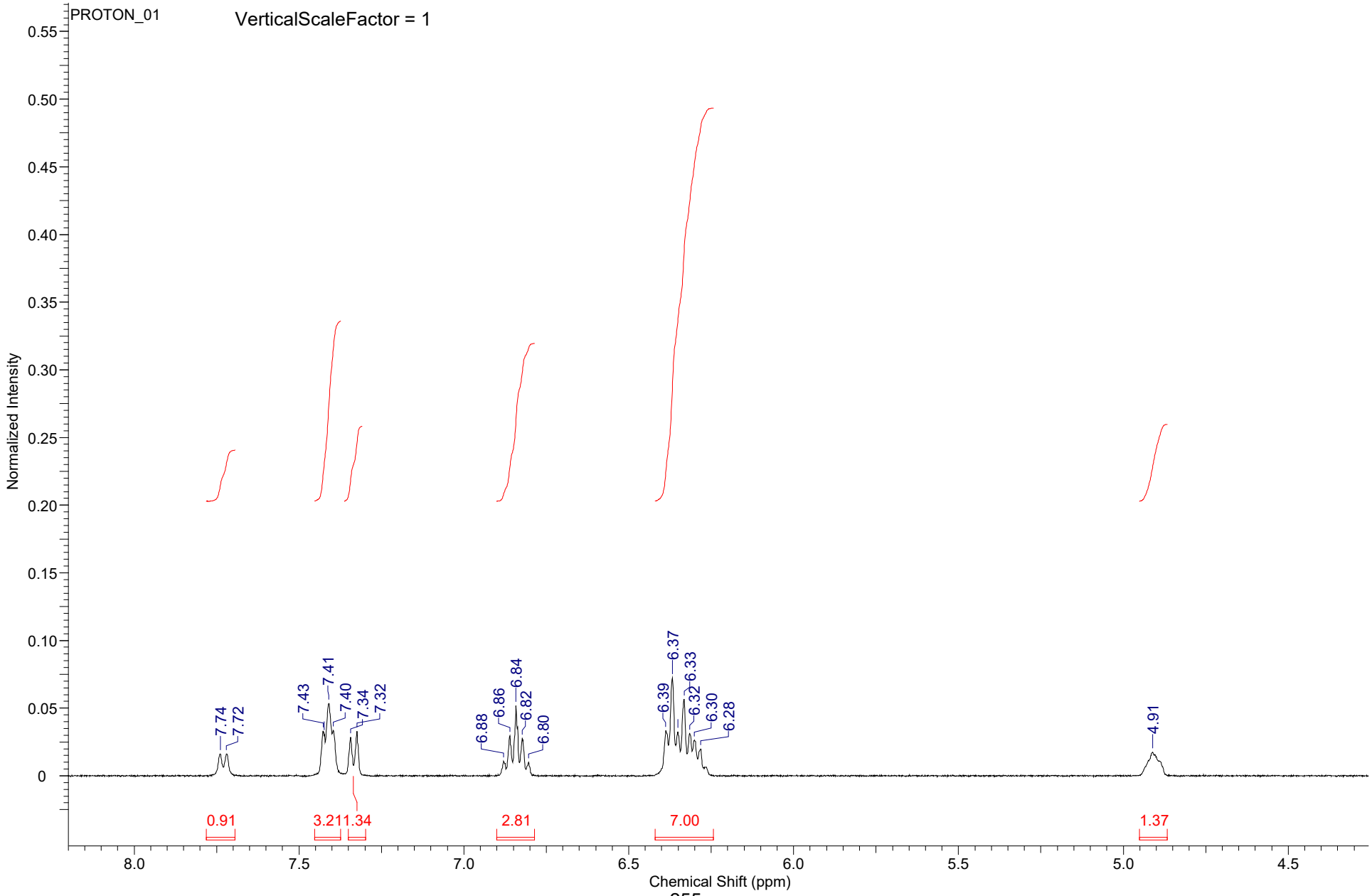
#	Meas. Ret	Area	Area %
1	3.45	2.002e6	100.00


MSD1 SPC, time=3.350:3.531 of APR2020\MM3123S.D ===> Peak at 3.45 min.
80000
60000
40000
20000
0
706.5
707.5
708.5
547.9
200
400
600
800
m/z

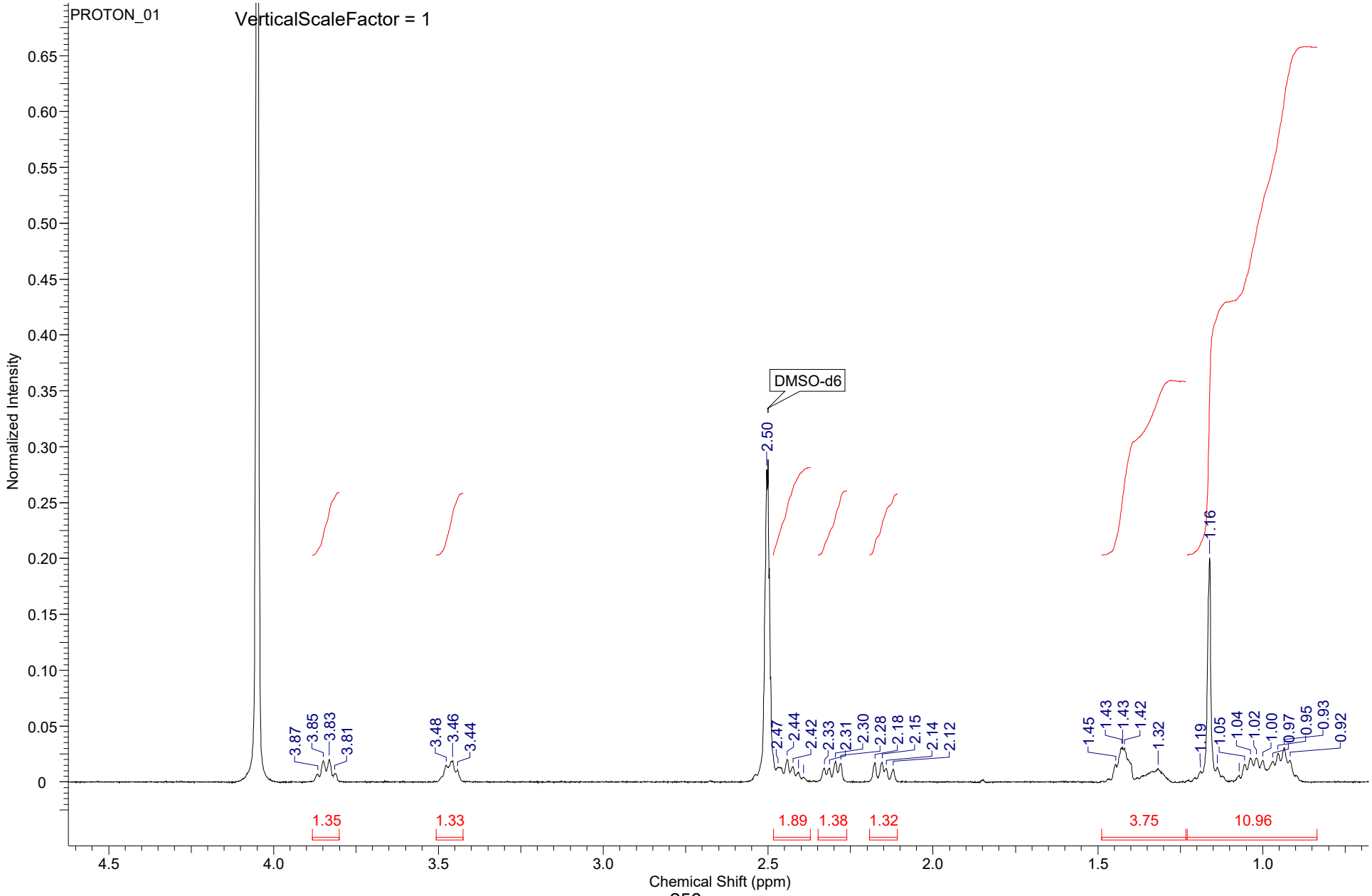
Acquisition Time (sec)	2.5559	***Comment***	MM3144 Ac-QFR-kbt			***Date***	Jun 17 2020
Date Stamp	Jun 17 2020	***File Name***	C:\Mahoney_Raw_NMR\MM3144_20200617_01\PROTON_01.fid\fid				
Frequency (MHz)	399.75	***Nucleus***	1H	***Number of Transients***	8	***Original Points Count***	16384
Points Count	16384	***Pulse Sequence***	s2pul	***Receiver Gain***	48.00	***Solvent***	DMSO-d6
Spectrum Offset (Hz)	2408.1594	***Spectrum Type***	STANDARD	***Sweep Width (Hz)***	6410.26	***Temperature (degree C)***	25.000



Acquisition Time (sec)	2.5559	***Comment***	MM3144 Ac-QFR-kbt			***Date***	Jun 17 2020
Date Stamp	Jun 17 2020	***File Name***	C:\Mahoney_Raw_NMR\MM3144_20200617_01\PROTON_01.fid\fid				
Frequency (MHz)	399.75	***Nucleus***	1H	***Number of Transients***	8	***Original Points Count***	16384
Points Count	16384	***Pulse Sequence***	s2pul	***Receiver Gain***	48.00	***Solvent***	DMSO-d6
Spectrum Offset (Hz)	2408.1594	***Spectrum Type***	STANDARD	***Sweep Width (Hz)***	6410.26	***Temperature (degree C)***	25.000



Acquisition Time (sec)	2.5559	***Comment***	MM3144 Ac-QFR-kbt			***Date***	Jun 17 2020
Date Stamp	Jun 17 2020	***File Name***	C:\Mahoney_Raw_NMR\MM3144_20200617_01\PROTON_01.fid\fid				
Frequency (MHz)	399.75	***Nucleus***	1H	***Number of Transients***	8	***Original Points Count***	16384
Points Count	16384	***Pulse Sequence***	s2pul	***Receiver Gain***	48.00	***Solvent***	DMSO-d6
Spectrum Offset (Hz)	2408.1594	***Spectrum Type***	STANDARD	***Sweep Width (Hz)***	6410.26	***Temperature (degree C)***	25.000



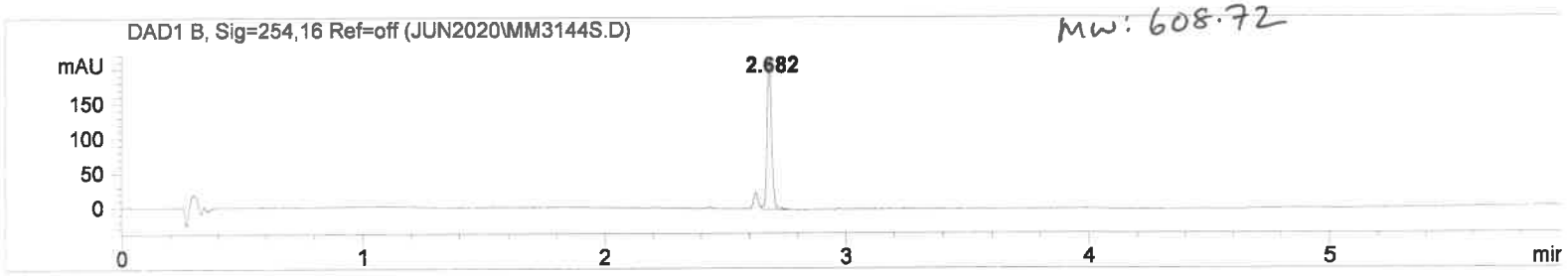
6.4.20

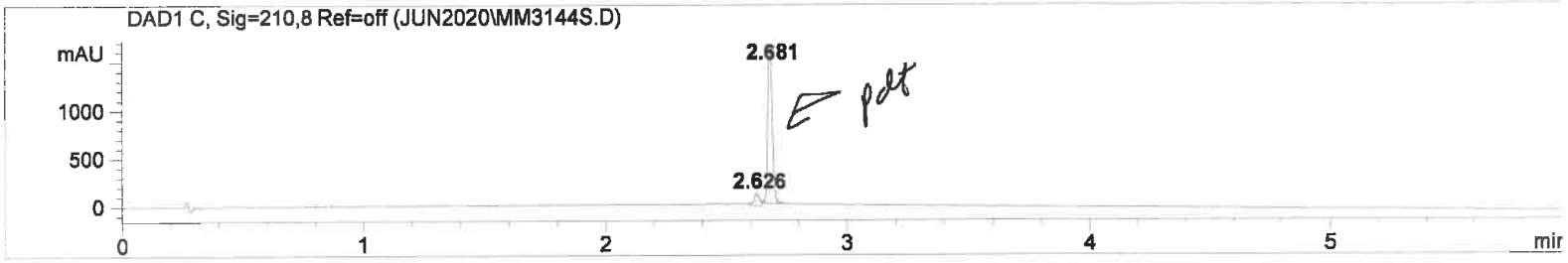
Injection Date : 6/4/2020 1:57:43 PM
Sample Name : MM3144S
Acq. Operator : Janetka
Acq. Instrument : Instrument 1
Acq. Method : C:\HPCHEM\1\METHODS\5P10H1.M
Last changed : 7/27/2010 4:44:37 PM

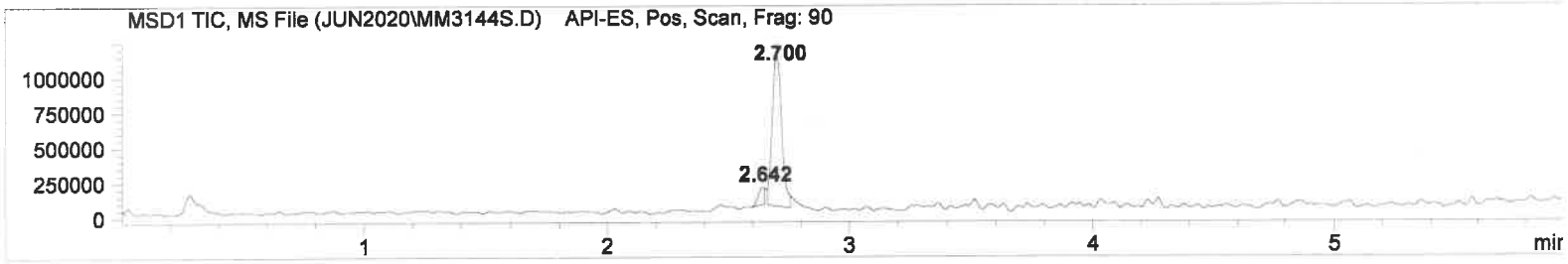
Location : Vial 18

Inj Volume : 5 ul

Ac-QFR-Kbt
MW: 608.72







DAD1 B, Sig=254,16 Ref=off

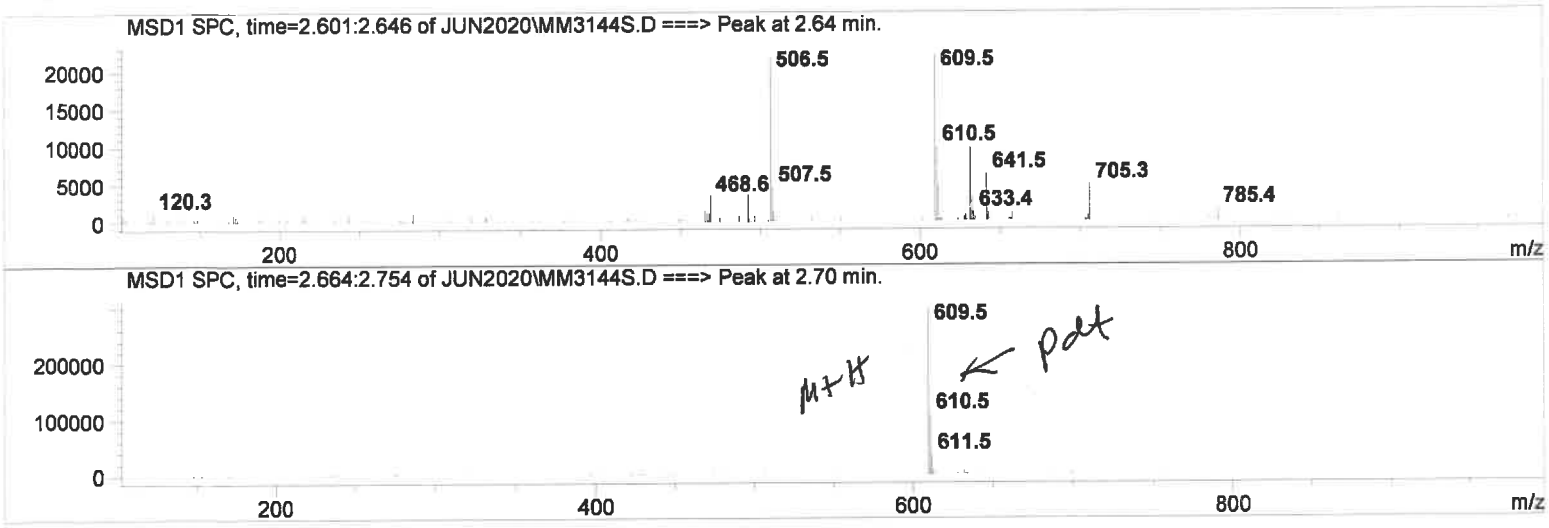
#	Meas. Ret	Area	Area %
1	2.68	2.802e2	100.00

DAD1 C, Sig=210,8 Ref=off

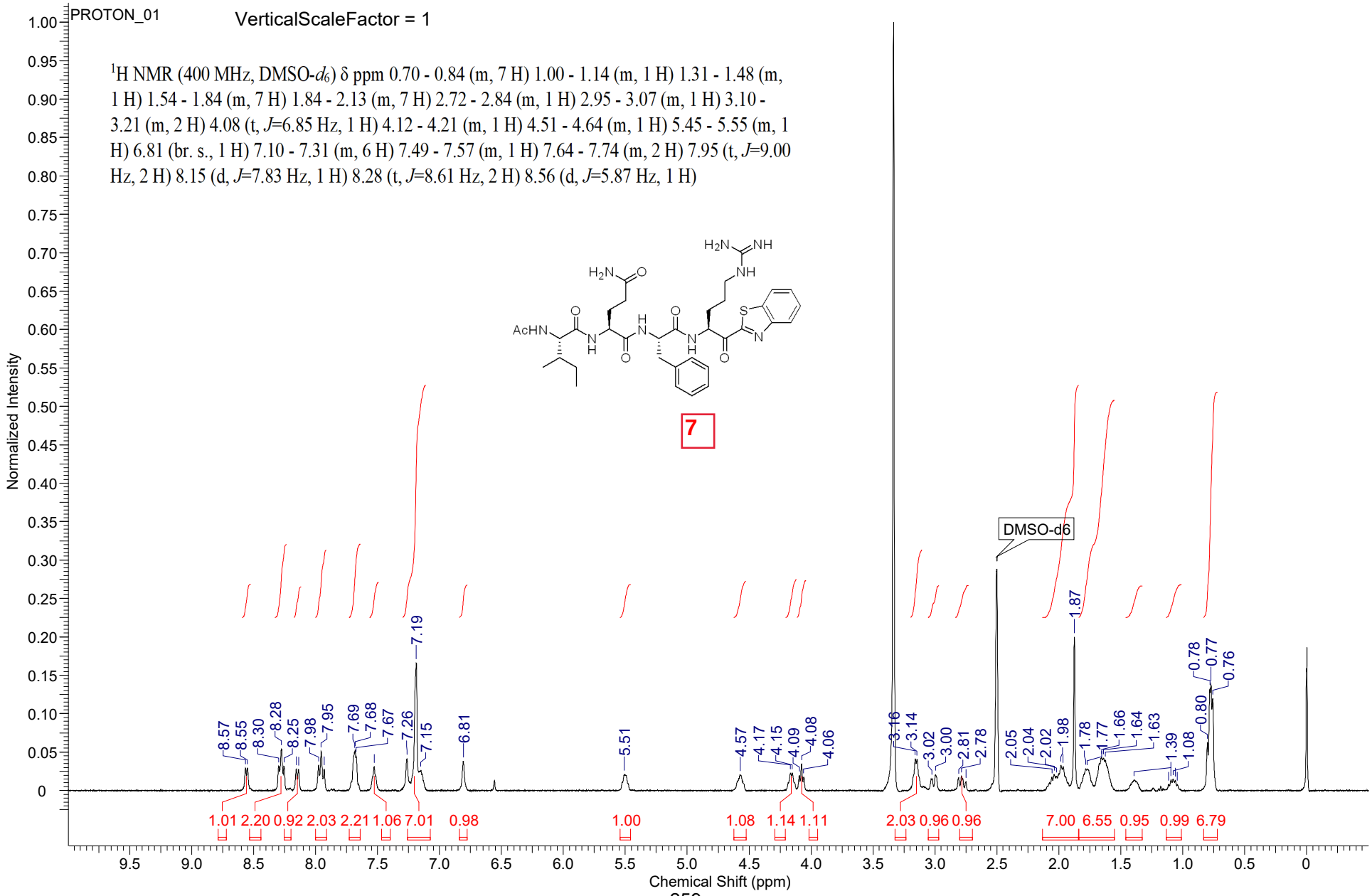
#	Meas. Ret	Area	Area %
1	2.63	2.039e2	8.85
2	2.68	2.101e3	91.15

MSD1 TIC, MS File

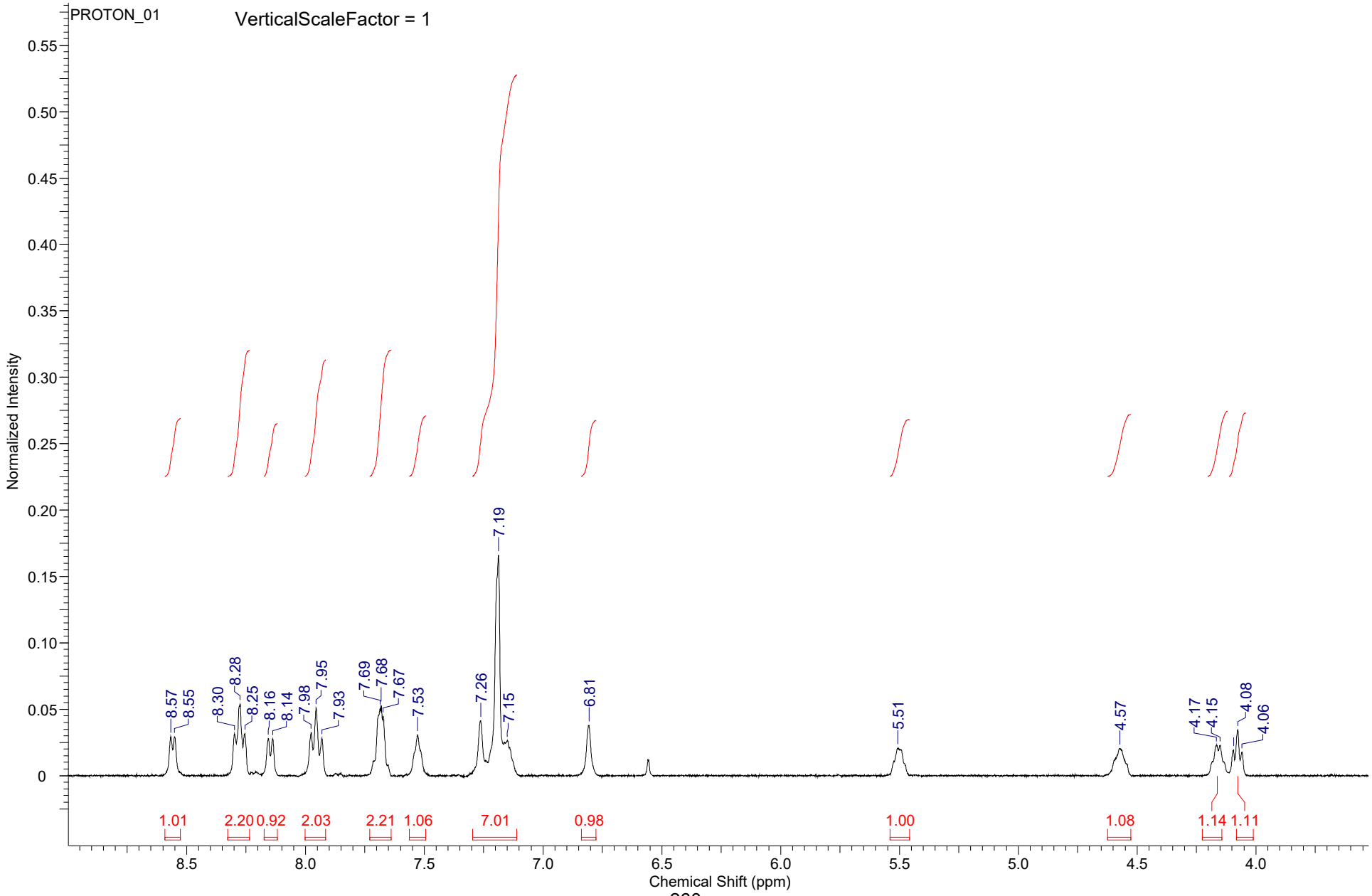
#	Meas. Ret	Area	Area %
1	2.64	2.272e5	7.19
2	2.70	2.931e6	92.81


MSD1 SPC, time=2.601:2.646 of JUN2020\MM3144S.D ===> Peak at 2.64 min.
20000
15000
10000
5000
0
120.3
468.6
506.5
507.5
609.5
610.5
641.5
633.4
705.3
785.4
200
400
600
800
m/z
MSD1 SPC, time=2.664:2.754 of JUN2020\MM3144S.D ===> Peak at 2.70 min.
200000
100000
0
M+H
609.5
pdt
610.5
611.5
200
400
600
800
m/z

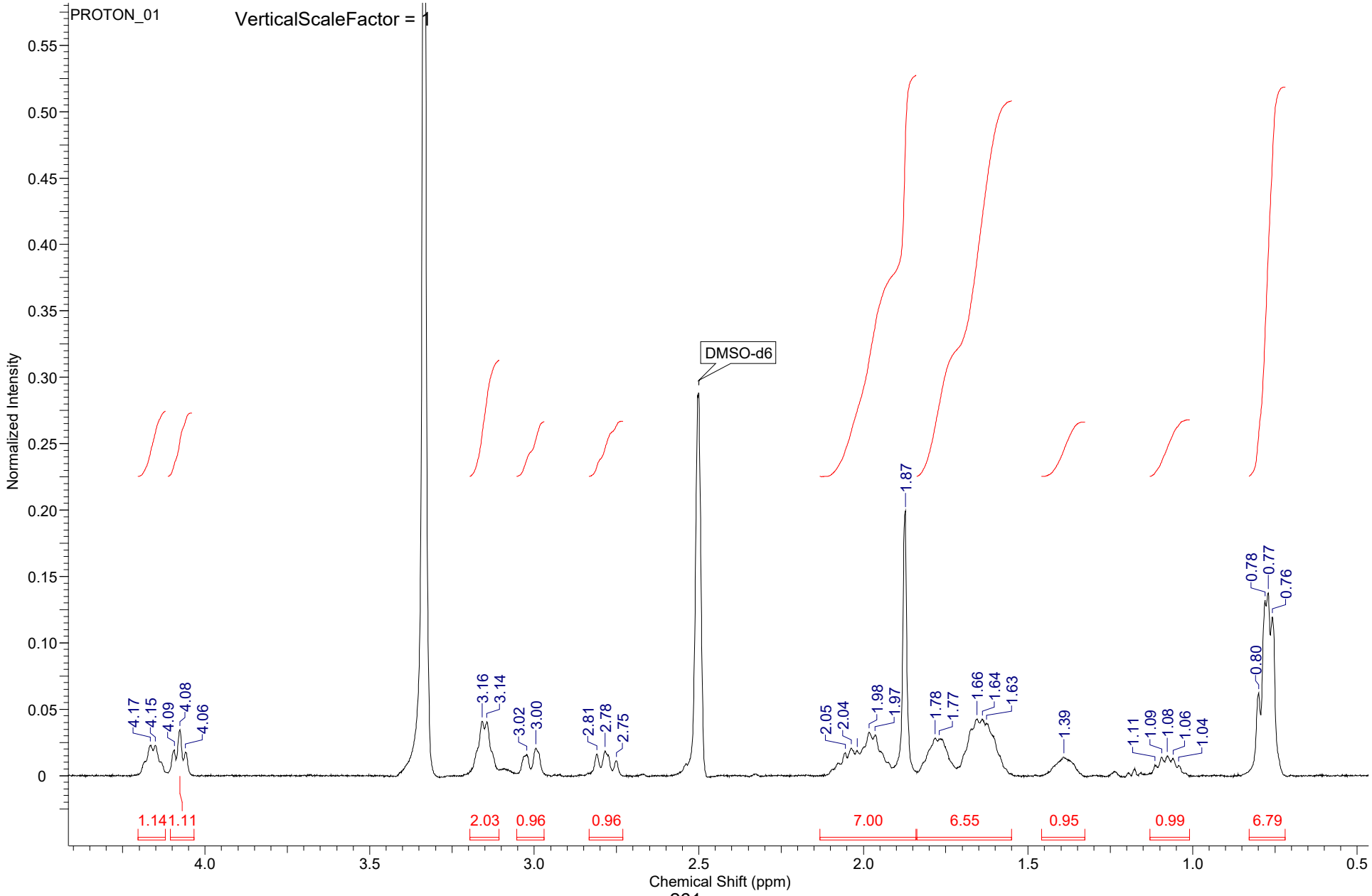
Acquisition Time (sec)	2.5559	***Comment***	MM3116 Ac-IQFR-kbt			***Date***	Jun 16 2020
Date Stamp	Jun 16 2020	***File Name***	C:\Mahoney_Raw_NMR\MM3116_20200616_02\PROTON_01.fid\fid				
Frequency (MHz)	399.75	***Nucleus***	1H	***Number of Transients***	8	***Original Points Count***	16384
Points Count	16384	***Pulse Sequence***	s2pul	***Receiver Gain***	46.00	***Solvent***	DMSO-d6
Spectrum Offset (Hz)	2412.4636	***Spectrum Type***	STANDARD	***Sweep Width (Hz)***	6410.26	***Temperature (degree C)***	25.000



Acquisition Time (sec)	2.5559	***Comment***	MM3116 Ac-IQFR-kbt			***Date***	Jun 16 2020
Date Stamp	Jun 16 2020	***File Name***	C:\Mahoney_Raw_NMR\MM3116_20200616_02\PROTON_01.fid\fid				
Frequency (MHz)	399.75	***Nucleus***	1H	***Number of Transients***	8	***Original Points Count***	16384
Points Count	16384	***Pulse Sequence***	s2pul	***Receiver Gain***	46.00	***Solvent***	DMSO-d6
Spectrum Offset (Hz)	2412.4636	***Spectrum Type***	STANDARD	***Sweep Width (Hz)***	6410.26	***Temperature (degree C)***	25.000



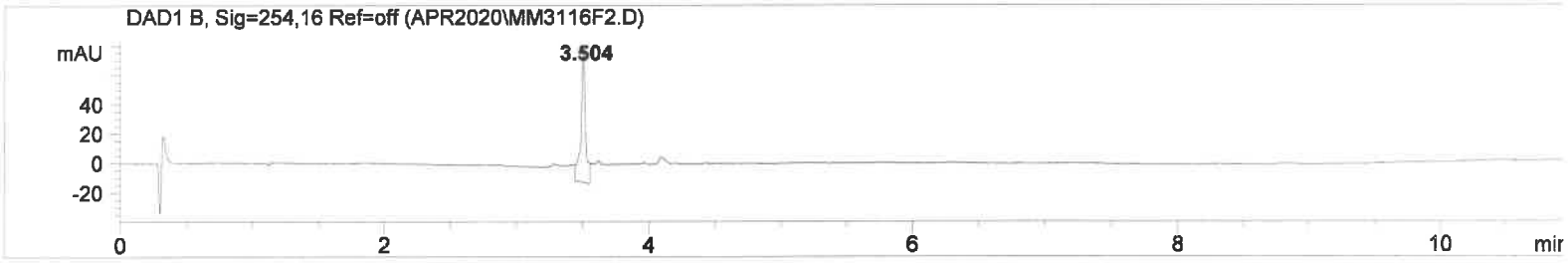
Acquisition Time (sec)	2.5559	***Comment***	MM3116 Ac-IQFR-kbt			***Date***	Jun 16 2020
Date Stamp	Jun 16 2020	***File Name***	C:\Mahoney_Raw_NMR\MM3116_20200616_02\PROTON_01.fid\fid				
Frequency (MHz)	399.75	***Nucleus***	1H	***Number of Transients***	8	***Original Points Count***	16384
Points Count	16384	***Pulse Sequence***	s2pul	***Receiver Gain***	46.00	***Solvent***	DMSO-d6
Spectrum Offset (Hz)	2412.4636	***Spectrum Type***	STANDARD	***Sweep Width (Hz)***	6410.26	***Temperature (degree C)***	25.000

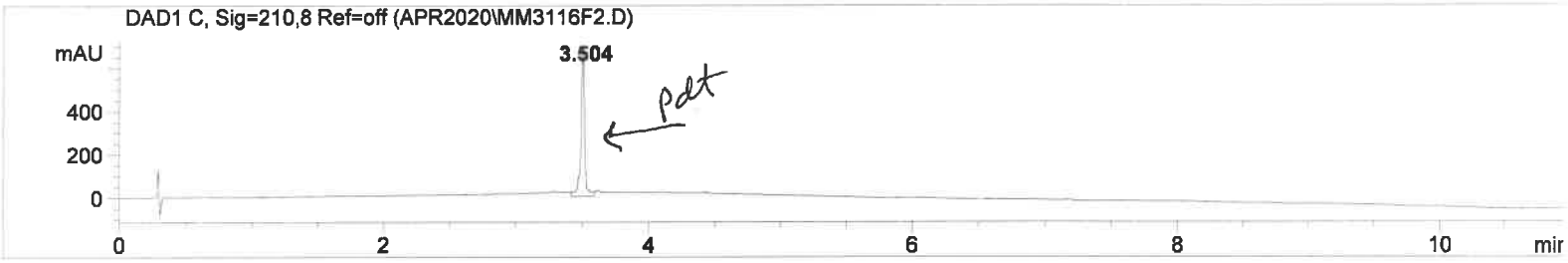


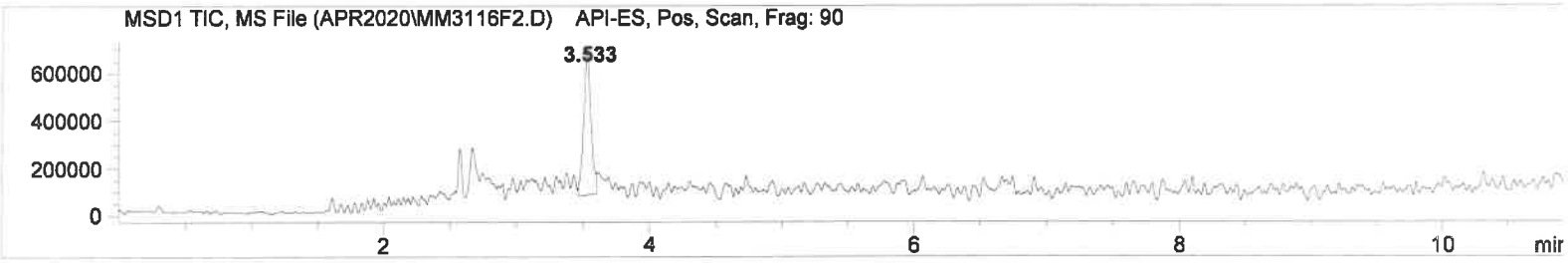
Ac-Ile-Gln-Phe-Arg-Klot ·TFA
MW: 721.88

Injection Date : 5/15/2020 12:10:40 PM
Sample Name : MM3116F2 Location : Vial 2
Acq. Operator : Janetka
Acq. Instrument : Instrument 1 Inj Volume : 5 ul
Acq. Method : C:\HPCHEM\1\METHODS\MM10H1MM.M
Last changed : 5/15/2020 12:07:07 PM

MM3116F2







DAD1 B, Sig=254,16 Ref=off

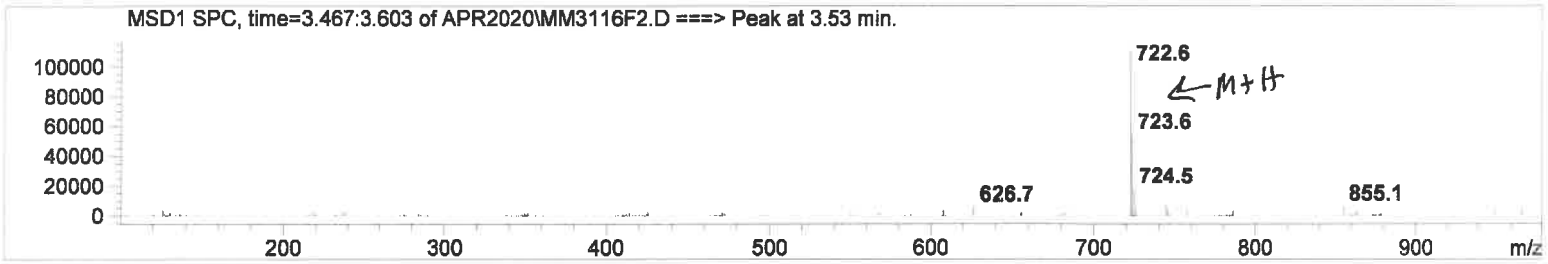
#	Meas. Ret	Area	Area %
1	3.50	2.120e2	100.00

DAD1 C, Sig=210,8 Ref=off

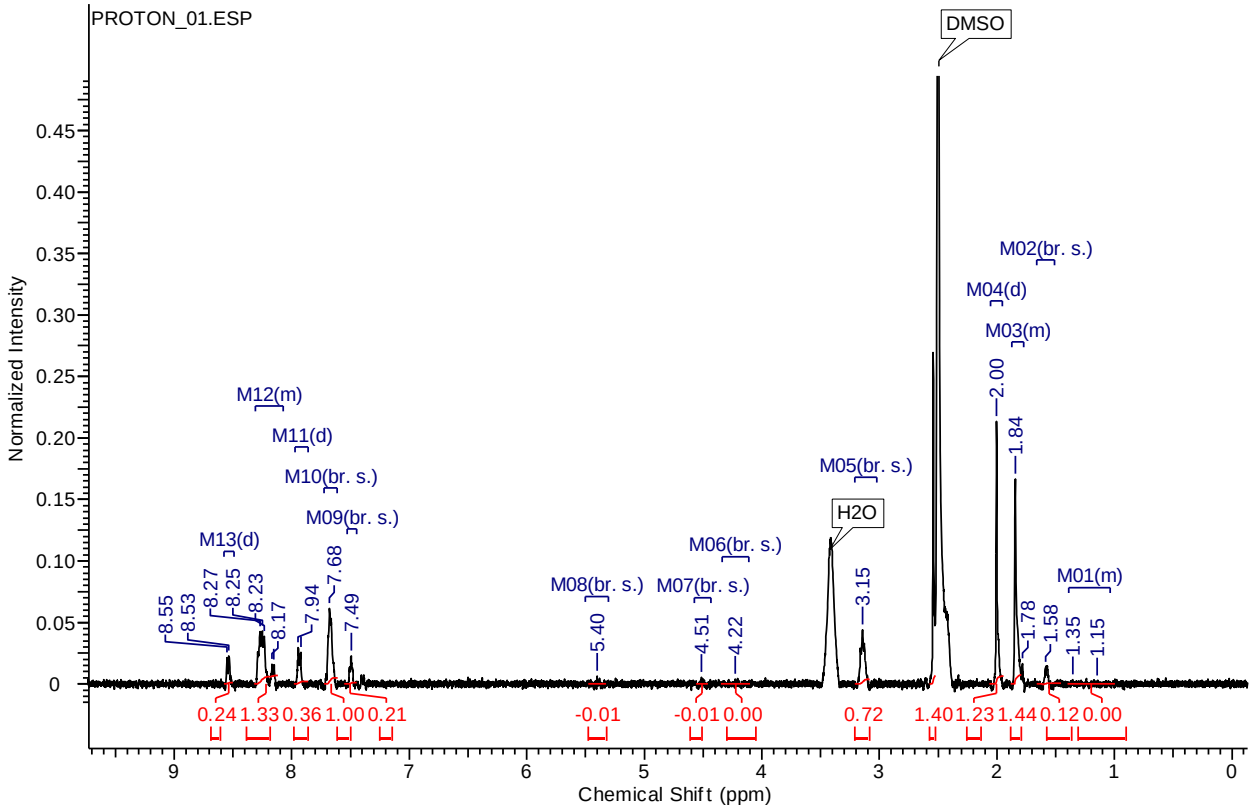
#	Meas. Ret	Area	Area %
1	3.50	1.305e3	100.00

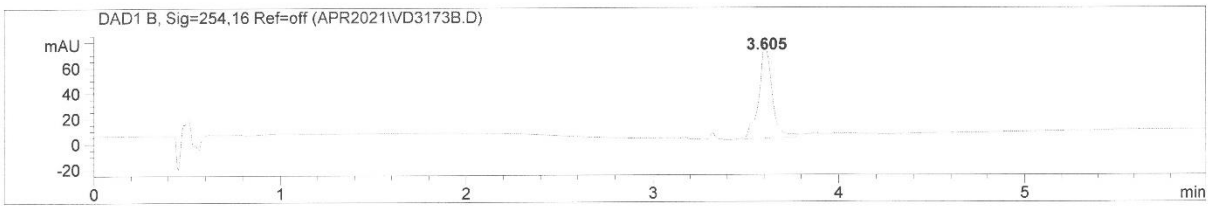
MSD1 TIC, MS File

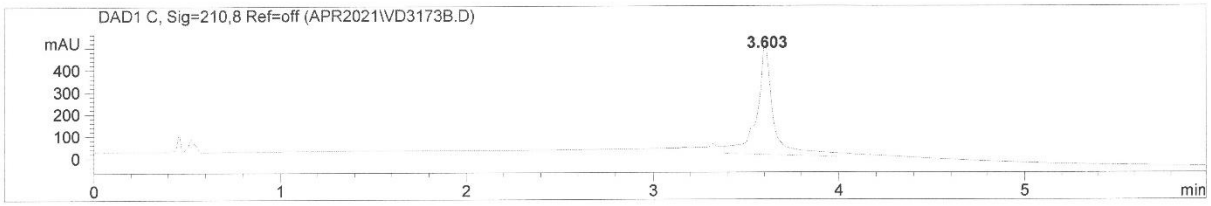
#	Meas. Ret	Area	Area %
1	3.53	2.295e6	100.00

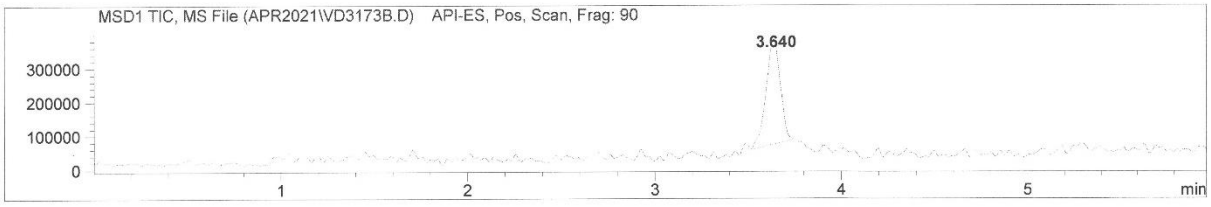

MSD1 SPC, time=3.467:3.603 of APR2020\MM3116F2.D ===> Peak at 3.53 min.
100000
80000
60000
40000
20000
0
626.7
722.6
M+H
723.6
724.5
855.1
200
300
400
500
600
700
800
900
m/z

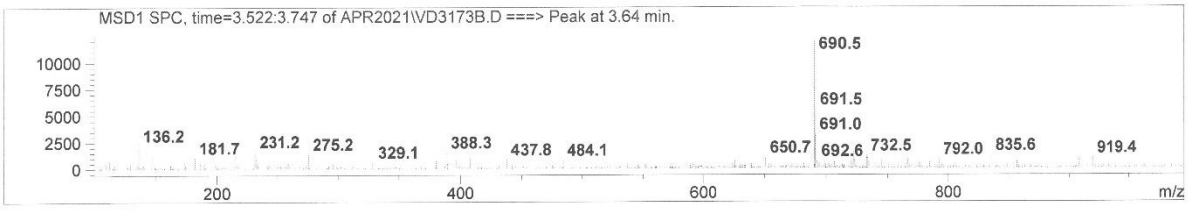
VD3173 (Compound 19)



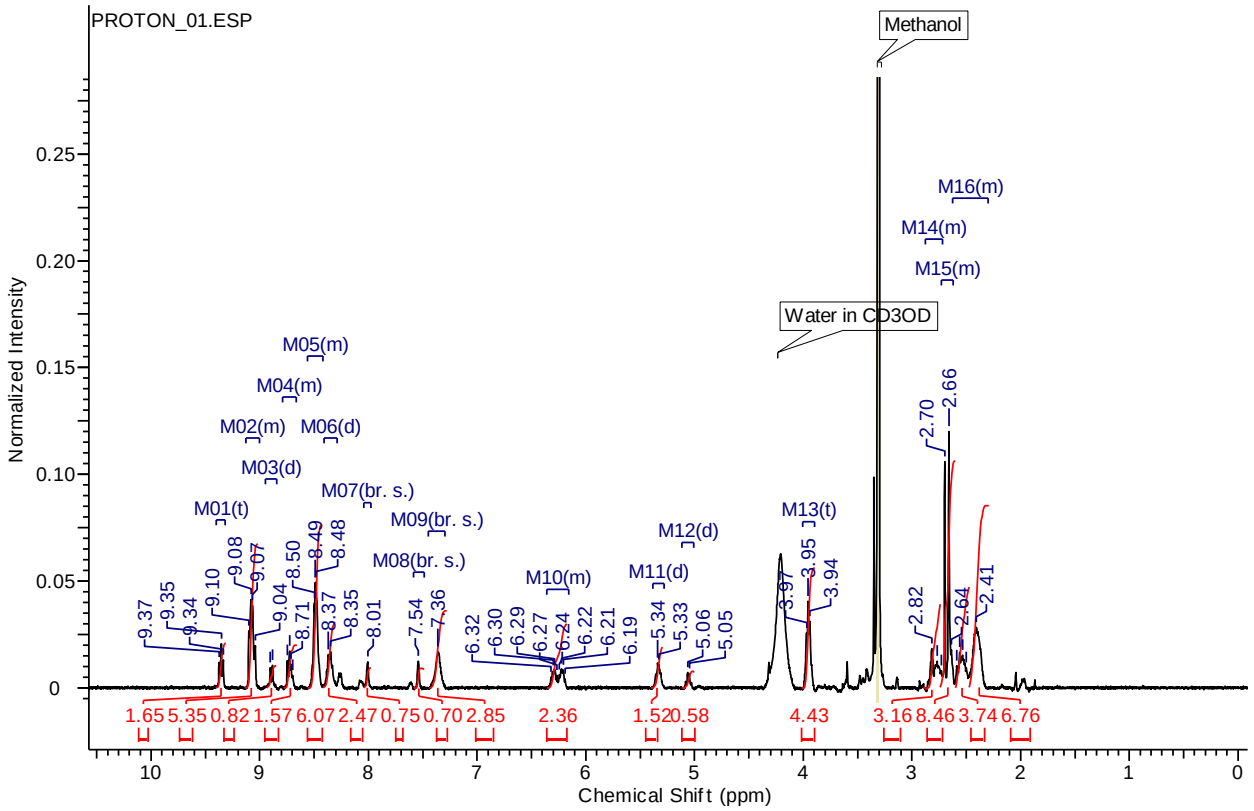


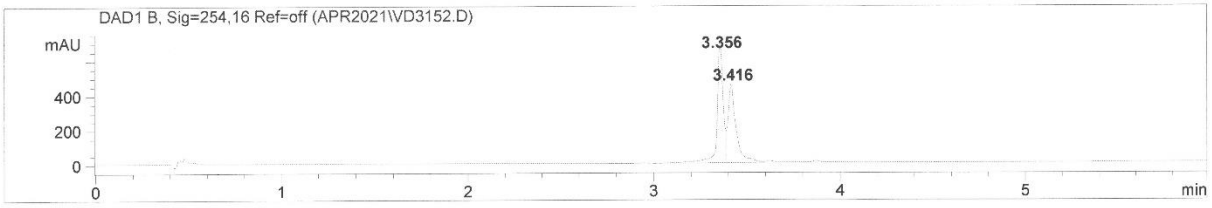


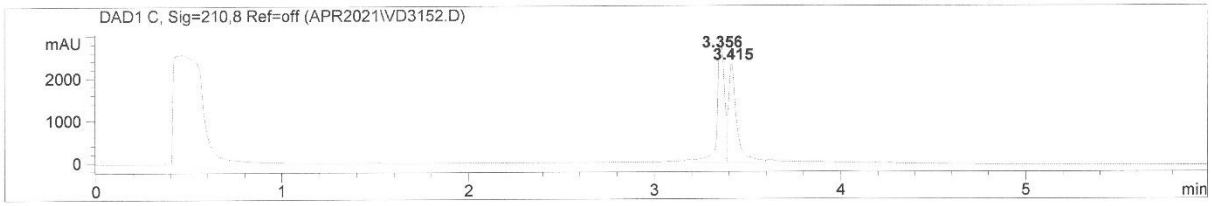


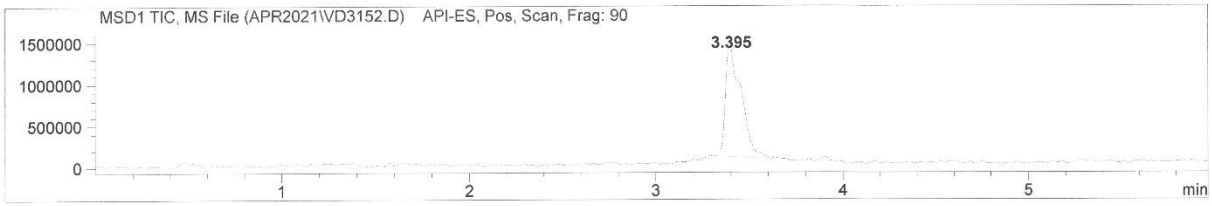


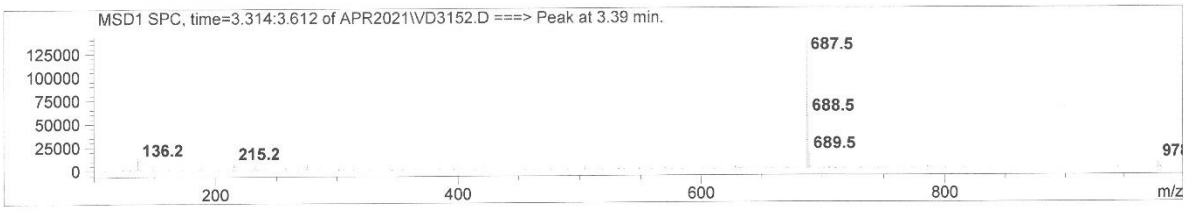
VD3152 (Compound 20)

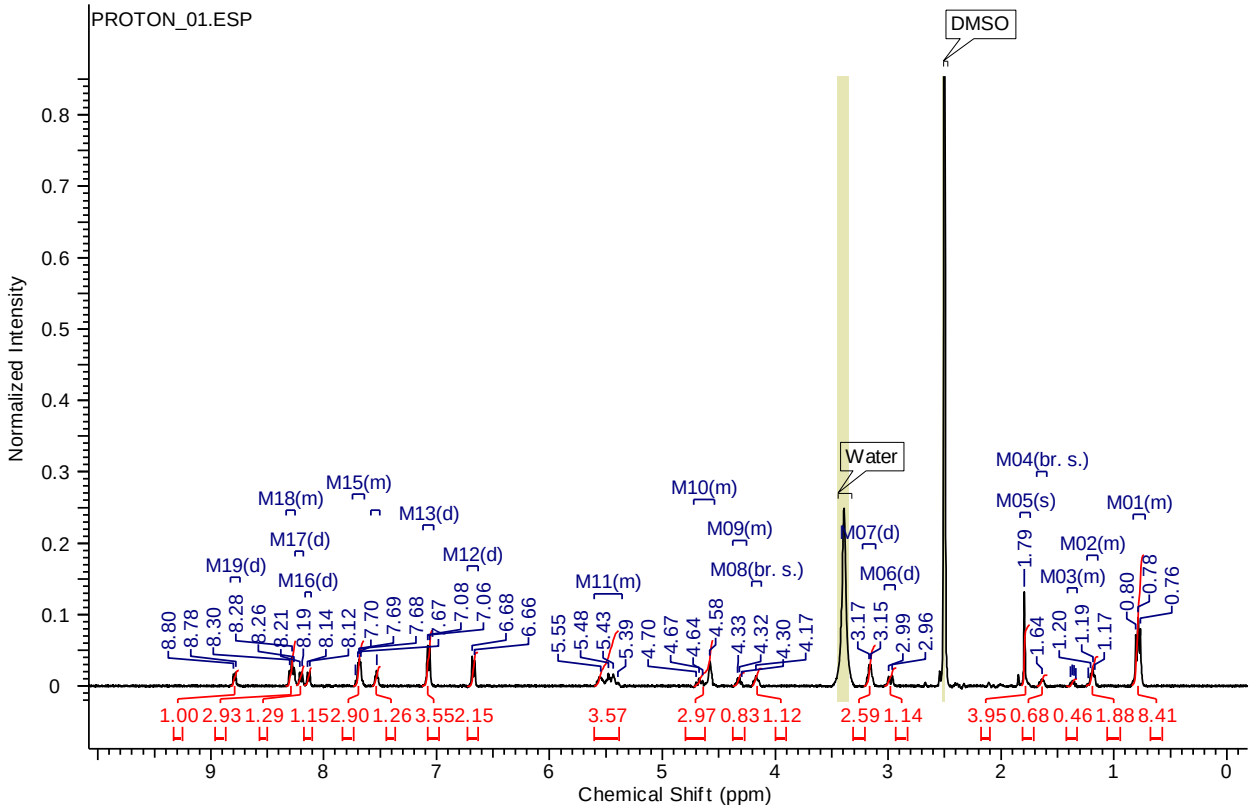


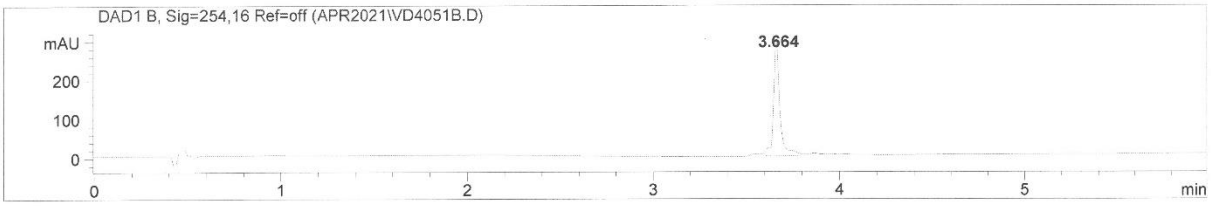


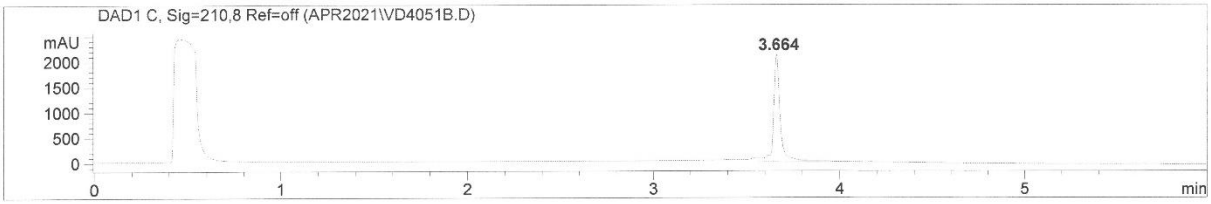


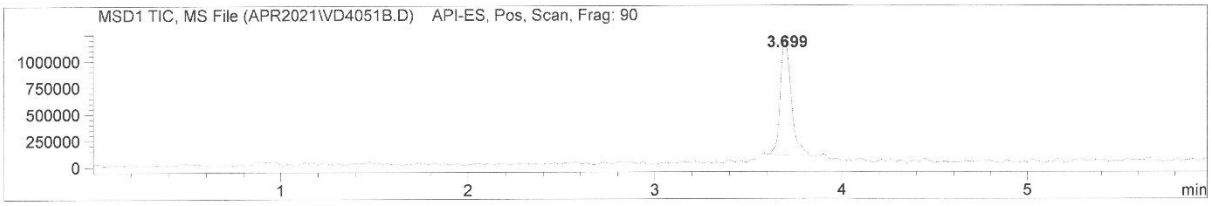


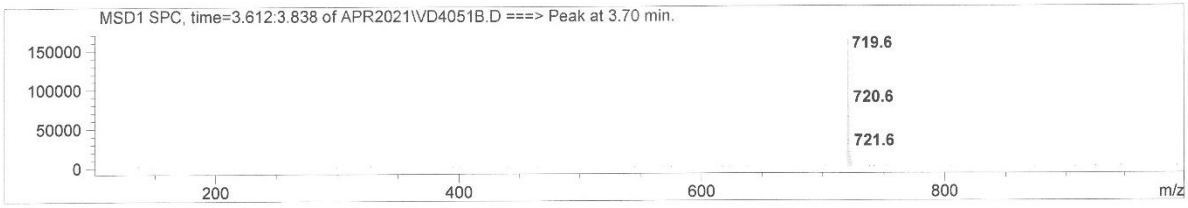



PROTON_01.ESP
DMSO
Water
Normalized Intensity
Chemical Shift (ppm)
M01(m)
M02(m)
M03(m)
M04(br. s.)
M05(s)
M06(d)
M07(d)
M08(br. s.)
M09(m)
M10(m)
M11(m)
M12(d)
M13(d)
M15(m)
M16(d)
M17(d)
M18(m)
M19(d)

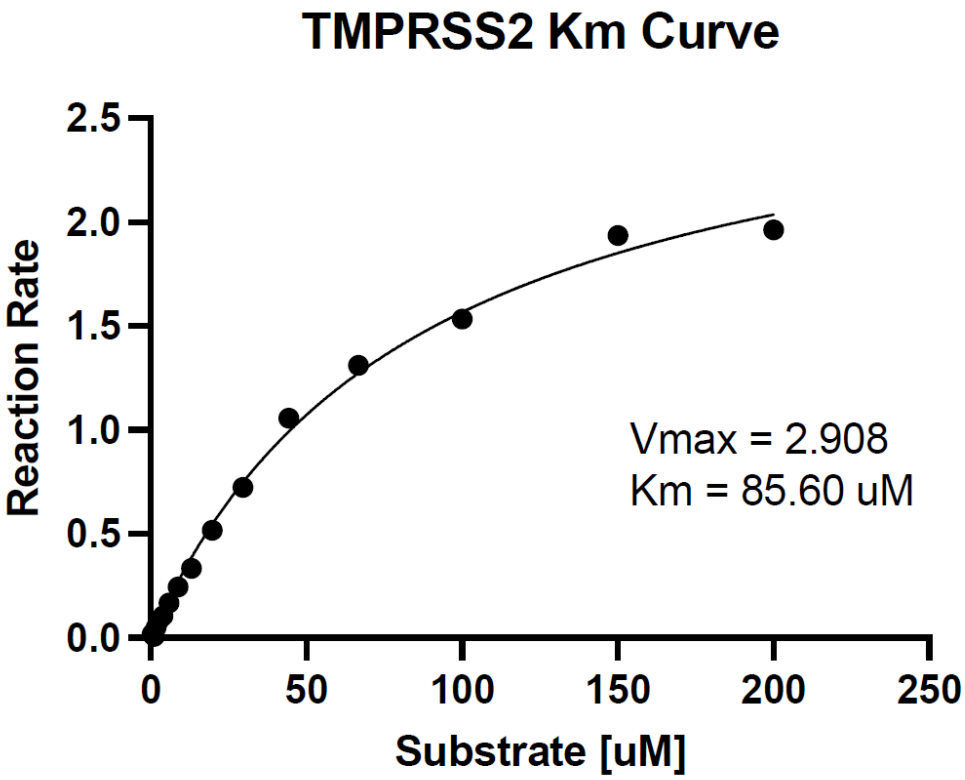

DAD1 B, Sig=254,16 Ref=off (APR2021\VD4051B.D)
mAU
3.664
min


DAD1 C, Sig=210,8 Ref=off (APR2021\VD4051B.D)
mAU
3.664
min

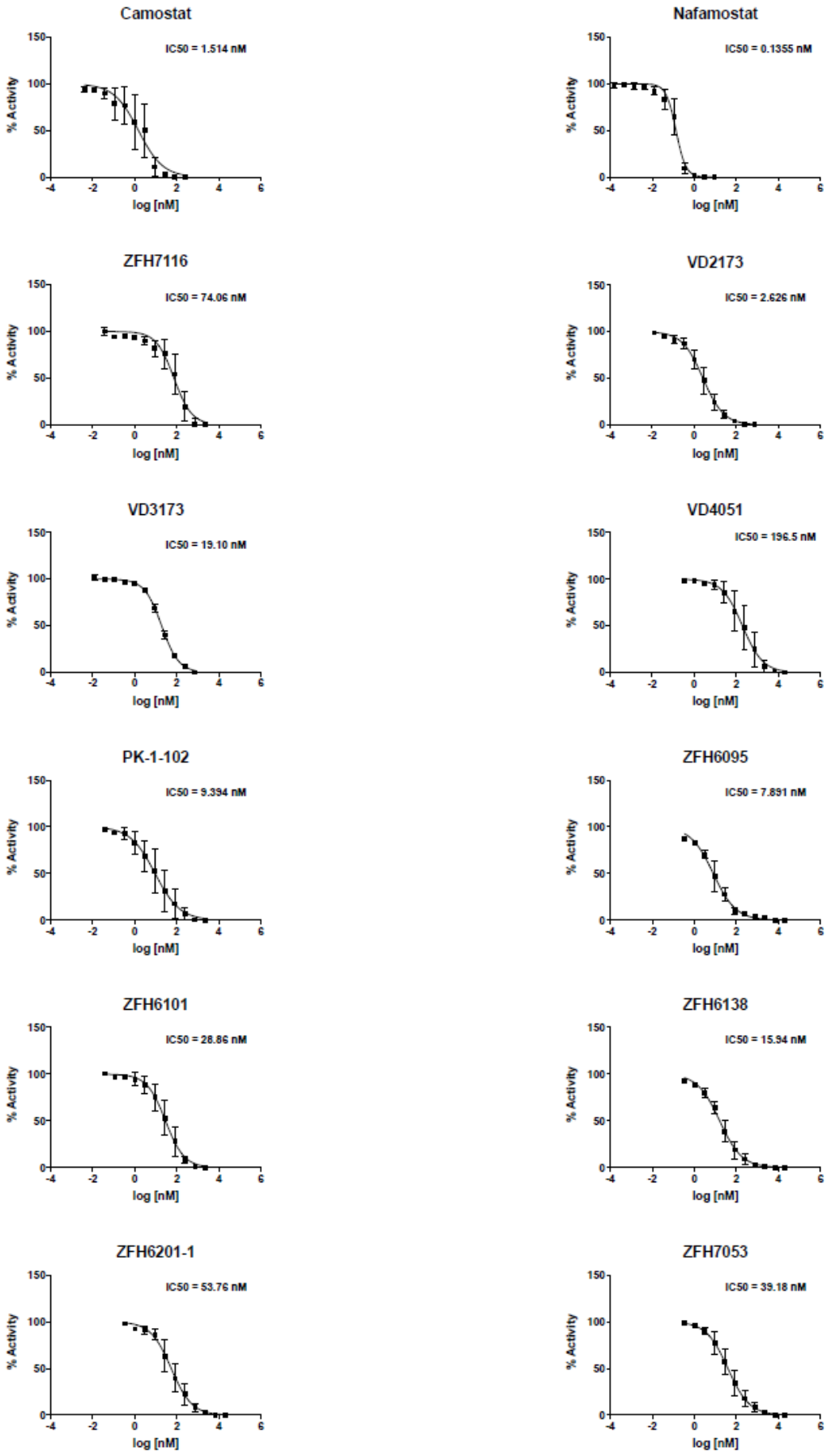

MSD1 TIC, MS File (APR2021\VD4051B.D) API-ES, Pos, Scan, Frag: 90
3.699
min


MSD1 SPC, time=3.612:3.838 of APR2021\VD4051B.D ===> Peak at 3.70 min.
719.6
720.6
721.6
m/z

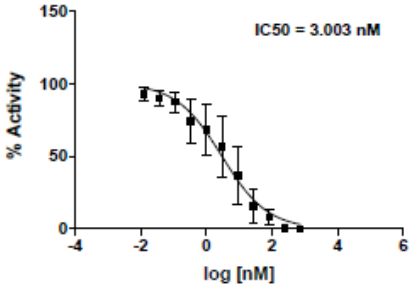
4. Figure A2.S2. K_m curve for Boc-QAR-AMC using full-length TMPRSS2 (Enzyme conc. 3 nM).



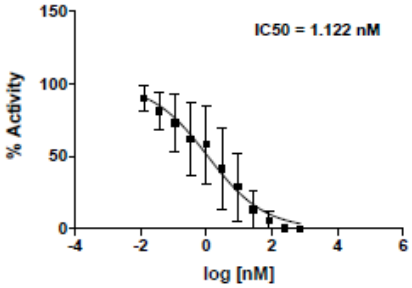
5. Figure A2.S3. IC_{50} inhibition curves of full-length TMPRSS2/Boc-QAR-AMC (Data in Table 1).



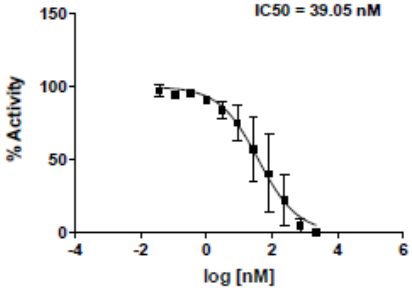
ZFH7063



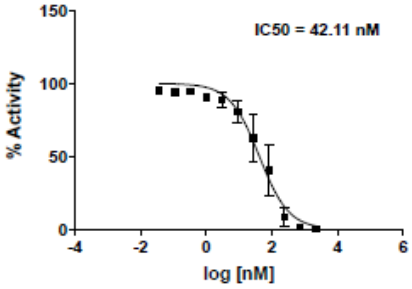
PK-1-89



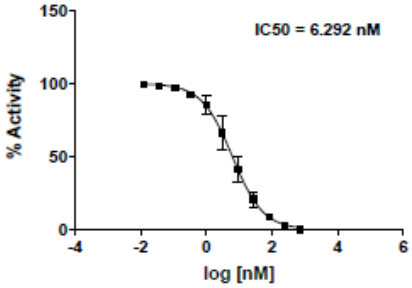
PK-1-18



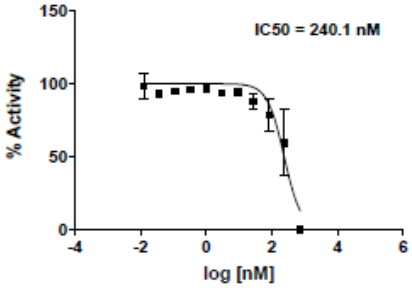
PK-1-45A1



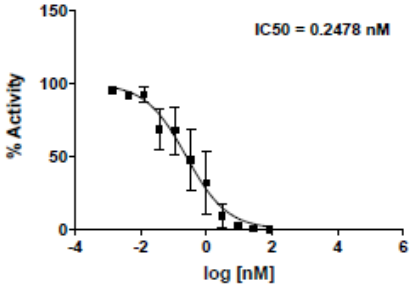
ZFH7064



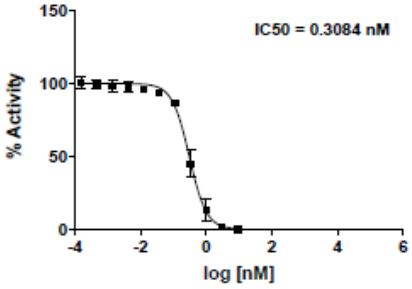
ZFH7006



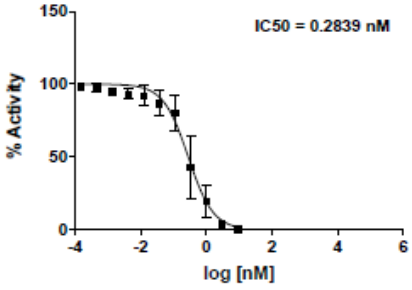
MM3116



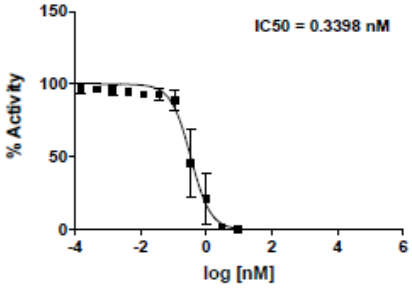
MM3144



MM3123



MM3122



6. Cell-based enzyme activity of compounds 1 and 2 in HEK-293 cells.

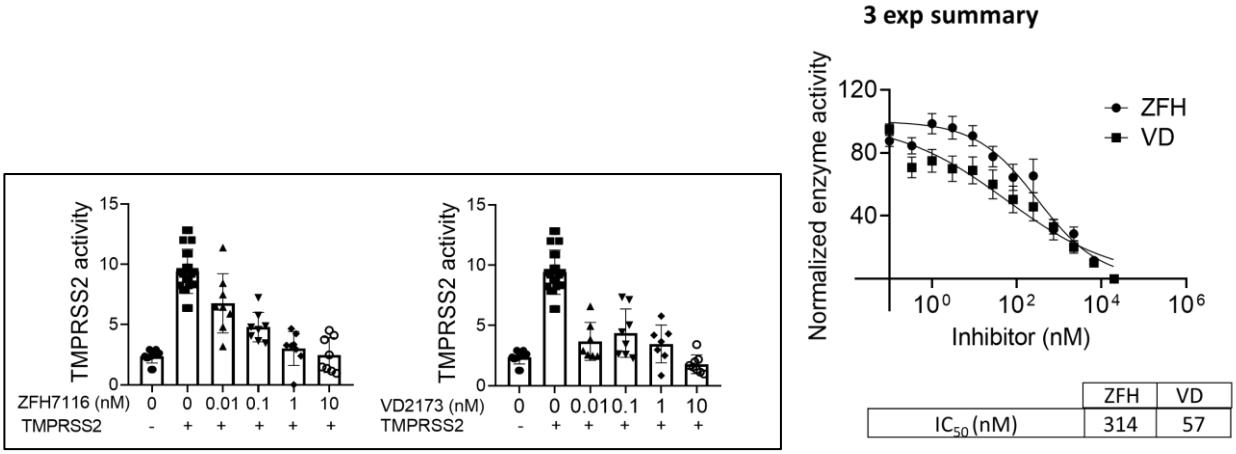


Figure A2.S4. Cell-based TMPRSS2 enzyme inhibition data of ZFH7116 (1) and VD2173 (2) in HEK-293 cells using Boc-QAR-AMC substrate.

7. Acute toxicity of MM3122 (4) data.

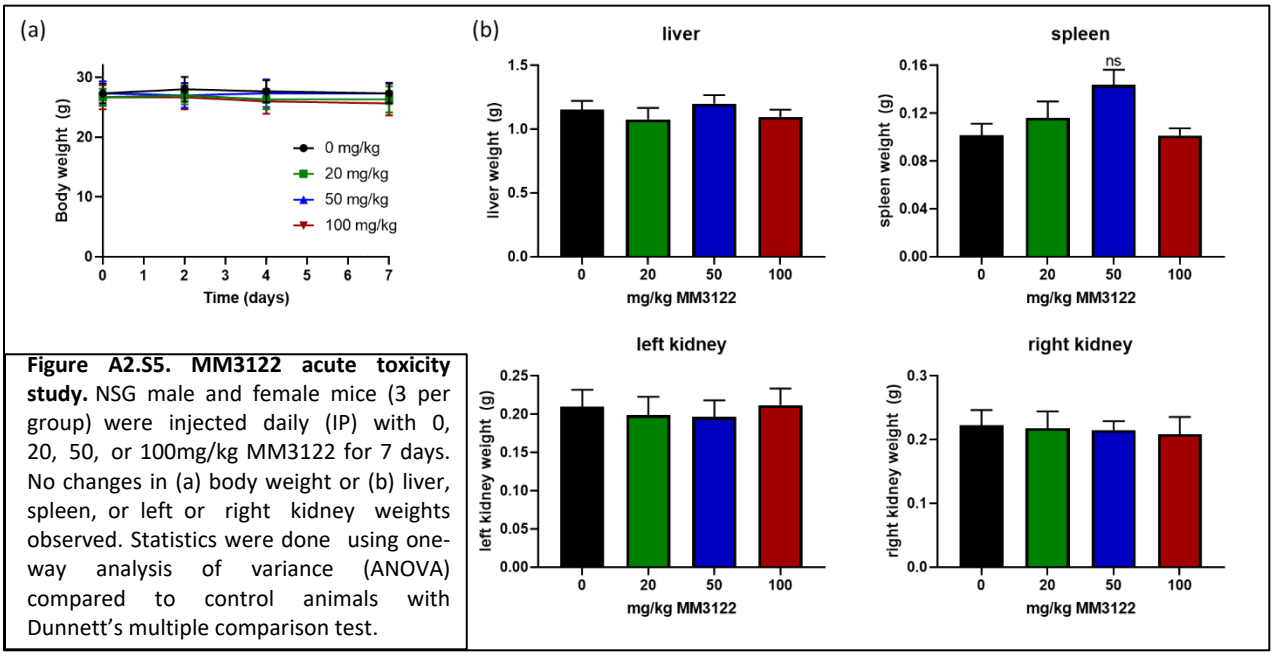


Figure A2.S5. MM3122 acute toxicity study. NSG male and female mice (3 per group) were injected daily (IP) with 0, 20, 50, or 100mg/kg MM3122 for 7 days. No changes in (a) body weight or (b) liver, spleen, or left or right kidney weights observed. Statistics were done using one-way analysis of variance (ANOVA) compared to control animals with Dunnett's multiple comparison test.

8. Activity of **1** and **2** and Camostat using Vero cells in pseudotype and chimeric VSV-SARS-CoV-2 viruses.

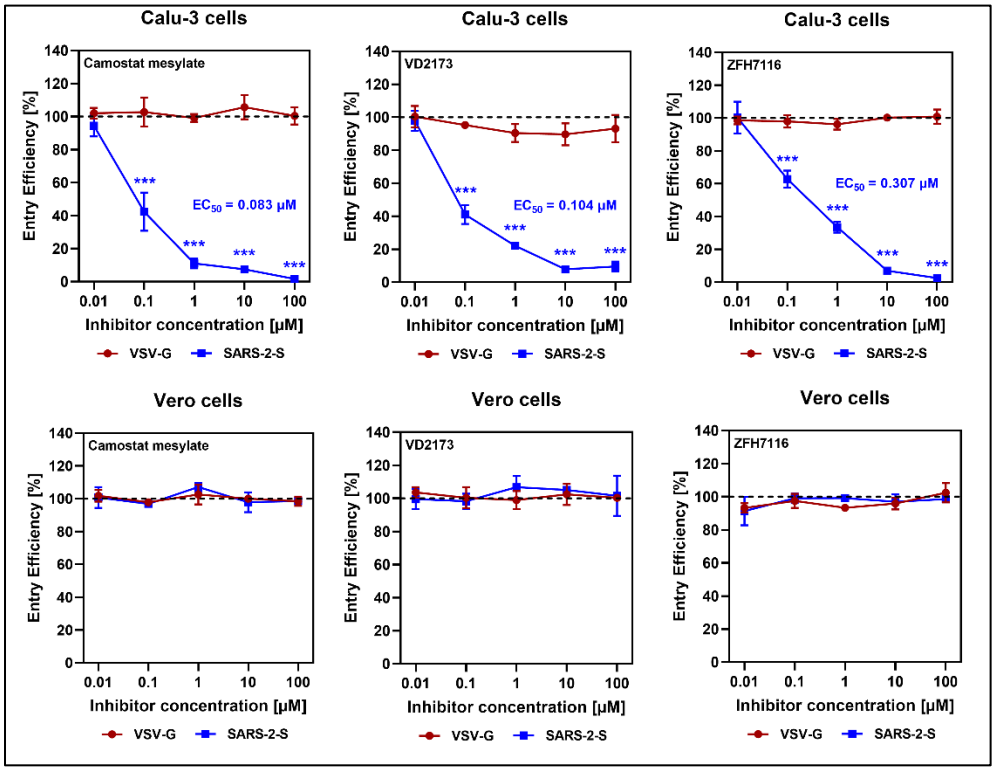


Figure A2.S6. Inhibition of SARS-CoV-2 cell entry into Calu-3 lung epithelial cells by ZFH7116 (1) and VD2173 (2) using VSV-SARS-CoV2-Spike protein Pseudotypes. EC_{50}s are calculated from an average of 3 separate experiments. Camostat, an irreversible serine protease inhibitor, was used as a positive control.

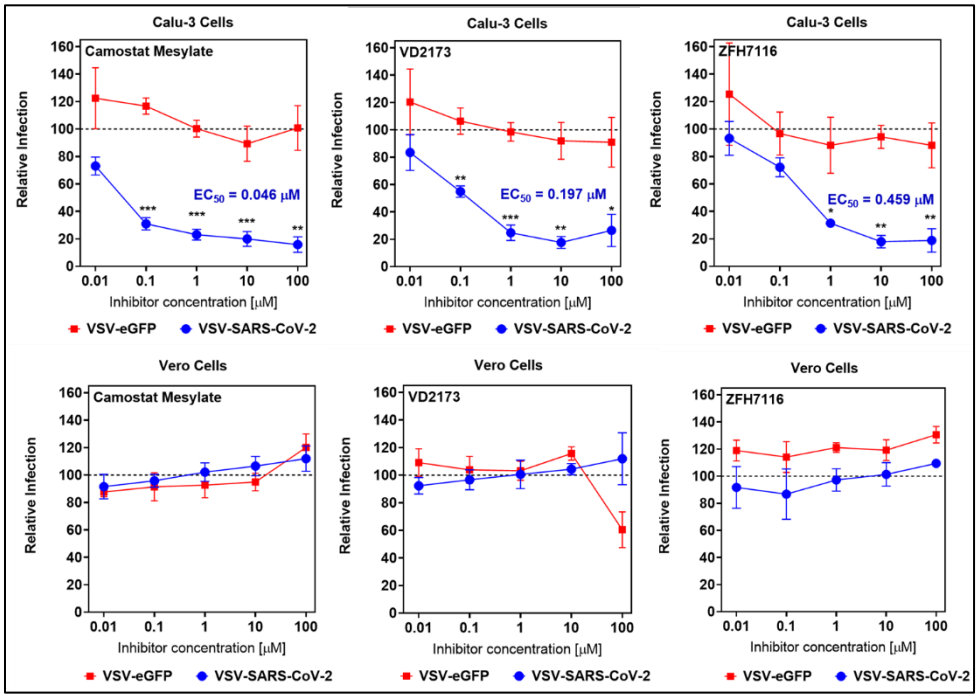


Figure A2.S7. Inhibition of SARS-CoV-2 cell entry into Calu-3 lung epithelial cells by ZFH7116 (1) and VD2173 (2) using VSV-SARS-CoV2-Spike protein Chimeras. EC_{50}s are calculated from an average of 3 separate experiments. Camostat, a non-selective protease inhibitor, was used as a positive control.

9. Expression and purification of TMPRSS2 protease domain.

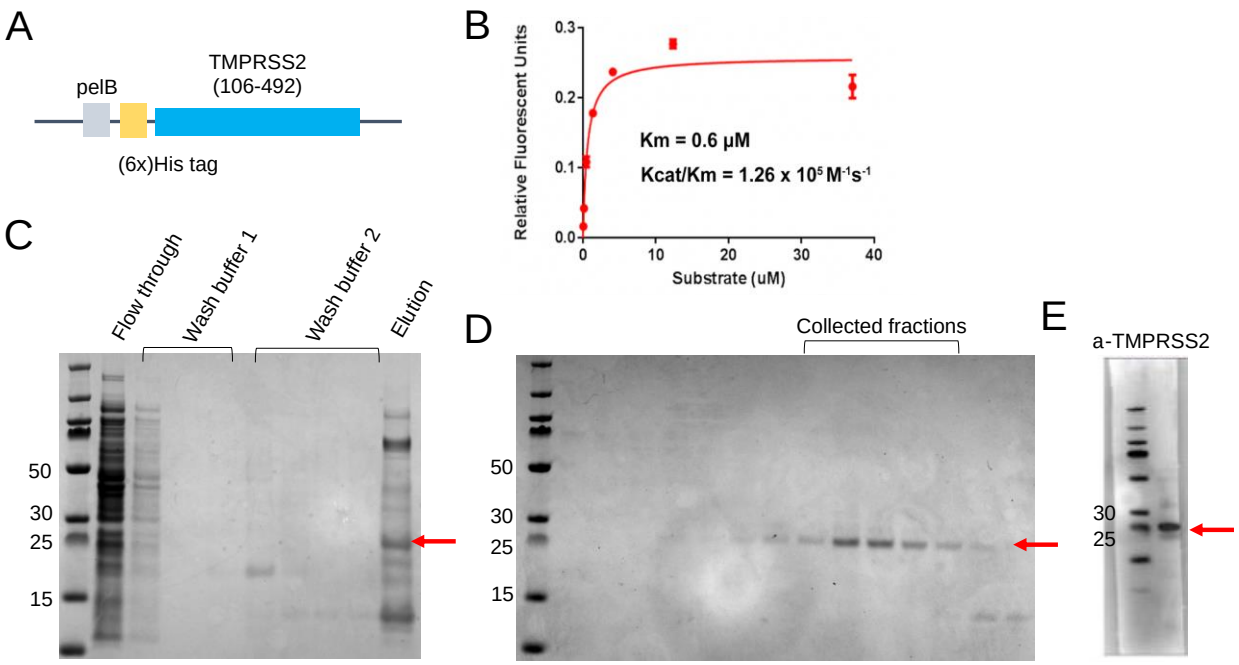


Figure A2.S8. Bacterial periplamic expression and purification of TMPRSS2-protease domain. (A) TMPRSS2-protease domain expression construct. N terminal fused pelB signalling peptide allows for secretion of the target protease to the periplasm allowing for correct folding and ease in purification downstream. (B) Michaelis-Menten curve of TMPRSS2-Protease domain against (MCA)-K-KARSAFA-K-(DnP). Initial velocities of peptide cleavage were plotted against substrate concentration. Kinetic values were calculated using GraphPad Prism. (C) Denaturing SDS-PAGE gel of Ni-NTA affinity purification of TMPRSS2-Protase domain. Correct band size for TMPRSS2-protease domain (~26kDa) shown in Red arrow. (D) SDS-PAGE gel of size exclusion chromatography fractions. Fractions only containing protease domain were collected and pooled. E) Western blot against pooled TMPRSS2-protease domain. Blot using monoclonal TMPRSS2-protease domain antibody (M05), clone 2F4 confirms purity of sample and no existence of degradation products.

Appendix 3:

Supplemental Material to Chapter 4

Fig A3.1: RNA-binding protein pulldown assay is specific for capped oligos. Lysates from A549 cells were bound to an A_m capped oligo in competition with the indicated 5' triphosphate or capped oligo lacking a 3' biotin. Bound protein was isolated by streptavidin pulldown, eluted by boiling in Laemmli buffer, run on a 5-15% SDS-PAGE gel, transferred to membrane, and western blot for eIF4E performed.

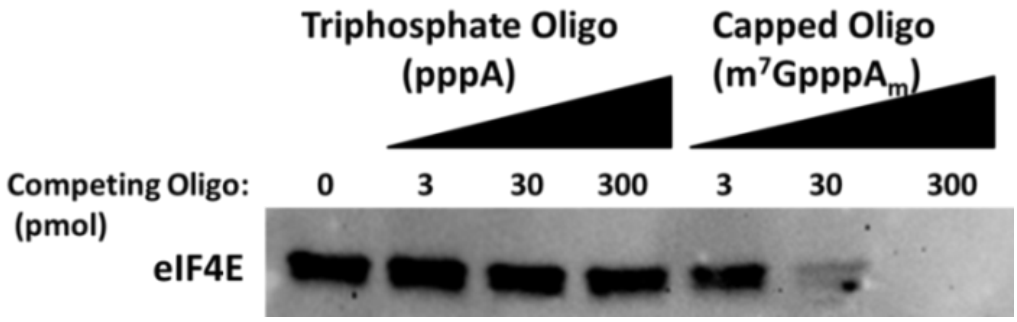


Table A3.1: Top 20 proteins bound to the indicated RNA oligos in the indicated cell lysate. PSC – protein spectrum counts.

HeLa cells:

WT m^6A_m	
Gene Name	**PSC**
PLEC	410
PYC	347
IF4G1	331
PTBP1	329
NONO	278
PRC2C	254
SFPQ	244
PABP1	228
GEMI5	198
IF4G3	180
FUBP2	177
ROA1	176
G3BP1	175
CELF1	173
ATX2L	162
PABP4	153
MATR3	150
YBOX1	145
FUBP1	141
Z3H7A	140

WT A_m	
Gene Name	**PSC**
PTBP1	440
IF4G1	417
PLEC	400
PYC	373
NONO	338
PRC2C	322
SFPQ	291
PABP1	287
G3BP1	235
GEMI5	228
IF4G3	212
EIF3A	204
ATX2L	185
PABP4	178
Z3H7A	143
SRRM2	138
NUFP2	136
ROA1	133
YBOX1	131
CSDE1	131

Table A3.1 (continued):

HeLa cells (continued):

KO m^6A_m	
Gene Name	**PSC**
PLEC	486
NONO	370
IF4G1	350
PTBP1	334
SFPQ	259
PRC2C	244
PABP1	210
FUBP2	204
FUBP3	192
PYC	187
GEMI5	179
IF4G3	176
CELF1	174
FUBP1	172
MATR3	161
G3BP1	160
ROA1	157
SND1	157
PABP4	149
ROA2	139

KO A_m	
Gene Name	**PSC**
PLEC	452
IF4G1	353
NONO	301
PRC2C	292
PTBP1	272
SFPQ	241
PABP1	236
IF4G3	199
GEMI5	192
PABP4	175
G3BP1	173
ATX2L	170
EIF3A	166
PYC	149
AHNK	147
MATR3	134
DDX5	132
SND1	131
YBOX1	115
PRC2A	112

Table A3.1 (continued):

A549 Cells:

WT m^6A_m	
Gene Name	**PSC**
PLEC	845
PTBP1	614
ROA1	578
FUBP2	424
HNRPQ	417
IF4G1	381
FUBP1	364
DYHC1	356
G3P	356
XRN2	333
PYC	324
PRKDC	318
ROA2	296
FLNA	290
GEMI5	290
ACLY	276
G3BP1	265
YBOX1	249
MAP1B	246
HNRL1	244

WT A_m	
Gene Name	**PSC**
PLEC	877
PTBP1	640
HNRPQ	510
IF4G1	398
DYHC1	383
PRKDC	322
PYC	312
FLNA	296
GEMI5	285
G3BP1	272
ACLY	256
AHNK	253
IFIT1	249
XRN2	239
FLNB	238
MAP1B	229
EIF3A	225
RL30	220
TRI25	215
PABP1	204

Table A3.1 (continued):

A549 cells (continued):

KO m^6A_m	
Gene Name	PSC
PLEC	915
ROA1	536
HNRPQ	361
PTBP1	319
G3P	310
IF4G1	306
DYHC1	305
PRKDC	301
ROA2	299
FUBP1	268
FLNA	267
PYC	263
RL30	243
XRN2	239
IFIT1	231
ACLY	224
G3BP1	218
FUBP2	215
MAP1B	211
GEMI5	210

KO A_m	
Gene Name	**PSC**
PLEC	1070
PTBP1	568
DYHC1	427
IF4G1	410
PRKDC	371
FLNA	325
PYC	324
HNRPQ	299
AHNK	299
RL30	272
FLNB	265
MAP1B	261
A0A0A0MTC1	255
ACLY	252
G3P	246
GEMI5	241
G3BP1	240
IFIT1	235
XRN2	233
DESP	232

Methods for Figures in Chapter 4: Discussion

Metabolic Radiolabeling of mRNA:

Radiolabeling was performed as previously described. Briefly, cells were infected with VSV at a MOI of 5 in serum/phosphate-free DMEM (Gibco #11971-025). 10 μg ml^{-1} Actinomycin D (Sigma #A5156) was added at 2.5 hpi, and at 3 hours post infection, 100 μCi ml^{-1} [^{32}P] phosphoric acid added to label newly synthesized RNA (Perkin Elmer #NEX053H). RNA was harvested in Trizol (Thermofisher #15596018) and poly(A)+ mRNA selected using the NEB Magnetic mRNA Isolation Kit (NEB S1550S).

Identification of methylated nucleotide levels by two-dimensional thin layer chromatography (2D-TLC):

2D-TLC was performed as previously described. Briefly, RNA was sequentially digested with 2 units of nuclease P1 (Sigma N8630) for 3 h at 37 ºC in water, then with 2 units of Cap-Clip Acid Pyrophosphatase for 3 h (Cell Script #C-CC15011H) in the manufacturer's buffer. Samples were spotted on PEI-F cellulose TLC plates. Plates were developed in the first dimension with 5 parts isobutyric acid (Sigma #I1754) to 3 parts 0.5 M ammonia (VWR #BDH153312K) for 14 h, and in the second dimension with a solvent of 70 parts isopropanol, 15 parts hydrochloric acid, and 15 parts water for 20 h. RNA species were positively identified by UV-shadowing (254 nm) of co-spotted (non-radioactive) commercially available standards (5' monophosphate forms).

Cap-Proximal Nucleotide Labeling of IFN-β induced mRNA:

Nascent mRNAs synthesized under mock or IFN-β treatment were isolated using a Click-iT Nascent RNA Capture Kit (Thermofisher #C10365). Wild type (parental) or

PCIF1 KO A549 cells were treated with vehicle (0.1% BSA) or 500 U ml-1 IFN-β for 8 hours, and at 1 hour post treatment, labeled with 2.5 mM 5-ethynyl uridine (5-EU). RNA was harvested in Trizol (Thermofisher #15596018) and poly(A)+ mRNA selected using the NEB Magnetic mRNA Isolation Kit (NEB S1550S). EU(+) mRNAs were then biotinylated and bound to streptavidin beads according to manufacturer's instructions. These mRNAs were then eluted from the beads by boiling at 95 °C in 95% formamide, 10 mM EDTA, and trizol extracted into nuclease-free water. Concentration was determined by Qubit RNA HS assay (Thermofisher #Q32855).

50 ng of this RNA was then used for cap-proximal nucleotide labeling. RNA was decapped with 200 units yDcpS (NEB #M0463S) for 1 h at 37 °C, heat inactivated at 70 °C for 15 minutes, dephosphorylated with shrimp alkaline phosphatase (NEB #M0371S), heat inactivated at 65 °C for 5 minutes, and phosphorylated with 10 μCi $^{[32]}$P γ-ATP and polynucleotide kinase (NEB M0201S). RNA was extracted in Trizol, unincorporated nucleotides removed by passing through a G50 column (Cytiva #27533002), digested to mononucleotides with nuclease p1, and 2D-TLC performed as above.

Nuclear Cytoplasmic Fractionation:

Cells were resuspended in cytoplasmic lysis buffer (30 mM HEPES pH 7.4m 2 mM MgOAc, 0.1% NP40, 5 mM DTT, Roche protease inhibitor cocktail) and lysed by passage through a 27G needle three times. This lysate was then centrifuged for 20 minutes at 500 x g at 4 °C. The supernatant (cytoplasmic fraction) and removed, and the pellet washed twice in cytoplasmic lysis buffer. The pellet was then lysed in RIPA buffer (25 mM Tris pH 7.6, 150 mM NaCl, 1% NP40, 1% sodium deoxycholate, 0.1%

SDS) and sonicated for 5 minutes. Cell debris was cleared by centrifugation at 12,000 xg for 10 minutes, and the supernatant was collected (nuclear fraction). Western blots were performed using antibodies against FLAG (mouse – Sigma F3165), GAPDH (mouse – Sigma G8795), and Lamin B1 (rabbit – Proteintech 12987).

Immunoflourescence Microscopy:

Cells were plated on coverslips, and infected with VSV at a MOI 3. At 6 hpi, cells were fixed in 2% paraformaldehyde for 15 minutes, and washed in PBS twice. Cells were stained with mouse anti-FLAG (Sigma F3165) and Alexa Flour Goat anti-mouse 488, as well as Hoeschst to stain nuclei. Coverslips were mounted on slides, and visualized on a Nikon Ti inverted microscope with W1 Yokogawa Spinning disk (50 um pinhole) in association with the Harvard Microscopy Resources on the North Quad (MicRON) facility.

RNA-seq Expression Analysis:

RNA was extracted in Trizol, and integrity verified by the Qubit RNA IQ assay kit (Thermofisher Q33221). RNA-seq was performed by the Washington University in St. Louis Genome Technology Access Center (GTAC) as follows after poly(A) selection (biological duplicate).

Samples were prepared according to library kit manufacturer's protocol, indexed, pooled, and sequenced on an Illumina NovaSeq 6000. Basecalls and demultiplexing were performed with Illumina's bcl2fastq software and a custom python demultiplexing program with a maximum of one mismatch in the indexing read. RNA-seq reads were then aligned to the Ensembl release 76 primary assembly with STAR version 2.5.1a1. Gene counts were derived from the number of uniquely aligned unambiguous reads by

Subread:featureCount version 1.4.6-p52. Isoform expression of known Ensembl transcripts were estimated with Salmon version 0.8.23. Sequencing performance was assessed for the total number of aligned reads, total number of uniquely aligned reads, and features detected. The ribosomal fraction, known junction saturation, and read distribution over known gene models were quantified with RSeQC version 2.6.24.

All gene counts were then imported into the R/Bioconductor package EdgeR5 and TMM normalization size factors were calculated to adjust for samples for differences in library size. Ribosomal genes and genes not expressed in the smallest group size minus one samples greater than one count-per-million were excluded from further analysis. The TMM size factors and the matrix of counts were then imported into the R/Bioconductor package Limma6. Weighted likelihoods based on the observed mean-variance relationship of every gene and sample were then calculated for all samples with the voomWithQualityWeights7. The performance of all genes was assessed with plots of the residual standard deviation of every gene to their average log-count with a robustly fitted trend line of the residuals. Differential expression analysis was then performed to analyze for differences between conditions and the results were filtered for only those genes with Benjamini-Hochberg false-discovery rate adjusted p-values less than or equal to 0.05. Graphs shown display CPM (counts per million reads mapped).

RNA-Binding Protein Pulldown and Liquid Chromatography - Mass Spectrometry Analysis:

RNA binding proteins bound to the VSV 5’ UTR were isolated as follows. An oligonucleotide corresponding to the first 30 nt of the M 5’ UTR and modified with A 5’ cap1 and 3’ biotin (biotin BB) was purchased from Tri-link

($m^7GpppA_mACAGAUAUCACAAUCUAAGUGUUAUCCCA$-BioBB). This oligo was then methylated or mock methylated with purified PCIF1 as previously described. This methylation was performed sequentially three times, with the second two reactions in the presence of SAH-hydrolase (Abcam 99326). The cap-proximal nucleotide of the final product was analyzed by 2D-TLC, and was found to be over 90% m^6A_m, while the mock methylated oligo was 100% A_m. A549 or HeLa cells were treated with IFN-β (500 units, 50 ng/ml) for 5 hours. These cells were then lysed in Pierce IP lysis buffer (Thermofisher #87787) according to manufacturer's instructions. These lysates were then used in conjuction with the capped oligos to identify RNA-binding proteins which bound to those oligos using a Pierce Magnetic RNA-protein Pulldown Kit (Thermofisher #20164), according to manufacturer's instructions.

Briefly, for each reaction, 50 pmol RNA oligo were pound to 50 ul of streptavidin beads. After washing, the RNA-bound beads were added to cell lysate in the following reaction (100 µl volume) for 1 hour at 4 °C: 150 µg (HeLa) or 100 µg (A549) lysate, 15% glycerol, 20 mM Tris (pH 7.5), 50 mM NaCl, 2 mM $MgCl_2$, 0.1% Tween-20. RNA-protein bound beads were washed 3 times with 20 mM Tris, 10 mM NaCl, and 0.1% Tween-20 as wash buffer. They were then washed once without detergent (20 mM Tris, 10 mM NaCl), resuspended in 30 µl of this same buffer. Following on-bead digestion in trypsin, liquid chromatography followed by mass spectrometry was performed by the Washington University Proteomics Core Laboratory (see below methods).

The expert technical assistance of Petra Erdmann-Gilmore, Dr. Yiling Mi and Rose Connors is gratefully acknowledged for these experiments. The proteomic experiments were performed at the Washington University Proteomics Shared

Resource (WU-PSR), R Reid Townsend MD.PhD., Director and Robert Sprung, PhD., Co-Director). The WU-PSR is supported in part by the WU Institute of Clinical and Translational Sciences (NCATS UL1 TR000448), the Mass Spectrometry Research Resource (NIGMS P41 GM103422; R24GM136766) and the Siteman Comprehensive Cancer Center Support Grant (NCI P30 CA091842).

Peptide preparation:

The peptides were prepared using a previously described method for on-bead tryptic digestion (Kim et al. PMID: 26912792). The beads were washed four times with 1 ml of 50 mM ammonium bicarbonate buffer (ABC; pH = 8.0) (ABC). The washed beads were resuspended in 40 µl of ABC buffer containing 8 M urea. The protein disulfide bonds were reduced using 2 µl of 0.5 M TCEP (Tris (2-carboxyethyl) phosphine) and incubation for 60 min at 30 ºC. The reduced proteins were alkylated using 4 µl of a 0.5 M solution of iodoacetamide with incubation for 30 min at RT in the dark. The urea was diluted to a concentration of 1.5 M by adding 167 µl of 50 mM ABC buffer. After addition of LysC (1mAU), the samples were incubated for 2 h at 30 °C in a Thermomixer with gyration at 750 rpm. Trypsin (1 µg) was added and the samples were incubated overnight at 30 °C in the Thermomixer gyrating at 750 rpm. The peptides were transferred to a 1.5 ml tube, the beads were washed with 50 µl of ABC buffer and the transfer and wash volumes were combined. Any residual detergent was removed by ethyl acetate extraction (Erde et al, PMID 28188519). The peptide samples were acidified with TFA (Trifluoroacetic acid) to a final concentration of 1% (vol/vol) TFA (pH < 2.0). The pH was checked with pH paper. The peptides were desalted using two micro-tips (porous graphite carbon, BIOMETNT3CAR) (Glygen) on a Beckman robot

(Biomek NX), as previously described (Chen Mol. Cell Proteomics PMID 22338125). The peptides were eluted with 60% (vol/vol) MeCN in 0.1% (vol/vol) TFA. After adding TFA to a final concentration of 5%, the peptides were dried in a Speed-Vac (Thermo Scientific, Model No. Savant DNA 120 concentrator). The peptides were dissolved in 20 µl of 1% (vol/vol) MeCN (acetonitrile) in water. An aliquot (10%) was removed for quantification using the Pierce Quantitative Fluorometric Peptide Assay kit (Thermo Scientific, Cat. No. 23290). The remainder was transferred to autosampler vials (Sun-Sri, Cat. No. 200046), dried and stored at -80 ºC.

UPLC-timsTOF Mass spectrometry:

The peptides were analyzed using a nano-Elute chromatograph coupled online to a hybrid trapped ion mobility-quadrupole time of flight mass spectrometer (timsTOF Pro, Bruker Daltonics, Bremen Germany) with a modified nano-electrospray source (CaptiveSpray, Bruker Daltonics). The mass spectrometer was operated in PASEF mode (PMID30385480). The samples in 1% (vol/vol) aqueous FA were loaded (2 µL) onto a 75 µm i.d. × 25 cm Aurora Series column with CSI emitter (Ionopticks) on a Bruker nano-ELUTE (Bruker Daltonics). The column temperature was set to 50 °C. The column was equilibrated using constant pressure (800 bar) with 8 column volumes of solvent A (0.1% (vol/vol) aqueous FA). Sample loading was performed at constant pressure (800bar) at a volume of 1 x sample pick-up volume plus 2 µL. The peptides were eluted using the one column separation mode with a flow rate of 400 nL/min and using solvents A and B (0.1% (vol/vol) formic acid/MeCN): solvent A containing 2%B increased to 15% B over 60 min, to 25% B over 30 min, to 35% B over 10 min, to 80% B over 10 min and constant 80% B for 10 min. The MS1 (mass spectra of peptide

precursors) and MS2 (fragmentation mass spectrum of peptide from precursor ion) spectra were recorded from m/z 100 to 1700. Suitable precursor ions for PASEF-MS/MS (parallel accumulation-serial fragmentation) were selected in real time from TIMS-MS survey scans by a PASEF scheduling algorithm (PMID 30385480). A polygon filter was applied to the m/z and ion mobility plane to select features most likely representing peptide precursors rather than singly charged background ions. The quadrupole isolation width was set to 2 Th for m/z < 700 and 3 Th for m/z > 700, and the collision energy was ramped stepwise as a function of increasing ion mobility: 52 eV for 0–19% of the ramp time; 47 eV from 19–38%; 42 eV from 38–57%; 37 eV from 57–76%; and 32 eV for the remainder. The TIMS elution voltage was calibrated linearly using the Agilent ESI-L Tuning Mix (m/z 622, 922, 1222).

Identification of Proteins:

Unprocessed data from the mass spectrometer were converted to peak lists using Proteome Discoverer (version 2.1.0.81, Thermo-Fischer Scientific). The MS2 spectra with charges +2, +3 and +4 were analyzed using Mascot software (Matrix Science, London, UK; version 2.5.1) {Perkins, 1999 #394}. Mascot was set up to search against a UniProt (ver July 2019) database of human proteins (20,667 entries), assuming the digestion enzyme was trypsin with a maximum of 4 missed cleavages allowed. The searches were performed with a fragment ion mass tolerance of 20 ppm and a parent ion tolerance of 20 ppm. Carbamidomethylation of cysteine was specified in Mascot as a fixed modification. Deamidation of asparagine, deamidation of glutamine, formation of pyro-glutamic acid from N-terminal glutamine, acetylation of protein N-terminus, oxidation of methionine, and pyro-carbamidomethylation of N-terminal

cysteine were specified as variable modifications. Peptides and proteins were filtered at 1% peptide-level false-discovery rate (FDR) in Scaffold (Proteome Software, Inc., Portland, OR, USA; v 4.11.1).

Capped RNA Competition:

A 27 nt oligonucleotide corresponding to the 5' VSV leader sequence containing a 5' triphosphate was capped and 2'O methylated using the vaccinia capping enzyme and 2'O methyltransferase (NEB #M2080S, #M0366). Remaining triphosphate oligos were digested by sequential treatment with shrimp alkaline phosphatase, polynucleotide kinase, and 5' monophosphate-dependent terminator exonuclease as previously described. RNA-binding protein pulldown was done as described above, but only with 15 pmol of the A_m capped oligo bound to the beads, and the binding reaction was supplemented with 3 pmol, 30 pmol, or 300 pmol competing triphosphate or capped oligo. RNA-binding proteins were eluted by boiling in 1X Laemmli-buffer, run on a precast 4-15% PAGE gel (Biorad #4561084), transferred to membrane, and western blot performed for eIF4E (Cell Signaling #9742).

Mouse:

Mouse experiments were performed by Konstantinos Boulias, Paul Rothlauf, and Michael Tartell. 4 week old C57/BL6 mice were infected intranasally with VSV in a volume of 10 µl of phosphate-buffered saline (PBS) after anesthesia with isoflurane. These mice were grouped according to sex and genetic background (wild type or PCIF1 knockout), and VSV infections at 10^2, 10^3, 10^4, 10^5, or 10^6 PFU were carried out in the aforementioned volume after dilution in PBS. Mice were monitored daily for weight and severity of disease for 14 days, and were immediately sacrificed if they showed greater

than 80% body weight loss or severe disease indicators (e.g. paralysis). Blood was taken from sacrificed mice by cardiac puncture, serum isolated by centrifugation at 2000 x g for 10 minutes, and stored at -80 °C. Dissected brains were pulverized through a 100 µm cell strainer in 0.5 ml PBS and stored at -80 °C. Plaque assays were performed on vero cells to identify viral titers in these samples by crystal violet staining.