Characterizing protein modifications and interactions in viral infections

by

Shu Yue Luo

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

University of Alberta

Abstract

Viral diseases are a constant threat that have profound impact on global health, economics and societies. Understanding virus-host interactions is key for advancing our knowledge of viral replication and pathogenesis, and thus for the development of effective antiviral strategies to better prepare us for future outbreaks. This thesis investigates the roles of post-translational modifications and protein-protein interactions during viral infections using advanced mass spectrometry techniques. Post-translational modifications such as phosphorylation, ubiquitination and proteolysis, significantly alter the function, stability, and localization of both viral and host proteins, thereby modulating the virus cycle and host immune responses. Virus-host protein-protein interactions form complex networks that viruses may exploit to manipulate host cellular machinery to their advantage, or the host utilizes for activating antiviral defense.

In Chapter 2, we identified potential human substrates targeted by the two SARS-CoV-2 viral proteases. We found that the main protease cleaved bromodomain-containing protein 2, while the papain-like protease cleaved splicing factor, proline- and glutamine-rich. Chapter 3 examines the ubiquitination landscape during vaccinia virus infection, highlighting critical host restriction factors that are ubiquitinated during viral infection, such as tripartite motif containing 25. Chapter 4 explores protein-protein interactions between the Mayaro virus and human host. We performed co-immunoprecipitation of viral proteins while simultaneously infecting the human cells with the virus, providing a comprehensive map of the viral protein interactome during infection within host cells.

Our findings underscore the importance of post-translational modifications and protein-protein interactions in virus-host dynamics, offering potential therapeutic targets and strategies for antiviral drug development.

Preface

This thesis encompasses findings that stem from collaborative research efforts.

Some sections of Chapter 1 have been published as:

> Luo, S. Y.; Araya, L. E.; Julien, O. Protease Substrate Identification Using N-terminomics. *ACS Chem Biol* **2019**, *14* (11): 2361-2371. DOI: 10.1021/acschembio.9b00398

I was one of the co-first authors and contributed to the literature research and manuscript composition equally with L.E.A.

Chapter 2 was adapted from a published manuscript:

> Luo, S. Y.; Moussa, E. W.; Lopez-Orozco, J.; Felix-Lopez, A.; Ishida, R.; Fayad, N.; Gomez-Cardona, E.; Wang, H.; Wilson, J. A.; Kumar, A.; Hobman, T. C.; Julien, O. Identification of Human Host Substrates of the SARS-CoV-2 M^{pro} and PL^{pro} Using Subtiligase N-Terminomics. *ACS Infect Dis* **2023**, *9* (4): 749-761. DOI: 10.1021/acsinfecdis.2c00458

I was responsible for experimental design and optimizations, mass spectrometry analysis, functional studies and manuscript composition. E.W.M. performed experiments related to PLpro. A.K., T.C.H. and O.J. contributed to conceptualization. Other authors assisted in experiments such as sample collection, viral infection and stable cell line generation.

Chapter 3 was adapted from a manuscript in preparation:

> Dong, J.; Luo, S. Y.; Smyth, S.; Melvie, G.; Julien, O.; Ingham, R. J. Characterizing changes in protein ubiquitylation during vaccinia virus infection.

I was responsible for method optimizations and sample preparation in mass spectrometry analysis, as well as statistical analysis. J.D. performed the viral infections, and J.D. and G.M. performed immunoblotting analysis in this Chapter. O.J. and R.J.I. contributed to conceptualization and experimental design. Other authors assisted in sample collection and viral infection.

Chapter 4 consists of unpublished work performed both by me and Joaquin Lopez-Orozco. I performed mass spectrometry analysis, statistical analysis and manuscript composition. J.L. assisted with many experiments such as electroporation and viral infection. Drs. Tom Hobman and Olivier Julien were involved with the experimental design.

Many figures were created using BioRender.

Acknowledgements

I would like to first express my sincere gratitude to my Ph.D. supervisor, Dr. Olivier Julien, for admitting me as one of the starting members of the Julien lab. Building a lab was hard, but it was also an unforgettable experience filled with both tears of frustration and joy. It was a transformative journey that significantly enhanced my skill set, impacted my career choice and shaped my character. Your guidance has been instrumental in structuring this thesis and fostering my growth as a scientist.

My heartfelt thanks also go to all the current and former Julien lab mates, especially my lab bench neighbour Kolden Van Baar, and the early members Erik Cardona-Gomez and Bridgette Hartley, who have been my comrades in navigating the ups and downs of research together. I will never forget our shared moments of brainstorming, troubleshooting experiments, and celebrating each other's successes. I am grateful for my peers in the department, especially Leila Pirayeshfard, Shelly Braun and Han Huang. It was great getting to know you all, and thank you for being there for me when I needed someone to talk to. Working with Jack Moore at the Alberta Proteomics and Mass spectrometry facility was always so great; thank you for being so approachable, knowledgeable and simply amazing.

Throughout my PhD program, I had the honour of meeting and collaborating with many great researchers and scientists in a variety of fields. I truly enjoyed all the in-depth scientific discussions over Zoom or in person and have learned so much from each of you. While I can’t name everyone due to length restriction, you guys are the reason why doing science is fun. I would like to thank Dr. Tom Hobman and all members of the Hobman lab, particularly Anil Kumar, Joaquín López-Orozco and Alberto Felix-Lopez, who have always generously assisted me in the virology side of this interdisciplinary research project since I first started. I am appreciative for Dr. Rob Ingham and members of the Ingham lab, particularly Jianing Dong, for

her outstanding contribution to Chapter 3 and proofreading parts of this thesis. Additionally, I am grateful for my committee member Dr. Joanne Lemieux and my undergraduate mentor Dr. Richard Fahlman, whom I will always look up to.

My deepest appreciation goes to my family, Mom, Dad, Aunt, Grandpa and Grandma, for your unwavering love and encouragement. I would not have been able to complete my degree without your constant support and your belief in me.

A special mention goes to BTS and particularly Jungkook, for their inspiring music and messages of hope and perseverance, and for keeping my mind refreshed during the demanding phases of my research.

Table of Contents

List of figures

List of tables

List of abbreviations

Abu	α-amino butyric acid
ACE2	Angiotensin-converting enzyme 2
ACN	Acetonitrile
AEBSF	4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride
ANK	Ankyrin repeat
AP-MS	Affinity-purification coupled with mass spectrometry
BET	Bromodomain and Extra-Terminal
BioID	Proximity-dependent biotin identification
BRD2	Bromodomain-containing protein 2
BRD4	Bromodomain-containing protein 4
BTB	Broad complex, tramtrack and bric-à-brac
COFRADIC	Combined fractional diagonal chromatography
CompPASS	Comparative Proteomic Analysis Software Suite
CPXV	Cowpox virus
CRL	Cullin-ring ubiquitin ligase
CRL3	Cullin-3 ubiquitin-ligase complex
DCM	Dichloromethane
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
DMEM	Dulbecco’s Modified Eagle Medium
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol

E	Envelope
EDTA	Ethylenediaminetetraacetic acid
EEV	Extracellular enveloped virus
ER	Endoplasmic reticulum
FA	Formic acid
FAIMS	Field Asymmetric Ion Mobility Spectrometry
HECT	Homologous to the E6-AP C-terminus
HMW	Higher-molecular-weight
HPLC	Reversed-phase liquid chromatography
IAM	Iodoacetamide
IFIT	Interferon Induced proteins with Tetratricopeptide repeats
IFNAR1	Interferon alpha and beta receptor subunit 1
IMV	Intracellular mature virus
ISG15	Interferon-stimulated gene 15 protein
LANA1	Latency-associated nuclear antigen 1
LC-MS/MS	Liquid chromatography coupled with tandem mass spectrometry
LMW	Lower-molecular-weight
m/z	mass-to-charge ratio
MAYV	Mayaro virus
MERS	Middle East Respiratory Syndrome
MHC	Major histocompatibility complex
MiST	Mass spectrometry interaction STatistics
MOI	Multiplicity of infection
Mpro	Main protease

NET	N-terminal extra terminal
NOTCH2	Neurogenic locus notch homolog protein 2
NSP	Nonstructural protein
ORF	Open reading frame
OTUD5	OTU domain-containing protein 5
pfu	Plaque forming unit
PLpro	Papain-like protease
PMSF	Phenylmethanesulfonyl fluoride
PPI	Protein-protein interaction
PRANC	Poxvirus protein repeat of ankyrin-C-terminal
PSM	Peptide-spectrum match
PSMD8	26S proteasome non-ATPase regulatory subunit 8
PTM	Post-translational modification
RBR	RING1-Between-RING2
RCR	RING-Cys-Relay
RIG-I	Retinoic acid-inducible gene-I
RING	Really Interesting New Gene
Rpb1	DNA-directed RNA polymerase II subunit A
RPMI	Roswell Park Memorial Institute
SAINT	Significance Analysis of INTeractome
SARS	Severe Acute Respiratory Syndrome
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SCF	Skp-1, cullin and F-box
SFPQ	Splicing factor, proline- and glutamine-rich

STAT2	Signal Transducer And Activator Of Transcription 1
TAILS	Terminal Amine Isotopic Labeling of Substrates
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
TFA	Trifluoroacetic acid
TFIIE2	Transcription initiation factor IIE subunit 2
TRIM25	Tripartite motif containing 25
TRIM28	Transcription intermediary factor 1-beta
UPS	Ubiquitin-proteasome system
VACV	Vaccinia virus
VACV-Cop	VACV strain was Copenhagen
VACV-WR	VACV Western Reserve strain
WT	Wildtype
z-VAD-fmk	N-Benzyloxycarbonyl-Val-Ala-Asp(O-Me) fluoromethyl ketone
ZAP	Zinc finger antiviral protein

Chapter 1: Introduction

Preface

Some sections of Chapter 1 have been published as:

> Luo, S. Y.; Araya, L. E.; Julien, O. Protease Substrate Identification Using N-terminomics. *ACS Chem Biol* **2019**, *14* (11): 2361-2371. DOI: 10.1021/acschembio.9b00398

I was one of the co-first authors and contributed to the literature research and manuscript composition equally with L.E.A.

1.1 Emerging viruses and viral diseases

In the last few decades, emerging viral diseases have posed significant strains to global public health systems. The increase in viral disease outbreaks is driven by several factors, such as transportation accessibility and environmental changes, and is expected to continue rising. (1, 2) Examples of such viral diseases include Severe Acute Respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS), COVID-19, monkeypox, Dengue and Zika, caused by various viral families including coronaviruses, poxviruses and mosquito-borne viruses. The rapid and often unpredictable spread of these viral diseases challenges the existing healthcare infrastructure worldwide and highlights the critical need for ongoing surveillance, research, and the development of new therapeutic and preventive measures.

Viruses are microscopic infectious agents that can infect all forms of life, from animals and plants to microorganisms. Unlike most living organisms, viruses lack the cellular machinery necessary for independent existence and rely on the host cell's machinery to replicate and propagate. Structurally, viruses are composed of genetic material—either DNA or RNA—encased in a protein coat, and sometimes surrounded by a lipid envelope. (3, 4) Their simplicity belies their profound impact on human health, societies and ecosystems as they co-exist and co-evolve with life. Understanding the complex virus-host interactions is a key step for formulating effective public health strategies and mitigating the impact of these diseases on the global population.

1.1.1 SARS-CoV-2

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19, has led to a global health crisis since its emergence in late 2019. (5) SARS-CoV-2 is an enveloped, positive-sense single-stranded RNA virus belonging to the family *Coronaviridae* and genus *Betacoronavirus*. The viral genome is approximately 30 kilobases

long, encoding for at least 29 proteins including structural proteins (spike glycoprotein, envelope, membrane, and nucleocapsid proteins), nonstructural proteins (NSP1-16), and accessory proteins (**Fig. 1.1**). (6)

The spike protein mediates viral entry into host cells by binding to the angiotensin-converting enzyme 2 (ACE2) receptor, followed by fusion of the viral and cellular membranes. Upon entry, the viral RNA genome is released into the cytoplasm, where it is translated into two large overlapping polyproteins, pp1a and pp1ab. These polyproteins are cleaved by the viral proteases NSP3pro (PLpro) and NSP5pro (3CLpro or Mpro) into individual NSPs, which form the replication-transcription complex **(Fig. 1.1**). The viral proteases also interact with cellular proteins, which I will discuss in Chapter 2. The replication-transcription complex then synthesizes negative-sense RNA intermediates, which serve as templates for the production of new positive-sense genomic RNA and subgenomic RNAs.

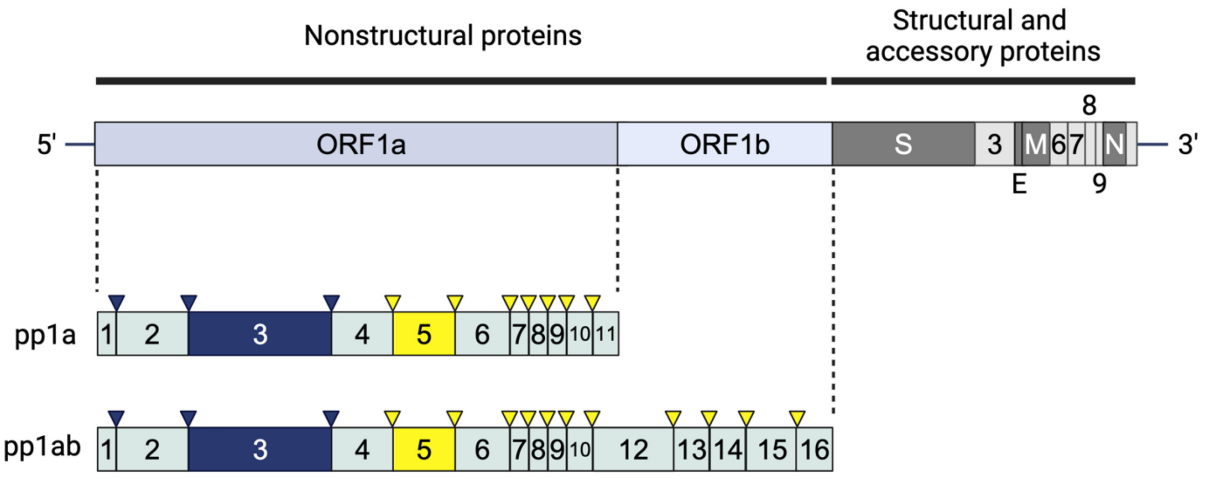


Figure 1.1. Proteolytic processing of SARS-CoV-2 polyproteins by NSP3pro (PLpro) and NSP5pro (Mpro). PLpro cleaves at three sites (dark blue triangles) and Mpro cleaves at eleven sites (yellow triangles) on the SARS-CoV-2 polyproteins pp1a and pp1ab, resulting in the release of nonstructural proteins essential for viral replication.

SARS-CoV-2 shares 80% genomic similarity with SARS-CoV, the causative agent of the SARS epidemic from 2002 to 2003. (6, 7) Many research findings on SARS-CoV were

applicable to studies on SARS-CoV-2, laying the groundwork and accelerating the development of vaccine and therapeutics for COVID-19. Moreover, foundational knowledge on coronaviruses, including their structure, replication mechanisms, and interactions with human hosts, was accumulated from years of research and played a crucial role in the resolution of the COVID-19 pandemic. For instance, structure-based design of prefusion-stabilized MERS-CoV and SARS-CoV spike proteins contributed to the generation of SARS-CoV-2 spike variant HexaPro, (8) which was the basis for COVID-19 vaccines including Spikevax (Moderna), (9) Comirnaty (Pfizer–BioNTech), (10) Jcovden (Janssen) (11) and Nuvaxovid/Covovax (Novavax). (12) Therefore, understanding the molecular mechanisms of viral replication and host evasion strategies on SARS-CoV-2 is not only crucial for combatting COVID-19, but also for deeper insights into coronavirus biology in better preparation for future outbreaks.

1.1.2 Vaccinia virus

Vaccinia virus (VACV) is a large double-stranded DNA virus belonging to the *Poxviridae* family and *Orthopoxvirus* genus. VACV is closely related to cowpox and horsepox viruses, (13, 14) and was used to eradicate smallpox in the global vaccination campaign in 1958-1977, (15) where the name *Vaccinia* was later used to recognize the serological distinction it has evolved from the cowpox virus. (16) Although closely related to the cowpox and variola virus (the causative agent of smallpox), VACV infection typically causes mild symptoms. Attenuated VACV strains are still used as a live-virus vaccine against smallpox, and it is the most extensively studied Orthopoxvirus in biomedical research to be genetically engineered for other viral disease vaccination, (17) cancer therapeutics (18) and cancer vaccines. (19) The first fully sequenced VACV strain was Copenhagen (VACV-Cop), and it is also the reference strain for other VACV variants. VACV-Cop has a linear 191 kbp genome encoding 256 proteins.

There are two infectious forms of VACV: the intracellular mature virus (IMV) and the extracellular enveloped virus (EEV) (**Fig. 1.2a**). (20) Although both forms enclose an identical DNA-containing core and lateral bodies, the EEV is surrounded by an additional membrane embedded with a unique set of membrane proteins. Upon infection, both IMV and EEV release virion cores into the cytosol. Early viral mRNA is synthesized within viral cores, encoding proteins required for immune evasion, virus uncoating, genome release, and DNA replication. After DNA replication, late genes are expressed, producing structural and nonstructural proteins for virion assembly. Spherical, non-infectious immature virions then acquire viral DNA to form fully infectious brick-shaped IMV. (21) Some IMVs are further enveloped with lipid bilayers in the trans-Golgi network and released as EEV (**Fig. 1.2b**). VACV replicates exclusively in cytoplasmic structures called factories, independent of the host nucleus. Therefore, its genome is required to encode for cytoplasmic transcription and DNA replication.

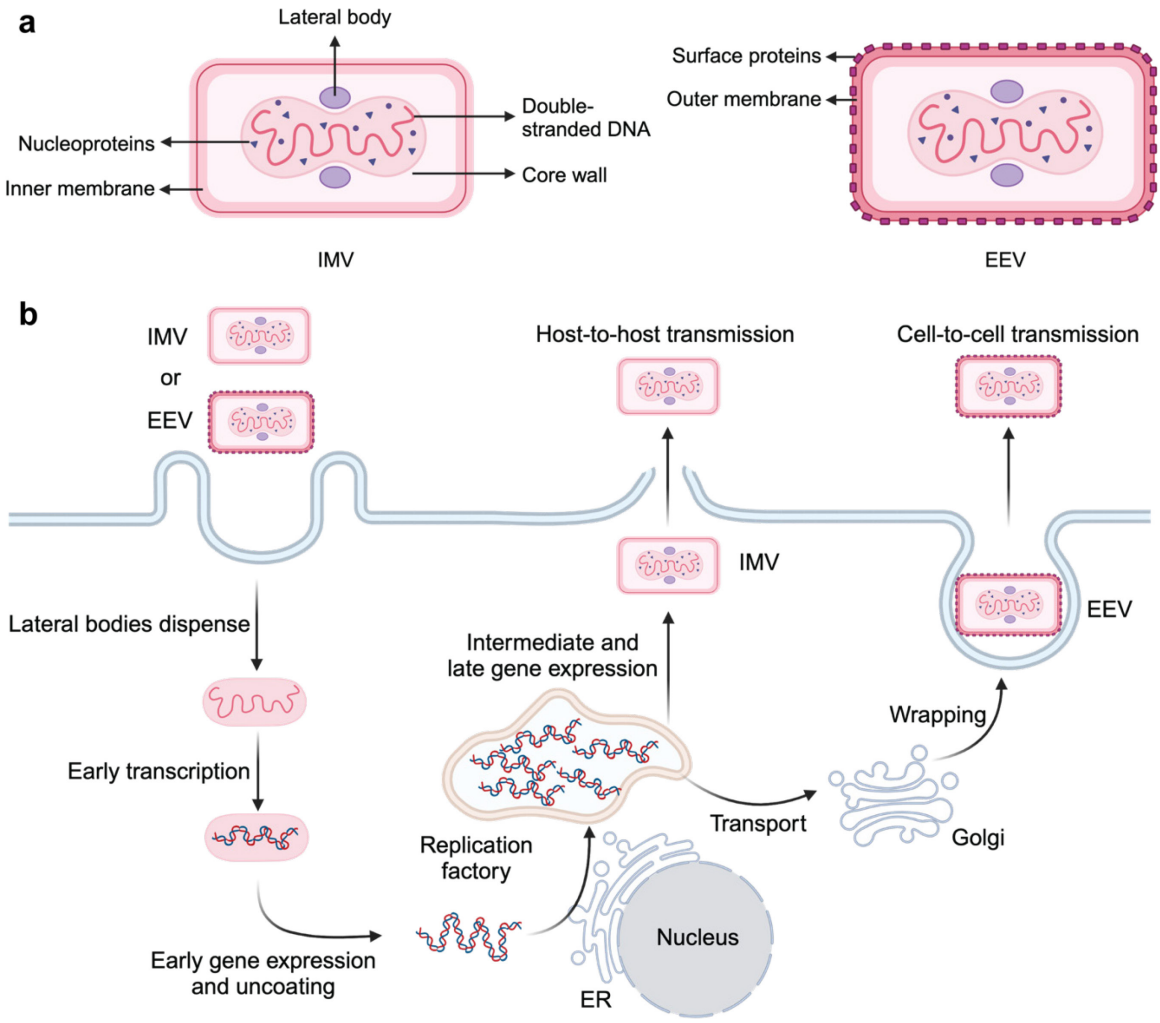


Figure 1.2. Morphology and replication cycle of VACV. a) The extracellular enveloped virus (EEV) form of VACV contains an additional outer membrane to the intracellular mature virus (IMV). Figure adapted from Shchelkunov and Shchelkunova. (22) **b**) IMV or EEV enters the host cell by direct fusion at the membrane or macropinocytosis, respectively. Upon entry, the lateral bodies dispense, and the core wall is activated as the site of early transcription. After early gene expression and core wall uncoating, the genome is released and surrounded by endoplasmic reticulum (ER) membranes, forming a replication factory with robust DNA replication. After intermediate and late gene expression, nascent virions mature into IMV and are released by cell lysis, or further processed in Golgi apparatus and exit by exocytosis. (23) Figure adapted from ViralZone, SIB Swiss Institute of Bioinformatics.

1.1.3 Mayaro virus

Mosquito-transmitted diseases infect up to 700 million people and lead to one million deaths worldwide every year. (24) With climate change, urbanization and international travel,

mosquito-borne diseases are becoming an increasing global threat. (25, 26) Researchers have coined a term “ChikDenMaZika syndrome,” (27) to describe the common symptoms observed in the recurrent epidemics of Chikungunya, Dengue, Mayaro and Zika virus. The scarcity of corresponding vaccines and therapies accentuated the significance of virus-host interactions research at the molecular level. The focus of Chapter 4 will be on one of these emerging mosquito-borne viruses, the Mayaro virus (MAYV).

MAYV is an emerging mosquito-transmitted virus that has caused endemics in Central and South America. The virus has a sylvatic cycle mainly in nonhuman primates, birds and reptiles as the natural reservoir, and can infect humans primarily through the mosquito species *Haemagogus* as the vector. (28) Infected patients develop symptoms characterized by abrupt onset fever and long-term joint pain, as well as other symptoms such as headache, muscle pain, joint swelling and rash. With the increasing number of urbanized MAYV infection cases, there is rising concern for large-scale global outbreaks. (29) As MAYV continues to impose a public health risk, there is currently no vaccine or treatment for the viral infection, as travelers rely on preventative measures such as mosquito repellents as protection.

MAYV is a positive-sense single-stranded RNA virus in the family *Togaviridae* and genus *Alphavirus.* Its genome is about 11 kilobases long, with its genomic 49S RNA encoding four non-structural proteins (NSP1-4) that possess enzymatic activities, while the subgenomic 26S RNA encodes five structural proteins: envelope glycoproteins (E1, E2, E3), capsid protein, a protein conjectured to be viroporin (6K) (**Fig. 1.3**).

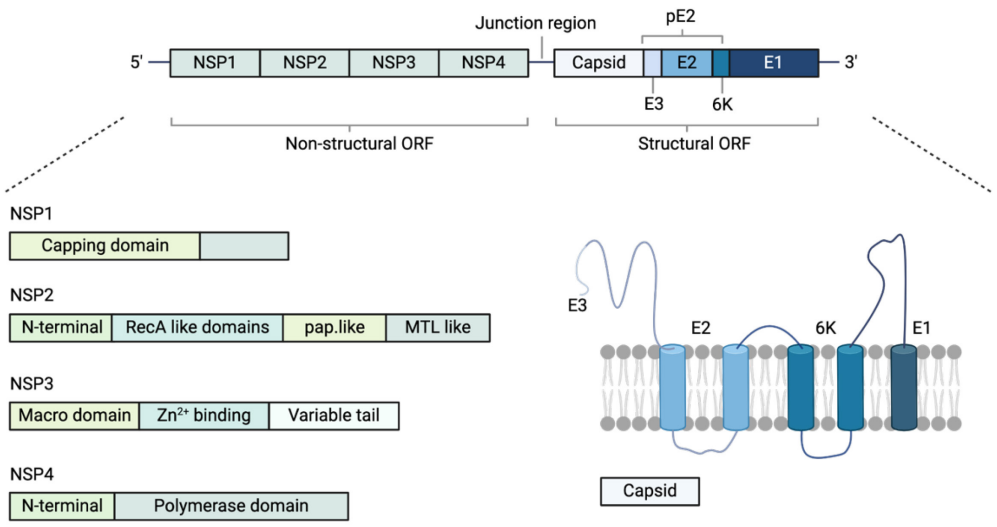


Figure 1.3. The MAYV genome and proteome. The MAYV genome contains two open reading frames (ORFs) encoding non-structural proteins NSP1-4 and structural proteins (capsid, E1, E2, E3, 6K). NSPs have enzymatic functions, while structural proteins are involved in viral entry and assembly. NSPs: non-structural proteins; Es: envelope proteins.

1.2 Host-virus interactions mediated by protein modifications and interactions

The human genome contains approximately 20,000 protein-coding genes. (30) The human proteome drives the molecular machinery in cellular processes and modulates essential biological pathways in a cell. As health and disease impact specific protein function, fate and turnover, the state of the proteome can in turn reflect the underlying mechanism of disease pathobiology. Specifically, one can monitor changes in protein abundances, post-translational modifications (PTMs) and protein-protein interactions (PPIs) to reveal dynamic alterations in cellular signaling pathways and molecular networks due to the pathological conditions (see (31, 32) for a review). More importantly, using advanced proteomics techniques, such as liquid chromatography with tandem mass spectrometry (LC-MS/MS), we can quantify differential protein levels, identify novel PTMs and map intricate PPI networks, ultimately leading to the discovery of potential biomarkers and therapeutic targets.

In the context of virology, proteomic analyses can elucidate virus-host interactions, (33–35) viral protein functions, (36) and host immune responses. (9) This can uncover viral strategies for establishing infection, immune evasion, and virion production, providing the foundation for host- or virus-directed antiviral therapeutic development. Thus, the comprehensive study of the proteome is an important part of understanding disease mechanisms and advancing clinical research and treatment.

1.2.1 Post-translational modifications (PTMs)

PTMs are the covalent processing of a protein after biosynthesis, including proteolysis, ubiquitination, phosphorylation and acetylation (for a review see (37, 38)). PTMs exponentially increase the diversities and complexities of protein functions, and form an integral part of protein signaling in biological systems. (39) It has been estimated that there are 100 proteoforms per protein-encoding gene, (39) and for highly modified proteins such as histone proteins, 287 non-redundant combinatorial PTMs were found for H3 in HeLa cells. (40) PTMs can act as activating or inhibitory switches, contributing to the dynamic nature of cellular signaling.

1.2.2 Proteolysis

Proteolysis is the irreversible hydrolysis of peptide bonds. The half-time of uncatalyzed proteolysis takes hundreds of years, (41) highlighting the indispensable role of proteases in protein catabolism by significantly accelerating this process. There are 473 known and 90 putative human proteases categorized into five families based on their catalytic mechanism: metallo-, cysteine, serine, threonine, and aspartyl proteases (reviewed by (42)). Cysteine, serine and threonine proteases hydrolyze peptide bonds by nucleophilic attack via the catalytic residues for which they are named, while metallo- and aspartyl proteases typically activate a water molecule to carry out the nucleophilic attack.

Protease substrate residues are numbered in relation to the site of backbone cleavage. For the substrate, residues N-terminal to the scissile amide bond are numbered sequentially starting with P1 (non-prime sites). Conversely, amino acid residues C-terminal to the scissile bond are numbered sequentially beginning with P1′ (prime sites) (**Fig. 1.4**). (43) The corresponding S numbering is used to designate the complementary regions of the protease active site.

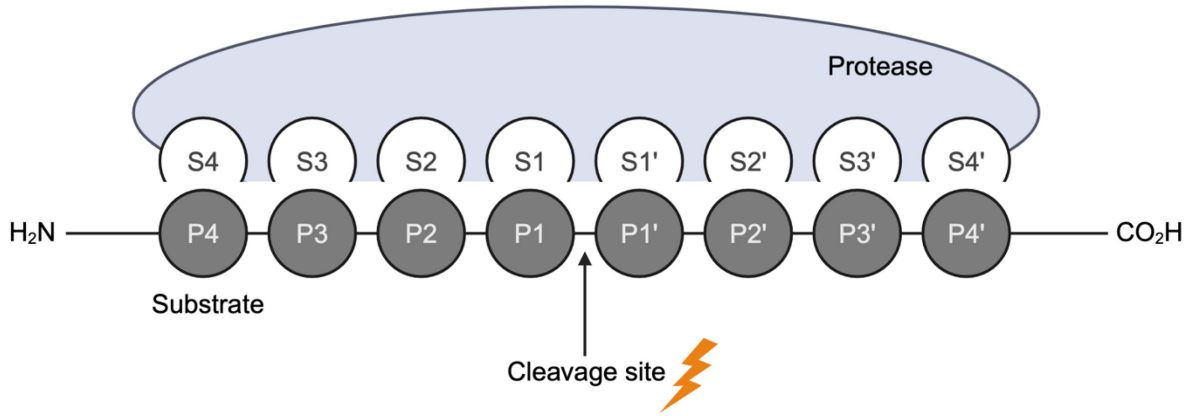


Figure 1.4. Nomenclature of cleavage sites on substrates and substrate-binding subsites on proteases. Substrate residues are labeled non-prime (P) and prime sites (P′) in increasing order, with P1 being the residue immediately before the cleavage site, and P1′ immediately after. Corresponding protease subsites are labeled non-prime (S) and prime (S′) binding the substrate at corresponding positions. The cleavage occurs between the P1 and P1′ sites.

Proteases can be found in plants, animals, bacteria, and viruses, regulating and conducting a diverse number of biological processes and complex signaling cascades (reviewed by (44)). Some proteases are fairly promiscuous, recognizing a single amino acid in any peptide chain, and thus causing proteolysis in a wide range of protein substrates. For example, trypsin is a commonly used enzyme for protein digestion, cleaving peptide bonds on the C-terminus of any lysine or arginine in a peptide sequence. In contrast, many proteases are highly specific to longer, distinct, amino acid motifs, reflecting their regulated function in complex physiological processes. A well-known protease in this category is the tobacco etch virus (TEV) protease, which has a more stringent substrate consensus sequence ENLYFQ|(G/S), where the

proteolysis occurs after glutamine. (45) As the TEV protease is required for the cleavage of the TEV viral precursor polyprotein into functional viral protein units, its high specificity plays a critical role in maintaining proper viral production and infectivity. Thus, the characterization of protease cleavage specificity (46–48) and identification of protease substrates is a fundamental step in the quest to understand the vital roles of proteases in various organisms.

Proteolysis can generate exposed, unstable neo-N-termini that can target the proteolytic fragment for proteasome-dependent degradation. (49) However, proteolysis can also alter protein functions and interactions by the precise removal of particular domains or signaling sequences, resulting in gain-of-function modifications. (50, 51)

1.2.3 Proteolysis in SARS-CoV-2

Viral proteases play a crucial role in the replication and pathogenesis of SARS-CoV-2 by processing the viral polyproteins required for the formation of the replication-transcription complex. In SARS-CoV-2, the two proteases, NSP3 and NSP5, have become prime targets for antiviral drug development.

NSP3pro, also known as papain-like protease (PLpro), is a cysteine protease with a catalytic triad consisting of Cys111, His272, and Asp286. (52) PLpro cleaves the viral polyproteins pp1a and pp1ab at three LXGG motifs, releasing NSPs 1, 2 and 3 (**Fig. 1.1**). (53, 54) PLpro also has deubiquitinating and deISGylating activities with high affinities for the ubiquitin-like interferon-stimulated gene 15 protein (ISG15). (55) Inhibitors such as GRL0617 have shown potential in blocking PLpro activity and suppressing viral replication *in vitro*. (56, 57)

NSP5pro, also known as 3C-like protease (3CLpro) or main protease (Mpro), is also a cysteine protease with a catalytic dyad consisting of Cys145 and His41. Mpro typically cleaves after a glutamine residue (P1=Q), cutting the viral polyproteins at eleven distinct sites, facilitating the release of NSPs essential for RNA transcription and replication (**Fig. 1.1**). This

specificity is crucial for the precise processing of NSPs. Mpro has been shown to cleave several host proteins to disrupt critical signaling pathways, such as NF-κB and interferon signaling, further aiding in immune evasion and viral replication. (58–63) In Chapter 2, I will present the systematic identification of SARS-CoV-2 Mpro and PLpro proteolysis targets in host cells.

Mpro is a well-validated target for antiviral drug development. Several inhibitors, including GC376 (64) and the FDA-approved drug Paxlovid (nirmatrelvir/ritonavir), (65–68) have demonstrated efficacy in inhibiting Mpro activity and preventing viral replication. Mpro inhibitors are still targets of on-going developments of second-generation oral drug with improved metabolic stability, (69) and pan-coronavirus antiviral therapy.

1.2.4 Ubiquitination

Ubiquitination is the reversible covalent attachment of ubiquitin on proteins. Ubiquitin is a 76-amino-acid protein, evolutionarily conserved among nearly all eukaryotic organisms, regulating many aspects of eukaryotic biology (**Fig. 1.5**).

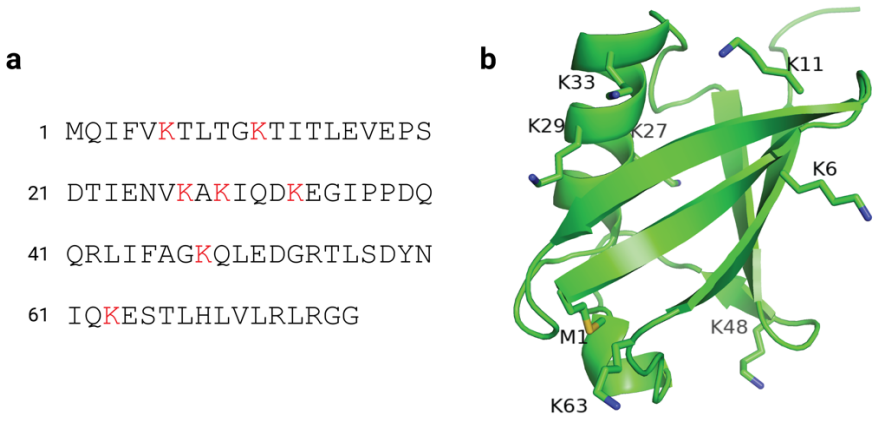


Figure 1.5. Sequence and structure of ubiquitin. a) Amino acid sequence of ubiquitin. Specific lysine residues involved in polyubiquitination (K6, K11, K27, K29, K33, K48, and K63) are highlighted in red. **b**) Ubiquitin structure in ribbon format (PDB: 1ubq, illustrated using Pymol), with lysine residues labeled and their positions indicated.

Ubiquitination is an ATP-dependent, highly ordered, and multistep enzymatic process. It involves 1) ubiquitin activating enzyme (E1) binding and activating ubiquitin in an ATP-

dependent manner, 2) transfer of ubiquitin onto ubiquitin-conjugating enzyme (E2) and 3) ubiquitin ligase (E3) recruitment of ubiquitin-linked E2, recognizing and catalyzing the ubiquitin transfer onto its substrates (**Fig. 1.6a**). Due to the increasing complexities of their functions, the human genome encodes two genes for E1, 30 to 50 for E2, and more than 600 for E3. (70)

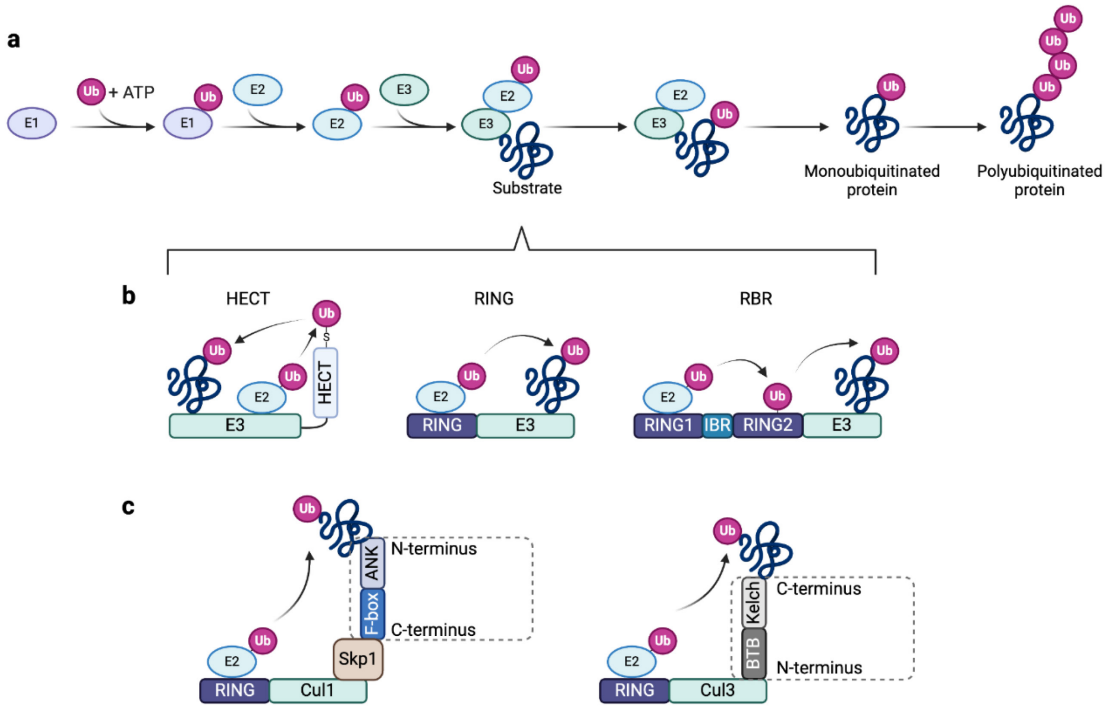


Figure 1.6. Schematic representation of the ubiquitination process and types of E3 ligases. a) The ubiquitination cascade involving E1, E2, and E3 that results in mono- or poly-ubiquitinated proteins. **b**) Three main classifications of E3 ligases: homologous to the E6-AP C-terminus (HECT) that form a thioester intermediate with ubiquitin before transferring it to the substrate, Really Interesting New Gene (RING) that facilitate direct transfer of ubiquitin from E2 to the substrate, and RING1-Between-RING2 (RBR) that have features of both HECT and RING, transferring ubiquitin via a two-step mechanism involving a thioester intermediate. **c**) Member of multi-subunit complex cullin-ring ubiquitin ligases (CRL): Skp-1, cullin and F-box (SCF) ubiquitin ligase complex and cullin-3 ubiquitin-ligase complex (CRL3). The dashed regions enclose VACV-encoded adaptor proteins. Figure adapted from Humphreys *et al.* and Cui *et al.* (71, 72)

E3 ligases are canonically categorized into three groups based on the type of domains they contain: 1) Homologous to the E6-AP C-terminus (HECT), 2) Really Interesting New Gene (RING), and 3) RING1-Between-RING2 (RBR) E3s (**Fig. 1.6b**). RING finger E3 ubiquitin ligases

are the largest E3 family. Among these, the largest subfamily is the cullin-ring ubiquitin ligases (CRL) that are responsible for ~20% for cellular ubiquitination. (73) The CRLs are protein complexes consisting of a cullin protein, a RING protein, and an adaptor protein or substrate recruiter. The cullin protein acts as a central scaffold, with its C-terminal and N-terminal domains binding to the RING protein and the adaptor protein respectively (**Fig. 1.6c**). (74) Each CRL subfamily has a distinct set of adaptor proteins, and they can influence E3 localization, catalytic activity and regulate E3-substrate interaction.

The Skp-1, cullin and F-box (SCF) ubiquitin ligase complex is the first identified multiunit CRL. (75–77) In SCF, the F-box protein is the substrate receptor that binds the linker protein Skp1, which is in contact with N-terminal cullin-1, for substrate ubiquitination by the RING E3 (**Fig. 1.6c**; left). (76)

BTB-Kelch proteins are also substrate-specific adaptors for the cullin-3 ubiquitin-ligase complex (CRL3), comprising an N-terminal BTB (broad complex, tramtrack and bric-à-brac) domain, and a C-terminal Kelch domain with tandem repeats of a 50-amino-acid Kelch motif (**Fig. 1.6c**; right). BTB-Kelch proteins are capable of binding both cullin-3 through the BTB domain, and the target substrate through the Kelch domain, without requiring linker proteins. (74) They can also dimerize to simultaneously and independently facilitate two distinct substrate ubiquitination. (78, 79)

E3 ligases catalyze the formation of an isopeptide bond between the ubiquitin C-terminal glycine and an amino group of the target protein, typically on the side chain of lysines. A target protein can also be mono- or poly-ubiquitinated, where polyubiquitination is the formation of ubiquitin chains through the linkage between multiple ubiquitin moieties. Ubiquitin has seven acceptor lysines (K6, K11, K27, K29, K33, K48, K63) and one methionine (M1) involved in

ubiquitin conjugation, resulting in the formation of ubiquitin chains with varying lengths, types and functional outcomes (**Fig. 1.5, 1.7**). (80)

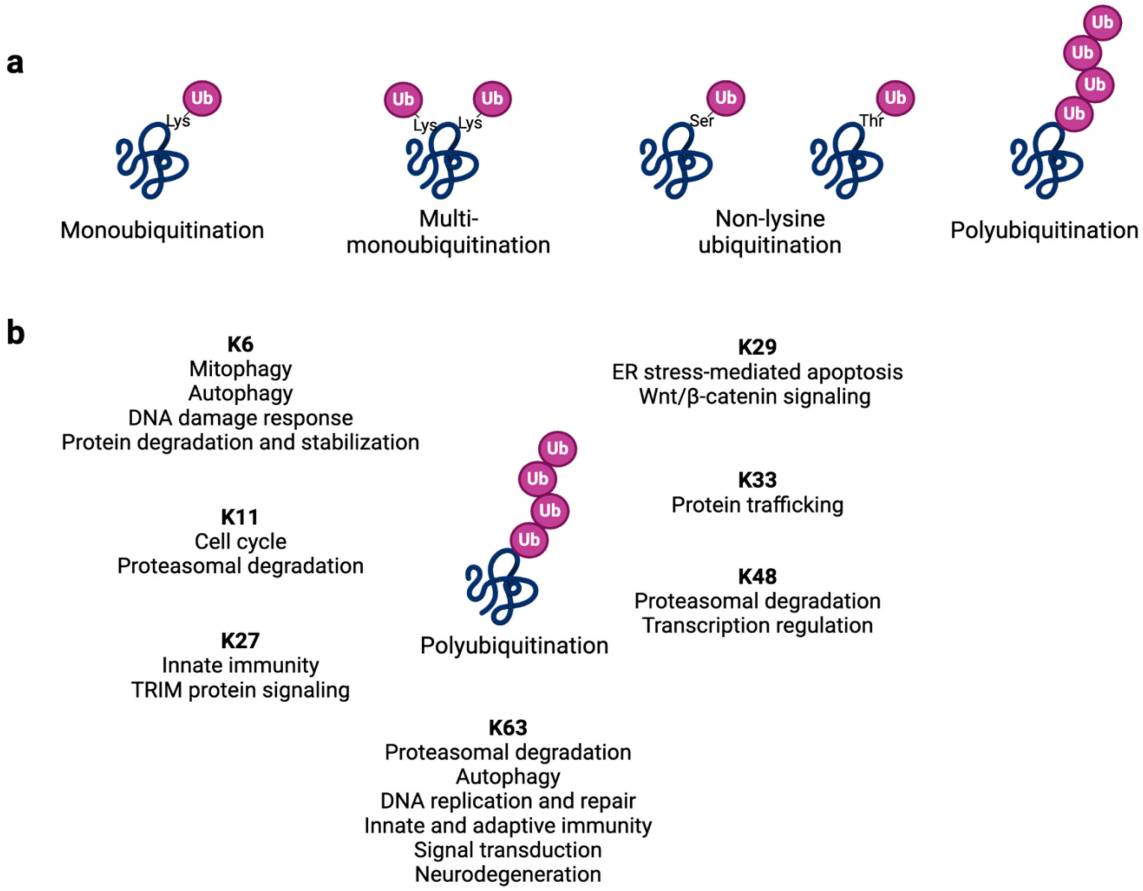


Figure 1.7. Ubiquitination occurs in multiple forms that result in distinct cellular functions. a) Types of ubiquitination include monoubiquitination that can occur on multiple sites and non-lysine residues, as well as polyubiquitination. b) Diverse functions of polyubiquitin chains linked through different lysine residues (K6, K11, K27, K29, K33, K48, and K63).

K48 and K63 are the most abundant linkages and account for ~80% of total linkages in mammalian cells. (81) Polyubiquitinated K48-linked and branched K48-K11 chains serve as the most potent signals for degradation by the ubiquitin-proteasome system (UPS). (82) It can be directly recognized by the intrinsic proteasomal ubiquitin receptors (26S proteasome regulatory subunits Rpn10 and Rpn13), or mediated by shuttle factors with both ubiquitin- and proteasome receptor-binding domains (e.g. UV excision repair protein RAD23 A and B, ubiquilin 1-4) (**Fig.**

1.8). (83) The substrate is then deubiquitinated by deubiquitinating enzymes, unfolded, and translocated to the 20S subunit core to come into contact with the proteolytic active sites for degradation (**Fig. 1.8**). (84)

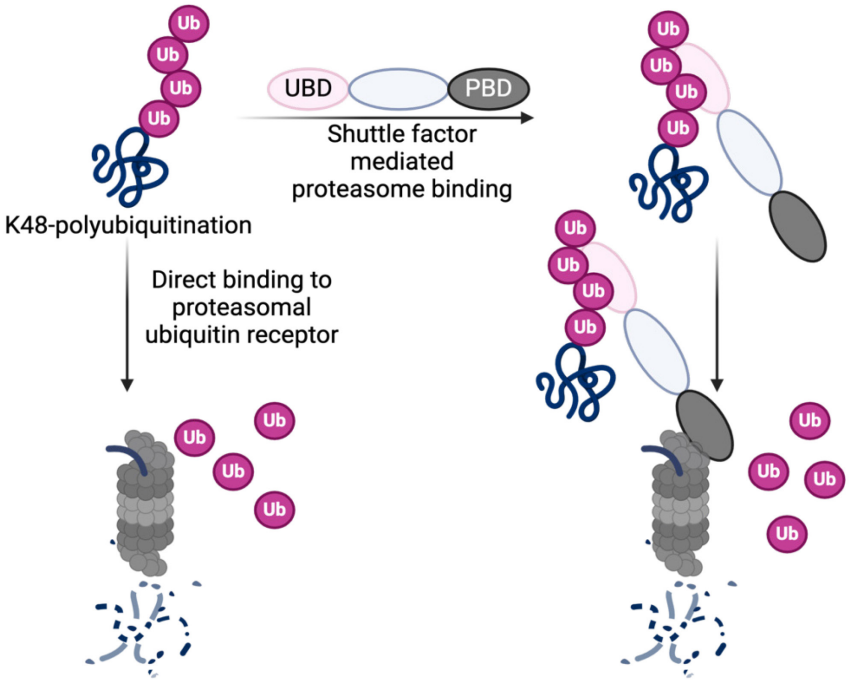


Figure 1.8. Ubiquitin proteasome degradation. K48-linked polyubiquitination canonically signals for proteasome-dependent protein degradation, where the substrate can directly bind proteasomal ubiquitin receptors or the binding is mediated by shuttle factors. UBD: ubiquitin-binding domain; PBD: proteasome-binding domain. Figure adapted from Pohl C. and Dikic I. (82)

Monoubiquitination is well-known for altering protein localization and protein interactions. For instance, membrane receptors like epidermal growth factor receptors can be monoubiquitinated for internalization to attenuate signal transduction, (85) while histone monoubiquitination regulates chromatin structure and gene expression. (86) Monoubiquitination can also target protein for UPS-dependent degradation. (87)

1.2.5 Ubiquitination in Vaccinia virus

The ability of poxviruses to undergo DNA replication and assembly exclusively within the cytoplasm is a unique feature among DNA viruses and requires extensive regulation and modulation of many cellular systems (reviewed by (88)). The UPS is a well-known pathway manipulated by the poxvirus, as it tightly regulates important cellular functions, such as the cell death, antigen presentation, signal transduction, and DNA repair (reviewed by (72, 89)) In Chapter 3, I will illustrate changes in the global ubiquitome during VACV infection and identify key host factors that are heavily ubiquitinated in response to VACV infection.

Several viruses have evolved mechanisms to exploit the ubiquitination system, by encoding ubiquitin ligases and deubiquitinases, or interacting with components of the ubiquitination machinery to target cellular components. (90, 91) During viral infection, both viral and host proteins can be ubiquitinated in pro- or anti-viral mechanisms. For instance, the virus can mediate the ubiquitination of host restriction factors for degradation by the UPS, while the host initiates signaling molecule ubiquitination leading to activation of immune response (**Fig. 1.9**). (92)

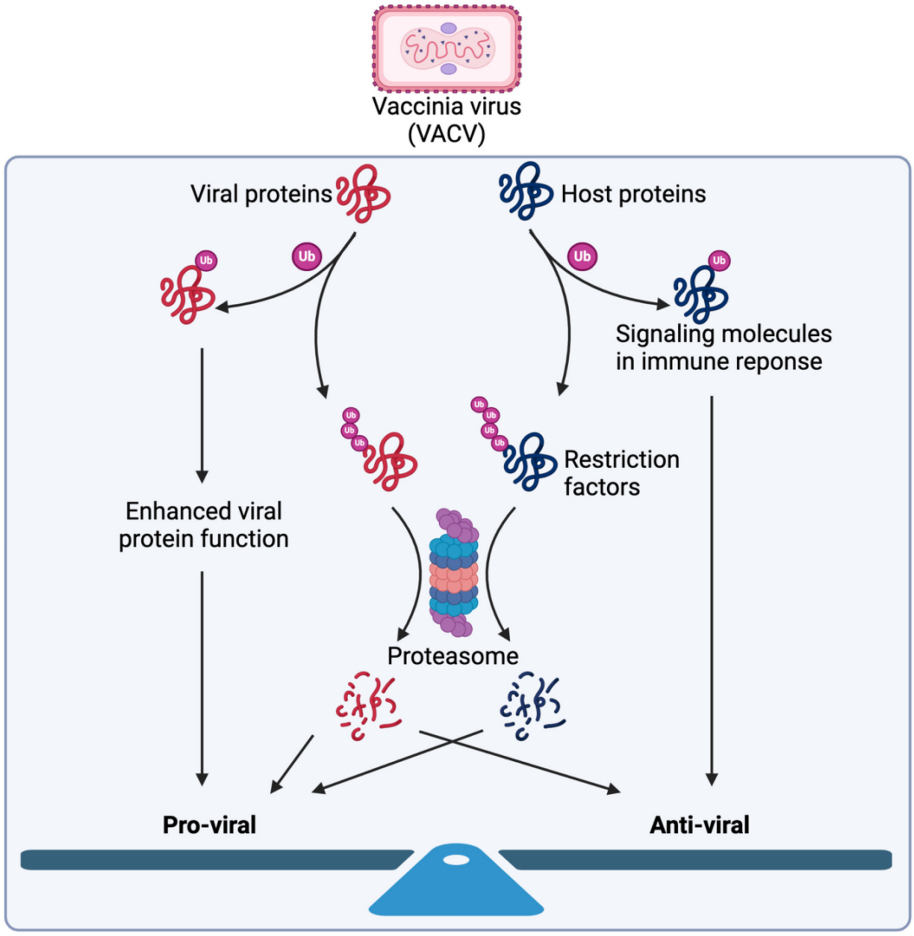


Figure 1.9. Role of ubiquitination in vaccinia viral pathogenesis. As the UPS can target proteins for degradation or activate protein functions, it is both exploited by the virus in viral pathogenesis and utilized by the host in immune response. Figure adapted from Luo H. *et al.* (92)

There are three classes of E3 ligase adaptors encoded by poxviruses (reviewed by (72)). VACV-Cop encodes three cullin-3 E3 ligase adaptors BTB-kelch proteins, A55, C2, and F3, one BTB-only protein C5, and one kelch-repeat-only protein B10 (**Fig. 1.6c**; right). Interestingly, among all viruses, only representative poxviruses contain BTB-kelch protein-encoding genes. (93)

Although the absence of the BTB-Kelch genes in other VACV isolates indicates that they are not essential for viral replication, the deletion of individual *C2L* and *A55R* genes in VACV-

Cop resulted in changes in the plaque morphology and decrease in virus-induced cytopathic effects. (94, 95) Intradermal infection of VACV *C2* or *A55R* deletion in the mouse model produced larger lesions, increased VACV-specific CD8+ T-cell proliferation, activation, and T cell-mediated cytotoxicity compared to the wild-type VACV. (94–96) Murine intradermal infection of VACV *F3L* significantly increased the NK cell population in lesions four days post infection. (97)

Additionally, VACV also expresses adaptor proteins involved in the SCF, known as the ANK-PRANC proteins (**Fig. 1.6c**; left). (98) These proteins contain ankyrin-repeat (ANK) domains and a diverged variant of the F-box protein called Poxvirus protein Repeat of Ankyrin-C-terminal (PRANC) domain. VACV-Cop encodes three full ANK-PRANC proteins (C9, B4, B18), two ANK domain only proteins (M1, K1), and a number of other truncated forms (C19L-C21L/B25R-B27R,C17L-C18L/B23R-24R, B20R, C15L/B21R). (99) These proteins were found to inhibit cellular including pathways NF-κB activation, (100) and interferon signaling, (101–103) and the deletion of ANK-PRANC genes such as *B4R* caused reduced virus spread. (104)

Furthermore, treatment with proteasome inhibitors such as MG132 pre- or post-infection drastically reduced viral titers, blocked viral replication and propagation. (105, 106) Specifically, in the presence of MG132, early gene expression was prolonged and poxviral factories failed to form, as the subsequent removal of MG132 rescued viral replication, factory-formation and late protein production. (105) RNAi screening has further shown that the formation of CRL3 is required for VACV genome replication. (107) However, whether this was due to lack of ubiquitination on viral and/or host proteins remained unknown.

Given that the E3 adaptor proteins and the addition of proteasome inhibitors resulted in aggravated virulence, it is evident that the UPS plays a significant role in VACV replication and infection. Furthermore, the function of ubiquitination modifications is beyond just targeting

protein for degradation as discussed in Chapter 1.2.4. The identification of ubiquitinated proteins and the associated modification sites during viral infection will greatly enhance our understanding of viral pathobiology. Although ubiquitination in cowpox virus-infected cell lysates has been profiled, (108) the global ubiquitination dynamics in the model poxvirus VACV infection has not been examined.

1.2.6 Protein-protein interactions (PPIs)

PPIs are physical contacts between two or more protein molecules, resulting from biochemical events driven by non-covalent interactions such as electrostatic forces, hydrogen bonding, and hydrophobic interactions that can be stable or transient. (109) PPIs are fundamental for orchestrating cellular activities, ranging from signal transduction to metabolic pathways. They govern the formation of multiprotein complexes essential for cellular function. By mediating interactions between key regulatory proteins, PPIs modulate enzyme activity, substrate specificity, and cellular response to environmental cues. Furthermore, PPI networks dynamically adapt to cellular conditions, influencing processes such as gene expression, protein trafficking, and immune response (reviewed by (32, 110)).

1.2.7 Protein-protein interactions in Mayaro virus

Intraviral and virus-host PPIs mediate and regulate essentially all steps of the MAYV replication cycle. Through interaction between E2 and cell surface receptors, MAYV enters the host cell via endocytosis. The low pH in the endosome induces fusion between E1 and endosomal membrane, causing the release of the nucleocapsid into cytoplasm, which disassembles and releases the viral genome. The open reading frame is first translated into a non-structural polyprotein precursor NSP1234, cleaved into NSP123 and NSP4 that interacts to form the initial replication complex for negative strand synthesis, followed by the sequential cleavage of NSP1|2 and NSP2|3 sites for the formation of the mature replication complex (**Fig.**

1.3, 1.10). (111) All three sites are processed by the viral protease activity residing in NSP2 or in non-structural polyproteins containing NSP2.

MAYV NSPs have distinct enzymatic functions that are vital for viral replication (**Fig. 1.10**). (111) NSP1 is a membrane-anchored mRNA-capping enzyme. NSP2 is a multifunctional protein that possesses protease, NTPase, RNA 5′-triphosphatase, and helicase activities. NSP3 contains a mono- and poly-ADP-ribose binding domain. NSP4 is an RNA-dependent RNA polymerase and terminal adenylyltransferase that is responsible for the synthesis of poly(A) tails of mRNA.

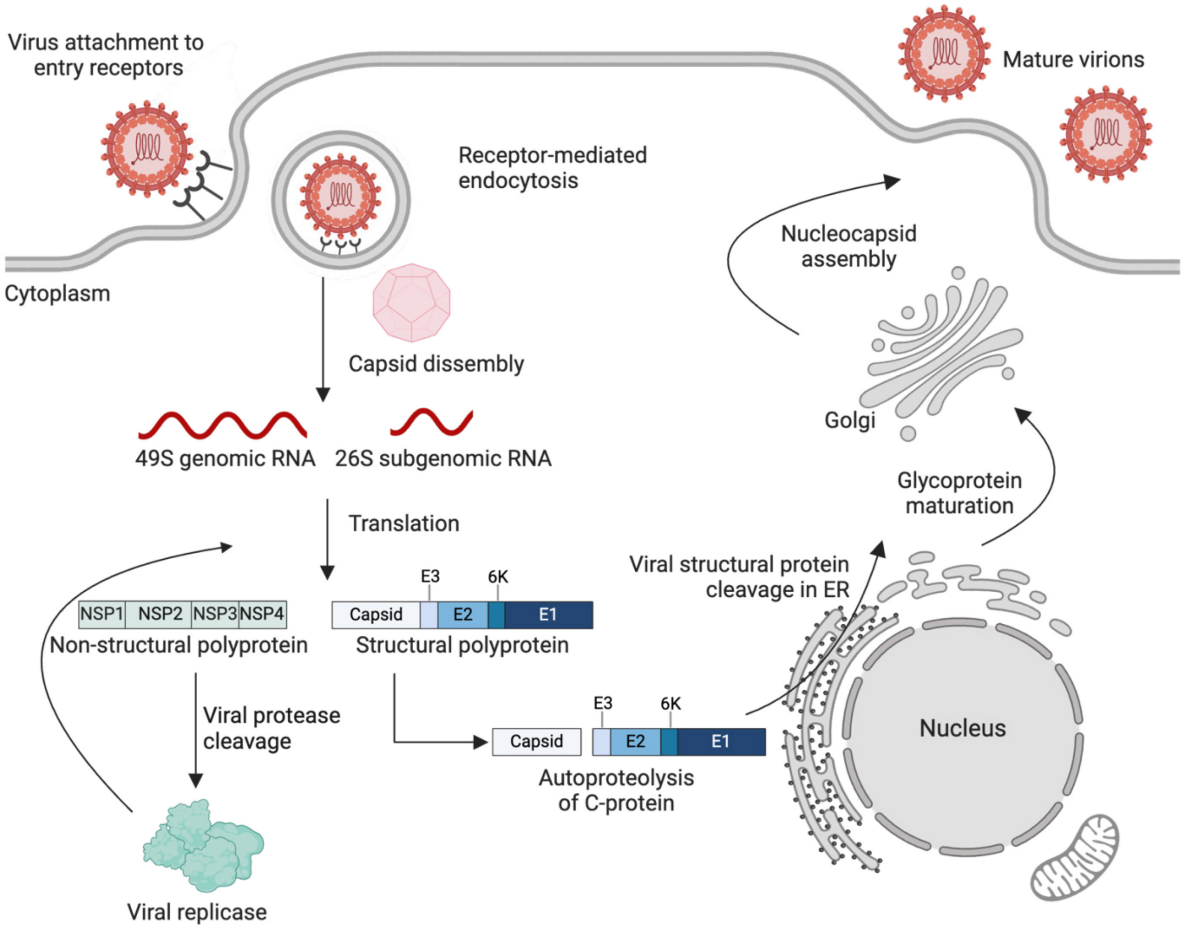


Figure 1.10. MAYV replication cycle. The MAYV replication cycle starts with the virus attaching to host cell receptors and entering via receptor-mediated endocytosis. The capsid disassembles, releasing 49S genomic RNA and 26S subgenomic RNA. The non-structural proteins (NSP1-4) are first translated to form the viral replicase, aiding in the synthesis of both genomic and negative-strand RNA. Structural proteins are then translated, cleaved in the ER, matured in the Golgi, and assembled into nucleocapsids. Mature virions are then released to infect other cells.

The subgenomic RNA is first translated into a single polyprotein precursor, and the capsid protein is rapidly cleaved off in the cytoplasm due to its C-terminal cis-autoproteolytic activity. The envelope structural polyprotein pE2-6K-E1 then translocates to the endoplasmic reticulum (ER) via an N-terminal signal sequence on E3, where it is first cleaved by cellular enzyme signalase for E1-pE2 heterodimer formation. The oligomerized heterodimer goes through the Golgi apparatus towards the cell membrane, where pE2 is further cleaved by

cellular furin to form mature E2 and E3. Finally, mature glycoproteins assemble into a complex and virion budding occurs at the plasma membrane (**Fig. 1.10**).

Both E1 and E2 are Type I transmembrane proteins. E1 contains a hydrophobic fusion peptide essential for the fusion of viral and cellular membranes, facilitating viral entry into host cells. The E2 protein is involved in receptor binding and is the main target of neutralizing antibodies. The E3 protein remains electrostatically bound to E2 until it is released at the late phase of infection. (111)

There has been considerable interest in therapeutic and vaccine development for alphaviruses. In 2023, a vaccine for a closely related alphavirus Chikungunya developed by Valneva was approved by the U.S. Food and Drug Administration, called Ixchiq, targeting individuals of age 18 years or older who are at increased risk of Chikungunya virus exposure. (112) However, no cross-protecting vaccines between Chikungunya and MAYV infection have been licensed. Due to the increased alphavirus co-circulation and risk for MAYV urbanization, increased efforts are needed for antiviral therapeutics. Establishing virus-host PPI networks will highlight hotspots where proteins or corresponding pathways can be targeted by existing drugs, and may be used to guide rational design of virus- and host-directed combination therapies. In Chapter 4, I will employ an improved approach to characterize virus-host interactions that better reflects the mechanism of MAYV infection.

1.3 Mass spectrometry

Bottom-up proteomics is a method used to study proteins by the fragmentation and identification of peptides. In combination with advanced LC-MS/MS instrumentation, bottom-up proteomics provides simple and reliable protein assignment and quantification, and can be used to characterize protein PTMs (113) and PPIs. (114)

In a typical bottom-up proteomics workflow, proteins are extracted from a biological sample, such as cells, tissues or bodily fluids, and digested into peptides using a digestion enzyme that is usually trypsin. The complex mixture of peptides is then separated by reversed-phase liquid chromatography (HPLC) based on their hydrophobicities. The sequentially eluting peptides are ionized to form precursor ions and introduced into the mass spectrometer. The precursor ions are transmitted to a mass analyzer, where their mass-to-charge ratios (m/z) are measured in the first mass scan (MS1). They are then fragmented along the peptide backbone using collisional dissociation, creating fragment ions, and their m/z are measured again by the mass analyzer in the second mass scan (MS2) to provide additional structural information on the peptide identity. The resulting mass spectra, along with their retention times in the LC, are analyzed using proteomics softwares and assigned to known protein sequences in the database (see (114) for a review).

In addition to providing information about peptide and protein quantity, bottom-up proteomics has been expanded for specialized aims such as mapping PTM species and sites (113, 115) and PPIs. (116) I will explore some of these approaches in greater depth below.

1.3.1 PTM enrichment methods

Since PTMs typically occur at low abundance, with most of the protein remaining unmodified, enrichment methods are necessary for systematic PTM profiling prior to mass spectrometry analysis. Some methods depend on antibody specificity for the PTM (e.g., ubiquitination and acetylation), others exploit the ionic charges of the modification (e.g., phosphorylation), and some require additional labeling tools for PTM enrichment (e.g., proteolysis). I will provide a review over the enrichment methods currently employed for identification of proteolysis and ubiquitination substrates in mass spectrometry.

1.3.2 N-terminomics

N-terminomics is the large-scale identification of protein neo-N-termini that can be used for the detection of proteolysis products. The N-terminomics methods consist of two main steps, 1) labeling of the proteolytic protein fragments, and 2) enrichment of the fragments from the complex mixture. The enrichment can be achieved by modifying the cleaved peptide termini with the addition of functional groups, or by taking advantage of differentially labeled isotopes. Since carboxyl groups are less reactive than primary amines, peptide C-terminal labeling methods have been difficult to develop and thus are not as established as N-terminal profiling approaches (N-terminomics).

There are many challenges to overcome in order to precisely identify cleavage sites. For example, the balance between protein expression and protein degradation leads to constant proteolysis within a cell. In addition, cellular heterogeneity between treated and untreated samples, or healthy and disease samples, makes it difficult to distinguish meaningful induced proteolytic activity from transient background events. Moreover, proteolytic fragments of interest are often low in abundance and can be targeted for subsequent degradation due to their instability. (49) Various degradomics methods were created over the years to overcome some of these issues, with the attempts of reliably detecting and resolving proteolytic peptides from protein substrates and background peptides. Recent advances in mass spectrometry have also greatly accelerated the progress and sensitivity of these proteomic methods (reviewed by (42)).

Negative selection approaches isolate native N-terminal peptides or target protease-mediated neo-N-termini, through the enrichment of other internal fragments (internal tryptic peptides). Popular negative enrichment methods such as combined fractional diagonal chromatography (COFRADIC) (117) and Terminal Amine Isotopic Labeling of Substrates (TAILS), (118) have been instrumental in mapping the degradome of many proteases, including caspases (119) and matrix metalloproteinases (**Fig. 1.11**). (120) However, these methods are

accompanied by limitations of potentially high background levels and thus must be used in combination with statistical analyses to distinguish the tagged and untagged N-termini in complex peptide mixtures. False negatives could still arise from the incomplete efficiency in blocking the N-termini of internal tryptic fragments, such as in steps of N-acetylation blocking in COFRADIC, (121) or reductive dimethylation by formaldehyde in chemical isotopic labeling in TAILS.

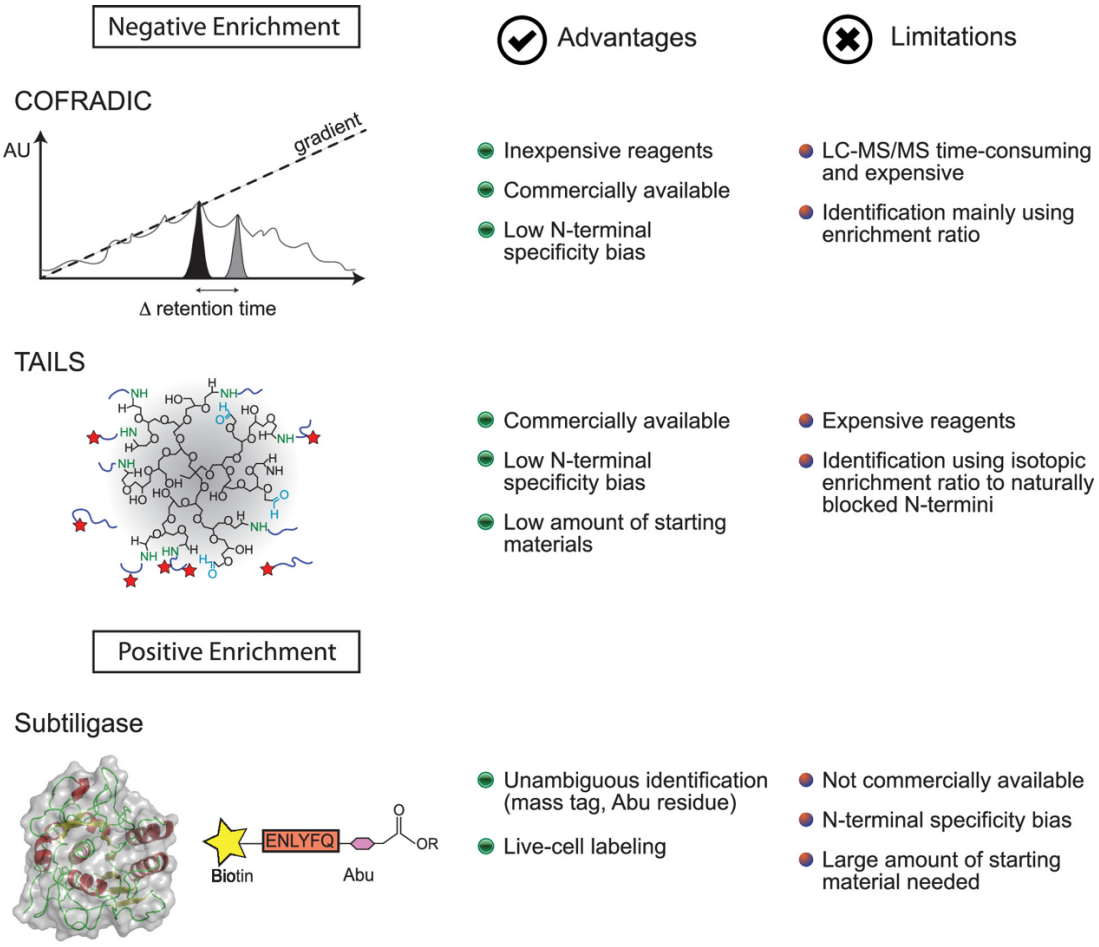


Figure 1.11. Mass spectrometry-based N-terminomics methods used for protease substrates identification. Advantages and limitations of current techniques are presented for COFRADIC, (117) TAILS (118) and subtiligase. The COFRADIC method takes advantage of differential gel electrophoresis migration of protein fragments. TAILS uses a polymer to capture all free amines. Subtiligase is a protein ligase used to enzymatically attach a biotin label to the backbone N-termini of protein fragments. COFRADIC: combined fractional diagonal chromatography; TAILS: terminal amine isotopic labeling of substrates.

Positive selection enables the direct labeling and capture of native N-termini and neo-N-termini on protein fragments. The difficulty arises from specifically labeling the backbone α-amines while leaving the protein ε-lysine side chains unmodified. An enzymatic N-terminomics approach was developed in this regard to selectively attach a biotin probe to N-terminal fragments generated by proteolysis. Subtiligase, an engineered peptide ligase generated from a protease called subtilisin, (122) is able to efficiently catalyze ligation between N-terminal α-amines with a peptide ester (or thioester tag; **Fig. 1.11**). In untreated or apoptotic Jurkat cells, proteolytic fragments generated by cellular protease activation are N-terminally ligated by subtiligase to a designed peptide ester substrate, containing a biotin molecule for immobilization on avidin and a TEV protease cleavage site for release and recovery. Therefore, after trypsin digestion, biotinylated peptides can be easily isolated and selected from internal and C-terminal tryptic fragments for the following LC-MS/MS analysis. Importantly, the recovered peptides also contain a unique non-naturally occurring amino acid (α-amino butyric acid, Abu) derivative tag upon TEV protease cleavage, such that labeled peptides can be unambiguously distinguished from background tryptic peptides.

Despite the robustness and low false discovery rate, this enzymatic subtiligase N-terminomics approach also has limitations. For example, a large amount of sample is often needed for effective enrichment and detection, often requiring milligrams of starting material due to the low efficiency of the labeling reaction (although recent advances in LC-MS/MS technology have reduced this requirement). Another major drawback is that subtiligase also has an intrinsic substrate specificity for ligation. (123) While subtiligase has broader specificity and higher catalytic efficiency than many peptide ligases, it still possesses prime-side preferences for specific residues. (122) This would limit its application in N-terminal bioconjugation, causing an under-representation of peptides. This issue was partly resolved by mutating subtiligase at different residues to alter its P1′ and P2′ specificity, allowing one to select from a cocktail of

subtiligase mutants tailored to broaden subtiligase specificity. (123) Finally, since subtiligase N-terminomics takes advantage of an enzymatic reaction as opposed to a chemical reaction, it is possible to perform labeling in live cells. Weeks *et al.* applied plasma membrane-targeted subtiligase variants to label cell surface N termini in living cells. (124)

1.3.3 Ubiquitin remnant enrichment

The study on protein ubiquitination has been historically challenging. Ubiquitin modification is large (~8 kDa if monoubiquitinated), with rapid, reversible turnover and low levels of steady-state conjugations. (125) Moreover, many ubiquitinated proteins are susceptible to UPS-mediated degradation, requiring the need for additional proteasome inhibitors. To identify ubiquitinated proteins at the proteome level, an affinity-tagged ubiquitin is usually overexpressed in cells, and ubiquitinated substrates are co-immunoprecipitated using the affinity tag. Alternatively, a ubiquitin-binding protein such as an anti-ubiquitin antibody or a protein containing a ubiquitin-binding domain can be used to purify endogenously ubiquitinated proteins.

Identifying ubiquitination sites takes additional steps. Conventional biochemical analysis requires site-directed mutagenesis on individual lysines to determine ubiquitination outcome and biological significance. Conveniently, the three C-terminal ubiquitin residues are arginine-glycine-glycine (RGG), and the protease trypsin cleaves any peptide bonds C-terminal to arginines or lysines. Trypsin digestion of a ubiquitin-linked protein results in a characteristic peptide at the ubiquitination site, later named ubiquitin remnant motif. This motif includes a two-residue glycine-glycine remnant from the ubiquitin C-terminal sequence, covalently bonded to the ε-amino group in the target lysine side chain via an isopeptide bond (**Fig. 1.12a**). (126) This lysine modification 1) is resistant to trypsin proteolysis at its C-terminus, resulting in a miss-tryptic peptide; 2) introduces a 114 Da mass increase. These peptides can therefore be

identified as a ubiquitination signature when searching for the modification (+114 Da on an uncleaved lysine) in a proteomic software. (125) Combining protein-level enrichment for ubiquitinated proteins and mass spectrometry-assisted ubiquitin remnant motif identification has enabled the successful identification of a select number of ubiquitination substrates and sites (**Fig. 1.12b**; left) (Appendix B). (125, 127–130) However, as a protein typically only contains one or a few ubiquitination sites, protein-level enrichment does not sufficiently reduce sample complexity. This resulted in the identification of most co-immunoprecipitated peptides that were not diGlycyl modified, hindering efficient and high-throughput endogenous ubiquitomics characterization.

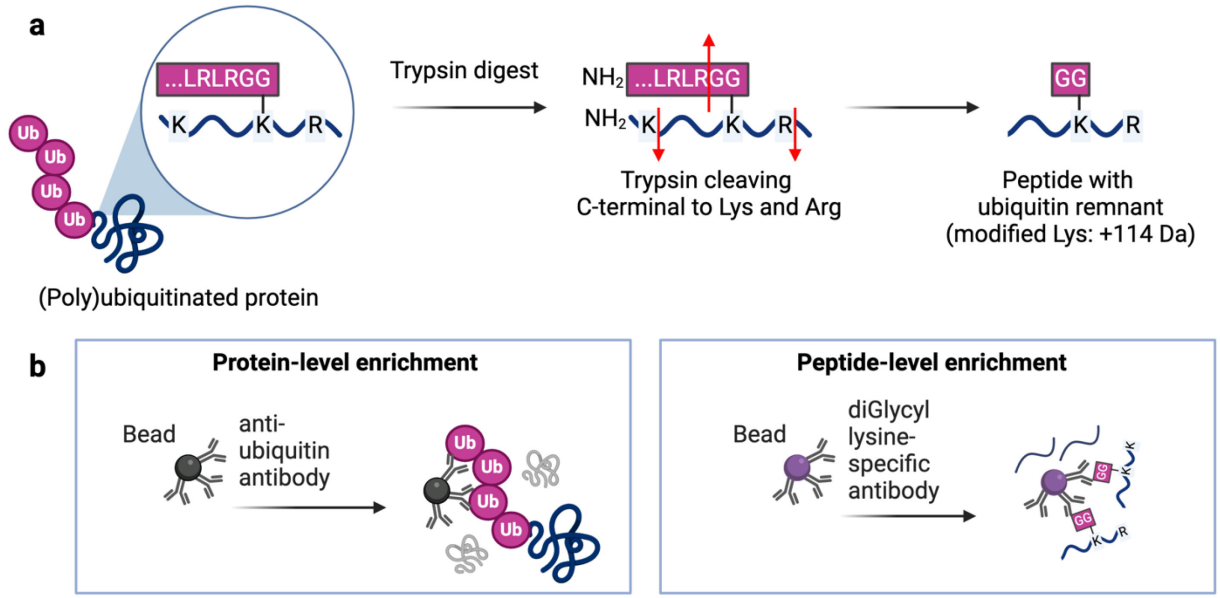


Figure 1.12. Identification of protein ubiquitination sites by mass spectrometry. a) Trypsin digestion of mono- or poly-ubiquitinated proteins leaves a diGlycyl modification on lysine side chains, also known as a ubiquitin remnant. **b**) Ubiquitinated proteins can be enriched using an anti-ubiquitin antibody, and digested to peptides for mass spectrometry identification. Alternatively, the proteome can be digested to peptides and only diGlycyl peptides are enriched for mass spectrometry analysis.

The revolutionary breakthrough in ubiquitomics arose from the anti-diGlycyl lysine antibody generated by Xu *et al.*, who immunized mice with diGly-modified histone and subsequently screened hybridoma lines for antibodies. (131) They found that the hybridoma line

GX41 generated monoclonal antibodies bound diGly lysines-containing proteins with high specificity. Further research demonstrated that the antibody could also immunoprecipitate diGlycyl-lysine containing peptides. (131) Cell Signaling commercialized this discovery named PTMScan ubiquitin remnant motif kit, which allowed peptide-level, diGlycyl-modified-lysine-specific immunoaffinity enrichment in a 1 mg cell lysate under a simple, robust and standardized workflow (**Fig. 1.12b**; right). Using this strategy, an increasing number of studies have profiled the cellular ubiquitome in biological processes, such as those associated with DNA damage, (132) protein degradation (133) and turnover (134), and cowpox virus infection. (108)

It is however important to note that the ubiquitin remnant profiling approach does not distinguish between ubiquitin and two other ubiquitin-like modifications, NEDD8 and ISG15. These ubiquitin-like proteins also end in C-terminal arginine-glycine-glycine and leave identical diGlycyl remnants on lysine side chains after trypsin digestion. Nonetheless, Kim *et al.* demonstrated that 94% of endogenous diGlycyl peptides were derived from ubiquitination versus neddylation or ISGylation in colon cancer epithelial cells HCT 116 by using ubiquitin-specific protease 2 and NEDD8-activating enzyme inhibitor. (135)

A ubiquitination-specific antibody, UbiSite, was generated to differentiate between ubiquitination and ubiquitin-like modifications. UbiSite recognizes the C-terminal 13 amino acids of ubiquitin (ESTLHLVLRLRGG) after proteolytic digestion with LysC that cleaves C-terminal to lysines. (136) Furthermore, UbiSite is not limited to modified lysines in its specificity, allowing for the identification of N-terminal and other non-canonical ubiquitination sites. Using UbiSite, Akimov *et al.* identified 63,000 unique ubiquitination sites on 9,200 proteins in Hep2 and Jurkat cells including 103 N-terminally ubiquitinated proteins. Although UbiSite is the only known antibody currently that allows distinct, unambiguous identification of ubiquitination sites, there are limitations associated with this technique. For instance, the 13-residue epitope can hinder peptide fragmentation efficiency and convolute sequence assignment. Relying on LysC only for

proteolysis will also reduce digestion efficiency, leading to generation of peptides that are too long for detection. Employing a combination of diGlycyl ubiquitin remnant and UbiSite can be helpful in pinpointing exclusive ubiquitination sites.

1.3.4 Immunoprecipitation and proximity labeling

For the high-throughput identification of protein dynamic interactions, advanced mass spectrometry-based methods are widely used. The most popular interactomics approaches are affinity-purification coupled with mass spectrometry (AP-MS), and enzyme-catalyzed proximity labeling such as BioID (proximity-dependent biotin identification).

In AP-MS, a protein of interest (“bait”) is cloned into a construct containing an affinity tag, and is overexpressed in mammalian cells. Agarose beads or magnetic agarose beads coupled with an antibody against the affinity tag are used to pull down the bait in the cell lysate, and therefore cellular proteins binding to the bait (“preys”) will be co-immunoprecipitated. The preys can be eluted using acidic buffers (e.g. 0.1 M glycine, pH 2.8) or SDS-containing buffers and heating, or directly digested to peptides for LC-MS/MS analysis (reviewed in (137)). The drawback of AP-MS is its dependence on the affinity of preys to the bait, and therefore fails to identify weak and transient protein interactions.

BioID, on the other hand, employs the overexpression of the bait fused with a biotin ligase BirA*, that allows promiscuous biotinylation of proteins in close proximity to the protein of interest over a radius of 10 nm. (138) Further improvements of BirA* mutants and truncated variants, such as TurboID, miniTurbo (139) and microID, (140) continue to reduce labeling time and ligase size, increasing reaction efficiency and specificity. However, proximity labeling does not distinguish between directly interacting and proximal proteins, and often leads to many false positives.

Computational scoring algorithms provide assessment of identified PPIs. The PPIs can be divided into three categories: 1) biologically relevant interactions; 2) biologically irrelevant interactions (physically existing but only due to sample preparation artifacts); and 3) false positives (physically non-existing interactions detected by error). (137, 141) The aim of different scoring algorithms is to reliably identify biologically relevant PPIs while reducing biologically irrelevant PPIs and filtering out false positives. The mostly used scoring pipelines includes Significance Analysis of INTeractome (SAINT), (142) Mass spectrometry interaction STatistics (MiST) (143) and Comparative Proteomic Analysis Software Suite (CompPASS) (**Table 1.1**). (144)

Algorithm	Developers	Pros	Cons
MiST	Nevan Krogan	● Highly used for virus-host AP-MS ● Configurable weights	● Not designed for intensity-level protein quantification calculation ● Default configuration penalizes common interactors unless trained
SAINT	Anne-Claude Gingras and Alexey I. Nesvizhskii	● Can integrate DIA-intensity level protein quantification ● Can boost scores based on interactions in literature	● Sometimes noisy
CompPASS	J. Wade Harper and Steven P. Gygi	● Can assess prey specificity by generating various scores	● Convoluted formula without weight customization

Table 1.1. Comparing popular AP-MS scoring pipelines. SAINT: Significance Analysis of INTeractome; (142) MiST: Mass spectrometry interaction STatistics; (143) CompPASS: Comparative Proteomic Analysis Software Suite. (144)

These algorithms leverage computational methods to analyze AP-MS data to distinguish biologically relevant interactions from noise, by calculating a composite score to assess prey specificity, reproducibility and abundance (MiST and CompPASS), or assigning scores based

on a probability distribution statistical model (SAINT). Each pipeline has advantages and limitations, and the implementation is dependent on the sample and data type.

The improved version of the SAINT algorithm, SAINTexpress, was extended to directly analyze intensity-based quantitative data. (145, 146) While the number of spectral counts is an efficient mode of protein quantification, the level of peptide intensities, especially measured by peptide fragment masses in data-independent acquisition, have been shown to be a more accurate mode of quantification. (147, 148) In addition, interaction scores can be boosted using known Gene Ontology associations between proteins, increasing confidence and our functional understanding in the identified PPI networks.

1.4 Objective of the thesis

At the molecular level, virtually all steps of viral infection and replication are reflected in changes of protein PTMs and PPIs. Identifying these changes will reveal mechanistic details of viral pathobiology and highlight key viral and cellular factors involved. Thus, the objective of the thesis was to employ specialized advanced mass spectrometry techniques to characterize PTMs and PPIs in viral infection. In Chapters 2 and 3, I describe the identification of SARS-CoV-2 protease substrates and changes in the global ubiquitome induced by VACV infection. In Chapter 4, I present a virus-host PPI network in MAYV infected cells. These findings underscore the critical functions of PTMs and PPIs in viral infections, offering new insights for the proteomic intricacies in the virus-host interplay.

Chapter 2: The role of viral protease-induced proteolysis in SARS-CoV-2

Preface

Chapter 2 was adapted from a published manuscript:

> Luo, S. Y.; Moussa, E. W.; Lopez-Orozco, J.; Felix-Lopez, A.; Ishida, R.; Fayad, N.; Gomez-Cardona, E.; Wang, H.; Wilson, J. A.; Kumar, A.; Hobman, T. C.; Julien, O. Identification of Human Host Substrates of the SARS-CoV-2 M^{pro} and PL^{pro} Using Subtiligase N-Terminomics. *ACS Infect Dis* **2023**, *9* (4): 749-761. DOI: 10.1021/acsinfecdis.2c00458

I was responsible for experimental design and optimizations, mass spectrometry analysis, functional studies and manuscript composition. E.W.M. performed experiments related to PLpro. A.K., T.C.H. and O.J. contributed to conceptualization. Other authors assisted in experiments such as sample collection, viral infection and stable cell line generation.

2.1 Abstract

The recent emergence of SARS-CoV-2 in the human population has caused a global pandemic. The virus encodes two proteases, Mpro and PLpro, that are thought to play key roles in suppression of host protein synthesis and immune response evasion during infection. To identify the host cell substrates of these proteases, active recombinant SARS-CoV-2 Mpro and PLpro were added to A549 and Jurkat human cell lysates, and subtiligase-mediated N-terminomics was used to capture and enrich protease substrate fragments. The precise location of each cleavage site was identified using mass spectrometry. Here, we report the identification of over 200 human host proteins that are potential substrates for SARS-CoV-2 Mpro and PLpro and provide a global mapping of proteolysis for these two viral proteases *in vitro* (**Fig. 2.1**). Modulating proteolysis of these substrates will increase our understanding of SARS-CoV-2 pathobiology and COVID-19.

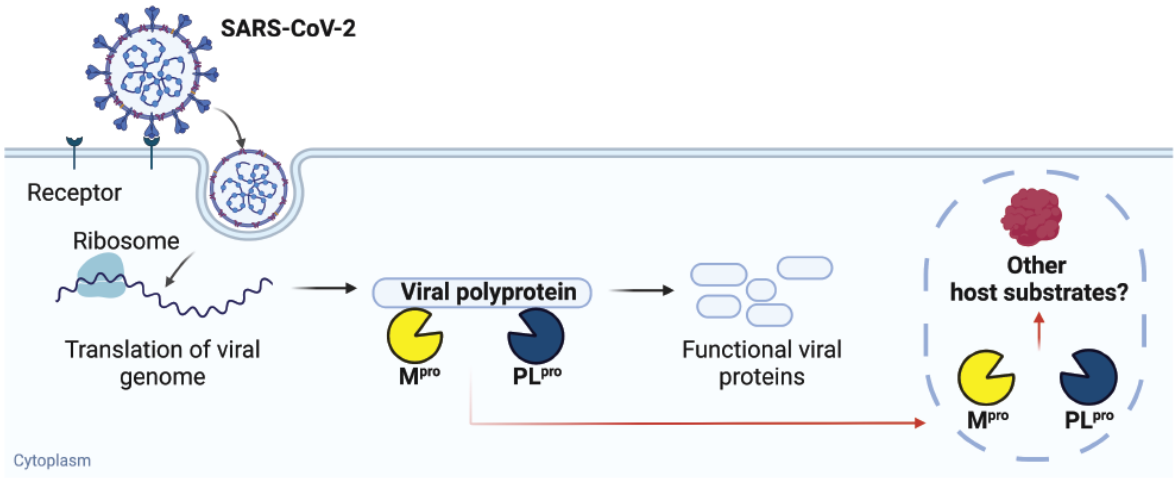


Figure 2.1. Graphical abstract of Chapter 2.

2.2 Introduction

SARS-CoV-2 is the causative agent of the COVID-19 pandemic. SARS-CoV-2 is an enveloped, positive-sense, single-stranded RNA virus in the family *Coronaviridae*, genus *β-coronavirus*. The genome of SARS-CoV-2 encodes at least 29 viral proteins including 4 structural proteins, 16 non-structural proteins (NSPs) and 9 accessory proteins.

Two of the viral proteins in SARS-CoV-2, NSP3 and NSP5, possess protease activity. They cleave two overlapping viral polyproteins (pp1a and pp1ab) translated in the major open reading frames ORF1a and ORF1b into the 16 NSPs (NSP1-16) in their active form. The NSPs possess essential enzymatic activities in viral replication, including helicase and RNA-dependent RNA polymerase. (see (149) for a review) Due to the critical role of the SARS-CoV-2 proteases, they are targets for antiviral drugs. GC376, a drug originally developed to treat feline coronavirus, also inhibits the main protease of SARS-CoV-2 and effectively blocks viral replication in cells. (64) Currently, Paxlovid (oral antiviral drug nirmatrelvir/ritonavir, Pfizer) is the only approved COVID-19 treatment targeting the SARS-CoV-2 viral protease. (65–68)

The two SARS-CoV-2 proteases are named according to their catalytic and structural similarities to other known enzymes. NSP3pro is also known as papain-like protease (PLpro) and cleaves at only three sites in the polyproteins pp1a and pp1ab. NSP5pro or picornaviral 3C-like protease (3CLpro), cleaves at eleven sites and is thus also referred to as the main protease (Mpro). Both SARS-CoV-2 proteases are cysteine proteases. The active site of SARS-CoV-2 Mpro contains a Cys145-His41 catalytic dyad. Based on its native cleavage sequence consensus in the polyproteins and its crystal structure, (150) Mpro preferentially cleaves after glutamine (P1 = Gln, Schechter and Berger nomenclature), (43) which allows stabilization in its S1 pocket by three hydrogen bonds. (151) Studies on SARS-CoV-1 Mpro show that cleavage can also occur after histidine but with lower frequency. (152) PLpro has a canonical cysteine protease catalytic triad Cys111-His272-Asp286, and is a multifunctional protein with both

proteolytic and mainly deubiquitinating activities. (153, 154) It cleaves almost exclusively after residues GlyGly at P1 and P2 positions, with high preference for hydrophobic residues in P4 (Leu in particular) and broader specificity in P3. (155)

In addition to proteolytic processing of viral polyproteins, viral proteases can cleave host substrates to modulate immune evasion and host gene expression shutoff. (156, 157) Although the interactomes of SARS-CoV-2 viral proteins have been well studied, (35, 36, 158) it is more challenging to characterize the entire range of substrates of viral proteases using conventional immunoprecipitation methods, since proteolysis can lead to substrate release and the subsequent degradation of protein fragments. Even with a catalytically-dead protease mutant, the protease-substrate interactions can be transient and difficult to detect.

A number of targeted studies have identified specific SARS-CoV-2 protease substrates in the human proteome. For example, Shin and co-workers hypothesized that high sequence homology between the SARS-CoV-1 and -2 proteases suggests that they may share common substrates and reported that ubiquitin-like interferon stimulated gene 15 protein is cleaved by SARS-CoV-2 PLpro. (55) Systematic screening of interferon stimulatory genes and human innate immune pathway proteins showed that SARS-CoV-2 Mpro cleaves the E3 ligase BRE1A (*RNF20* gene), (58) NLRP12, TAB1, (59), and CARD8, (60) and PLpro also cleaves IRF3 to dysregulate the host innate immune response. (59) Other researchers examined disrupted cellular apoptosis and autophagy pathways. Wenzel *et al.* found that Mpro cleaves NEMO, an essential modulator of NF-kappa-B signaling in brain endothelial cells, (61) while Mohamud *et al.* reported that PLpro cleaves the protein kinase ULK1. (159) Another method to identify potential viral protease targets is to search for short stretches of homologous host-pathogen sequences in the human proteome. Using this technique, Reynolds *et al.* showed that PLpro cleaves cardiac myosin proteins (MYH7 and MYH6), FOXP3, HER4 and PROS1 *in vitro*, (160) and Miczi *et al.* showed that Mpro cleaves C-terminal-binding protein 1. (161)

N-terminomics profiling of Mpro and PLpro can facilitate identification of human proteins potentially cleaved during SARS-CoV-2 infection on a greater scale. Meyer *et al.* characterized proteome-wide viral cleavage events occurring in both SARS-CoV-2 infected African green monkey kidney cells (Vero E6) and human lung carcinoma cells overexpressing the virus entry receptor (A549-ACE2). (162) Refining the cleavage sites to match viral protease specificities, they identified 14 putative Mpro and PLpro substrates. Further biochemical analysis confirmed Mpro cleavage of pinin, phosphoribosylaminoimidazole carboxylase (*PAICS* gene) and golgin A3 (*GOLGA3* gene), whereas PLpro cleaves the protein kinase Src.

When a purified recombinant viral protease is incubated with human cell lysates, N-terminomics methods, such as terminal amine isotopic labeling of substrates (TAILS), (163) can be used to identify protease cleavage sites. Using this method, Koudelka *et al.* identified 318 unique protein substrates of SARS-CoV-2 Mpro in lung epithelial carcinoma cells and pulmonary microvascular endothelial cells, but did not validate these substrates in infected cells. (62) Also using TAILS, Pablos *et al.* profiled 101 Mpro substrates in human embryonic kidney cells and lung epithelial cells treated with antiviral type I interferons. (63) They further characterized and performed functional studies on several of these Mpro substrates such as PTBP1 and the RNA polymerase, RPAP1, which are proteins involved in host transcription and translation. They confirmed the Mpro cleavage of proteins in the Hippo signaling pathway: the transcriptional coactivator YAP1, protein kinase MAP4K5, CREB1 and ATF-1, as well as proteins involved in the antiviral response, such as galectin-8 and FYCO1.

Here, we employed an enzyme-mediated N-terminomics approach for the comprehensive identification of potential substrates of SARS-CoV-2 Mpro and PLpro in the human proteome. Using subtiligase-mediated N-terminomics in cell lysates, we identified 191 and 16 putative substrates of SARS-CoV-2 Mpro and PLpro, respectively. The enzymatic-labeling approach presented here is unique and complementary to the known SARS-CoV-2

degradome reported by other groups. By comparing our results to previous studies, we have generated a list of all current SARS-CoV-2 protease substrates reported thus far, thereby filling the gap of uncharacterized Mpro and PLpro interactomes. There is still a need for additional antivirals for COVID-19 patients, and the characterization of SARS-CoV-2 protease cellular targets will help us better understand the fundamental virology of SARS-CoV-2.

2.3 Materials and methods

2.3.1 SARS-CoV-2 Mpro expression and purification

The recombinant His6-GST-dual-tagged SARS-CoV-2 Mpro expression plasmid in the pGEX-6P-1 vector was cloned and kindly gifted by Dr. Rolf Hilgenfeld's lab. (150) The plasmid was transformed into *Escherichia coli* strain BL21-Gold (DE3) cells (Novagen). Cells were grown in LB media supplemented with 100 µg/mL ampicillin at 37°C to an OD_{600} at 0.8. Protease expression was induced with 0.5 mM IPTG at 37°C for 5 hours. Cells were harvested by centrifugation at 6,000 x g and lysed by EmulsiFlex (Avestin). The cell lysates were clarified by centrifugation, and the soluble fraction was purified by HisTrap FF column (5 mL; Cytiva). The eluants were pooled and dialyzed with 10 units of PreScission protease (Cytiva) per mg of target protein. The cleaved proteins were applied to connected GSTrap FF (1mL; Cytiva) and Talon (1mL; Cytiva) columns. The flow-through was collected and concentrated using Amicon Ultra 15 centrifugal filters (10 kDa). The purified untagged proteins were diluted with glycerol, flash frozen in liquid nitrogen and stored at -80˚C.

2.3.2 SARS-CoV-2 PLpro expression and purification

The GST-tagged SARS-CoV-2 PLpro expression plasmid in the pGEX-6P-1 vector was graciously gifted by Dr. Shaun K. Olsen's lab. (155) The plasmid was transformed in the *Escherichia coli* strain BL21(DE3)pLysS. Cells were grown in LB media supplemented with 100 µg/mL ampicillin and 25 µg/mL chloramphenicol at 37°C with 250 rpm shaking to an OD_{600} at

0.8. The media was supplemented with 0.1 mM zinc sulfate to increase protein stability, and protein expression was induced with 0.5 mM IPTG 18°C with 200 rpm shaking for 16 hours. The cells were harvested by centrifugation at 6,000 x g for 15 minutes at 4˚C and subsequently lysed in binding buffer (50 mM Tris pH 7.5, 300 mM NaCl, 2 mM 2-mercaptoethanol) by EmulsiFlex (Avestin). The lysates were clarified by centrifugation and the soluble fraction was purified GSTrap HP column (5 mL; Cytiva). The eluants in elution buffer (50 mM Tris pH 7.5, 300 mM NaCl, 2 mM 2-mercaptoethanol, 100 mM reduced glutathione) were pooled and dialyzed for 12 hours with 10 units GST-PreScission protease (Cytiva) per mg of target protein, or 1 mg protease per 50 mg target protein, in dialysis buffer (50 mM Tris pH 7.5, 300 mM NaCl, 1 mM EDTA, 1 mM DTT). The cleaved proteins were purified by GSTrap HP column (5 mL; Cytiva). The flow-through and wash fractions were pooled and concentrated by Amicon Ultra 15 centrifugal filters (10 kDa). The purified untagged protease was diluted to 10% glycerol, flash frozen in liquid nitrogen and stored at -80˚C.

2.3.3 Synthesis of coumarin fluorescent probe

A total of 200 mg of Rink Amide AM resin (0.89 mmol/g) were transferred to the reaction cartridge (Poly-Prep Chromatography Column, Bio-Rad) and 6 mL of dichloromethane (DCM) were added to the resin for swelling (30 min with constant mixing). DCM was removed by vacuum filtration and the resin was washed three times with dimethylformamide (DMF), one time with methanol, one time with DCM and a final wash with DMF (6 mL per wash).

The Fmoc-group was removed with 6 mL of 20% (v/v) Piperidine in DMF. The resin and deprotection solution were gently agitated for 30 min. After that, the solution was removed by vacuum filtration and the resin was washed five times with DMF (6 mL per wash). After the final wash, a Kaiser test (ninhydrin test) was performed to confirm the removal of the Fmoc-group. The Kaiser test reagents were prepared according to manufacturer’s recommendations

(AAPPTec). For Reagent A, 16.5 mg of potassium cyanide (KCN) were dissolved in 25 mL of distilled water. A 1:50 dilution was made with 1 mL of the KCN solution and 49 mL of pyridine. For Reagent B, 1 g of ninhydrin was dissolved in 20 mL of butanol. The Reagent C contains 20g of phenol in 10 mL of n-butanol. A few beads were transferred to a 1.5 mL Eppendorf tube. Three drops of each reagent were added. The mixture was heated for 3 min at 95°C in a heating block. The presence of a blue color indicates deprotection of the resin. Addition of the Fmoc-ACC group was carried out according to Poreba *et al.* with a few modifications. (164) A mixture of Fmoc-ACC-OH (0.35 mmol, 2 eq), HATU (0.35 mmol, 2 eq) and Collidine (0.53 mmol, 3 eq) in 3 mL of DMF was added to the resin. The cartridge was protected from light with aluminum foil and incubated with gentle agitation for 24h. Next day, the mixture was removed by vacuum filtration and the resin was washed five times with DMF (6 mL per wash). Two extra ACC additions were carried out using the same conditions. Kaiser test was performed at the end to confirm coupling completion.

The sequence used for the probe corresponds to the most preferred substrate for Mpro, Ac-Abu-Tle-Leu-Gln-ACC. (151) Each step addition was done for 2h with constant mixing using the Fmoc-protected version of each residue (1.75 mmol, 10 eq), HATU (1.75 mmol, 10 eq) and Collidine (1.75 mmol, 10 eq) in 5 mL of DMF. Followed by Fmoc-group removal for 15 min with 20% (vol/vol) Piperidine in DMF (1.2 mL of Piperidine in 4.8 mL of DMF). Five DMF washes at the end of addition and deprotection steps were done (6 mL each). The completion of the reaction was monitored with the Kaiser test. After the final deprotection, capping of the N-termini was done with 6 mL of the acetylation mixture (Acetic anhydride, Pyridine and DMF in a 20:20:60% v/v/v) for 30 min with constant mixing. Once Kaiser test was negative (yellow color in solution and beads), the resin was washed five times with 6 mL of DMF and three times with 6 mL of DCM. The resin was dried by vacuum filtration for 1h. Cleavage of the final product was carried out for 2h with constant mixing with 5 mL of the cleavage solution,

TFA/H2O/Tripropylsilane (95:2.5:2.5% v/v/v). The solution was recovered and precipitated in 40 mL of cold Diethyl ether for 1h. Tube was spun down at 8000xg for 20 min. The pellet was resuspended in ACN/H2O (50/50% v/v) and lyophilized until fully dry. The purity of the substrate was confirmed by MALDI-TOF (Autoflex speed MALDI-TOF, Bruker). The final ACC probe was dissolved in DMSO at a final 10 mM concentration and stored at -80°C. Similarly, the probe of sequence Ac-Leu-Arg-Gly-Gly-ACC was synthesized for PLpro, (155) and stored at -80°C in DMSO at a final concentration of 1 mM.

2.3.4 SARS-CoV-2 Mpro activity assay

Activity assays were performed in 96-well standard opaque plates using microplate reader (SpectraMax M3; Molecular Devices) in assay volumes of 100 µL. In cell-free assays, 20 mM Tris-HCl, pH 7.5, 1 mM ethylenediaminetetraacetic acid (EDTA) and 10 mM dithiothreitol (DTT) was used as the assay buffer. To assay in cell lysates, cells were lysed by probe sonication in lysis buffer (20 mM Tris-HCl, pH 7.5, 0.1% Triton x-100 and 10 mM DTT with protease inhibitors [5 mM EDTA, 1 mM 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), 1 mM phenylmethanesulfonyl fluoride (PMSF), and 20 µM z-VAD-fmk (N-Benzyloxycarbonyl-Val-Ala-Asp(O-Me) fluoromethyl ketone]. Cell lysates were clarified by centrifugation and the soluble fraction was taken as the assay buffer. Final concentrations of 0.5 µM purified SARS-CoV-2 Mpro and 2 µM of coumarin probe Ac-Abu-Tle-Leu-Gln-ACC dissolved in DMSO were added to the buffer, with a final [DMSO] of 0.2%. The PMT gain was set to low with reads in 30 s intervals for 1 hour, at λexcitation of 355 nm and λemission of 460 nm.

2.3.5 SARS-CoV-2 PLpro activity assay

Activity assays were performed in 96-well standard opaque plates using a microplate reader (SpectraMax M3; Molecular Devices) in assay volumes of 100 µL. In cell-free assays, 20

mM Tris-HCl, pH 8.0, 5 mM NaCl and 5 mM DTT was used as the assay buffer. To assay in cell lysates, cells were lysed by probe sonication in lysis buffer (20 mM Tris-HCl, pH 8.0, and 0.1% Triton x-100 with protease inhibitors [5 mM EDTA, 1 mM AEBSF, 1 mM PMSF, and 4 mM iodoacetamide (IAM) with 30 min incubation in the dark]. IAM was quenched with 20 mM DTT then the cell lysates were clarified by centrifugation and the soluble fraction was taken as the assay buffer. Final concentrations of 5 μM purified SARS-CoV-2 PLpro and 10 μM of coumarin probe Ac-Leu-Arg-Gly-Gly-ACC dissolved in DMSO were added to the buffer, with a final [DMSO] of 1%. The PMT gain was set to low with reads in 30 s intervals for 3 hours, at λexcitation of 355 nm and λemission of 460 nm.

2.3.6 Cell culture

A549 and Jurkat (ATCC) were cultured respectively in Dulbecco's Modified Eagle Medium (DMEM) (Gibco #11995-065) and Roswell Park Memorial Institute (RPMI) 1640 medium (Gibco #11875-093), supplemented with 10% fetal bovine serum, 100 U/mL penicillin and 100 μg/ml streptomycin, and 2 mM L-glutamine.

2.3.7 N-Terminal labeling and enrichment

The expression constructs for subtiligase expression (WT and M222A mutants) were a gift from Jim Wells (University of California San Francisco) and Amy Weeks (University of Wisconsin-Madison). (123) Jurkat (5×10^9) and A549 (2.5×10^9) cells were respectively used in each corresponding replicate. Cells were harvested by centrifugation and lysed by gentle probe sonication in lysis buffer to maintain native protein fold (20 mM Tris-HCl, pH 7.5, 0.1% Triton x-100 and 10 mM DTT with protease inhibitors [5 mM EDTA, 1 mM AEBSF, 1 mM PMSF] for Mpro and 20 mM Tris-HCl, pH 8.0, and 0.1% Triton x-100 with protease inhibitors [5 mM EDTA, 1 mM AEBSF, 1 mM PMSF, 4 mM IAM subsequently quenched with 20 mM DTT prior to addition of PLpro] for PLpro). In Jurkat cell lysates, we also added 20 μM z-VAD-fmk to

irreversibly inhibit endogenous caspases prior to adding SARS-CoV-2 Mpro. Cell lysates were clarified by centrifugation. For PLpro, 10X assay buffer (200 mM Tris-HCl, pH 8.0, 50 mM NaCl, 50mM DTT) was added 1:10 to clarify the lysate. 0.5 μM of purified SARS-CoV-2 Mpro or 5 μM of purified SARS-CoV-2 PLpro was added to the soluble cell lysates for 2 h incubation, with aliquots taken out to monitor protease activity as a function of time. N-terminal labeling was then performed with 1 μM stabiligase WT, 1 μM subtiligase M222A, and 1 mM TEVest6 (123) for 1 h. Tagged protein fragments were precipitated using acetonitrile, then denatured (8 M Gdn-HCl) and reduced (5 mM TCEP), and thiols were alkylated (10 mM IAM), before ethanol precipitation. Biotinylated N-terminal peptides were then captured with NeutrAvidin agarose beads (ThermoFisher) for 24 h. The beads were washed using 4 M Gdn-HCl, trypsinized, and peptides were released from the beads using TEV protease. The TEV protease was precipitated using 2.5% TFA, and the peptides were desalted with using C18 Ziptips (Rainin).

2.3.8 Mass spectrometry analyses

Peptides were separated using a nanoflow-HPLC (Thermo Scientific EASY-nLC 1200 System) coupled to Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific). A trap column (5 μm, 100 Å, 100 μm × 2 cm, Acclaim PepMap 100 nanoViper C18; Thermo Fisher Scientific) and an analytical column (2 μm, 100 Å, 50 μm × 15 cm, PepMap RSLC C18; Thermo Fisher Scientific) were used for the reverse phase separation of the peptide mixture. Peptides were eluted over a linear gradient over the course of 120 min (or 90 min for the PLpro Jurkat dataset) from 3.85% to 36.8% acetonitrile in 0.1% formic acid. 2 replicates of Mpro N-terminomics in Jurkat lysates were injected on the MS with and without the installation of FAIMS Pro interface (Field Asymmetric Ion Mobility Spectrometry) to broaden protein coverage. Data were analyzed using ProteinProspector (v5.22.1) against the concatenated database of the human proteome (SwissProt.2017.11.01.random.concat), with maximum false discovery rate 1% for peptides. The peptides were searched at a maximum of 3 missed trypsin

cleavages with TrypsinPro digest specificity relaxed at peptide N-termini. Search parameters included a precursor mass tolerance of 15 ppm, a fragment mass tolerance of 0.8 Da, precursor charge range of 2-5, with the constant modification carbamidomethylation (C), and variable modifications of and Abu (N-term), deamidated (N/Q), and oxidation (M). The maximum number of variable modifications was set to 3. MS data are available through MassIVE: MSV000088583 and MSV000088584 (Mpro), and MSV000090124 and MSV000090125 (PLpro).

2.3.9 *In vitro* cleavage assays of putative substrates

HEK293T and HEK293T-ACE2 cells were transiently transfected with plasmid GFP-BRD2 (Addgene #65376) or FLAG-SFPQ (Addgene #166960) using Polyplus jetOPTIMUS DNA transfection reagent, lifted using 0.5 mM EDTA and pelleted. Jurkat and A549 cells were cultured, lifted using 0.5 mM EDTA, pelleted by centrifugation, and lysed in the same lysis buffer used in the N-terminomics protocol above. The cell lysates were incubated with or without the active recombinant SARS-CoV-2 Mpro and PLpro, with activity assays to monitor protease activity in parallel. Aliquots of the cell lysates were taken at time points 0, 1, and 2 h, and reactions were quenched by boiling with the 5X Laemmli buffer for 5 min. The GFP-BRD2 Q206A plasmid was generated using site-directed mutagenesis with the forward oligo: GCCAAGTTGGCAGCGCTCGCGGGCAGTGTTACCAGTG and reverse oligo: CACTGGTAACACTGCCCGCGAGCGCTGCCAACTTGGC to mutate codon CAG to GCG (oligos purchased from IDT). The thermocycle was performed on 50 ng GFP-BRD2 (Addgene, #65376) and pfu (Truin Science Ltd., #ETS4020) with 5 min initial denaturation at 95˚C, 17 cycles of 50 s denaturation at 95˚C, 50 s of annealing at 50˚C, and 16 min of extension of 68˚C, and final extension at 68˚C for 10 min. The PCR product was incubated with DpnI, and transformed in DH5α cells. The final extracted plasmid was Sanger sequenced.

2.3.10 Stable cell line generation and viral infection

Stable cell line generation and viral infection. HEK293T-ACE2, A549-ACE2 and H23-ACE2 stable cell lines and SARS-CoV-2 infection were performed as described previously. (165) SARS-CoV-2 (hCoV-19/Canada/ON-VIDO-01/2020; GISAID accession no. EPI_ISL_425177) was kindly provided by Darryl Falzarano (Vaccine and Infectious Disease Organization, Saskatoon, Canada). HEK 293T-ACE2 and A549-ACE2 cells were developed by electroporating a human ACE2 encoding plasmid (Addgene #1786; a gift from Hyeryun Choe). The cells were passaged six times in culture, surface-stained for ACE2 (goat anti-ACE2; AF933-SP; R&D Systems), and the highest 2% of cells expressing ACE2 were sorted from the bulk population. Virus culture and experiments were performed according to level 3 containment procedures. Virus stocks were generated and titrated (by plaque assay) in Vero E6 cells, and HEK293T-ACE2, A549-ACE2 and H23-ACE2 cells were infected using MOI = 1.

2.3.11 Immunoblot

SARS-CoV-2 infected cell lysates and cell lysates incubated with SARS-CoV-2 Mpro and PLpro were loaded on 7.5% or 10% SDS-PAGE gels. After separation, proteins were transferred onto 0.45 µm nitrocellulose membranes (BioRad), blocked in 2.5% fish skin gelatin in TBS at RT for 1 h, then incubated with primary antibodies diluted in 2.5% fish skin gelatin in TBST at 4˚C overnight. The membrane was washed 3X with TBS for 5 min, and incubated with secondary antibodies diluted in 2.5% fish skin gelatin in TBST at RT for 1 h. The membrane was washed again 2X with TBST and 1X with TBS for 5 min before viewing on the LI-COR Odyssey imaging system. Antibodies and mammalian plasmids used in this study are presented in **Table 2.1**. For full uncropped immunoblot images see Appendix A.

Acc #	Gene Name	Sources	Catalog #	Plasmids for overexpression studies
Q13263	TRIM28	R&D Systems	MAB7785	https://www.addgene.org/45569/
Q7Z2W4	ZC3HAV1	Proteintech	16820-1-AP	https://www.addgene.org/45907/
		GeneTex	GTX120134	https://www.addgene.org/45906/
P52948	NUP98	Wozinak Lab		-
P25440	BRD2	Abcam	ab139690	https://www.addgene.org/65376/
Q53GL7	PARP10	LSBio	LS-C747885	-
P23246	SFPQ	Thermo Fisher	PA519663	https://www.addgene.org/166960/
P0DTD1	NSP3	-	-	https://www.addgene.org/165108/ https://www.addgene.org/165131/
P42212	GFP	Abcam	ab6673	-
P04406	GAPDH	Cell Signaling	2118	-
P68363	α-tubulin	Cell Signaling	3873	-
Q93H4B7	β-tubulin	Sigma	T5293	-
P60709	β-actin	Abcam	ab8224	-

Table 2.1. Antibodies and plasmids used in the study.

2.4 Results

To identify the host substrates of SARS-CoV-2 Mpro and PLpro and their corresponding cleavage sites, we used a subtiligase-mediated N-terminomics approach to positively enrich the newly generated N-termini from cleaved proteins in human cell lysates (123, 166) (**Fig. 2.2**). Nascent N-termini were enzymatically labeled with a biotinylated peptide ester using subtiligase, allowing for the subsequent positive enrichment of biotinylated proteins on immobilized neutravidin beads. The proteins were further digested by trypsin, and the bound N-terminal peptides were released from the beads by cleavage at a tobacco etch viral (TEV) site engineered into the biotin ester tag. This leaves a unique N-terminal α-aminobutyric acid (Abu) modification on the peptides allowing for unambiguous and precise identification of SARS-CoV-2 protease cleavage sites using LC-MS/MS.

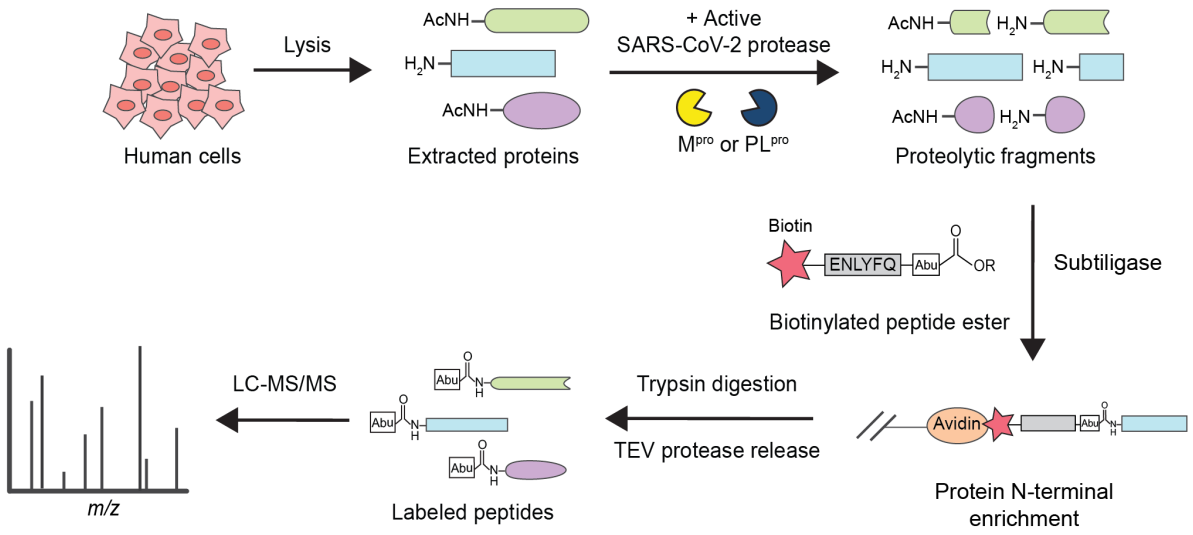


Figure 2.2. Identification of SARS-CoV-2 Mpro and PLpro host substrates in *in vitro* subtiligase N-terminomics. Active recombinant Mpro or PLpro was added to human cell lysate, generating protein cleavages that were labeled with a designed biotinylated peptide ester by subtiligase. After enrichment by neutravidin, trypsin and TEV protease were added for the release of labeled peptides with a unique N-terminal mass tag Abu (α-aminobutyric acid), allowing for identification of viral protease cleavage sites in LC-MS/MS.

2.4.1 Activity assay in cell lysates

We used two different cell lines, A549 (adenocarcinomic human alveolar basal epithelial cells) and Jurkat (human T lymphocyte cells) to compare the results across different cell origins and increase substrate identification in human proteomes. To ensure the purified recombinant protease was active in cell lysates, we monitored its proteolytic activity using a fluorescence activity assay. The optimal P4-P1 substrates of SARS-CoV-2 Mpro and PLpro were previously identified via substrate specificity screening. (151, 155) Coumarin probes based on these sequences were used to test recombinant protease activities: Ac-Abu-Tle-Leu-Gln-ACC was incubated with purified Mpro, and Ac-Leu-Arg-Gly-Gly-ACC with PLpro in both cell-free environment and in cell lysates (**Fig. 2.3, 2.4**). Following optimization, the viral proteases were able to cleave the coumarin probes in the complex cellular environment, demonstrating comparable fluorescence signals to cell-free assays.

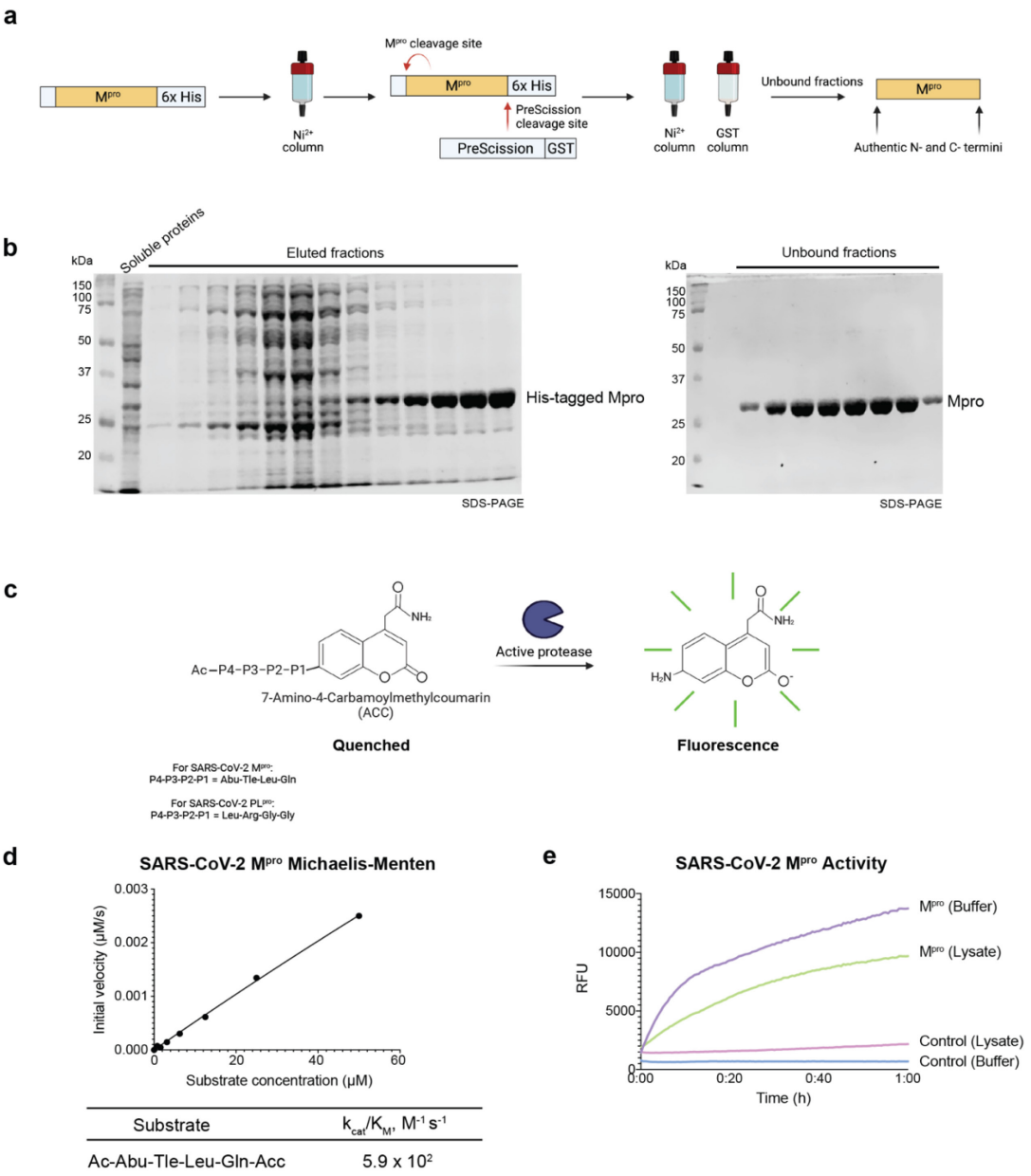


Figure 2.3. Plasmid construct, protein expression and purification of SARS-CoV-2 Mpro. **a**,**b**) SARS-CoV-2 Mpro was expressed and purified by affinity purification with authentic N- and C-termini. **c**) The fluorescence activity assay was carried out using the optimal coumarin substrate Ac-Abu-Tle-Leu-Gln-ACC. **d**) The enzyme kinetics assay was performed using 0.09 μM Mpro, and 0.78 to 50 μM coumarin substrate in 100 μL total assay volume. The kcat/KM was calculated using the linear region of the Michaelis-Menten curve kcat/KM = slope / [E], (155) and is consistent with previously reported value. (151) **e**) The protease activity was monitored in parallel with the reverse N-terminomics, in cell-free conditions and cell lysates with 0.5 μM Mpro and 2 μM coumarin substrate, showing that it was proteolytically active.

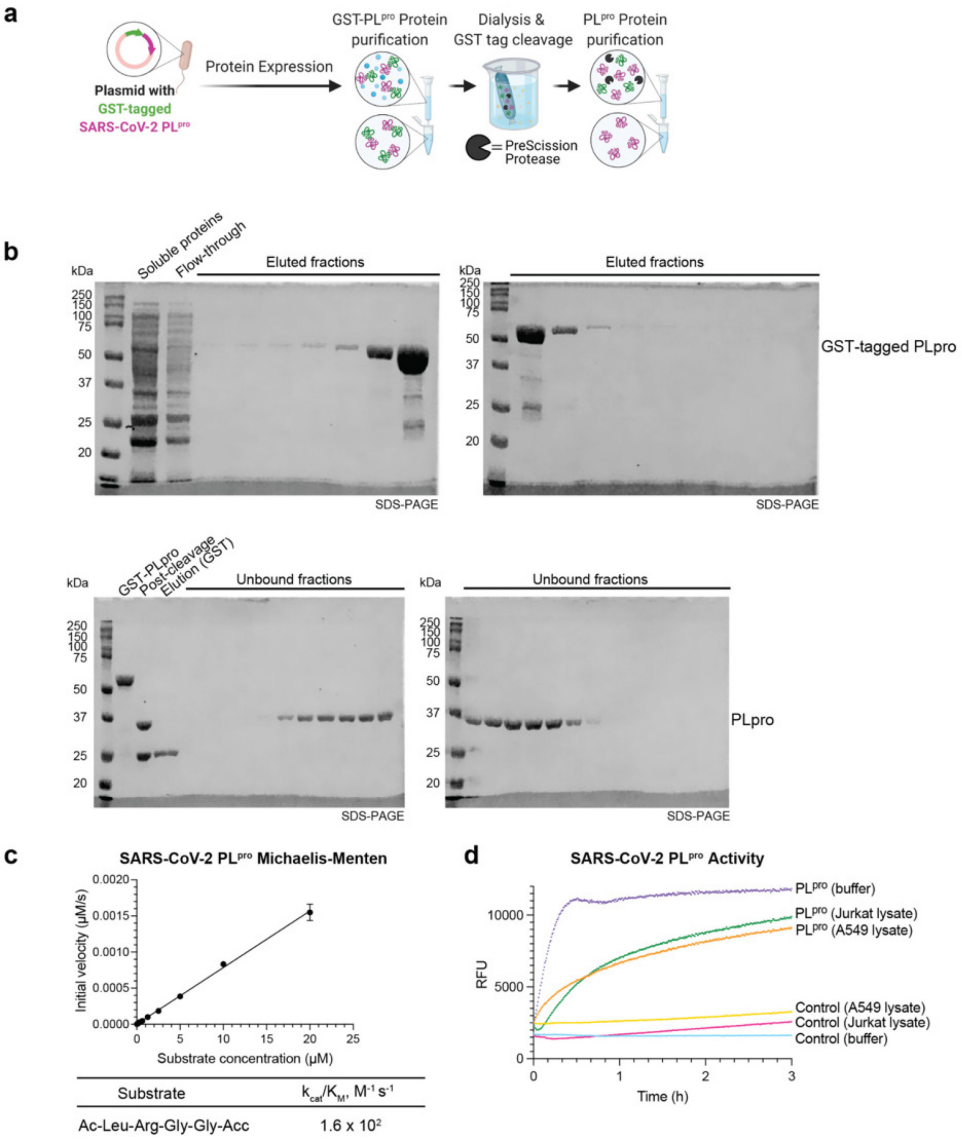


Figure 2.4. Protein expression, purification and activity of PLpro. a) PLpro expression and purification was conducted using a plasmid encoding for the protease domain of Nsp3 with an N-terminal GST tag. GST-PLpro is purified from *E. coli* lysates using a glutathione sepharose column. The GST tag is removed in an overnight dialysis using a PreScission protease. **b**) Prior to removal of the GST tag, GST-PLpro is collected in the elution fractions. Following GST tag removal, the PreScission protease and GST-tag remain bound to the column and collected in the second elution while PLpro is obtained in the unbound fractions. **c**) The fluorescence activity assay was carried out using the optimal coumarin substrate Ac-Leu-Arg-Gly-Gly-ACC (see **Fig. 2.2**). The enzyme kinetics assay was performed using 0.5 μM PLpro, and 0.78 to 20 μM coumarin substrate in 100 μL total assay volume. The kcat/KM was calculated using the linear region of the Michaelis-Menten curve kcat/KM = slope / [E]. (155) **d**) The activity of 5 μM PLpro was measured using 10 μM of the coumarin substrate in buffer, A549 and Jurkat cell lysates. Performed by Eman W. Moussa.

2.4.2 Identification of Mpro substrates

Since primary T lymphocytes have been previously reported to be infected by SARS-CoV-2, (167) and Jurkat cells were a convenient suspension cell model for cell expansion, we first performed two N-terminomics replicates in Jurkat cell lysates to identify human host substrates of the Mpro. We discarded labeled peptides with N-termini located within the first four residues from the start of a protein sequence in order to focus our analysis on endoproteolytic sites (e.g. to avoid protein start sites and methionine removal). In the Jurkat proteome, we were able to identify 746 labeled unique cleavages in 600 host substrates. We then searched for cleavage sites with Gln and His residues at P1 position only (P1=Q/H), which correspond to Mpro specificity. (151) This yielded 154 unique cleavages in 146 substrates, exhibiting a 20.6% enrichment at P1=Q/H from background protease activity (**Fig. 2.5a**). To expand the host proteome targeted by Mpro, we performed two experimental replicates with lung epithelial cells (A549) and identified 2283 unique labeled cleavage sites. Of these, 210 cleavage sites in 196 substrates contained P1=Q/H, corresponding to an enrichment rate of 9.0% (**Fig. 2.5a**). Interestingly, the enrichment rate of Mpro-specific cleavage sites in A549 was lower than in Jurkat, but still higher than untreated lysates with endogenous proteases that typically showed a P1=Q/H at 3.9% (2.5% and 1.4%, respectively). (168) This suggested that the added Mpro was active in cell lysates and cleaved human substrates. Combining labeled cleavages with P1=Q/H in A549 and Jurkat, we found 334 unique sites (**Fig. 2.6**). However, we hypothesized that a new emerging virus such as SARS-CoV-2, would cleave host substrates at new sites i.e. sites not identified in previous N-terminomics experiments. Therefore, we used the DegraBase (169) to eliminate sites previously observed in healthy and apoptotic cells in subtiligase-based N-terminomics experiments. These included background proteolysis in human cells due to the incomplete inhibition of endogenous proteases. In addition, we could narrow down the list of Mpro substrates by identifying cleavage sites found in the PLpro dataset matching Mpro specificity, and vice versa. In total, we estimated that 39% of cleavage sites featuring P1=Q/H

may not be directly attributed to the Mpro activity (**Fig. 2.6a**, 131/334 cleavage sites). While these cleavages may result from background proteolysis of the host cells, we also cannot rule out that the viral proteases may be targeting the same sites as the host proteins. Ultimately, we focused our analysis on the 157 cleavage sites where P1=Q and 46 where P1=H, found in 148 and 46 host substrates, respectively (for a total of 191 substrates as three substrates contain both P1=Q and P1=H cleavage sites). Overall, we observed 203 unique cleavage sites (P1=Q/H) in 191 human substrates cleaved by the Mpro of SARS-CoV-2 *in vitro*.

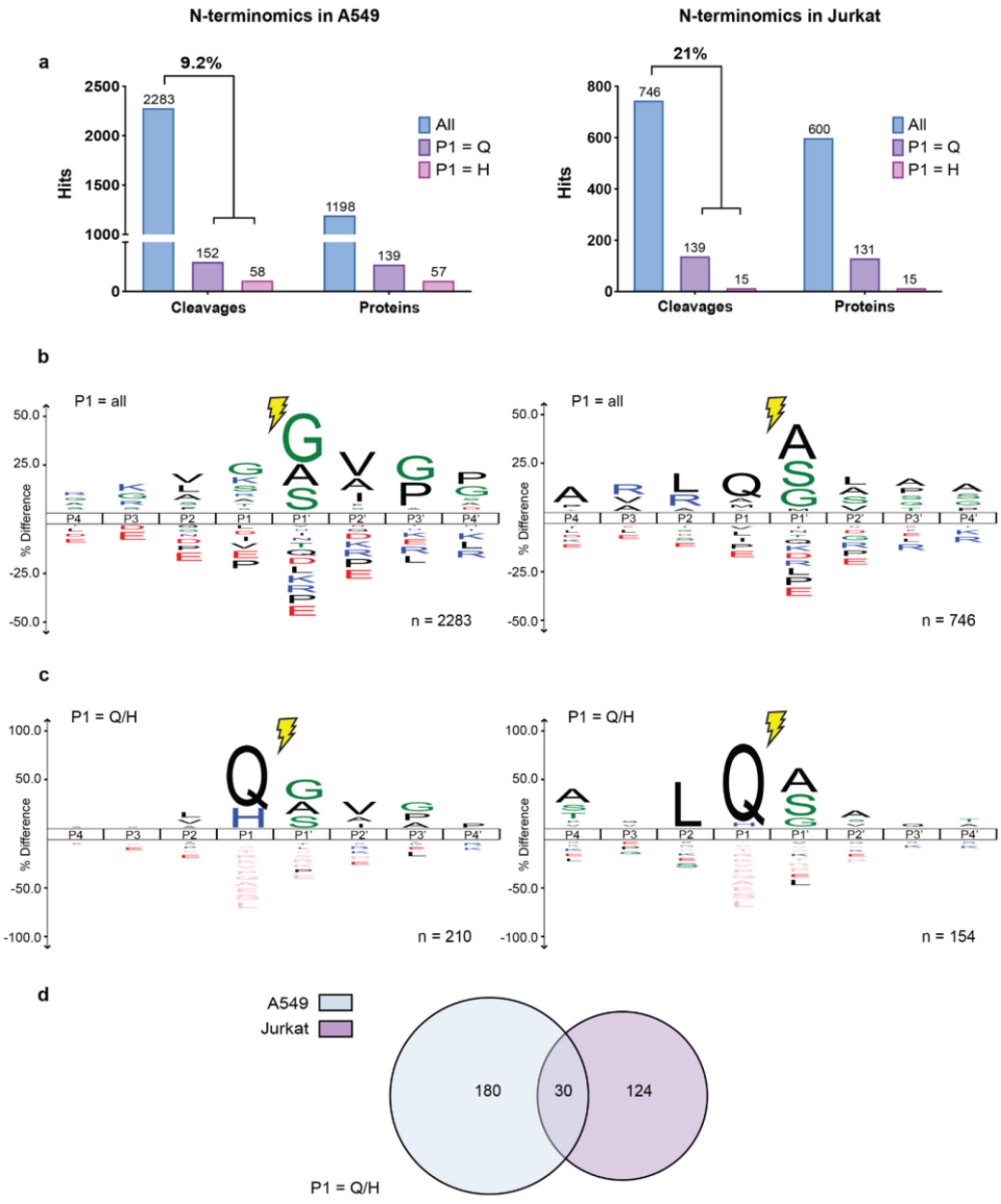


Figure 2.5. Identification of SARS-CoV-2 Mpro substrates. a) N-terminomics statistics of two A549 replicates (left) and of two Jurkat replicates (right). In A549, 2283 unique cleavages were labeled, and 210 sites in 196 host proteins correspond to SARS-CoV-2 Mpro specificity with Gln or His at P1 residue (P1=Q/H) at 9.2% enrichment rate. In Jurkat, 746 unique labeled cleavages were identified with 154 sites at P1=Q/H in 146 proteins, showing an enhanced enrichment at 21%. **b**) IceLogo showing P4-P4′ residue enrichment in all labeled cleavage sites in A549 (left) and Jurkat (right), and **c**) in sites where P1=Q/H only. **d**) Venn diagram showing the overlap in P1=Q/H cleavages between the A549 and Jurkat proteomes.

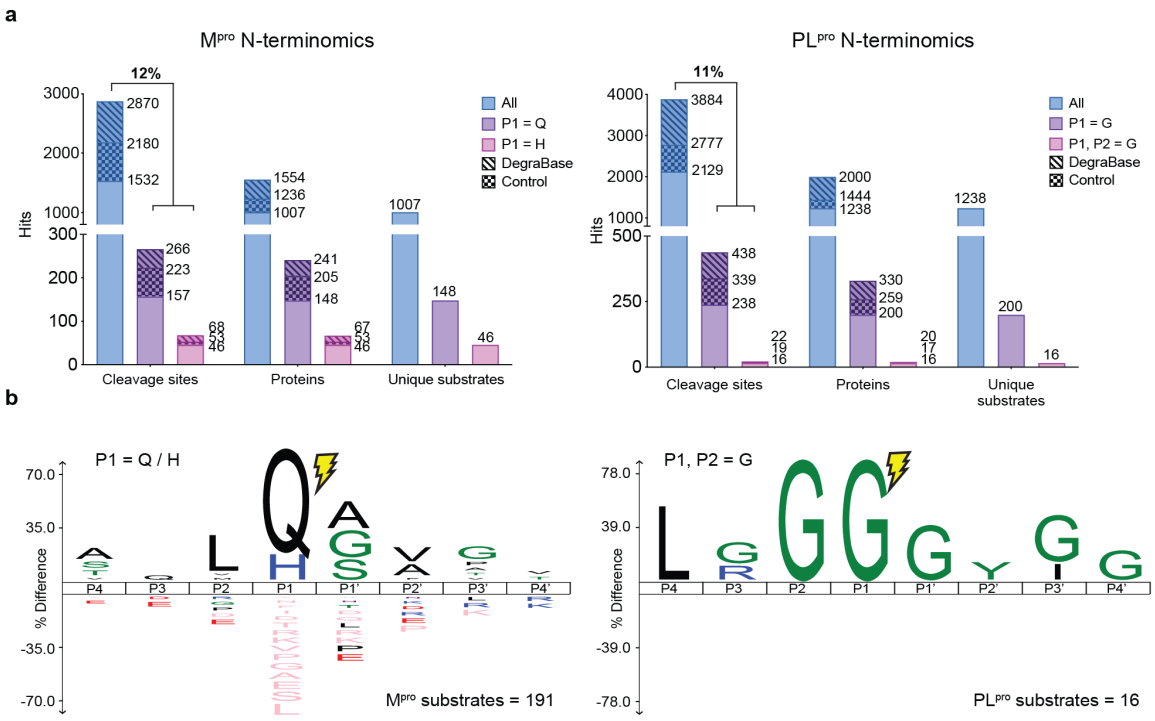


Figure 2.6. Identification of SARS-CoV-2 Mpro and PLpro substrates using subtiligase N-terminal labeling. a) N-terminomics statistics of two Jurkat and two A549 replicates for each of Mpro (left) and PLpro (right). For Mpro, 2870 unique labeled cleavages were identified with 334 sites at P1=Q/H in 308 proteins, showing an enhanced enrichment at 12%. For PLpro, 3884 unique labeled cleavages were identified with 438 sites at P1=G and 22 sites at P1,P2=G in 330 and 20 proteins, respectively, showing enrichment for P1=G at 11%. For each viral protease, the number of Mpro and PLpro substrates unique to each protease is also reported. **b**) IceLogo showing P4-P4′ residue enrichment for Mpro with P1=Q/H (left) and for PLpro with P1,P2=G (right).

2.4.3 Identification of PLpro substrates.

We conducted similar experiments with PLpro. We performed two replicates in A549 and Jurkat cell lysates and identified 3884 labeled unique cleavage sites in 2000 human proteins (**Fig. 2.6**). We then looked for cleavage sites with Gly in the P1 position only or with Gly in both P1 and P2 positions, corresponding to the known PLpro specificity. (155) We identified 438 unique cleavages in 330 host proteins corresponding to a 11.2% enrichment of P1=G, and 22 unique cleavages in 20 proteins corresponding to a 0.65% enrichment of P1,P2=G (**Fig. 2.6**). Additionally, by comparing these results to the DegraBase and removing any cleavage sites with P1,P2=G detected in the Mpro N-terminomics experiments, we identified 16 unique

cleavage sites that have not been previously observed in healthy and apoptotic cells by subtiligase-based N-terminomics (**Table 2.2**). Of particular interest, 11 of these featured a Leu at P4 (LxGG motif). By comparison, only one LxGG cleavage site was observed in the Mpro dataset. Overall, we identified 16 new cleavage sites at P1,P2=G in 16 putative substrates in SARS-CoV-2 PLpro *in vitro* N-terminomics, with 11 featuring a LxGG motif.

Acc #	P4-P1 | P1′-P4′	Gene	Protein Name	Subcellular Localization
O00487	$LGGG_{10}$ |MPGL	PSMD14	26S proteasome non-ATPase regulatory subunit 14	Cytosol, extracellular region, nucleoplasm, nucleus
Q9NVZ3	$AVGG_{211}$|SLVQ	NECAP2	Adaptin ear-binding coat-associated protein 2	Cytoplasmic vesicle (clathrin-coated vesicle membrane), cell membrane
P04632	$LKGG_{11}$ |GGGG	CAPNS1	Calpain small subunit 1	Cytoplasm, cell membrane
P22626	$NQGG_{281}$|GYGG	HNRNPA2B1	Heterogeneous nuclear ribonucleoproteins A2/B1	Nucleus, nucleoplasm, cytoplasm, cytoplasmic granule, secreted (extracellular exosome)
P05787	$YAGG_{422}$|LSSA	KRT8	Keratin, type II cytoskeletal 8	Cytoplasm, nucleus (nucleoplasm), nucleus matrix
Q6PKG0	$LPGG_{12}$ |ATLL	LARP1	La-related protein 1	Cytoplasm, cytoplasmic granule
Q9UJU2	$LSGG_{7}$ |GGGG	LEF1	Lymphoid enhancer-binding factor 1	Nucleus
P22059	$LGGG_{23}$ |GAGP	OSBP	Oxysterol-binding protein 1	Cytoplasm (cytosol, perinuclear region), golgi apparatus membrane, endoplasmic reticulum membrane, golgi apparatus (trans-Golgi network)
O14908	$LGGG_{38}$ |GSGG	GIPC1	PDZ domain-containing protein GIPC1	Cytoplasm, membrane
Q7L014	$LRGG_{884}$|TILA	DDX46	Probable ATP-dependent RNA helicase DDX46	Nucleus speckle, nucleus (Cajal body), membrane
O15234	$LRGG_{33}$ |GSCS	CASC3	Protein CASC3	Cytoplasm (perinuclear region, stress granule, cytoplasmic ribonucleoprotein granule), nucleus, nucleus speckle, cell projection (dendrite)
O60610	$LPGG_{624}$|VCIS	DIAPH1	Protein diaphanous homolog 1	Cell membrane, cell projection (ruffle membrane), cytoplasm (cytoskeleton, microtubule organizing center, centrosome, spindle), nucleus

Acc #	P4-P1 | P1′-P4′	Gene	Protein Name	Subcellular Localization
A0A0B4J2F0	IAGG$_{21}$ |VYIF	PIGBOS1	Protein PIGBOS1	Mitochondrion outer membrane
P35637	GSGG$_{192}$|GYGN	FUS	RNA-binding protein FUS	Nucleus
P23246	LGGG$_{637}$|GGIG	SFPQ	Splicing factor, proline- and glutamine-rich	Nucleus speckle, nucleus matrix, cytoplasm
P62987	LRGG$_{76}$ |IIEP	UBA52	Ubiquitin-60S ribosomal protein L40	Cytoplasm, nucleus

Table 2.2. Putative substrates of SARS-CoV-2 PL^{pro}.

2.4.4 SARS-CoV-2 Mpro cleaves BRD2

The Bromodomain and Extra-Terminal (BET) domain family of proteins is known to 1) regulate gene expression by interacting with acetylated histones and 2) facilitate RNA polymerase II transcription (see ref. (170) for a review). In SARS-CoV-2, multiple studies reported that BET proteins can have both pro- and anti-viral effects. (35, 171, 172) In our N-terminomics experiment, we observed bromodomain-containing protein 2 (BRD2) cleavage by Mpro after Q206 (AALQ|GSVT) in Jurkat cell lysates. We investigated BRD2 cleavage by immunoblot in both Jurkat cell lysates and lysates of HEK293T-ACE2 cells overexpressing GFP-tagged BRD2. We observed the appearance of a cleavage product matching the molecular weight of the N-terminal fragment of BRD2 after cleavage at Q206 following a 2h incubation with Mpro (23 kDa, 50 kDa with GFP) (**Fig. 2.7b,c**). To confirm this cleavage site, we overexpressed the mutant GFP-BRD2 Q206A and did not observe the cleavage product (**Fig. 2.7c**). In addition, no cleavage product was observed for GFP-BRD2 when incubated with Mpro in the presence of the Mpro inhibitor GC376 (**Fig. 2.8**). (64) Collectively, these results confirmed that BRD2 is a Mpro substrate. To examine the effect of viral infection on the level of host BRD2, we infected three cell lines (A549-ACE2, HEK293T-ACE2 and H23-ACE2) with SARS-CoV-2, and saw a decrease in full-length BRD2 levels compared to uninfected control (**Fig. 2.7d,e**, **Fig. 2.9**), and the presence of a cleavage product in infected HEK293T-ACE2 and H23-ACE2 (**Fig. A.1d, Fig. 2.9b**).

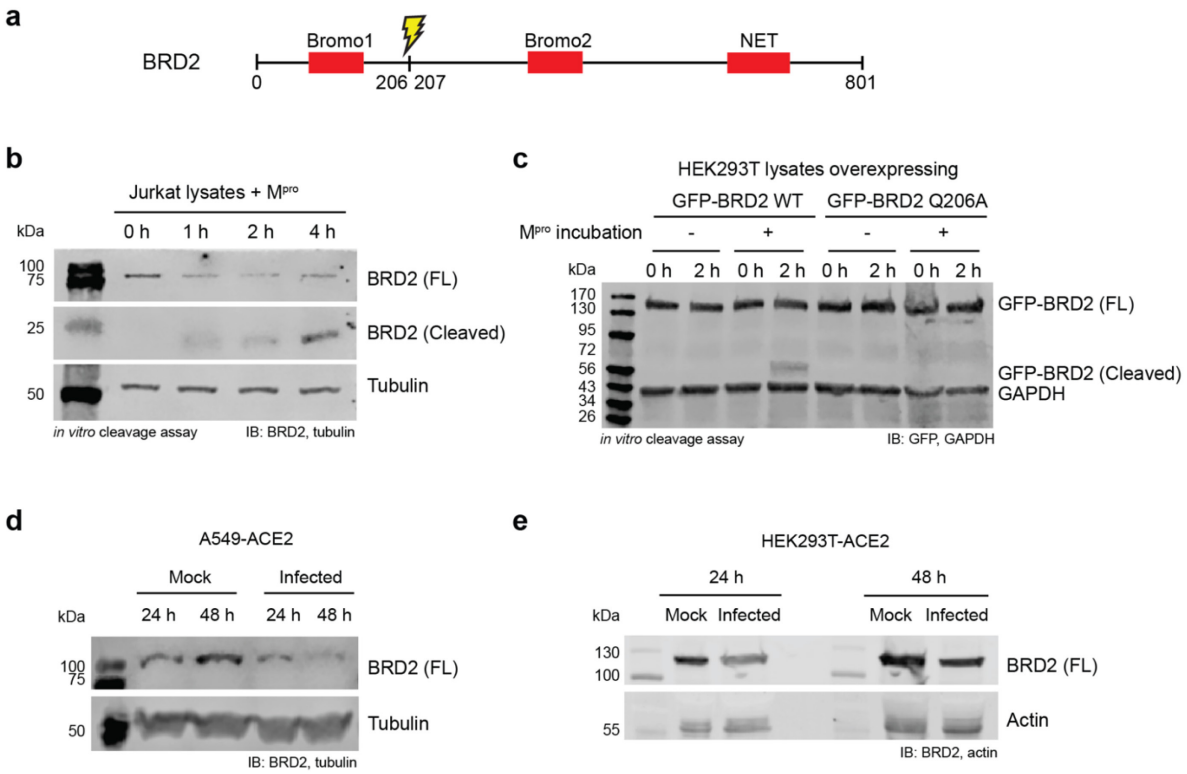


Figure 2.7. Proteolysis of BRD2 by Mpro *in vitro* and in SARS-CoV-2 infected cells. a) BRD2 contains Bromo 1, Bromo 2, and N-terminal extra terminal (NET) domains. Our N-terminomics study identified Mpro cleavage site in BRD2 after Q206, cleaving off the Bromo 1 domain. **b**) BRD2 was cleaved by recombinant Mpro in Jurkat cell lysates. Jurkat cell lysates were incubated with recombinant Mpro for 0-4 hours, and immunoblotted against BRD2. A cleavage product appeared with incubation time as the full length BRD2 level decreased. **c**) GFP-BRD2 WT and Q206A mutant overexpression in HEK293T-ACE2 and *in vitro* cleavage by recombinant SARS-CoV-2 Mpro. HEK293T-ACE2 cells overexpressing GFP-BRD2 were lysed, and the cell lysates were incubated with Mpro for 2 hours and immunoblotted against GFP. Cleavage was only observed with GFP-BRD2 WT. Depletion of full-length BRD2 was also observed in SARS-CoV-2 infected **d**) A549-ACE2 and **e**) HEK293T-ACE2 at 24 and 48 h.p.i.

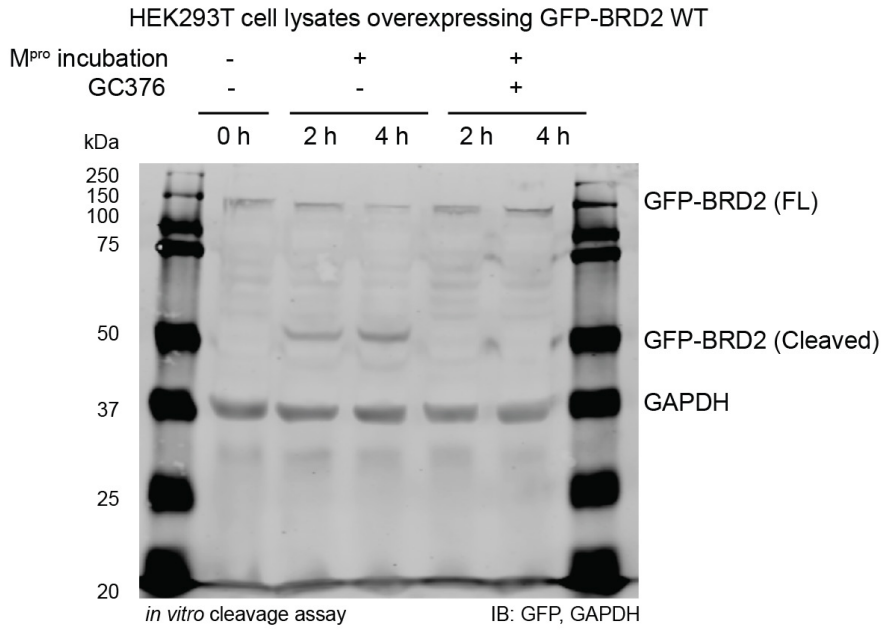


Figure 2.8. BRD2 is cleaved by SARS-CoV-2 Mpro (n=2, biological replicates)**.** HEK293T lysates overexpressing GFP-BRD2 were incubated with 0.5 μM Mpro for 0, 2, and 4 h in the absence or presence of 8 μM Mpro inhibitor GC376 (Selleck Chemicals, #S0475, dissolved in DMSO). GFP-BRD2 cleavage product was only observed without GC376 in the assay.

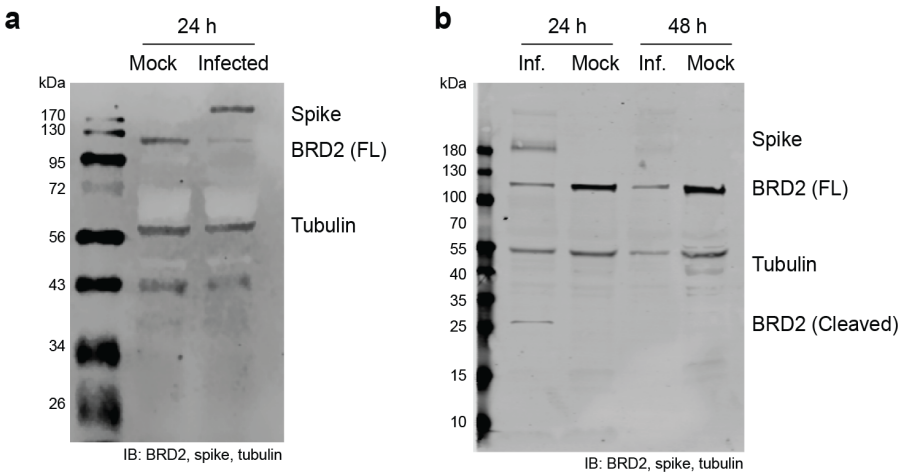


Figure 2.9. Immunoblots of SARS-CoV-2 infected H23-ACE2 lysates in two biological replicates. a,b) A decrease in full-length BRD2 level was observed at 24 h.p.i., as most Infected cells underwent cell death at 48 h, causing the decrease in spike detection. A band corresponding to the apparent cleavage product of BRD2 at ~23 kDa was present in **b**).

2.4.5 SFPQ is cleaved in SARS-CoV-2 infected cells

Among the 16 putative substrates identified for SARS-CoV-2 PLpro, we further investigated the splicing factor, proline- and glutamine-rich (SFPQ). SFPQ is a DNA- and RNA-binding protein found in paraspeckles, which are subnuclei compartments located in the interchromatin space of mammalian cell nuclei and regulate gene expression. SFPQ was shown to play a proviral role in Influenza A virus transcription, with its downregulation resulting in reduced viral replication. (173) SFPQ is also utilized as a pro-viral host dependency factor by several RNA viruses during infection, including the encephalomyocarditis virus (174) and hepatitis delta virus, (175) and even targeted for proteolysis in human rhinovirus A16 to promote viral replication. (176) Immunoblotting of SARS-CoV-2 infected A549-ACE2 lysates showed the full-length SFPQ at 76 kDa in both the mock and infected cells at 24 and 48 h.p.i., and the expected 69 kDa SFPQ cleavage product only in infected lysates (**Fig. 2.10**). This was consistent with our *in vitro* studies and suggests that PLpro cleaves SFPQ during infection. However, when uninfected Jurkat and A549-ACE2 cell lysates were incubated with SARS-CoV-2 PLpro, no corresponding cleavage product was observed on immunoblot (**Fig. 2.10c**, **Fig. A.4a**). Additionally, no clear *in vitro* cleavage product was observed in PLpro incubation with overexpressed FLAG-SFPQ in HEK293T-ACE2 cells (**Fig. A.4b**), even after immunoprecipitation (**Fig. A.4c**). Expression of full length NSP3 in HEK293T-ACE2 cells (24 h post-transfection) also did not induce detectable endogenous SFPQ proteolysis by immunoblot (**Fig. A.4d**). While SFPQ is clearly cleaved during viral infection, it is also likely involving host proteases.

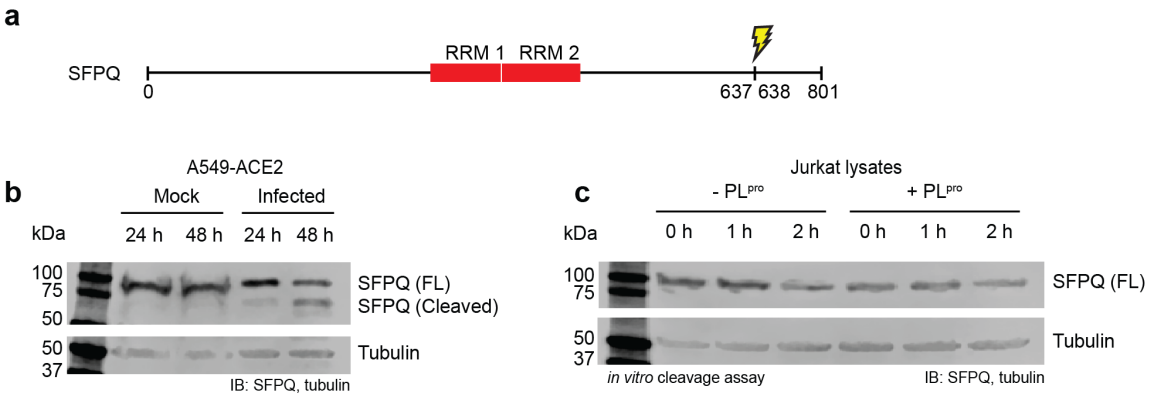


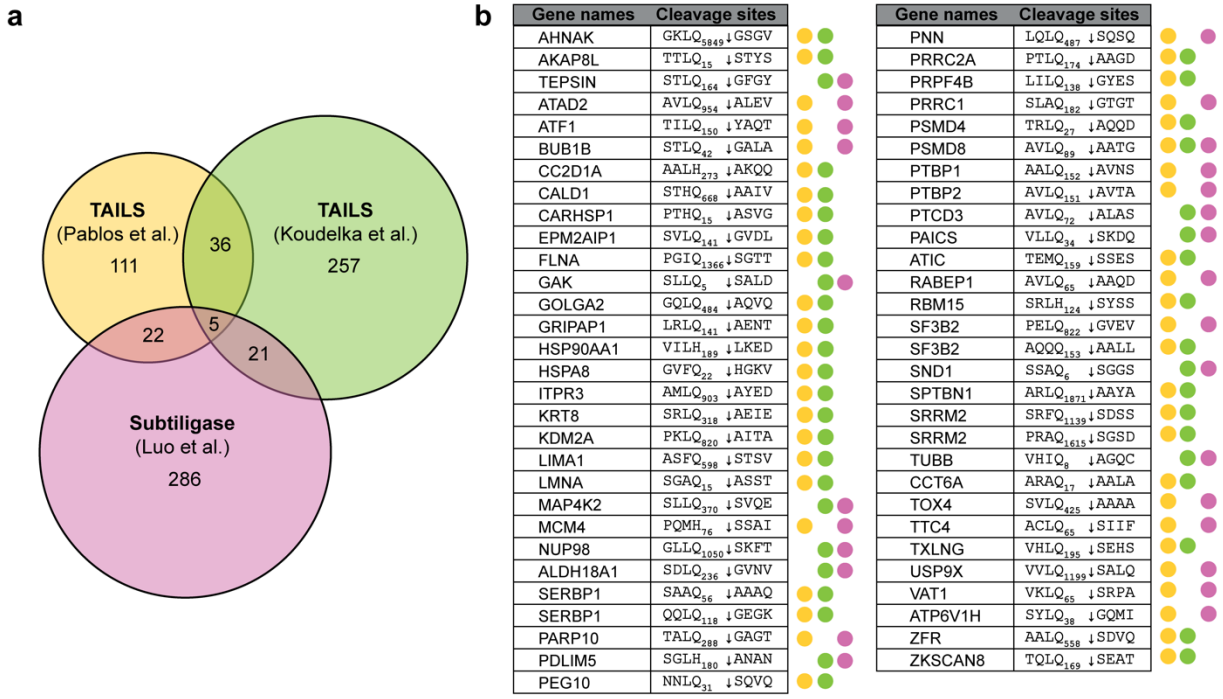
Figure 2.10. Proteolysis of SFPQ in SARS-CoV-2 infected cells. a) PLpro cleaves SFPQ after G637, C-terminal to the RNA recognition motifs. **b**) SFPQ was cleaved in A549-ACE2 cells infected with SARS-CoV-2. c) SFPQ cleavage by PLpro could not be detected using immunoblotting in Jurkat cell lysates. Performed by Eman W. Moussa.

2.4.6 Comparative analysis of all known Mpro substrates

We compared our data with SARS-CoV-2 Mpro substrates reported using TAILS or TMT labeling (**Fig. 2.11**). (62, 63, 162) TAILS and subtiligase-based labeling are complementary N-terminomics methods used to identify the N-termini of proteins; the protease-induced neo-N-termini are identified by negative and positive enrichment, respectively (see review (166)). The two methods each have their own advantages, such as the ability to identify cleavages in low-abundance proteins and different biases in P1′ sites. Thus, a compilation of all N-terminomics data on the SARS-CoV-2 Mpro can expand our understanding of how viral proteases function to regulate the host cell environment. We provide here a global analysis of all reported cleavage sites of SARS-CoV-2 Mpro identified: our subtiligase-mediated N-terminomics study, the TAILS experiments from Pablos *et al.* and Koudelka *et al.*, and the N-terminomics study in SARS-CoV-2 infection-induced proteolysis from Meyer *et al.* (**Fig. 2.11a**). For consistency across the datasets, cleavage sites with P1=Q/H in any cell lines, that have passed the authors' statistical evaluation in the case of negative enrichment techniques, and are not within the first 4 residues of the start of a protein sequence, are included in this comparative analysis. In total, there are 742 unique cleavage sites in 604 human proteins attributed to SARS-CoV-2 Mpro activity. Of

these, 59 new substrate cleavages were identified by two or more studies (**Fig. 2.11b**). Interestingly, one protein called NUP107 was identified in all four datasets, with the cleavage site between residues Q35 and A36 (VLLQ35|ASQD). NUP107 is a nucleoporin and a member of the nuclear pore complex that mediates the transport between the cytoplasm and nucleus. Many nuclear pore complex proteins are cleaved during RNA viral infections, such as picornaviruses including polioviruses and rhinoviruses, (177) to impair the transport of transcription factors involved in the immune response or mRNA maturation, or to block mRNA transport, promoting the translation of viral mRNAs or enabling access to nuclear factors required for viral replication (reviewed by (178)). However, the NUP107 cleavage site was also found in the PLpro dataset, suggesting that perhaps this substrate is targeted by both Mpro and an unidentified human host protease. Similarly, stathmin and XRCC1 were also identified as putative Mpro substrates by three groups but were found in our PLpro dataset. Stathmin regulates the cell cycle by re-organization of the microtubule cytoskeleton. Down-regulation of phosphorylation on stathmin and other cytoskeleton assembly proteins was also observed in SARS-CoV-2 infected cells. (179) The DNA repair protein XRCC1 is required for the repair of DNA single-strand breaks, and has been found to interact with proteins encoded by DNA viruses such as human papillomavirus, (180) as well as hepatitis B and C viruses. (181) There are also two new protein cleavages that were commonly identified in three groups but not observed in the DegraBase or in the PLpro N-terminomics experiments: the 26S proteasome non-ATPase regulatory subunit 8 (PSMD8) cleaved after Q89, and a bifunctional enzyme involved in *de novo* purine biosynthesis called PAICS that is cleaved after Q34. PSMD8 is a part of the 26S proteasome complex and is a host restriction factor in HIV-1. (182) PAICS is identified as an oncogene in several tumor types, but it has also been shown to bind to influenza A virus nucleoprotein. (183) Although the link between virology and the two proteins have not been well studied, their cleavages could imply a significant role in the SARS-CoV-2 pathobiology. For instance, cellular proteins involved in purine biosynthesis have been explored

as potential antiviral targets, (184, 185) and many viruses developed mechanisms to utilize the host ubiquitin-proteasome systems, which will be discussed more in-depth in Chapter 3.



Gene names	Cleavage sites
AHNAK	$GKLQ_{5849}$↓GSGV
AKAP8L	$TTLQ_{15}$ ↓STYS
TEPSIN	$STLQ_{164}$ ↓GFGY
ATAD2	$AVLQ_{954}$ ↓ALEV
ATF1	$TILQ_{150}$ ↓YAQT
BUB1B	$STLQ_{42}$ ↓GALA
CC2D1A	$AALH_{273}$ ↓AKQQ
CALD1	$STHQ_{668}$ ↓AAIV
CARHSP1	$PTHQ_{15}$ ↓ASVG
EPM2AIP1	$SVLQ_{141}$ ↓GVDL
FLNA	$PGIQ_{1366}$↓SGTT
GAK	$SLLQ_{5}$ ↓SALD
GOLGA2	$GQLQ_{484}$ ↓AQVQ
GRIPAP1	$LRLQ_{141}$ ↓AENT
HSP90AA1	$VILH_{189}$ ↓LKED
HSPA8	$GVFQ_{22}$ ↓HGKV
ITPR3	$AMLQ_{903}$ ↓AYED
KRT8	$SRLQ_{318}$ ↓AEIE
KDM2A	$PKLQ_{820}$ ↓AITA
LIMA1	$ASFQ_{598}$ ↓STSV
LMNA	$SGAQ_{15}$ ↓ASST
MAP4K2	$SLLQ_{370}$ ↓SVQE
MCM4	$PQMH_{76}$ ↓SSAI
NUP98	$GLLQ_{1050}$↓SKFT
ALDH18A1	$SDLQ_{236}$ ↓GVNV
SERBP1	$SAAQ_{56}$ ↓AAAQ
SERBP1	$QQLQ_{118}$ ↓GEGK
PARP10	$TALQ_{288}$ ↓GAGT
PDLIM5	$SGLH_{180}$ ↓ANAN
PEG10	$NNLQ_{31}$ ↓SQVQ

Gene names	Cleavage sites
PNN	$LQLQ_{487}$ ↓SQSQ
PRRC2A	$PTLQ_{174}$ ↓AAGD
PRPF4B	$LILQ_{138}$ ↓GYES
PRRC1	$SLAQ_{182}$ ↓GTGT
PSMD4	$TRLQ_{27}$ ↓AQQD
PSMD8	$AVLQ_{89}$ ↓AATG
PTBP1	$AALQ_{152}$ ↓AVNS
PTBP2	$AVLQ_{151}$ ↓AVTA
PTCD3	$AVLQ_{72}$ ↓ALAS
PAICS	$VLLQ_{34}$ ↓SKDQ
ATIC	$TEMQ_{159}$ ↓SSES
RABEP1	$AVLQ_{65}$ ↓AAQD
RBM15	$SRLH_{124}$ ↓SYSS
SF3B2	$PELQ_{822}$ ↓GVEV
SF3B2	$AQQQ_{153}$ ↓AALL
SND1	$SSAQ_{6}$ ↓SGGS
SPTBN1	$ARLQ_{1871}$↓AAYA
SRRM2	$SRFQ_{1139}$↓SDSS
SRRM2	$PRAQ_{1615}$↓SGSD
TUBB	$VHIQ_{8}$ ↓AGQC
CCT6A	$ARAQ_{17}$ ↓AALA
TOX4	$SVLQ_{425}$ ↓AAAA
TTC4	$ACLQ_{65}$ ↓SIIF
TXLNG	$VHLQ_{195}$ ↓SEHS
USP9X	$VVLQ_{1199}$↓SALQ
VAT1	$VKLQ_{65}$ ↓SRPA
ATP6V1H	$SYLQ_{38}$ ↓GQMI
ZFR	$AALQ_{358}$ ↓SDVQ
ZKSCAN8	$TQLQ_{169}$ ↓SEAT

Figure 2.11. Comparative analysis of SARS-CoV-2 Mpro putative substrates. a) Overlap in identified cleavage sites among SARS-CoV-2 Mpro N-terminomics *in vitro* studies. (62, 63) **b**) 59 common cleavage sites were identified by different Mpro studies, and have not been observed in the PLpro N-terminomics dataset nor in the DegraBase. (169)

2.5 Discussion

2.5.1 The possible roles of BRD2 cleavage in SARS-CoV-2

BET proteins interact with many viral proteins and modulate viral infections. In particular, bromodomain-containing protein 4 (BRD4) forms a complex with E2 for transcriptional silencing in human papillomaviruses, (186) and BRD2 interacts with latency-associated nuclear antigen 1 (LANA1) in Kaposi's sarcoma-associated herpesvirus. (187) BRD2 binds the SARS-Cov-2 envelope E protein (35) and is required for ACE2 transcription which likely benefits SARS-CoV-2 replication in human lung epithelial cells. It also acts as a host antiviral factor by promoting the transcription of genes involved in type I interferon response. (171) In another recent study,

BRD2, 3 and 4 inactivation was shown to increase viral infection in cells and mice overexpressing ACE2. (172) Our N-terminomics and immunoblot studies showed BRD2 cleavage by Mpro after Q206 (AALQ|GSVT). This viral protease cleavage removes the bromodomain I (BDI) (**Fig. 2.7a**), potentially disrupting BRD2 binding to the acetylated histones and thereby affecting host gene transcription. (188) Our N-terminomics study also detected proteolysis of BRD4, which is another member of the BET family that binds to SARS-CoV-2 E protein. (35, 172) The cleavage of BRD4 after Q1077 (SQFQ|SLTH) in the C-terminal region could interfere with the formation of the P-TEFb transcriptional complex, preventing the activation of interferon-stimulated genes. (189) Thus, the BET proteins have a sophisticated role during SARS-CoV-2 viral pathogenesis, that may interact with multiple viral proteins and fine-tune the gene expression of key proteins involved in biological pathways.

2.5.2 Noncanonical specificity of Mpro

Previous biochemical analyses and N-terminomics studies on SARS-CoV-2 Mpro placed a rather stringent selection filter for Mpro substrates, where only cleavage sites with a P1=Q are considered as potential Mpro targets. Indeed, based on the crystal structure of both SARS-CoV-1 and -2, Gln can occupy the S1 pocket by stable interactions with His163, Phe140 and Glu166. (150, 151, 190) However, the ability for SARS-CoV-1 Mpro to recognize other residues at the P1 site, specifically His and Met, and incorporate them into its active site has been reported by peptide library screening. (152) In Koudelka *et al.*'s *in vitro* SARS-CoV-2 Mpro N-terminomics experiment, a strong enrichment of His in P1 position was also observed in the identified cleavage sites. (62) Pablos *et al.* explored the noncanonical cleavage sequences of Mpro in detail using peptide libraries derived from N-terminomics substrates combined with molecular docking simulations, and showed that Mpro can cleave after P1=G/H/M. (63) Similarly, while Leu is the preferred residue at the P2 position, Mpro can also recognize other hydrophobic residues at P2 such as Val, Phe, Met, Ala and Ile, and has even broader specificities at

positions P3 and P4. Hence, we selected potential Mpro substrates featuring a Gln or His at the P1 position in our *in vitro* N-terminomics (P1=Q/H only) to allow some selectivity at the P1 site, but relaxed restrictions on other sites. We did not include cleavage sites P1=M for further investigation (3.8% of total labeled cleavages), as these were not enriched over typical background proteolysis. (168) We also looked at the secondary structure of the Mpro cleavage site locations, where we found that 10 cuts occurred in α-helices, 6 in β-strands, 1 in turn and the rest in uncharacterized or disordered regions.

2.5.3 Potential activation of other cellular proteases in cell lysates

The putative SARS-CoV-2 Mpro and PLpro cellular targets were subjected to pathway analyses using Metascape (191) to reveal how viral proteases potentially disrupt cellular processes during infection. Mpro cleavage of host substrates is predicted to affect the cell cycle and cellular gene expression (**Fig. 2.12a**), and the enriched processes of PLpro substrates highlight metabolism of RNA (**Fig. 2.12b**). In our *in vitro* N-terminomics experiments, there are also many labeled cleavages in the human proteome that do not fall under the specificity profiles of the viral proteases Mpro and PLpro. While we cannot rule out exogenous co-purified protease activity from *E. coli*, we believe that by adding a cocktail of protease inhibitors (targeting metallo- and serine proteases) and focusing on substrates matching Mpro and PLpro specificities, we have minimized the identification of non-related protease substrates. In addition, we can exclude substrates found in the PLpro dataset matching Mpro specificity and vice versa, using each dataset to identify unique cleavage sites to the viral proteases. However, the non-selective inhibition of other proteases is not 100% efficient. The observed cleavage sites that do not fall under the specificity profiles of Mpro and PLpro may be due to cellular protease activation and may still be of interest. The activity of the host proteases can be attributed to a few possibilities, such as the direct activation by the viral proteases to initiate proteolysis of other proteins, or indirectly resulted from the viral protease incubation in the

cellular proteome. Therefore, we searched all labeled cleavage sites from our N-terminomics datasets on TopFind 4.1, (192) to investigate which endogenous proteases account for those cleavages (**Fig. 2.13**). A majority of the cleavage sites correspond to granzyme M specificity (P1=L/M). (193) As the viral-infection induced activation of granzyme M is characteristic in cytotoxic T lymphocytes, it is interesting to find its activation by viral proteases in an *in vitro* environment.

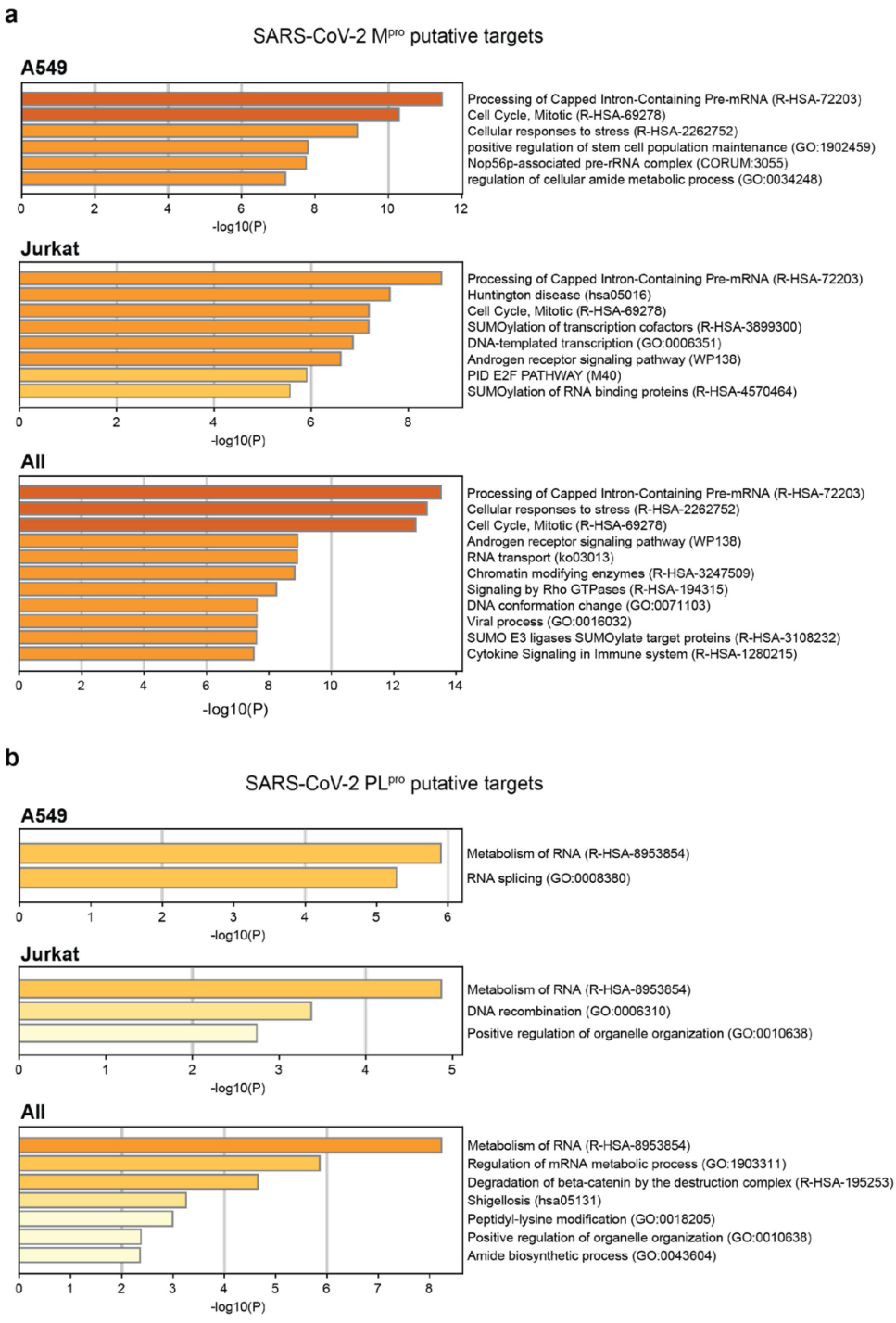


Figure 2.12. Gene Ontology analysis of **a**) M^{pro} and **b**) PL^{pro} putative substrates in A549, Jurkat, and both A549 and Jurkat cell lysates using Metascape. (191)

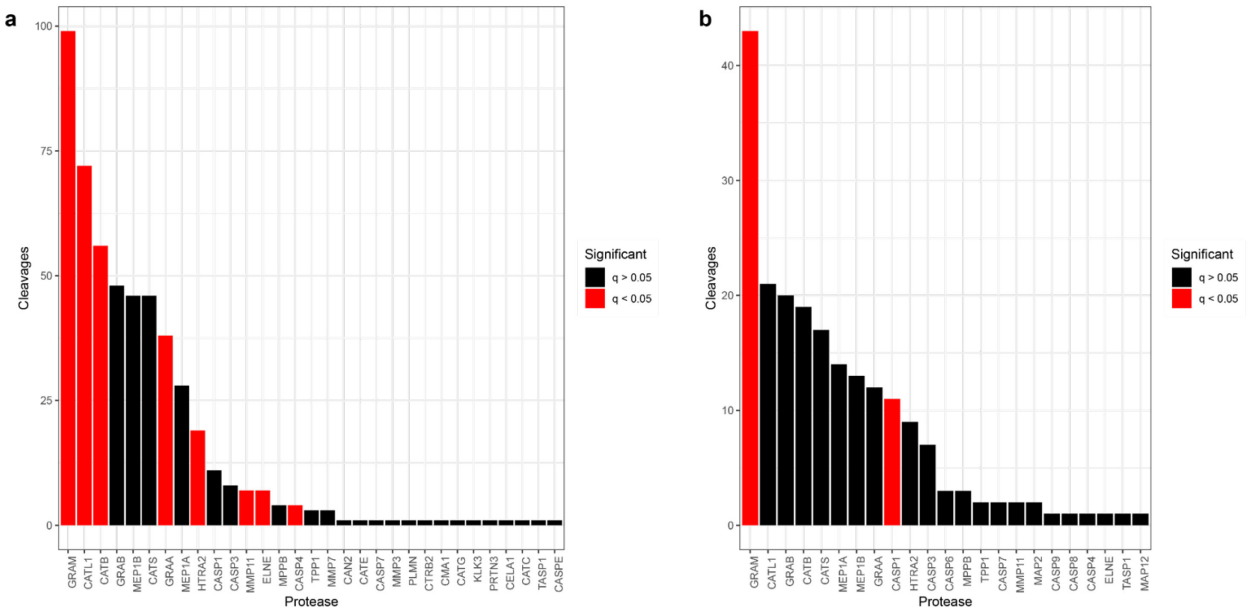


Figure 2.13. **TopFind analysis of all labeled cleavage sites** in **a**) M^{pro} and **b**) PL^{pro} subtiligase N-terminomics experiments.

2.5.4 Up- and down-regulation of viral protease substrates during viral infection

We initially hypothesized that viral proteases would cleave host restriction factors to improve replication efficiency, and that the cleaved host protein fragments, due to their low stability, could be subsequently targeted for degradation by the host cell machinery. However, when we used the list of all putative substrates identified by N-terminomics and compared it to the reported proteome changes during SARS-CoV-2 infection (36, 194) or CRISPR screens, (195, 196) we found that the protease substrates we identified did not correlate with lower protein levels in infected cells. It is possible that these were not targeted for degradation by the cell. Alternatively, some of these proteolytic fragments could potentially lead to a gain-of-function, such as is the case of SFPQ, where a proviral factor can be cleaved by a viral protease. (176)

2.5.5 Limitations of the study

We acknowledge that a large number of substrates identified in our *in vitro* N-terminomics may not actually be cleaved during infection and could be bystanders. We performed studies on the protein substrates known to play a role in host antiviral defense, such as transcription intermediary factor 1-beta (TRIM28) and the zinc finger antiviral protein (ZC3HAV1, also known as ZAP). However, we did not see depletion of TRIM28 and ZAP in the *in vitro* cleavage assays or in infected cells using immunoblot (**Fig. A.5**). Similarly, proteolysis in protein mono-ADP-ribosyltransferase PARP10 and nuclear pore complex protein Nup98-Nup96 could not be detected via immunoblot. This could be in part due to inability of the proteases to access substrates during infection and/or the fact that high concentrations of viral proteases were used in the *in vitro* studies (0.5 μM and 5 μM, for Mpro and PLpro respectively). It can also be challenging to precisely detect substrate proteolysis or degradation via immunoblot of infected cell lysates, where depending on the cell line, only a fraction of the cell population is infected, and a subfraction of those infected cells has only low levels of proteolysis in the corresponding host proteins. Many commercial antibodies also failed to detect protease-cleaved substrates in immunoblots. There are many reasons that could explain this discrepancy: substrate degradation, proteolysis by host proteases or *E. coli* protease contaminants, sub-optimal time points, subcellular localization, and IFN-induced protein expression. It is also possible that the epitope could also be damaged by proteases in the lysate, or that post-translational modifications (ubiquitination, phosphorylation, etc.) of the substrates could prevent antibody recognition. Furthermore, the overexpression of ACE2 receptor improves cellular susceptibility to viral infection in human cell lines, such as HEK293T and A549, but since many host proteins are involved in ACE2-mediated pathways, the constitutively overexpressed ACE2 might affect the degradation of these substrates *in vivo*, such as TRIM28 (197) and BRD2. (171) When we compared the results from this study to other subtiligase-based N-terminomics studies

on human proteases such as caspase-3 and -9, most identified substrates showed robust cleavage by immunoblot in *in vitro* cleavage assays and in apoptotic cells. (198) The drastic difference in detection between studies demonstrates that proteolysis in host proteins by viral proteases may occur only at very low levels. As a host cell is infected, even though many cellular pathways are disrupted, the virus prevents cell death in order to sustain viral replication. Hence, the low level of cellular protein proteolysis by viral proteases can be interpreted as a mechanism for the virus to maximize replication efficiency, while maintaining cell viability.

Chapter 3: The role of ubiquitination in VACV infection

Preface

Chapter 3 was adapted from a manuscript in preparation:

> Dong, J.; Luo, S. Y.; Smyth, S.; Melvie, G.; Julien, O.; Ingham, R. J. Characterizing changes in protein ubiquitylation during vaccinia virus infection.

I was responsible for method optimizations and sample preparation in mass spectrometry analysis, as well as statistical analysis. J.D. performed the viral infections, and J.D. and G.M. performed immunoblotting analysis in this Chapter. O.J. and R.J.I. contributed to conceptualization and experimental design. Other authors assisted in sample collection and viral infection.

3.1 Abstract

Ubiquitination is a post-translational modification that is required for poxviruses to replicate their genomes and evade the host immune response. In this study, we used diGlycyl ubiquitin remnant enrichment to profile global ubiquitinated sites and quantified changes in protein ubiquitination states early after Vaccinia virus (VACV) infection. Our study demonstrated that many viral and host protein ubiquitination states were significantly altered during VACV infection. This included examples of degradative and non-degradative ubiquitination of host proteins, and revealed novel ways the host ubiquitin-proteasome system (UPS) is engaged during infection. In particular, we found that the cellular E3 Ub/ISG15-ligase TRIM25 was heavily ubiquitinated and subsequently degraded early after infection of cells with VACV. This may represent a novel host restriction factor that VACV utilizes the UPS to overcome.

3.2 Introduction

Poxviruses are large double-stranded DNA viruses including important human pathogens such as variola virus (causative agent of smallpox) that killed hundreds of millions of people before its eradication, and monkeypox virus that is responsible for the recent global outbreak. Moreover, some poxviruses such as cowpox virus are important livestock pathogens. Although closely related to other poxviruses, Vaccinia virus (VACV) infection typically causes mild symptoms, and was in fact used for smallpox vaccination in the global campaign. Therefore, VACV has been the most extensively studied poxvirus model for development in vaccines and cancer therapies.

Ubiquitination is the reversible, covalent post-translational modification of proteins with the 76 amino acid protein ubiquitin. The process starts with the ATP-dependent activation of ubiquitin by an E1 ubiquitin-activating enzyme, followed by its transfer to an E2 ubiquitin-conjugating enzyme. The E2 enzyme then facilitates the ubiquitin transfer to substrates, directly

or indirectly, through binding an E3 ubiquitin-ligase protein which brings both E2 and substrate together. Proteins can be modified by ubiquitin by monoubiquitination, where a single ubiquitin molecule is added to a lysine residue in a target protein. Alternatively, chains of ubiquitin, formed by covalent ubiquitin conjugations through lysine residues, can also form on substrates. Other than acting as a canonical signal for proteasome-mediated protein degradation, both mono- and polyubiquitination are involved in a variety of biological processes, such as cell cycle, autophagy, protein trafficking, and innate immunity. Thus, ubiquitination is an extremely versatile modification that serves numerous functions in the cell. Not surprisingly, poxviruses have acquired several mechanisms to co-opt the host UPS for their benefit.

Poxviruses require the 26S proteasome (105, 106, 108) and cellular ubiquitination components (e.g. E1 ubiquitin-activating enzyme (105, 107) and Cullin3 (107)) to productively infect cells. In addition, these viruses encode within their genomes several E3 ubiquitin-ligases and substrate adapters for multi-subunit cellular E3 ubiquitin-ligases. Particularly, VACV encodes a number of cullin-3 E3 ligase adaptors BTB-kelch proteins that have been implicated in impairing the innate and adaptive immune responses, (95–97, 102, 103) blocking necroptotic cell death (199), and facilitating virus spread. (104)

While poxviruses utilize the UPS to successfully establish infection, this system is also important for the host immune response to infection. For example, innate antiviral signaling pathways, such as those involved in the activation of RIG-I, (200, 201) and NF-κB, (202, 203) utilize degradative and non-degradative ubiquitination. In addition, the UPS is an important part of the adaptive immune response by processing viral peptides for presentation in major histocompatibility complex (MHC) class I. (204).

As the UPS can target proteins for degradation or activate protein functions, it is both leveraged by the virus in viral pathogenesis and employed by the host in antiviral defense.

However, how the VACV infection engages in the UPS has never been systematically studied. To explore the intricate ubiquitome in VACV infection, we infected HeLa cells with VACV and enriched for diGlycyl peptides in the infected and control cell lysates using a diGlycyl-lysine specific antibody. We performed statistical analysis to identify proteins containing diGlycyl peptides with significantly altered abundances in the VACV infected sample, including 42 host proteins and 77 viral proteins. One of the interesting substrates was the tripartite motif containing 25 (TRIM25), an E3 ligase known to play a role in innate immune defense against viruses. We found that TRIM25 higher-molecular-weight (HMW), likely ubiquitinated, species were evident as early as 1 h post-infection. Our study established the ubiquitome of both viral and host proteins during VACV infection, and identified the ubiquitination of potential host restriction factors.

3.3 Materials and methods

3.3.1 Cell culture and virus

HeLa cells were obtained from Dr. Jim Smiley (University of Alberta) and maintained in Dulbecco's Modified Eagle Medium (Sigma, #D5796-500ML) supplemented with 10% fetal bovine serum (Sigma-Aldrich, F1051-500ML) and 1% antibiotic-antimycotic solution (Gibco, #15240062). All cells were incubated at 37˚C and 5% CO_2. VACV Copenhagen (VACV-Cop) was obtained from Dr. Michele Barry (University of Alberta).

3.3.2 DiGlycyl peptide enrichment

HeLa cells (2×10^7) were pretreated with 10 μM MG132 (Enzo, #BMLPI1020005) for 1 h. Cells were then inoculated with VACV-Cop viruses at a multiplicity of infection (MOI) of 3 for 1 h and incubated for a further 4 h with 10 μM MG132 in an incubator at 37°C and 5% CO_2. Cells were lysed in 8 M urea, 40 mM Tris-HCl pH 8.5 and gentle probe sonication (amplitude = 20%, alternate between on and off every 2 s, for a total of 1 min), and the lysates were clarified by

spin centrifugation at 13,000 rpm in benchtop centrifuge for 10 min at 4˚C. Proteins were reduced with 10 mM dithiothreitol (DTT) for 45 min at 37˚C, alkylated with 30 mM iodoacetamide for 45 min at room temperature, followed by addition of 20 mM DTT. Proteins were then digested with trypsin in 1:100 enzyme:substrate ratio overnight at 37˚C, and trypsin was precipitated in 2% trifluoroacetic acid (TFA). Peptides were desalted using SOLA HRP SPE cartridge (Thermo Fisher, #60109-001). Desalted and dried peptides were subjected to automated diGlycyl enrichment using the PTMScan ubiquitin remnant motif kit (Cell Signaling, #59322) on a KingFisher Duo Prime. Briefly, 1 mg peptides were resuspended in 1 mL HS (high degree of specificity and sensitivity) bind buffer #1 provided in the kit, and peptide binding with 20 µL of the Ab-bead slurry was done at room temperature for 2 h. The beads were washed 4x with 1 mL HS IAP (immunoaffinity purification) wash buffer provided in the kit and 2x with 1 mL H2O. Peptides were eluted 2 times with 50 µL 0.15% TFA at room temperature.

3.3.3 Mass spectrometry analysis

Peptides were desalted with C18 ziptips (#ZTC18S960, Millipore) and recovered in buffer A (0.1% formic acid) prior to mass spectrometry analysis. The samples were analyzed using a nanoflow-HPLC (Thermo Scientific EASY-nLC 1200 System) coupled to an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific). Reverse phase separation of the peptides was done with an Aurora Ultimate analytical column (25 cm x 75 µm ID with 1.7 µm media, IonOpticks). Peptides were eluted with a solvent B gradient (80% ACN, 0.1% FA) for 120 min. The gradient was run at 400 nL/min with analytical column temperature set at 45°C. Data were analyzed using Proteome Discoverer (v2.4.1.15) against the concatenated database of the human proteome (UP000005640) combined with VACV-Cop proteome (UP000008269), with relaxed false discovery rate set at 5% and restricted at 1%. Search parameters included a maximum of two missed trypsin cleavages, a precursor mass tolerance of 15 ppm, a fragment mass tolerance of 0.8 Da, with the constant modification carbamidomethylation (C), and variable

modifications of acetyl (protein N-term), deamidated (N/Q), oxidation (M), and GlyGly (uncleaved K). The maximum number of variable modifications was set to 4. Statistical analysis was performed by Proteome Discoverer using background-based t-test and protein abundance ratios were calculated using pairwise peptide ratios. Mass spectrometry data files are available through MassIVE Respository #MSV000094020.

3.3.4 Metascape Analysis

diGlycyl peptides enriched in either uninfected or infected cells were analyzed by Metascape (v3.5.20230501) using express analysis and the human species setting. Categories with a significant adjusted p value ($q<0.05$) are shown.

3.3.5 Immunoblot analysis

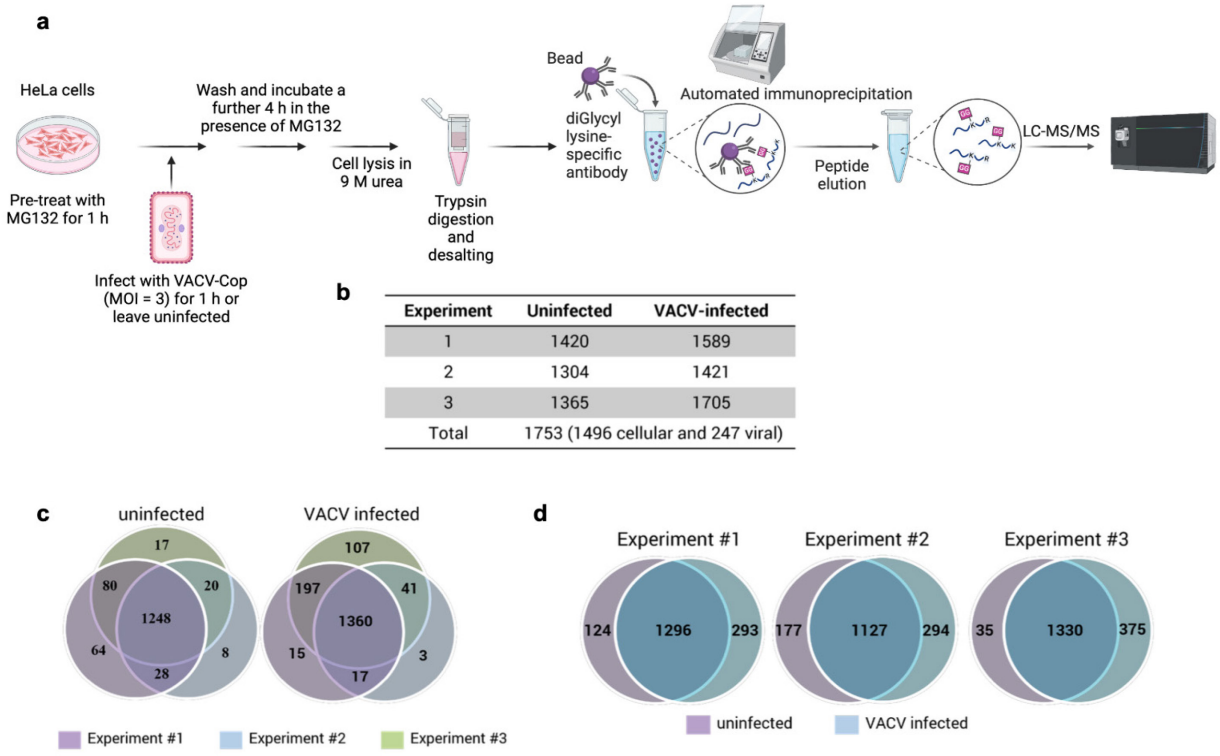
Protein lysates or immunoprecipitations in SDS-PAGE sample buffer were separated by SDS-PAGE (Hoeffer) before being transferred to nitrocellulose membranes (1620115, Bio-Rad) using a semi-dry transfer apparatus (Bio-Rad). Membranes were then blocked in 5% milk in PBS overnight at 4℃ before being incubated with the indicated primary antibodies. Membranes were then incubated with secondary antibodies, goat-anti-rabbit 680 or goat-anti-mouse 800, for 45 min at room temperature in the dark and imaged using an Odyssey scanner (LI-COR Biosciences).

3.4 Results

3.4.1 Identification of diGlycyl peptides from uninfected and VACV-Cop-infected HeLa cells

HeLa cells pre-treated with the proteasome inhibitor, MG132 for 1 h were either left uninfected or infected with VACV-Cop at a MOI of 3 for 1 h followed by an additional 4 h incubation in the presence of MG132 (**Fig. 3.1a**). Lysates were then subjected to trypsin

digestion and diGlycyl peptides were enriched before being analyzed by mass spectrometry. diGlycyl enrichment utilizes an antibody (Ab) that recognizes the diGlycyl remnant present on ubiquitinated lysine residues after trypsin digestion. (131) Three independent experiments identified 1753 (1496 cellular and 257 viral) distinct diGlycyl peptides (**Fig. 3.1b**). The majority of peptides (≥ 78%) were identified in both uninfected and infected cells (**Fig. 3.1c**), albeit their abundance was not necessarily the same. Moreover, there was considerable overlap in diGlycyl peptides identified between the independent experiments (**Fig. 3.1d**).



Experiment	Uninfected	VACV-infected
1	1420	1589
2	1304	1421
3	1365	1705
Total	1753 (1496 cellular and 247 viral)	

Figure 3.1. Identification of diGlycyl peptides in VACV-Cop-infected HeLa cells. a) Outline of the experimental workflow used for the purification and identification of diGlycyl peptides. **b**) Summary of diGlycyl peptides identified in the three independent experiments. **c**) Venn diagrams illustrating the number of the peptides identified in uninfected cells, VACV-Cop-infected cells, or both in each independent experiment. **d**) Venn diagrams illustrating the number of the peptides identified in uninfected or infected cells in common between the three independent experiments.

We focused our analysis on the 1720 diGlycyl peptides found in at least 2 independent experiments. Quantifying average peptide abundance over the three independent experiments revealed 304 peptides enriched for (≥ 2-fold change and $p \leq 0.05$) in VACV-Cop-infected cells. The majority of these diGlycyl peptides (247; 81%) were from viral proteins, and 57 peptides were associated with 42 cellular proteins (Fig. 3.2) (Table 3.1). Furthermore, we identified 37 cellular diGlycyl peptides enriched for (≥ 2-fold change and $p \leq 0.05$) in uninfected cells (**Fig. 3.2**).

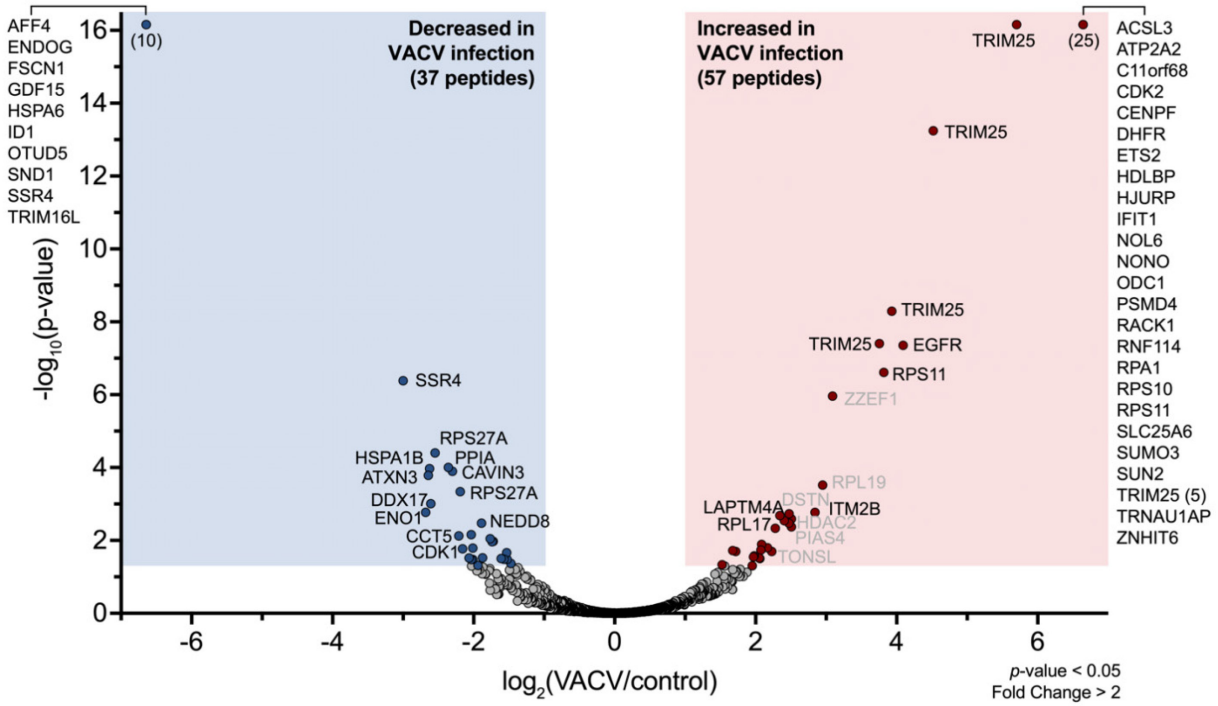


Figure 3.2. Volcano plot of diGlycyl peptides derived from host proteins found in at least two independent experiments. Peptides in blue were enriched ≥ 2-fold (p-value ≤ 0.05) in uninfected cells, whereas peptides in pink were enriched ≥ 2-fold (p-value ≤ 0.05) in VACV-Cop-infected cells. Peptides in grey were considered of lower confidence given their prominent enrichment in only one of the replicates.

Gene Name	Acc #	Protein Descriptions	diGlycyl lysines
ACSL3	O95573	Long-chain-fatty-acid--CoA ligase 3	K509
ASPM	Q8IZT6	Abnormal spindle-like microcephaly-associated protein	K1875, K2618
ATP2A2	P16615	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	K481
ATP5F1A	P25705	ATP synthase subunit alpha, mitochondrial	K175
C11orf68	Q9H3H3	UPF0696 protein C11orf68	K200
CDK2	P24941	Cyclin-dependent kinase 2	K20
CENPF	P49454	Centromere protein F	K223
DHFR	P00374	Dihydrofolate reductase	K64
DSTN	P60981	Destrin	K45
EEF1G	P26641	Elongation factor 1-gamma	K294
EGFR	P00533	Epidermal growth factor receptor	K716

Gene Name	Acc #	Protein Descriptions	diGlycyl lysines
ELP2	Q6IA86	Elongator complex protein 2	K459
ETS2	P15036	Protein C-ets-2	K144
HDAC2	Q92769	Histone deacetylase 2	K75
HDLBP	Q00341	Vigilin	K453
HJURP	Q8NCD3	Holliday junction recognition protein	K495
HNRNPM	P52272	Heterogeneous nuclear ribonucleoprotein M	K716
IFIT1	P09914	Interferon-induced protein with tetratricopeptide repeats 1	K259
ITM2B	Q9Y287	Integral membrane protein 2B	K16
LAPTM4A	Q15012	Lysosomal-associated transmembrane protein 4A	K7
MYL3	P08590	Myosin light chain 3	K142
NOL6	Q9H6R4	Nucleolar protein 6	K149
NONO	Q15233	Non-POU domain-containing octamer-binding protein	K371
ODC1	P11926	Ornithine decarboxylase	K115
PIAS4	Q8N2W9	E3 SUMO-protein ligase PIAS4	K69
PRPF8	Q6P2Q9	Pre-mRNA-processing-splicing factor 8	K1392
PSMD4	P55036	26S proteasome non-ATPase regulatory subunit 4	K365
RACK1	P63244	Receptor of activated protein C kinase 1	K175
RNF114	Q9Y508	E3 ubiquitin-protein ligase RNF114	K112
RPA1	P27694	Replication protein A 70 kDa DNA-binding subunit	K167
RPL13A	P40429	60S ribosomal protein L13a	K188
RPL19	P84098	60S ribosomal protein L19	K126
RPL23A	P62750	60S ribosomal protein L23a	K115
RPS10	P46783	40S ribosomal protein S10	K139
RPS11	P62280	40S ribosomal protein S11	K20, K45
RPS17	P08708	40S ribosomal protein S17	K49
SLC25A6	P12236	ADP/ATP translocase 3	K33, K245

Gene Name	Acc #	Protein Descriptions	diGlycyl lysines
SUMO3	P55854	Small ubiquitin-related modifier 3	K41
SUN2	Q9UH99	SUN domain-containing protein 2	K502, K1263
TRIM25	Q14258	E3 ubiquitin/ISG15 ligase TRIM25	K112, K283, K320, K320/K332, K335, K425, K439
TRNAU1AP	Q9NX07	tRNA selenocysteine 1-associated protein 1	K166, K310
ZZEF1	O43149	Zinc finger ZZ-type and EF-hand domain-containing protein 1	K2501

Table 3.1. Cellular proteins containing significantly enriched diGlycyl peptides in infected cells, shown in **Fig. 3.2** (pink shaded region.) The "/" under the Column "diGlycyl lysines" indicates the identification of a peptide with a single diGlycyl modification, but with insufficient MS2 fragmentation data to determine the exact position of the modified lysine on the peptide.

3.4.2 Differentially enriched human diGlycyl peptides associated with antiviral signaling

Metascape analysis was performed to determine whether any cellular pathways or processes were overrepresented from proteins associated with diGlycyl peptides preferentially found in uninfected or infected cells. Translation and the cell cycle were categories associated with proteins with peptides enriched in infected cells (**Fig. 3.3a**), whereas processes linked to proteins with peptides enriched in uninfected cells included deubiquitination, viral infection pathways, DNA replication, and necroptosis (**Fig. 3.3b**). However, since these analyses may not include all known functions attributed to these proteins, we examined the literature to identify additional activities regulated by these proteins. We found that several of the peptides were from proteins associated with antiviral signaling, Ub/Ub-like signaling, and DNA replication/repair (**Fig. 3.3c**). Moreover, several peptides were from proteins previously implicated in poxvirus infection including IFITs, (103) WDR6, (205) RACK1, (206) DDX5, (207) and EGFR. (208)

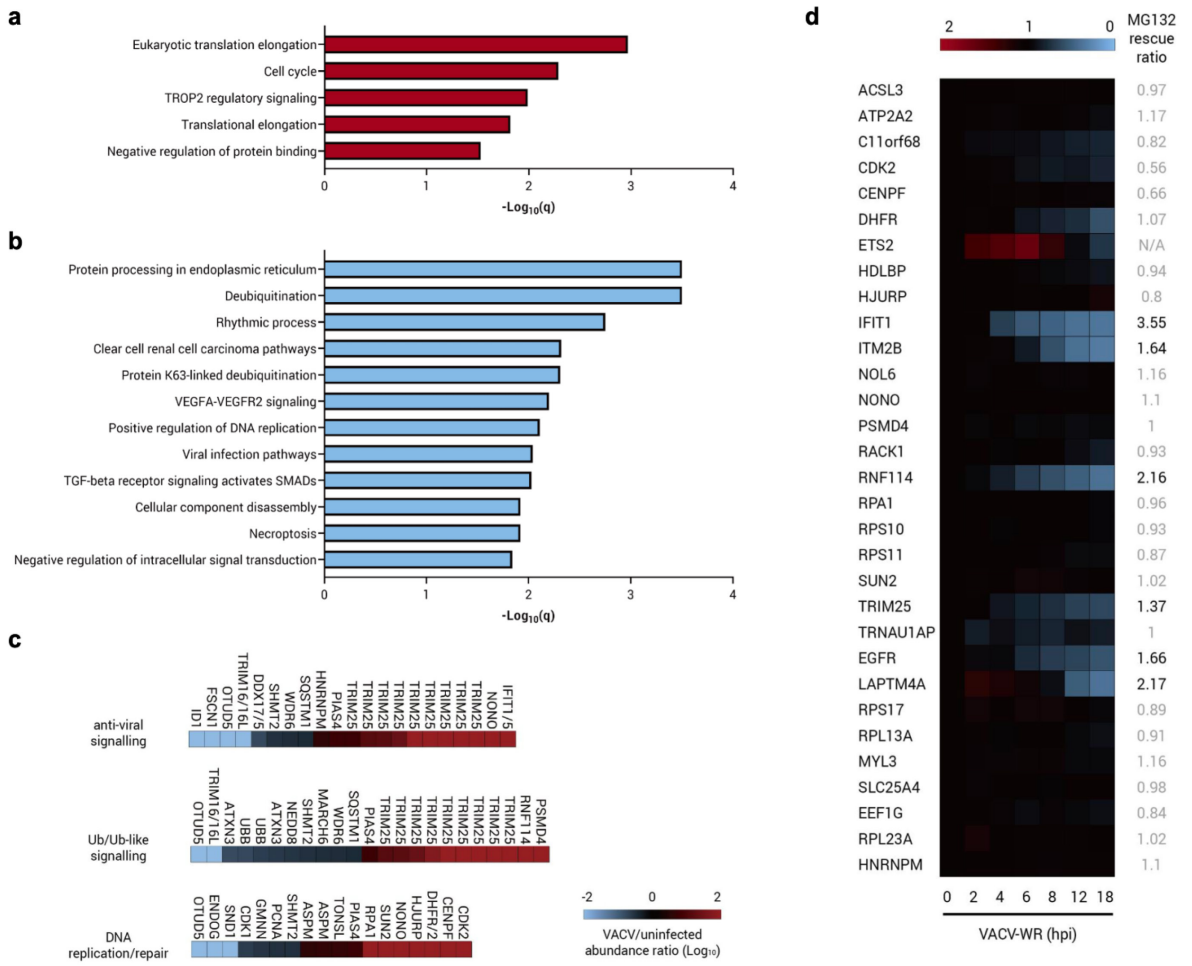


Figure 3.3. Analysis of the protein functions and abundances with enriched diGlycyl peptides in VACV infection. Metascape analysis of cell signaling pathways and processes of proteins associated with diGlycyl peptides enriched in **a**) VACV-Cop-infected or **b**) uninfected HeLa cells. Categories with significant adjusted p-value ($q \leq 0.05$) are shown. **c**) Heat map of identified diGlycyl proteins associated with antiviral signaling, Ub or Ub-like signaling, or DNA replication/repair was determined from the literature. The abundance ratio (VACV infected/uninfected; Log10 scale) is indicated. Peptides with an abundance ratio of -2 were exclusively found in uninfected cells, whereas those with an abundance ratio of 2 were found exclusively in infected cells. **d**) For proteins with enriched diGlycyl peptides in VACV-Cop infected HeLa cells, we examined their overall protein-level abundance changes using the dataset generated by Soday *et al.* in VACV-WR infected HFFF-TERT cells. (209) The heat map shows the changes in protein levels during infection relative to the uninfected sample which was arbitrarily set at 1. Proteins whose abundance decreased during infection are highlighted with asterisks. MG132 rescue ratio, as determined by Soday et. al. is indicated. Those with a rescue ratio greater ≥ 1.25 are indicated in black.

The identification of a diGlycyl lysine residue on a peptide is indicative of the site being ubiquitinated. However, it does not provide any information about the consequence of this

ubiquitination. Therefore, we compared our data to published work that analyzed changes in protein levels in HFFF-TERT cells infected with VACV Western Reserve strain (VACV-WR). (209) Despite this study using a different cell line and VACV strain, several of the proteins with diGlycyl peptides enriched in VACV-Cop-infected HeLa cells were decreased over the course of infection (**Fig. 3.3c**). Furthermore, this study showed that protein levels for most of the down-regulated proteins could be partially rescued by treatment with MG132 (**Fig. 3.3c**). This indicates these proteins are likely degraded in a proteasome-dependent manner in VACV-infected cells which has been shown for the IFIT proteins. (103) Intriguingly, the levels of many proteins with diGlycyl peptides enriched in infected cells did not decrease over the course of infection (**Fig. 3.3c**). This suggests that ubiquitination of these proteins may serve a non-degradative function.

3.4.3 Identification of diGlycyl VACV peptides

Although VACV replication is highly dependent on the cellular ubiquitination system, ubiquitination on viral proteins during VACV infection has never been systematically examined. In our study, we detected 222 unique diGlycyl peptides in 77 proteins in the VACV-Cop proteome (**Table 3.2**). We found diGlycyl modifications on known ubiquitinated viral proteins, such as N1 protein, an inhibitor of the nuclear factor NF-κB and apoptosis that contributes to virulence, (210) and E5 protein, which facilitates the ubiquitination and degradation of the DNA sensor cyclic GMP-AMP synthase. (211)

Of note, two viral E3 ligase adaptor BTB-Kelch proteins A55 and C2 were identified with five and one diGlycyl sites, respectively. Three VACV ANK-PRANC adaptors, B20, C9 and M1 were also found with diGlycyl sites. C9 protein binds IFITs and resists type I interferon-induced state. (102, 103) All of the diGlycyl modifications were located in the substrate binding domains (C-terminal Kelch domain in BTB-Kelch or N-terminal ankyrin domain in ANK-PRANC). This

suggests that the ubiquitination of these adaptor proteins can play a key role in viral infection by regulating viral and/or host substrate recognition in addition to targeting unfolded viral proteins in viral translation.

VACV Gene	Acc #	Protein name	# of diGlycyl modifications	diGlycyl lysines
A3L	P20643	Major core protein 4b	2	K164, K471
A18R	P20534	Transcript termination protein A18	3	K47, K88, K290
A20R	P20995	DNA polymerase processivity factor component A20	3	K248, K310, K349
A31R	P21096	Protein A31	1	K13
A35R	P21058	Protein A35	2	K56, K62
A37R	P21060	Protein A37	3	K126/K132, K233, K243
A42R	P68695	Profilin	1	K109
A44L	P21097	3 beta-hydroxysteroid dehydrogenase/Delta 5-->4-isomerase	3	K188, K198, K296
A47L	P21067	Protein A47	1	K75
A49R	P21068	Protein A49	1	K143
A51R	P21069	Protein A51	3	K23/29, K38, K236
B2R	P20999	Poxin	3	K59, K187/K198, K198
B8R	P21004	Soluble interferon gamma receptor B8	3	K211, K212, K239
B12	P21098	Pseudokinase B12	2	K5, K52
B13R	P20841	Putative serine proteinase inhibitor 2 homolog first part	1	K77
B14R	P20842	Putative serine proteinase inhibitor 2 homolog second part	1	K208
B19R	P21077	Surface antigen S	8	K88/K93, K111, K123, K123, K150, K189, K242, K255
B20R	P21078	Protein B20	1	K29
B29R	P21090	Inactive chemokine-binding protein	1	K98
C2L	P21037	Kelch repeat protein C2	1	K486/K491
C9L	P21042	Ankyrin repeat protein C9L	1	K5
C10L	P21043	Protein C10	1	K126

VACV Gene	Acc #	Protein name	# of diGlycyl modifications	diGlycyl lysines
C16L	P21100	Protein C16/B22	1	K16
D1R	P20979	mRNA-capping enzyme catalytic subunit	5	K84, K383, K559, K724, K730
D5R	P21010	Primase D5	14	K36, K124, K291, K291, K317/K325, K325, K377, K435, K484, K513, K567, K576, K576/K591, K591
D8L	P20508	Cell surface-binding protein	1	K32
D12L	P20980	mRNA-capping enzyme regulatory subunit	4	K6, K88, K95, K214
DUT	P68634	Deoxyuridine 5'-triphosphate nucleotidohydrolase	1	K51
E2L	P21080	Protein E2	2	K475, K707/K708
E3L	P21081	RNA-binding protein E3	1	K45
E5R	P21046	Protein E5	6	K67, K70, K87, K223, K229, K310
F1L	P68450	Protein F1	1	K76
F4L	P20493	Ribonucleoside-diphosphate reductase small chain OS	2	K144, K295
F6L	P68601	Protein F6 OS	1	K9
F11L	P21052	Protein F11	3	K30, K31, K235
F12L	P21053	Protein F12	5	K70, K150, K361, K365, K480
F15L	P21020	Protein F15	2	K31, K49
F16L	P21021	Protein F16	1	K92
F17R	P68454	Phosphoprotein F17	2	K17, K74
G1L	P21022	Metalloendopeptidase G1	1	K7
G5R	P21026	Putative nuclease G5	1	K213
G7L	P21028	Assembly protein G7	2	K268, K292/K295
H1L	P20495	Dual specificity protein phosphatase H1	2	K8, K165
H5R	P20538	Late transcription elongation factor H5	1	K16
HR	P20632	Interferon antagonist K1L	1	K282
I3L	P20499	Protein I3	3	K182, K189, K211

VACV Gene	Acc #	Protein name	# of diGlycyl modifications	diGlycyl lysines
I4L	P20503	Ribonucleoside-diphosphate reductase large subunit	8	K5, K10/K17, K655, K675, K679, K687, K691, K740
I6L	P68463	Telomere-binding protein I6	3	K6, K274, K365
I7L	P20501	Core protease I7	1	K410/K414
K5L	P21084	Protein K5	4	K80, K82, K91, K102
K6L	P68465	Protein K6	1	K34
KBTB1/A55R	P21073	Kelch repeat and BTB domain-containing protein 1	5	K369, K430, K514/K516, K516, K535
L4R	P20981	Core protein VP8	1	K204
LIG	P20492	DNA ligase	9	K9, K21, K177, K221, K241, K241, K260, K470/473, K477
M1L	P20640	Ankyrin repeat protein M1	6	K45, K184, K197/K200, K244, K229, K452
M2L	P21092	Protein M2	1	K180
N1L	P21054	Protein N1	1	K78
N2L	P20641	Protein N2	1	K103
O1L	P21093	Protein O1	1	K498
PAPL	P21079	Poly(A) polymerase catalytic subunit	3	K35, K75, K327
PAPS	P21033	Cap-specific mRNA (nucleoside-2'-O-)-methyltransferase	4	K132, K241, K282, K303
POL	P20509	DNA polymerase	8	K21, K82, K359, K514, K532, K685, K810, K844
RPO7	P68315	DNA-directed RNA polymerase 7 kDa subunit	1	K29
RPO18	P21034	DNA-directed RNA polymerase 18 kDa subunit	3	K33, K43, K121
RPO19	P68610	DNA-directed RNA polymerase 19 kDa subunit	1	K70
RPO22	P68608	DNA-directed RNA polymerase 22 kDa subunit	2	K7, K11
RPO30	P21082	DNA-directed RNA polymerase 30 kDa polypeptide	2	K54, K126

VACV Gene	Acc #	Protein name	# of diGlycyl modifications	diGlycyl lysines
RPO132	P68694	DNA-directed RNA polymerase 133 kDa polypeptide	12	K139, K452, K455, K762, K777, K790, K800, K882, K1002, K1108, K1117, K1125
RPO147	P20504	DNA-directed RNA polymerase 147 kDa polypeptide	15	K50, K54, K75, K109, K663, K710, K1095, K1097, K1119, K1202/K1210, K1210, K1218, K1218/K1227, K1227, K1228
SPI-1	P20531	Serine proteinase inhibitor 1	3	K162, K233, K249
SPI-3	P20532	Protein K2	1	K181
TK	P68564	Thymidine kinase	1	K149
TMK	P68693	Thymidylate kinase	2	K14, K47
UNG	P20536	Uracil-DNA glycosylase	3	K86, K90, K169
VETFL	P20635	Early transcription factor 82 kDa subunit	1	K513
VITF3L	P20998	Intermediate transcription factor 3 large subunit	6	K124, K139, K153, K153/K154, K173/K177, K357
VPK1	P20505	B1 kinase	2	K204, K238

Table 3.2. List of the identified 77 diGlycyl-modified VACV proteins and sites. Grey-shaded rows indicate the VACV BTB-Kelch proteins; blue-shaded rows indicate the ANK-PRANC proteins. The "/" under the Column "diGlycyl lysines" denotes the identification of a peptide with a single diGlycyl modification, however with insufficient MS2 fragmentation information to assign the exact modified lysine position on the peptide.

We also compared our list to the reported cowpox viral (CPXV) protein ubiquitination profiled using the same diGlycyl ubiquitin remnant approach. (108) Specifically, the authors revealed 137 conserved ubiquitination sites in 54 viral proteins shared among five CPXV strains. They found that these highly conserved proteins were also encoded by VACV-Western Reserve strain (VACV-WR), although they did not characterize the ubiquitome in VACV-WR infection. However, we observed limited overlap between the diGlycyl modifications in diGlycyl

viral proteins between CPXV and VACV-Cop (<50%), even though all these viral proteins are also encoded by VACV-WR (**Figure 3.4**). Additionally, when we compared the viral proteins with the highest diGlycyl sites in CPXV and VACV, there was a more pronounced contrast. For instance, major core protein 4a precursor was most heavily modified in CPXV, but no associated diGlycyl sites were found in VACV. Similarly, many VACV DNA-directed RNA polymerases were found to contain several diGlycyl sites, with some of them presenting over ten modifications, but they were far less modified in CPXV. This might suggest that CPXV and VACV have different ubiquitination mechanisms despite shared lineage; however, a comprehensive sequence alignment is necessary to confirm this hypothesis.

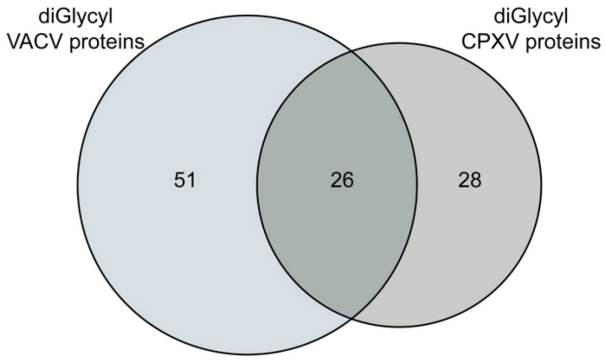


Figure 3.4. Overlap between diGlycyl viral proteins in VACV and CPXV viral proteins. A previous study identified conserved ubiquitination sites in 54 viral proteins among five CPXV strains. (108) Although all were also encoded by VACV-WR, many were not ubiquitinated during VACV infection.

3.4.4 K6-, K11- and K63-linkages were significantly decreased

There are seven acceptor lysines on ubiquitin that can serve as linkage anchors for polyubiquitination. We wanted to use the abundance of diGlycyl modified lysines on ubiquitin, as an indicator for the relative abundance of ubiquitin linkages and their corresponding biological contexts during infection. We found diGlycyl peptides corresponding to ubiquitin-ribosomal protein eS31 fusion protein RPS27A, specifically on K6, K11, K27, K48 and K63. To our surprise, we found three of the ubiquitin diGlycyl peptides were significantly decreased in VACV

infection, with modifications located on K6 (PSM; peptide-spectrum matches = 5), K11 (PSM = 16) and K63 (PSM = 6). No significant changes in abundance were observed for other sites.

The most well-known role for K11-linked ubiquitination is cell cycle regulation. Interestingly, VACV has also been long reported to arrest and inhibit cell cycle progression and mitosis. (212–215) More recently, VACV B1 kinase and/or B12 pseudokinase were shown to mediate the degradation of cell cycle effectors. (216) B12 was identified with two diGlycyl modifications in our study (**Table 3.2**). Additionally, many significantly increased diGlycyl host peptides are involved in the cell cycle regulation (**Fig. 3.3a**). This further implies that VACV might disrupt cell cycle regulation by decreasing K11-linked polyubiquitin and by ubiquitination of viral proteins, providing mechanistic details for a well-established phenomenon.

K6- and K63-linked ubiquitin conjugations are both highly involved in DNA damage response and repair. VACV was also shown to utilize cellular DNA damage response to recruit host proteins for viral genome replication. (217) This ties in with the heavy ubiquitination of viral DNA-directed RNA polymerases, along with that of the viral replisome components H3 and I3 we observed in our study (**Table 3.2**). K63-linked deubiquitination was also highly enriched in the significantly decreased diGlycyl host peptides (**Fig. 3.3b**), further indicating DNA damage response signaling was involved in VACV infection.

Collectively, our data identified diGlycyl modifications on known ubiquitinated VACV proteins, and novel viral proteins that dysregulate biological processes such as cell cycle and DNA damage response. This is further supported by the significant decrease of K6, K11 and K63 ubiquitin linkages also involved in cell cycle and DNA damage response, showing the pivotal role of ubiquitination in the virus cycle.

3.4.5 DiGlycyl modifications on TRIM25

The most striking observation from the diGlycyl enrichment data was the identification of nine diGlycyl peptides, that could be definitively assigned to six lysine residues, from TRIM25 that were enriched for or exclusively found in VACV-Cop-infected HeLa cells (**Fig. 3.5a,b**). Moreover, TRIM25 levels were found to decrease ~40% in a proteasome-dependent manner over the course of infection of HFFF-TERT cells with VACV-WR (**Fig. 3.3c**). (209)

TRIM25 is a really interesting new gene (RING) domain-containing E3 Ub-ligase that also functions as an E3 ligase for the Ub-related molecule, ISG15 (**Fig. 3.5c**). (218) TRIM25 performs many different functions, (219) but is perhaps most well-known for introducing K63-linked ubiquitin chains onto retinoic acid-inducible gene-I (RIG-I). (220, 221) RIG-I is a pattern recognition receptor that recognizes viral RNA and initiates signals that lead to the production of type I interferons (reviewed in (222)). Rather than promoting RIG-I degradation, ubiquitination of RIG-I by TRIM25 promotes RIG-I activation. (220) It is not surprising that many viruses have acquired mechanisms to interfere with the ability of TRIM25 to ubiquitinate RIG-I. (223–225) However, no studies have examined whether poxviruses interfere with TRIM25 signaling. Therefore, we decided to further investigate the putative ubiquitination of TRIM25 in VACV-infected cells.

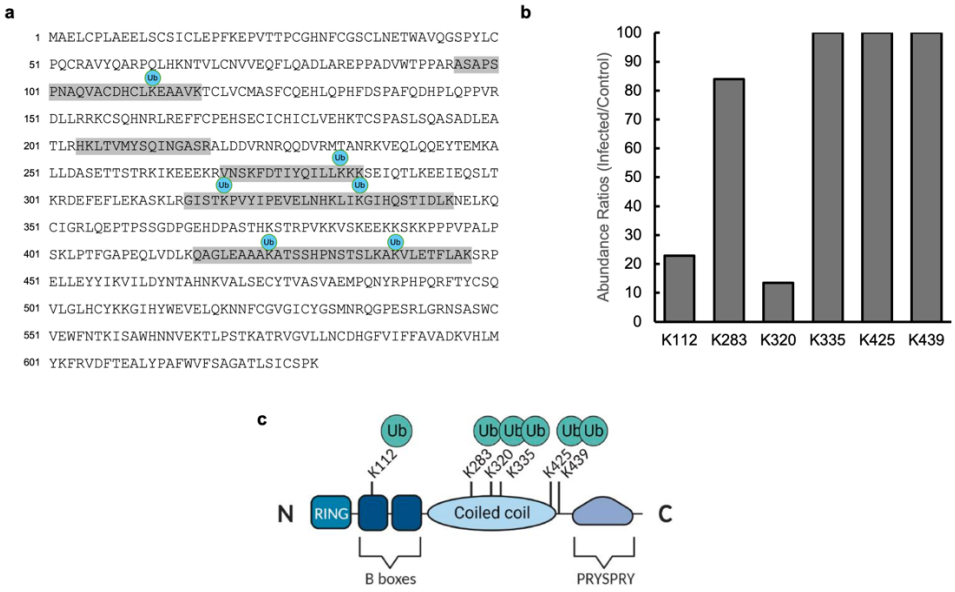


Figure 3.5. Extensive diGlycyl modifications found on TRIM25 using ubiquitin remnant profiling in VACV-Cop-infected HeLa cells. a) TRIM25 protein sequence and coverage in diGlycyl-peptide enriched fractions. Grey shaded region denotes detected peptides. **b**) Abundance ratios between VACV-Cop-infected and control for the diGlycyl peptides containing the six lysines. An arbitrary 100 was set as the maximum. **c**) The locations of identified TRIM25 diGlycyl peptides relative to its domain. Four diGlycyl sites were located in the coiled-coil region.

3.4.6 TRIM25 modifications were likely due to ubiquitination

We confirmed that TRIM25 was highly modified by immunoblot analysis. Anti-TRIM25 immunoblot analysis revealed the appearance of HMW, potentially ubiquitinated, TRIM25 species and a decrease in the lower-molecular-weight (LMW) TRIM25 band over the course of infection (**Fig. 3.6a**). HMW TRIM25 proteins were evident as early as 1 hour post-infection (hpi), persisted throughout infection, but were reduced at 24 hpi (**Fig. 3.6a**). Quantification of total anti-TRIM25 immunoreactive bands in uninfected and infected cells showed a decrease in anti-TRIM25 levels over the course of infection (**Fig. 3.6b**).

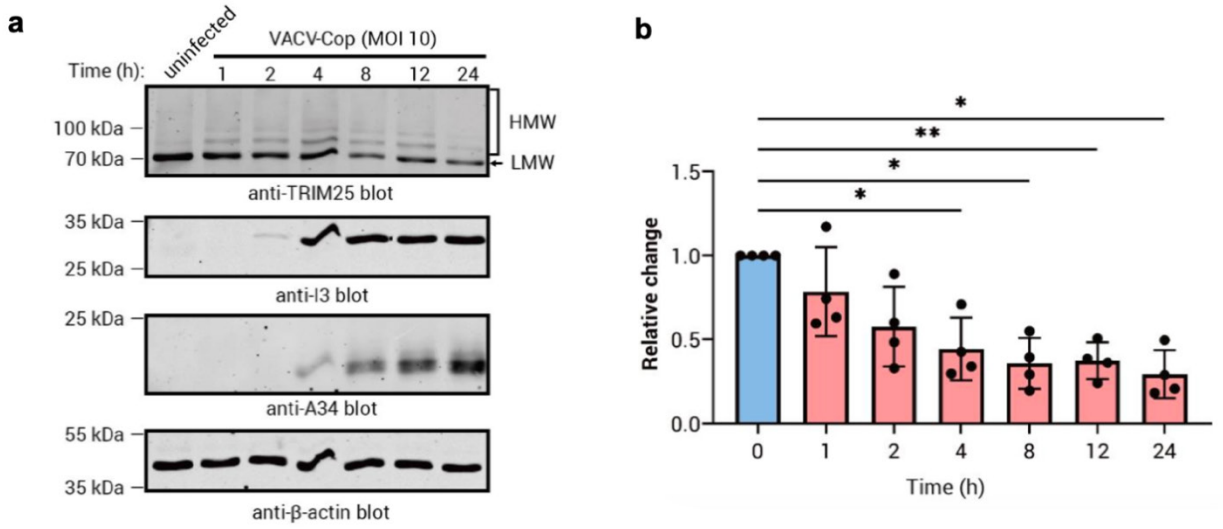


Figure 3.6. Immunoblot analysis showed modified TRIM25 species at as early as 1 hpi. a) Lysates of HeLa cells uninfected or VACV-Cop-infected (MOI of 10) for the indicated times were immunoblotted with antibodies against the indicated proteins. I3 and A34 are viral early and late proteins, respectively, and immunoblots of these were included to show the course of infection. The anti-β-actin blots were included to show protein loading. Molecular mass markers are indicated to the left of blots. **b**) Quantitative analysis of TRIM25 immunoreactive bands in VACV-Cop-infected cells (mean and standard deviation) relative to uninfected cells from 4 independent experiments. A one-way ANOVA was used to calculate statistical significance between uninfected cells (0 h time point) and the indicated time points post-infection. *; $p \leq 0.05$, **; $p \leq 0.01$. Performed by Jianing Dong in the Ingham laboratory.

To further investigate whether the HMW TRIM25 species were ubiquitinated forms of the protein, we pre-treated HeLa cells with the E1 inhibitor, TAK-243, (226) and infected cells with VACV-Cop in the presence of the inhibitor (**Fig. 3.7a**). A dose-dependent decrease in TRIM25 HMW species was observed in infected cells treated with TAK-243 (**Fig. 3.7b**; upper panel), although this was associated with a decrease in I3 protein expression (**Fig. 3.7b**; middle panel). An examination of ubiquitinated proteins in lysates from TAK-243-treated cells revealed the inhibitor was effective (**Fig. 3.7c**). Taken together, these results show that E1 activity is required for the appearance of HMW TRIM25 species at the time of infection, and identified TRIM25 diGlycyl modifications by mass spectrometry were derived from ubiquitination.

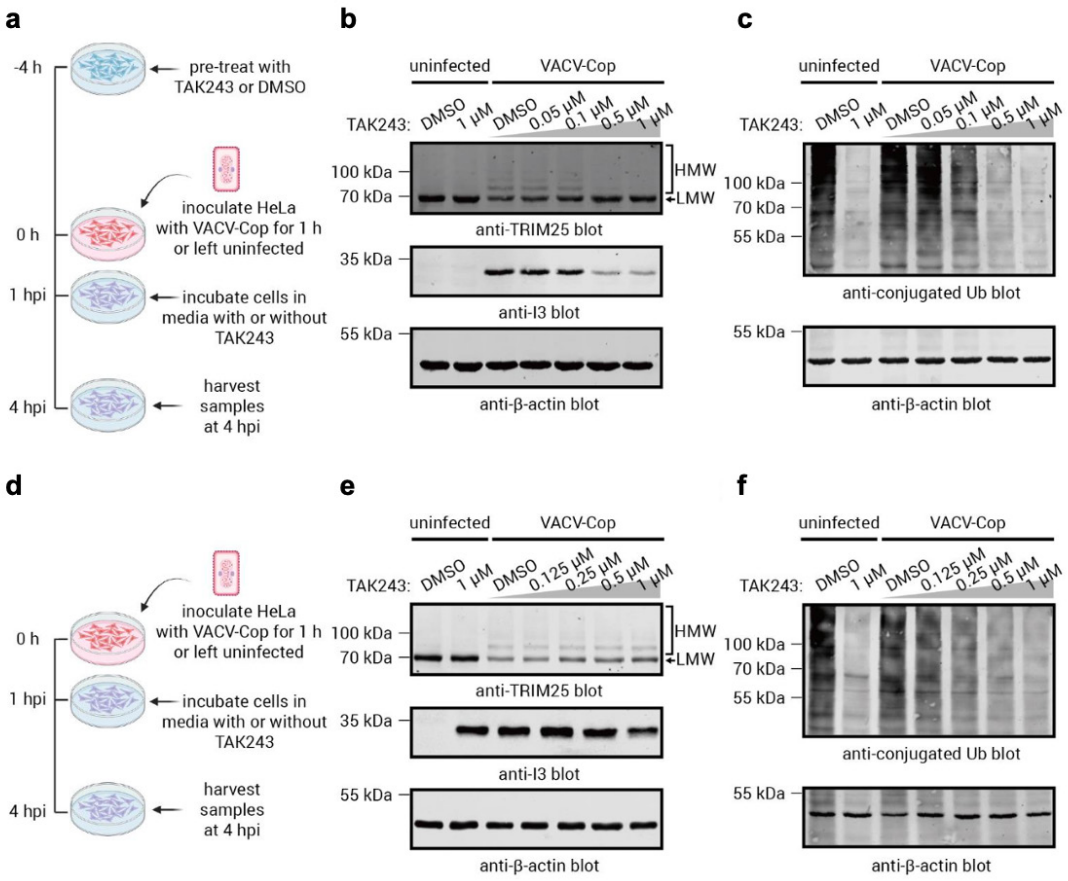


Figure 3.7. TRIM25 ubiquitination in VACV-Cop-infected cells treated with E1 inhibitor. a) Schematic of the experimental workflow used to examine the effect of the E1 inhibitor TAK243 treatment on TRIM25 HMW species. **b,c**) Lysates from uninfected cells or cells with VACV-Cop for 4 h (MOI of 10) treated with the indicated concentrations of inhibitor were immunoblotted with antibodies against the indicated proteins. The anti-I3 blot shows infection of the cells, and the anti-β-actin blots show protein loading. The conjugated ubiquitin antibody recognizes mono- and poly-ubiquitinated proteins. The anti-β-actin blots were included to show protein loading. Molecular mass markers are indicated to the left of blots. Performed by Jianing Dong in the Ingham laboratory.

Additionally, we examined whether the formation of HMW TRIM25 species was affected by treatment with the neddylation E1 protein inhibitor, MLN4924 (**Fig. 3.8a**). (227) No changes in HMW TRIM25 species were observed in cells treated with MLN4924 (**Fig. 3.8b**; upper panel). However, neddylation of a known NEDD8 modified protein, Cullin1, (228) was affected by MLN4924 treatment (**Fig. 3.8c**; upper panel). Taken together, these experiments support the notion that TRIM25 is ubiquitinated, but not neddylated, in VACV-Cop-infected cells.

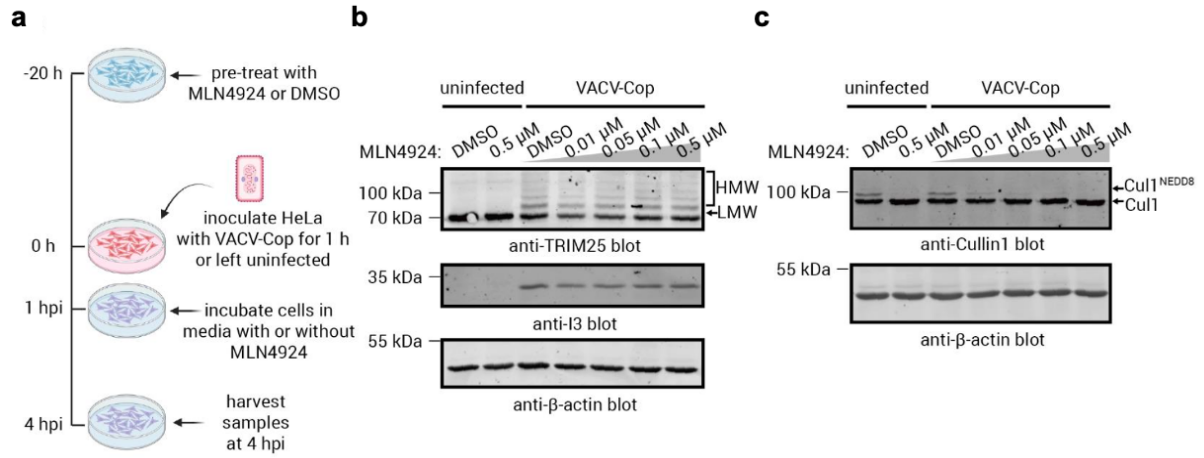


Figure 3.8. TRIM25 ubiquitylation in VACV-Cop-infected cells treated with neddylation E1 inhibitor. a) Schematic of the experimental workflow used to examine the effect of the neddylation E1 inhibitor, MLN4924, on TRIM25 HMW species. **b**, **c**) HeLa lysates from uninfected cells or cells with VACV-Cop for 4 h (MOI of 10) treated with the indicated concentrations of inhibitor were immunoblotted with antibodies against the indicated proteins. The anti-I3 blot shows infection of the cells, and the anti-β-actin blots show protein loading. The anti-Cullin1 Ab recognizes both unmodified Cullin1 (Cul1) and NEDD8-modified Cullin1 ($Cul1^{NEDD8}$). The anti-β-actin blots were included to show protein loading. Molecular mass markers are indicated to the left of blots. Performed by Grace Melvie in the Ingham laboratory.

3.4.7 Effect of proteasome inhibition on TRIM25

Proteomics studies suggest that TRIM25 is degraded in a proteasome-dependent manner in VACV infection (**Fig. 3.3d**). (209) To directly test this, HeLa cells were treated before and after inoculation with the proteasome inhibitor, MG132, and TRIM25 levels were examined at 4 and 8 hpi (**Fig. 3.9a**). Total TRIM25 levels were modestly increased in cells treated with MG132 (**Fig. 3.9b**), more so 8 hpi, but this difference was not statistically significant (**Fig. 3.9c,d**). Similar observations were made for the LMW TRIM25 species (**Fig. 3.9b,e,f**). MG132 treatment blocked expression of the B5 viral late protein (**Fig. 3.9b**) which was consistent with previous studies showing late gene expression requires proteasome activity. (105–107) Thus, the reduction in TRIM25 protein levels is, at least in part, due to proteasomal degradation.

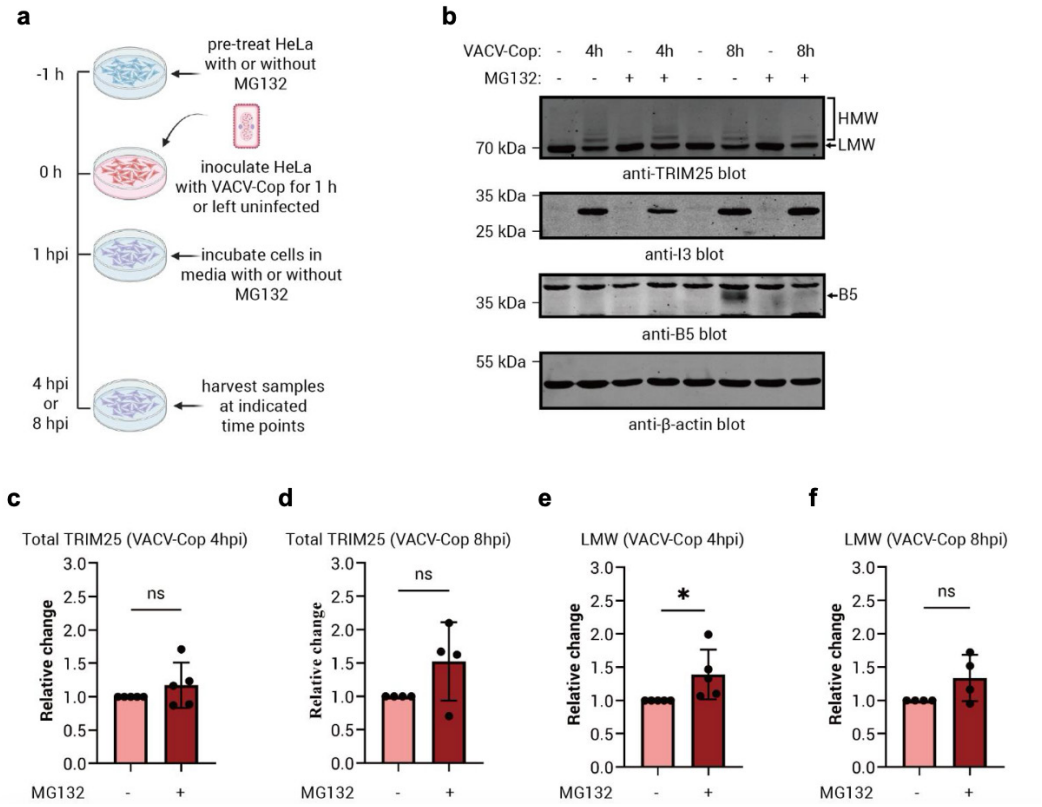


Figure 3.9. MG132 treatment modestly rescues TRIM25 levels in VACV-Cop-infected HeLa cells. a) Outline of the experimental workflow used in this experiment. **b**) Lysates were immunoblotted for Abs against the indicated proteins. I3 and B5 are viral early and late proteins, respectively. Immunoblots for these proteins were included to show the course of virus infection. The anti-β-actin blots were included to show protein loading. Molecular mass markers are indicated to the left of blots. Quantitative analysis of total anti-TRIM25 immuno-reactive bands 4 (**c**) or 8 (**d**) hpi in cells infected in the presence or absence of MG132 are shown. Quantification is from 5 (**c**) or 4 (**d**) independent experiments and expressed relative to the samples not treated (-) with MG132. The abundance of the LMW TRIM25 band 4 (**e**) or 8 (**f**) hpi are shown. Quantification represents the mean and standard deviation from 5 (**e**) or 4 (**f**) independent experiments and expressed relative to the samples not treated (-) with MG132. A paired, one-tailed Student t test was used to calculate statistical significance. ns; not significant, *; $p \leq 0.05$. Performed by Jianing Dong in the Ingham laboratory.

3.5 Discussion

3.5.1 Mono- vs poly-ubiquitination

The use of the ubiquitin remnant antibody targeting diGlycyl lysines enables high-throughput profiling of the cellular ubiquitome, but it does not provide information about the

number or the type of conjugated ubiquitin chains on proteins. For instance, there were at least six diGlycyl lysines identified on TRIM25 in our study, which corresponded to the accumulation of HMW species detected by immunoblot analysis. The modified forms of TRIM25 were most predominant between 80 to 90 kDa (unmodified mass = 71 kDa) and appeared more as a smear higher up on the molecular weight scale, suggesting that the most abundant species was a mono- or di-ubiquitinated form of TRIM25 based on the size of ubiquitin (8 kDa) (**Fig. 3.6**). However, we could not definitively attribute all the HMW formation to multi-monoubiquitination or polyubiquitination or a mixture of both on the identified six sites.

We found that the proteasome inhibitor MG132 treatment modestly increased unmodified TRIM25 levels (**Fig. 3.8**) consistent with the previous study reporting proteasome-dependent TRIM25 degradation during VACV infection (**Fig. 3.3d**). (209) However, there was no statistical difference in total TRIM25 levels (including HMW forms) between treatments with and without MG132, suggesting that not all ubiquitinated forms of TRIM25 were targeted for proteasome-dependent degradation (**Fig. 3.8c,d**). We also could not rule out that the changes in anti-TRIM25 immunoreactive bands in infected cells were due to the anti-TRIM25 antibody exhibiting differential recognition or binding capability to the modified protein.

Collectively, the HMW TRIM25 species persisted through infection with and without MG132 treatment, and likely represented mono- and di-ubiquitinated forms of TRIM25 that could serve a non-degradative function.

We aimed to monitor the abundances of diGlycyl lysines on ubiquitin as an indicator for changes in the types of ubiquitin linkages during VACV infection. Interestingly, we found that three of the ubiquitin diGlycyl lysines (K6, K11 and K63) were significantly decreased in VACV-infected HeLa. No significant changes in abundance were observed for other sites. This suggested there was at least decrease in some forms of polyubiquitination.

3.5.2 Ubiquitin-like modifications

The trypsin digestion of proteins modified by the ubiquitin-like proteins, ISG15 and NEDD8, also generates diGlycyl motifs. Although studies have shown that 94% of cellular diGlycyl motifs were derived from ubiquitin, (135) additional experiments were needed to confirm the diGlycyl modifications of interest specifically resulted from ubiquitination. We verified that the formation of HMW TRIM25 species was fully inhibited by pretreatment of the E1 inhibitor TAK243 (**Fig. 3.7b**), and was unaffected by the neddylation E1 inhibitor MLN4924 (**Fig. 3.8b**). Our results demonstrated these HMW species were not neddylated but ubiquitinated forms of the protein. However, further research is required to fully eliminate the possibility of ISGylation. Immunoblot experiments to determine whether TRIM25 was ISGylated were inconclusive, and specific ISGylation inhibitors are not available to our knowledge. Thus, we cannot rule out that the HMW TRIM25 species may also represent ISGylated forms of TRIM25.

Alternatively, the use of the UbiSite antibody, which recognizes the C-terminal 13 amino acids of ubiquitin (ESTLHLVLRLRGG) after LysC digestion, can provide definitive assignment of the ubiquitination sites on proteins (reviewed in Chapter 1.3.3). (136) However, LysC only cleaves C-terminal to unmodified lysine residues. Based on the protein sequence surrounding the identified TRIM25 diGlycyl lysines (**Fig. 3.5a**), LysC digestion on ubiquitinated TRIM25 will generate peptides that are too long for LC-MS/MS detection and thus may be ineffective in this case.

3.5.3 Viral infection-induced deubiquitination

In our study, we focused our analysis on proteins with significantly increased diGlycyl modifications in VACV infection. However, a generally overlooked field is viral-infection induced deubiquitination. Many viruses encode deubiquitinases. For instance, SARS-CoV-2 PLpro is a deubiquitinase and deISGylase. (153, 154) Although the VACV genome does not encode any

known deubiquitinases, viral proteins can interact with host deubiquitinases to regulate immune signaling pathways and induce deubiquitination. One of the proteins we detected with significantly decreased diGlycyl sites in VACV infection was OTU domain-containing protein 5 (OTUD5; also known as deubiquitinating enzyme A, DUBA). OTUD5 is a deubiquitinase that negatively regulates immune response by suppressing type I interferon production. (229) We observed substantially reduced ubiquitination on OTUD5 K385 during VACV infection (**Fig. 3.2**). As OTUD5 is activated by phosphorylation, (230) the effect of ubiquitination or deubiquitination on OTUD5 activity has never been examined. Lastly, the significant decrease of K6, K11 and K63 ubiquitin linkages also implied less ubiquitination or even enhanced deubiquitination in VACV infection. Further investigations on deubiquitination in viral infection will provide a more comprehensive understanding of the viral replication strategies and mechanisms. For instance, the same protein can undergo ubiquitination and deubiquitination at distinct lysine residues during viral infection, serving different functions. Specifically, we found three cellular proteins containing both significantly increased and decreased diGlycyl modifications (PEG10, RPL10 and RPS10), although their roles in VACV infection were not explored in our study.

Chapter 4: Virus-host protein-protein interactions in MAYV infection

Preface

Chapter 4 consists of unpublished work performed both by me and Joaquin Lopez-Orozco. I performed mass spectrometry analysis, statistical analysis and manuscript composition. J.L. assisted with many experiments such as electroporation and viral infection. Drs. Tom Hobman and Olivier Julien were involved with the experimental design.

4.1 Abstract

Mayaro virus (MAYV) infection is an emerging mosquito-transmitted viral disease with high expansion and urbanization risks. Building a virus-host protein-protein interaction (PPI) network will be a significant step in understanding the molecular mechanisms of viral infection in support of effective vaccine and therapeutic development. One of the conventional mass spectrometry-based interactomics approaches is affinity-purification coupled with mass spectrometry (AP-MS). Although substantial success has been achieved in mapping virus-host PPIs using AP-MS, there are some limitations associated with this approach especially in virology applications. To enhance the biological significance of interactomics in virology studies, we used AP-MS to characterize virus-host PPIs in the context of viral infection. We have identified host proteins that are specific to viral protein subcellular locations and biological pathways, as well as key common interactors that are shared among multiple viral proteins that could otherwise be missed using the conventional approach.

4.2 Introduction

Found in parts of South and Central America, MAYV is transmitted to humans through the bite of an infected mosquito. (231) MAYV infection can cause fever, headache, muscle and joint pain. The virus has caused many epidemics, and has been shown to be transmitted by mosquito species *Aedes aegypti* (232) and *Aedes albopictus* (233) that have spread to many countries, posing a risk for urbanization and future global outbreaks. To date, there are no vaccines or treatments for MAYV disease. Although a few host protein interactors are identified for some MAYV non-structural proteins (NSPs), (234) the viral protein interactome has not been systematically characterized under infection conditions.

To gain insights on the mechanism of MAYV pathobiology, we employed affinity-purification coupled with mass spectrometry (AP-MS) to map the interaction network between

virus and host proteins. Interactomics using LC-MS/MS characterizes protein-protein interactions, and in the context of virology studies, can be used to characterize virus-host protein interactions. After viral entry in host cells, the viral genome is released into the cytoplasm, and translated into non-structural and structural proteins. The viral proteins interact with the host to compete for molecular resources, and viral replication is dependent on these protein-protein interactions. Building a virus-host protein interaction network is a significant step in understanding the molecular mechanisms of viral infection and in defining the viral pathobiology. Furthermore, characterizing viral protein interactome can reveal dysregulated cellular pathways in the host, along with key viral interactors that might be restriction and dependency factors. These viral interactors can be modulated for antiviral therapeutics.

AP-MS has been widely used in mapping virus-host protein-protein interactions in many studies. (33–36, 235) However, there are some limitations associated with this approach when characterizing virus-host PPIs, as the cells are expressing a viral protein but not infected by the virus. Viral infection inherently alters the host proteome and signaling pathways, and for instance, protein-protein interactions downstream of elevated interferon response or activated apoptotic signaling may be potentially missed. Secondly, the viral proteins are usually expressed in the cell one at a time. This overlooks the potential formation of the viral replication complex in the virus cycle, where the binding of host proteins may only be mediated by a stable viral protein complex. We hypothesized that by performing AP-MS in virus-infected cells, we would better define the virus-host interactome of MAYV (**Fig. 4.1**).

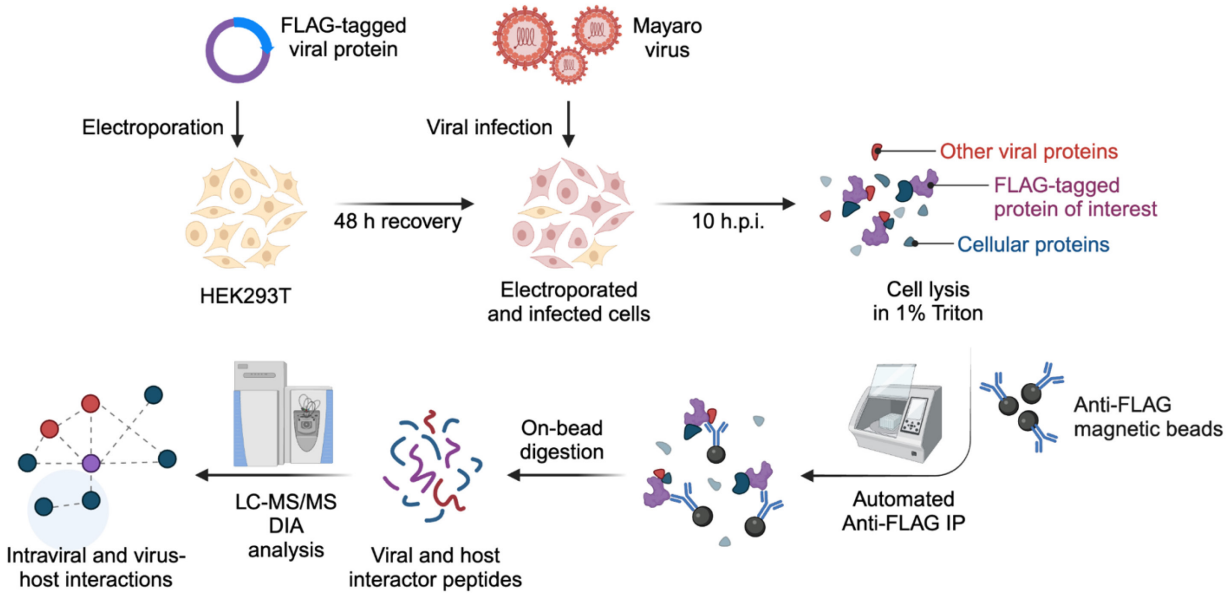


Figure 4.1. Overview of the AP-MS in infected cells experimental design. HEK293T cells were transfected with a FLAG-tagged viral protein construct by electroporation, recovered, and then infected by MAYV, generating a pool of infected cells overexpressing a FLAG-tagged viral protein. The binding proteins (viral and cellular) were enriched using automated anti-FLAG co-immunoprecipitation and digested to peptides for LC-MS/MS analysis.

4.3 Methods

4.3.1 Plasmids

Plasmids used in this study were pcDNA 3.1(−) 3×FLAG constructs generated and described previously(234). In brief, 3×FLAG tag was cloned C-terminal to MAYV viral protein-specific gene in the pcDNA 3.1(−) plasmid. (236) Due to the cis-autoproteolysis activity of capsid protein, which cleaved off the C-terminal 3×FLAG tag, a 3×FLAG sequence was added in frame to the N-terminal region of the protein. The NSP2 C480A construct was cloned by site-directed mutagenesis using overlapping oligos AAAGCTAAAGTGGCCTGGGCCAAATGC and GCATTTGGCCCAGGCCACTTTAGCTTT. The integrity of all constructs was confirmed by Sanger sequencing.

4.3.2 Cell culture

HEK293T cells (ATCC) were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Gibco, #11995-065), supplemented with 10% fetal bovine serum (Sigma, #F1051-500ML) and 20 mM HEPES (Gibco, #15630-080). Cells were maintained at a 37˚C incubator with 5% CO_2.

4.3.3 Electroporation

Sub-confluent HEK293T cells were harvested by trypsinization, and collected by centrifugation at 500 x g for 5 min. Cells were resuspended in 1.5 x 10^7 cells/mL of Cytomix buffer [10 mM potassium phosphate buffer, pH 7.6, 25 mM HEPES, pH 7.6, 120 mM KCl, 0.15 mM $CaCl_2$, 2 mM EGTA, 5 mM $MgCl_2$ (all pHs adjusted to 7.6 using KOH)]. 100 µL of plasmid DNA at 100 ng/µL and 400 µL of cell suspension were transferred to an electroporation cuvette (Bio-Rad, 0.4 cm gap width, #1652088), supplemented with freshly prepared 2 mM ATP and 5 mM glutathione. Cells were pulsed once with 975 µF and 300 V (Bio-Rad, Gene Pulser II). Time constants were typically between 17-20 ms. Electroporated cells were transferred to 10 mL of pre-warmed medium (described under 4.3.2), where 1 mL was added to a well in the 6-well plate in preparation for cell counting for viral infection, and the rest (~10 mL) was added to a T75 for cell recovery.

4.3.4 Viral infection

MAYV-mNeonGreen reporter virus was developed by Dr. Anil Kumar in the Hobman lab. The virus strain used in the study was Mayaro Venezuela 16A (Genebank KP842794), (237) which was kindly gifted by Brandy Russell at the Centre for Disease Control and Prevention (Fort Collins, CO, USA). Virus stocks were generated in C6/36 cells [kindly provided by Dr. Sonja Best, NIH Rocky Mountain laboratories (Hamilton, MT, USA) and cultured in Dr. Tom Hobman's lab] (234) and titered using HEK293T cells. 24 h post electroporation, HEK293T cells in the T75 were replaced with fresh medium. 48 h post electroporation, HEK293T cells in the 6-

well plate were lifted by trypsinization for counting using trypan blue and Countess II FL Automated Cell Counter (ThermoFisher). According to the cell count, the required viral particles were added to the corresponding T75 flask to yield a multiplicity of infection (MOI) of 1.

For MOI = 1:

$$volume\ of\ virus\ needed\ = \frac{cell\ count\ in\ the\ 6-well\ plate\ (in\ cells/mL)\ \times\ 10}{viral\ titre}$$

10 hours post infection (h.p.i.), cells were lifted using 0.5 mM EDTA and collected by centrifugation at 800 x g for 5 min. Cell pellets were frozen upright in the -80˚C.

4.3.5 Cell viability assay

Cell viability assay was performed in a cell culture 96-well opaque-wall microplate (Greiner bio one, #655083) using CellTiter-Glo Luminescent Cell Viability Assay (Promega, #G7571).

4.3.6 Automated anti-FLAG affinity purification and on-bead trypsin digestion

Cells were resuspended in 500 µL lysis buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl and 1% Triton X-100) supplemented with 1X Halt protease and phosphatase inhibitor (ThermoFisher, #PI78443) lysed by gentle sonication, and clarified by centrifugation at 13,000 rpm for 10 min. 20 µL of the clarified lysate was saved as “input”. The rest proceeded to automated anti-FLAG affinity purification using KingFisher Duo Prime (ThermoFisher) and anti-DYKDDDDK magnetic agarose beads (Pierce, #A36797) according to the manufacturer’s instructions with minor modifications. In brief, 50 µL of magnetic beads were pre-washed with lysis buffer, incubated with cell lysates for 2 h at 4 ˚C, washed twice with lysis buffer, and washed twice with 100 mM ammonium bicarbonate. The beads were then reduced with 10 mM

DTT, alkylated with 50 mM iodoacetamide, and directly digested to peptides with 0.6 µg of trypsin for 5 h at 37˚C. Digested peptides were dried using GeneVac (EZ 2 4.0).

4.3.7 Mass spectrometry analysis

For input lysates prior to anti-FLAG pulldown, sample cleanup was performed using S-trap micros (ProtiFi) with the addition of 5% SDS. Peptides were desalted with C18 ziptips (Millipore, #ZTC18S960) and recovered in buffer A [0.1% formic acid (FA)] prior to mass spectrometry analysis. The samples were analyzed using an IonOpticks Aurora TS 25 cm coupled to an Orbitrap Exploris 480 Mass Spectrometer (ThermoFisher). Reverse phase separation of the peptides was done with an Aurora Ultimate analytical column (25 cm x 75 µm ID with 1.7 µm media, IonOpticks). Peptides were eluted with a solvent B gradient (80% ACN, 0.1% FA) for 90 min. The gradient was run at 400 nL/min with analytical column temperature set at 45°C. For both input and IP samples, DIA analysis was used using 37.5 m/z isolation window widths, and data were searched using Spectronaut (v18.5) with directDIA library-free workflow and factory settings against the concatenated database of the human proteome (UP000005640, v2021-08-30) and MAYV Venezuela 16A proteome (Genebank KP842794) (237) with false discovery of peptides restricted at 1%. Search parameters included a maximum of two missed trypsin cleavages, with the constant modification carbamidomethylation (C), and a maximum of five variable modifications of acetyl (protein N-term) and oxidation (M). Only tryptic peptides were searched, and MS1 intensities were normalized using total peptide amount. Protein abundances were calculated using the sum of peak areas of peptide MS2 quantities.

4.3.8 Analysis of protein-protein interactions

Protein interactions were scored based on DIA MS2-level protein abundances. The data report files generated by Spectronaut were reformatted as required input files for executing SAINT and MiST analysis.

For SAINT analysis, SAINTexpress-int program (146) was run with default settings and controls scored as independent experiments. To incorporate known interaction data to boost interaction scores, the Gene Ontology Annotation database for human was downloaded from EMBL-EBI (v2023-11-10), and the corresponding protein associations for all preys detected in the study were extracted. SAINT scores were visualized using dot plot. (ProHits-viz(238))

MiST scores were calculated for each viral protein individually against control, with log2 transformed MS2-level protein abundances and default fixed weights (reproducibility = 0.30853, specificity = 0.68551, abundance = 0.00596). (141)

4.4 Results

4.4.1 Optimization of transfection and infection level

To perform AP-MS in infected cells, one of the key steps was to generate and maximize the cell population that were both 1) expressing the target affinity-tagged viral protein and 2) infected by the virus of interest. We selected electroporation to improve plasmid uptake efficiency, and allowed cells to recover for 48 h. Optimizing viral infection post-transfection was also necessary to infect a large cell population at an early stage of infection without inducing cell death or excessive protein degradation. We tested a series of MOIs and earlier infection times prior to 24 h.p.i., at which point we noticed infected cells detach by microscopy. Using cell viability assay and flow cytometry with cells infected by MAYV-mNeonGreen reporter virus, we found MOI of 1 and harvesting 10 h.p.i. yielded the highest percentage of both viable and infected cells (**Table 4.1**). Therefore, we were able to establish a protocol with highest viral protein expression and infection.

	10 h.p.i.		12 h.p.i.	
	Viability	**Infection**	**Viability**	**Infection**
Mock	100%	0	100%	0
MOI 1	89%	86%	45%	90%
MOI 3	67%	89%	31%	96%
MOI 5	50%	93%	35%	94%
MOI 10	36%	91%	42%	94%

Table 4.1. The highest percentage of viable and infected cells was observed at MOI=1 by MAYV 10 h.p.i. HEK293T cells were infected with MAYV encoding a GFP reporter protein at MOI 1, 3, 5, and 10, along with 10 and 12 h.p.i. Numbers of viable cells were determined by cell viability assay, and infection efficiency was determined by flow cytometry. Percentages were generated using mock cells as the denominator.

4.4.2 Overexpression of viral protein in infected cells

The viral protein constructs used in this study were designed and described previously. (234) The expression of different viral proteins affects cell proliferation rates and can cause cytotoxicity, resulting in variations in cell counts. For instance, NSP2 causes host transcriptional shutoff and its expression is highly cytotoxic to cells. (234, 239) To partially mitigate the cytotoxicity caused by viral-protease induced proteolysis, and also to better capture NSP2 protease substrates in the otherwise transient proteolysis interactions, we generated a catalytically dead mutant for the viral protease encoded in the NSP2 (NSP2 C480A, NSP2-dead). Additionally, for all cells expressing various viral proteins to be treated with MOI=1 equivalence of plaque forming units (pfu) of viral particles, we counted the number of living cells in each condition and replicate prior to infection (**Figure 4.2**). We noticed substantial cytotoxicity in cells transfected with NSP2, and even further reduced cell count with NSP2 C480A. This demonstrates that the viral protease catalytic activity alone did not contribute to NSP2 cytotoxicity. The observed variation in cell proliferation and death caused by viral protein

expression underscores the importance of accurate cell counts for the downstream infection step.

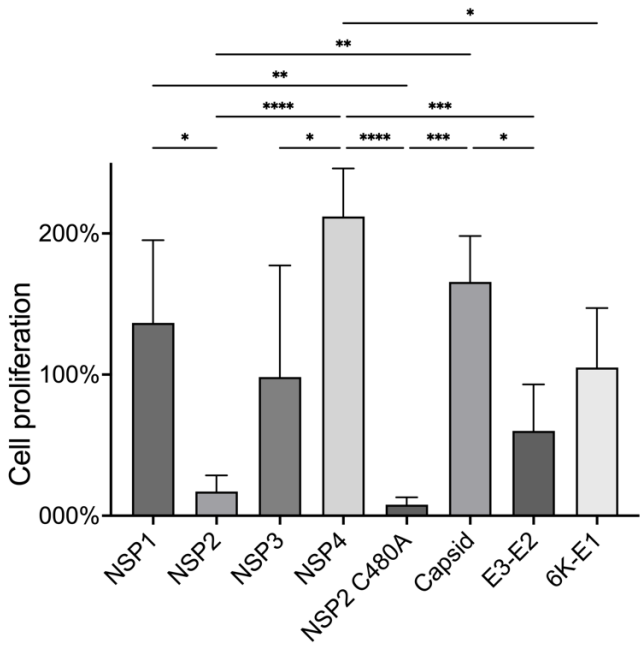


Figure 4.2. Cell proliferation and death post-electroporation. Average numbers of living cells were normalized to the cells electroporated with CMV-FLAG vector (n = 4). Error bars denote standard deviations (SDs). One-way ANOVA was performed (assuming Gaussian distribution and equal SDs), where the means were compared pairwise with Tukey corrections. *; $p \leq 0.05$, **; $p \leq 0.01$, ***; $p \leq 0.001$, ****; $p < 0.0001$. All other unplotted comparisons were not significant ($p > 0.05$).

4.4.3 Viral protein detection in infected input and IP

After HEK293T cells were transfected with FLAG-tagged viral protein constructs, recovered for 48 h, and then infected with MAYV for 10 h (MOI = 1) according to cell counts, the cells were lysed under gentle conditions. At this stage, it was critical to inactivate viruses while maintaining protein complexes. We opted for a lysis buffer containing 1% Triton X-100 and intermediate salt concentration (150 mM NaCl), and further used probe sonication at low amplitude (15%) for complete cell lysis. Aliquots of lysate input were taken before automated anti-FLAG AP-MS. Co-immunoprecipitated proteins were subjected to on-bead trypsin

digestion, and both input and IP were analyzed by mass spectrometry using DIA for more in-depth proteome coverage and improved mass accuracies.

We were able to detect >7000 and >3000 proteins for input and IP, respectively (**Fig. 4.3**), with the exception of NSP2 and NSP2 C580A that were lower than other viral proteins, which was consistent with cell counts (**Fig. 4.2**). This showed the extensive profiling of the transfected and infected global cell proteomes, and robust enrichment of the viral protein co-immunoprecipitation.

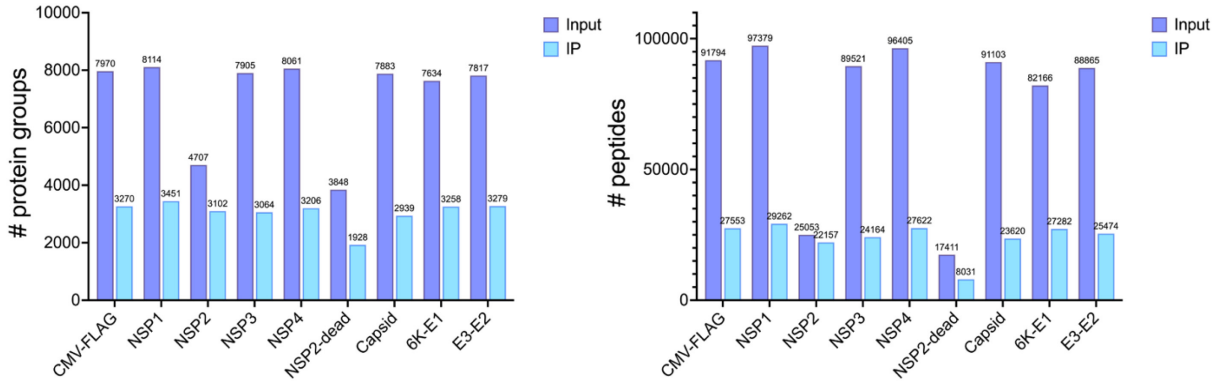


Figure 4.3. Data summary of protein groups and peptides detected in input and IP. IP samples showed an enrichment ratio at ~50% (n = 4).

Additionally, using input data, we can assess the efficiency of transfection and infection by monitoring the viral protein abundance levels. We used two parameters: MS2 quantity that reflects protein intensities, and the number of precursors corresponding spectral counts (**Fig. 4.4**). We saw high levels of viral protein expression and enrichment in both the input and IP.

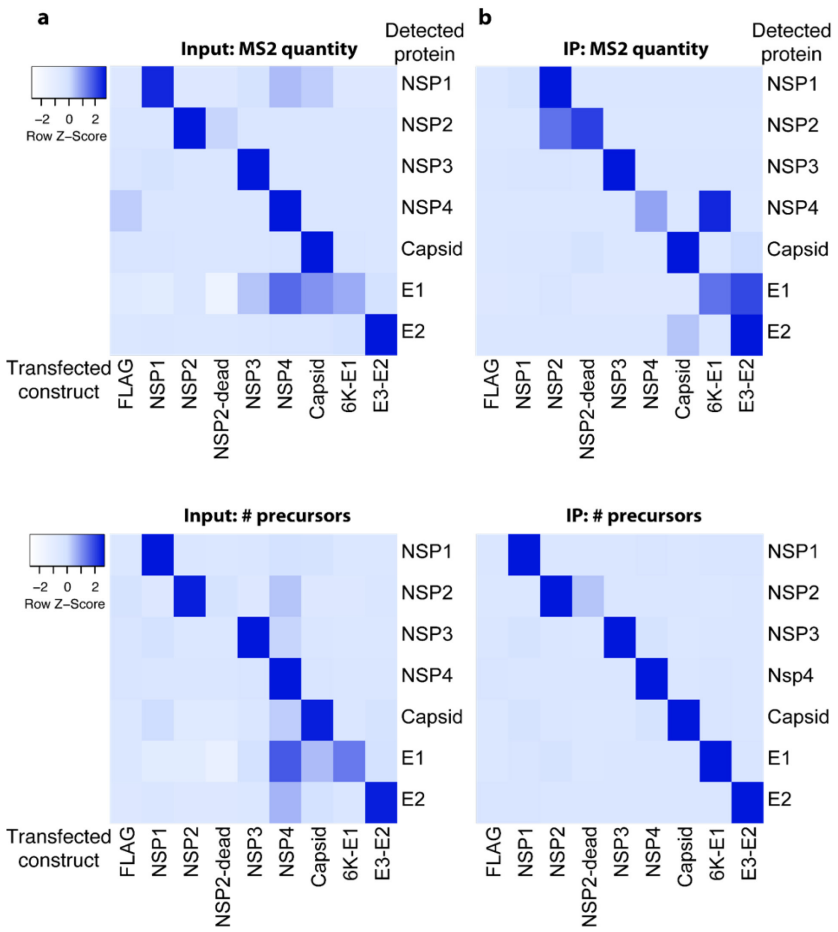


Figure 4.4. Heatmaps showing the MS2 quantity and number of precursors of the overexpressed viral protein in corresponding samples in both input and IP.
Overexpressed viral proteins searched using protein abundance and number of precursors were identified in the infected lysate input and particularly enriched in the IPs.

The overexpression of viral protein prior to infection can alter the viral replication rate, especially with structural proteins that boost viral entry to the cells. Interestingly, we saw strong association between E1 and NSP4, where we detected high levels of E1 in NSP4-transfected cell lysates (**Fig. 4.4a**), and high levels of NSP4 in 6K-E1 transfected IP (**Fig. 4.4b**; upper panel). This suggested that 1) the overexpression of the NSP4, RNA-dependent RNA polymerase, boosted viral replication; 2) E1 interacted with NSP4. Although the interaction between NSP4 and E1 has not been previously reported, all positive-strand RNA viruses replicate in association with the cytoplasmic membranes in infected cells, (240) and the RNA

replicases, such as MAYV NSP4, are anchored to the membrane and therefore can very likely interact with E1.

4.4.4 Scoring virus-host protein-protein interactions using MiST

To identify host protein interactors specific to viral proteins, we first used MiST, the most commonly used computational algorithm to score AP-MS data in virus-host PPIs. (141) We scored the dataset as a whole using default fixed weights, where all baits and the corresponding prey MS2 quantities were included. We then visualized the MiST scores using Cytoscape (**Fig. 4.5**). (241) This generated a narrow, restricted network including seven baits.

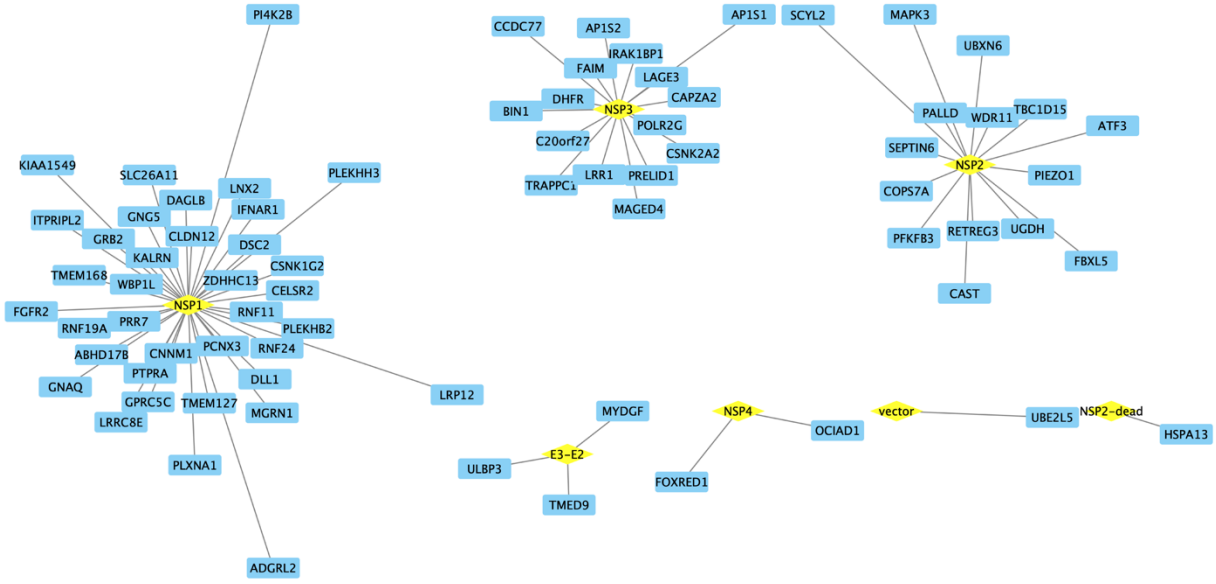


Figure 4.5. Complete MiST analysis of the MAYV protein interactome in infected cells. This analysis emphasized unique interactions specific to each viral protein. Blue rectangles: host proteins “preys”; yellow diamonds: viral proteins “baits”; edges diffused according to MiST scores (closer to the bait = more significant).

Of note, MiST does not define any conditions as negative controls. Using default fixed weights for prey specificity, reproducibility and abundances, the algorithm treats all conditions as equally independent experiments, and penalizes common interactions as nonspecific background proteins. While this is a robust strategy to filter for highly specific PPIs that are

unique to each bait protein, it is not best suited for our study. In our experiment, the cells were both transfected and infected, and the overexpressed affinity-tagged viral proteins were involved in the viral replicase formation, resulting in the potential identification of complex-bound proteins shared among viral proteins. Furthermore, multiple viral proteins can interact with the same key cellular protein to enhance a single function or to modulate different functions during infection. For instance, SARS-CoV-2 envelope protein interacts with bromodomain-containing protein 2 (BRD2) to increase the transcription of cellular receptors for viral entry, but the viral protease Mpro cleaves BRD2 to reduce cellular receptor expression at later stage of infection. (35, 171, 172, 242) Therefore, common virus-host PPIs remain biologically relevant and should be included in our analysis.

We then scored the dataset pairwise using MiST, where we compare each viral protein bait to the control vector exclusively, enabling a binary comparison between the IPs in experimental condition and the negative control (infected cell lysates overexpressing CMV-FLAG) (**Fig. 4.6**). We applied a more stringent filter for MiST score ≥ 0.9 to gate the number of interactions at higher confidence, and incorporated the prey abundances in the IP experiment by border thickness. Additionally, for each identified interactor, we looked for its corresponding protein abundance in the input lysates. The overexpression of viral proteins can induce proteomic changes especially in the context of viral infection. Thus, whether a viral protein interactor was elevated or depleted in abundance would further indicate its mechanistic role in viral pathobiology, and guide how we could modulate its level to counter infection. We coloured the prey nodes according to its protein abundance in input lysates in comparison to the vector control, where shades of blue denoted decreased abundances and orange indicated increased. We observed that the border thickness did not correspond to orange shading of the nodes, showing that the increased prey abundances in the IP could not be simply attributed to cellular proteomic changes after viral protein transfection, but resulted from virus-host interaction. In

fact, many high-confidence preys were decreased in protein abundance at the input level, suggesting that the viral protein interaction caused this decline.

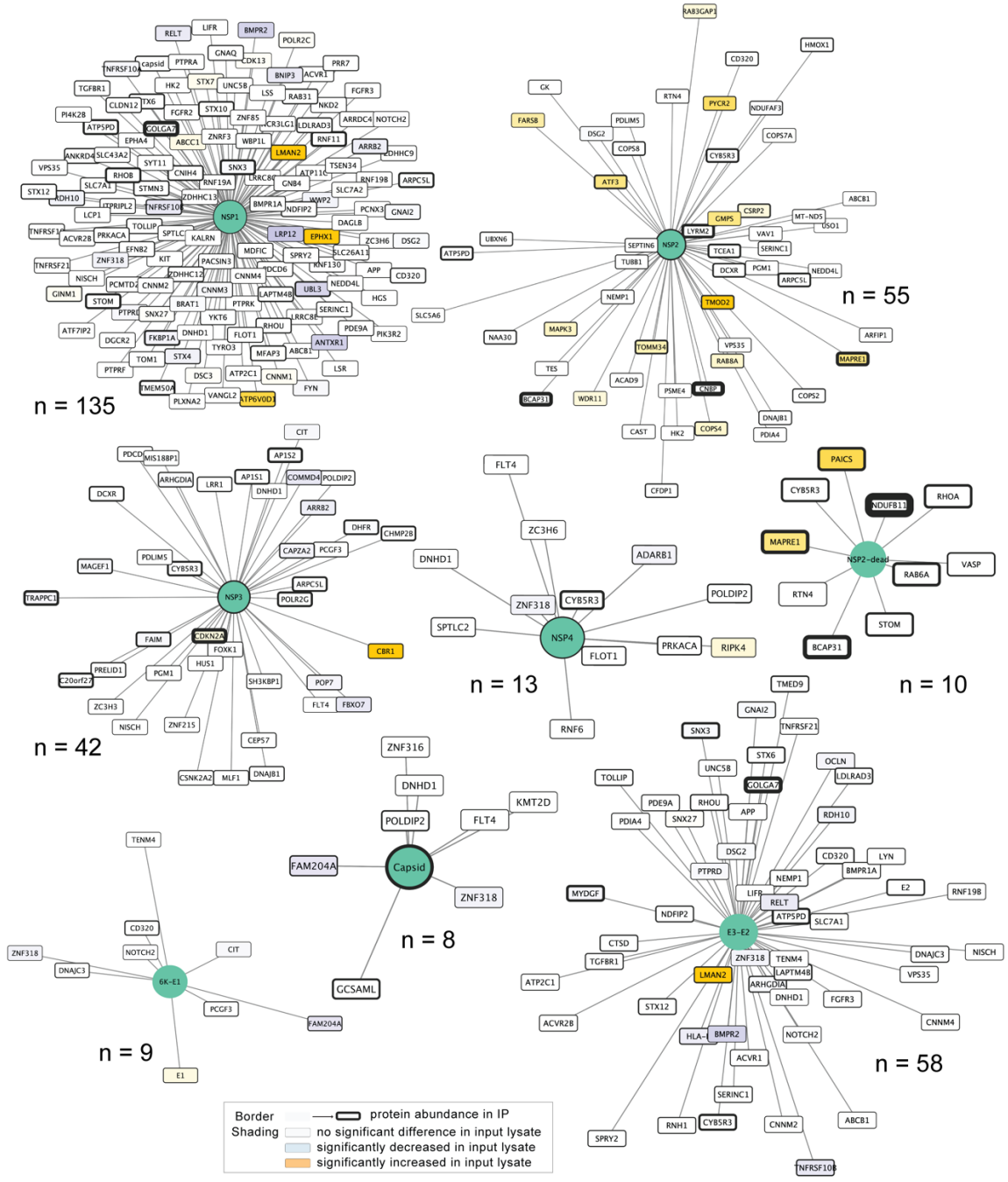


Figure 4.6. MiST analysis of individual MAYV bait proteins against control (MiST score ≥ 0.9). Cyan circles: viral proteins "baits"; rectangles: host proteins "preys", shaded from blue to orange = from decreased to increased in input lysates compared to infected cells transfected with control vector (CMV-FLAG); border thickness = abundance in IP. Edges diffused according to MiST scores (closer to the bait = more significant).

Although the binary MiST scoring approach created a more comprehensive visualization of MAYV-host PPI network, it did not illustrate the interconnectedness of viral protein interactomes. For example, neurogenic locus notch homolog protein 2 (NOTCH2), a type I transmembrane protein that interacted with Epstein-Barr viral protein, (243, 244) was found to interact with both MAYV transmembrane glycoproteins 6K-E1 and E3-E2 (**Fig. 4.7**). Attempts to visualize all the interactomes using Cytoscape resulted in a cluttered, convoluted network that was difficult to interpret (**Fig. 4.7**).

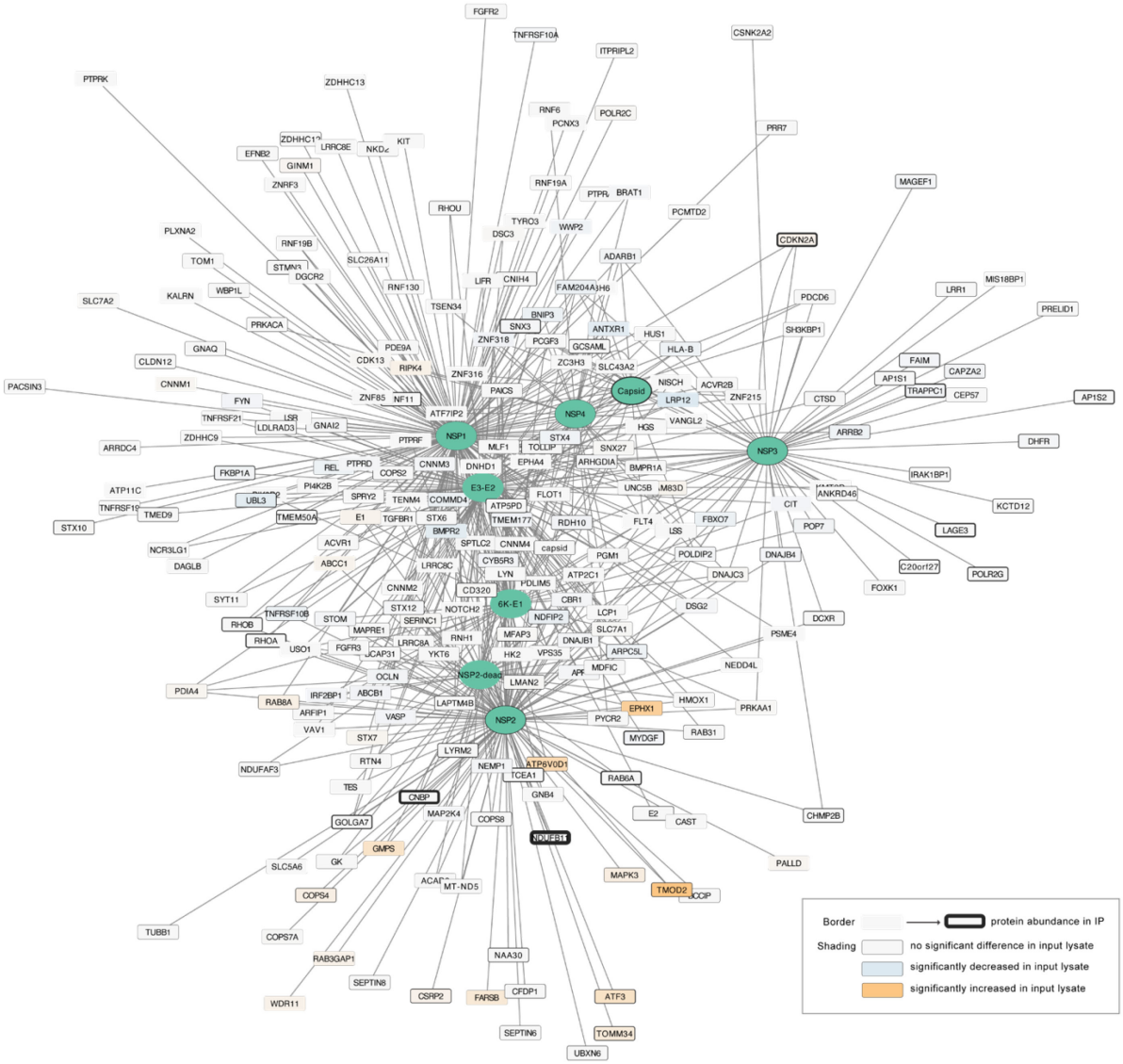


Figure 4.7. Merged MiST analysis of individual MAYV bait proteins against control (MiST score ≥ 0.9). Cyan circles: viral proteins “baits”; rectangles: host proteins “preys”, shaded from blue to orange = from decreased to increased in input lysates compared to infected cells transfected with control vector (CMV-FLAG); border thickness = abundance in IP. Edges diffused according to MiST scores (closer to the bait = more significant).

4.4.5 Scoring virus-host protein-protein interactions using SAINT

Alternatively to MiST, SAINT is also highly used for scoring PPIs in AP-MS and proximity labeling. (245) SAINT uses a statistical model to assign either true or false interactions based on explicitly defined negative control purifications, placing a stronger emphasis on high prey abundance relative to control. Its enhanced implementation, SAINTexpress, was able to accommodate common interactors at varying quantitative levels with different baits in a single integrated model. (146)

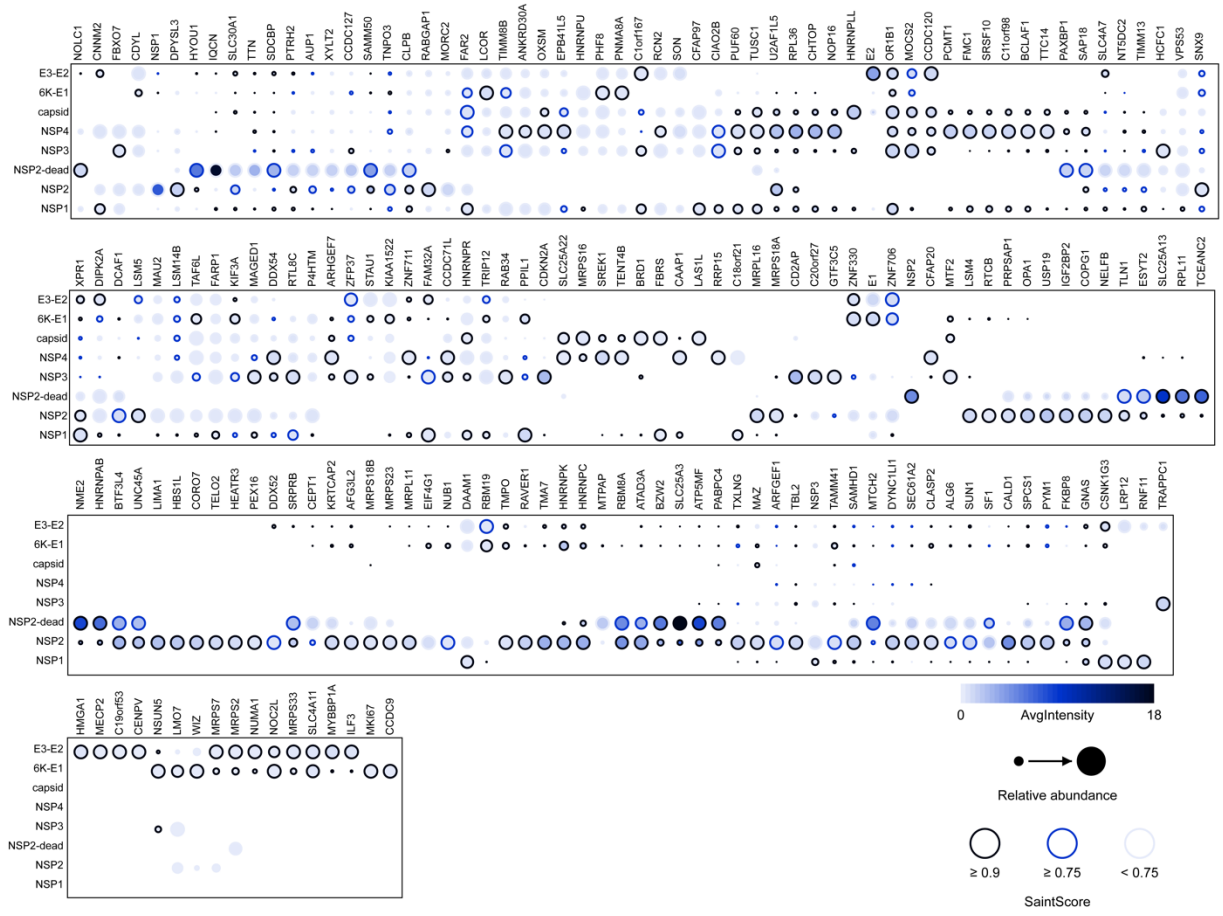


Figure 4.8. SAINT analysis of the MAYV viral protein interactome. Dot plot showing the enrichment of viral protein interactors over FLAG control, scored using SAINT analysis boosted

with reported interactions in Gene Ontology annotations. Three perspectives were represented: raw prey abundances (average intensity) shown by the dot colour scaling from blue to black; relative prey abundances across baits shown by the dot size; prey confidence indicated by the dot outline (with SaintScore ≥ 0.75 considered significant). Virus-host protein-protein interactions in infected cells show clustered patterns specific for NSP2 (Row 3) and Envelope proteins (Row 4).

We used SAINTexpress to score MAYV-host PPIs based on peptide fragment intensities measured by data-independent acquisition, and visualized the results using dot plot (**Fig. 4.8**). (238) This analysis highlighted many common interactions shared among viral proteins. This observation was consistent with the mechanisms of viral protein processing and in MAYV replication, where viral polyproteins are sequentially cleaved, and mature viral proteins form stable complexes in infected cells (reviewed in Chapter 1.2.6). Particularly, interactors could be grouped into distinct clusters on the dot plot, that were specific to envelope proteins (Row 4), NSP2 (Row 3), and others shared among most viral proteins. There were proteins unique to NSP2 WT and those that were more abundant in NSP2-dead, which possibly resulted from interactions due to the NSP2 protease catalytic activity.

4.4.6 Interactor enrichment analysis

We next evaluated the subcellular localization of the MAYV viral protein interactors. To do this, we used the lists of interactors generated from the binary MiST analysis (**Fig. 4.6**), and lowered the minimum MiST scores to 0.75. The expanded interaction networks encapsulated all potential viral protein interactors, and allowed a more comprehensive analysis of the viral protein interactomes. The viral protein interactors demonstrated different subcellular localization profiles (**Fig. 4.9**). For instance, ER was enriched in E3-E2 interactors, which was expected as the glycoproteins were synthesized and processed in the ER (**Fig. 1.9**). Cytoplasm was also highly enriched in the localization of NSP2-, 3- and 4-binding proteins, but more proteins were membrane-associated for NSP1 interactors, suggesting the unique involvement of NSP1 in

membrane protein signaling. Furthermore, both NSP2 and NSP2 mutant interactors showed highly similar compartmentalization, despite differences in the composition of their interactomes.

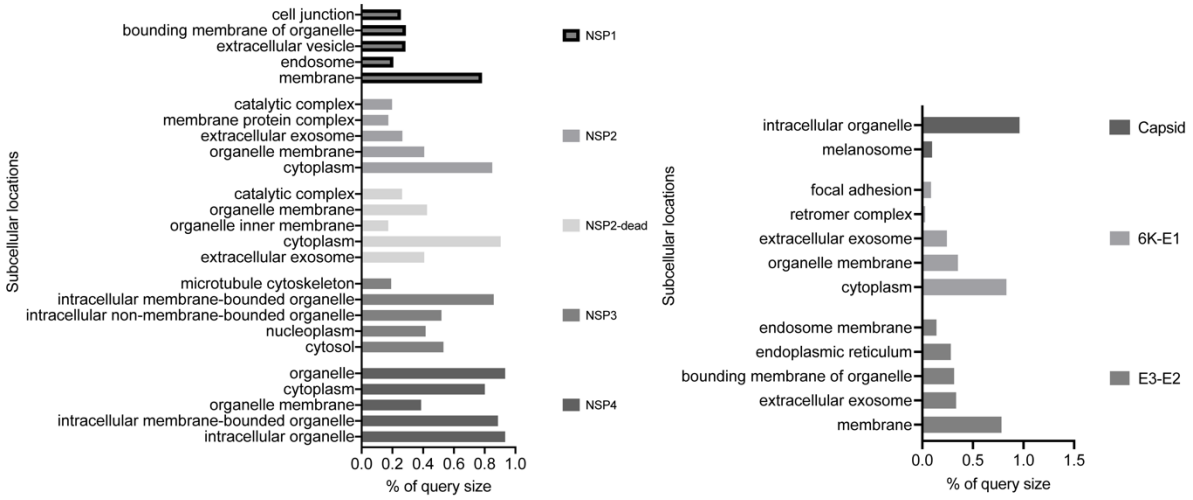


Figure 4.9. Top five enriched subcellular locations of MAYV protein interactors (p-value < 0.05)**.** Enrichment analysis was performed using the stringAPP on the Cytoscape.

Many of the enriched biological processes of the viral protein interactors corresponded to the subcellular locations (**Fig. 4.10**). For instance, NSP1 interactors were both membrane-associated and uniquely involved in the cell surface receptor signaling pathway. We also observed the enrichment in oxidative phosphorylation only in NSP2-dead but not NSP2, which indicated the protease substrates might be involved in MAYV infection-induced oxidative stress. (246) Interestingly, both NSP3 and capsid interactomes exhibited strong enrichment in subcellular locations, yet were weakly associated in biological processes. These analysis tools could confirm the proper subcellular localization of the overexpressed protein constructs, and strengthen our understanding of roles of the MAYV proteins during infection.

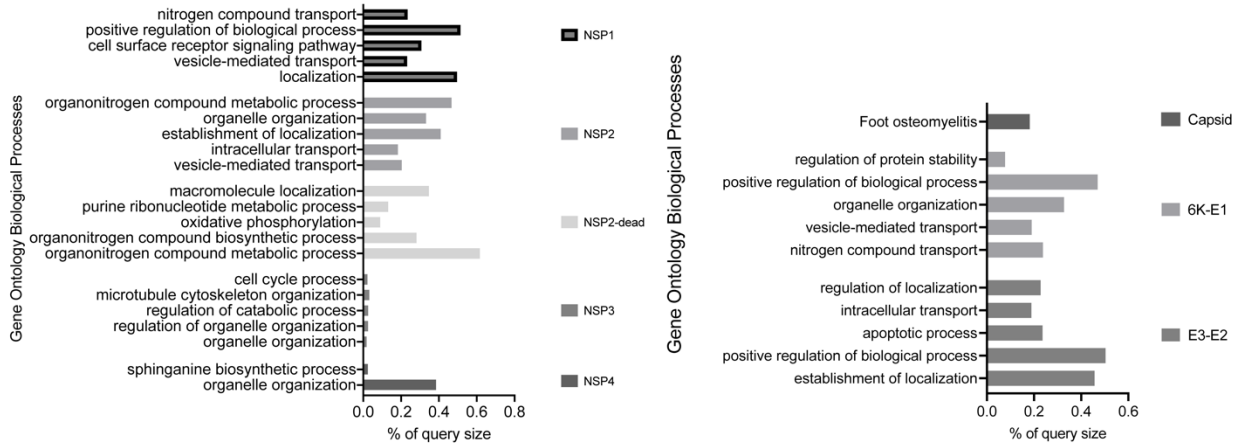


Figure 4.10. Top five enriched biological processes of MAYV protein interactors (p-value < 0.05)**.** Enrichment analysis was performed using the stringAPP on the Cytoscape.

4.4.7 Integrating SAINT and MiST scores

Using various strategies to score and analyze the identified virus-host PPIs indeed expanded the interactome landscape from different perspectives. The next step was to focus on key interactors with high confidence across multiple scoring algorithms. Therefore, we integrated SAINT and binary MiST scores of the MAYV interactome, by setting both scores to a minimum of 0.75, and searching for the overlap between the two pools of putative interactors. We then searched for their protein abundances at the input level and visualized the result using Cytoscape (**Fig. 4.11**). This resulted in a rather simplified but interconnected network that can direct immediate targets for follow-up studies.

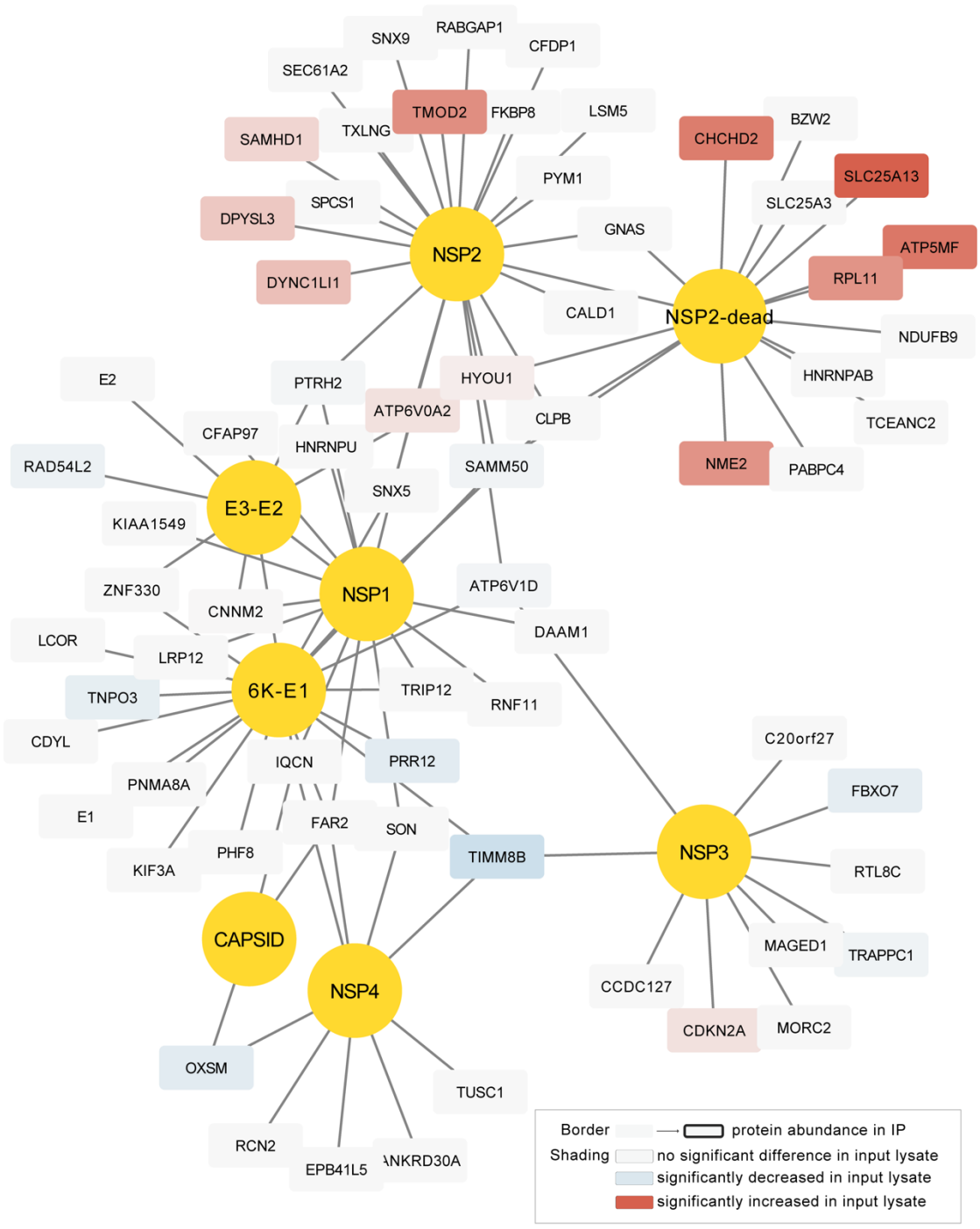


Figure 4.11. MAYV-host interaction networks with proteins scoring ≥ 0.75 in both binary MiST and SAINT analysis. Yellow circles: viral protein baits; rectangles: host proteins preys, shaded from blue to red = from decreased to increased in input lysates compared to infected cells transfected with control vector (CMV-FLAG). Edges diffused according to MiST scores (closer to the bait = more significant).

4.5 Discussion

4.5.1 Limitations of the study

The interaction of the tagged viral proteins with viral proteins produced during infection, requires the ability to form complexes in trans. We acknowledge that this may not be the case for all proteins. We assumed that the overexpressed viral protein participated in process of viral replication despite being translated from different mRNA molecules (trans-interactions). Overexpressing one viral protein during infection could also lead to potential changes in viral replication, such as the increased viral E1 abundance observed in infected cells with NSP4 overexpression (**Fig. 4.4**), suggesting higher viral titre due to increased level of NSP4 replicase. This altered stoichiometry could introduce artifactual PPIs that do not occur in native viral infections. Furthermore, to allow cell recovery from electroporation, the viral protein was constitutively expressed for 48 h before infection, where many PPIs between the bait and cellular proteins might have already taken place. We assumed that the later infection could induce changes in the PPIs significant for viral infection and replication, which would be captured in the subsequent anti-FLAG pull down. For instance, the envelope proteins E1 and E2 are only released from the ER upon heterodimerization (reviewed in Chapter 1.2.7). We designed the protein constructs in the study for this to only occur with infection, and the overexpressed envelope proteins are otherwise retained in the ER. While AP-MS in infected cells has certain limitations that require further evaluation, it marks a considerable step forward as a virus-host interactomics screening approach compared to conventional AP-MS.

4.5.2 Infection significantly alters viral protein interactome

Our group has previously conducted AP-MS with MAYV proteins under conventional uninfected cell conditions, also scored using binary MiST analysis. (234) To evaluate PPIs only induced by infection, we compared the viral protein interactors, both scored by MiST binary

analysis with a threshold of 0.75, with and without infection. Strikingly, we observed very limited overlap, ranging from no to three common hits. The only exception was NSP2-dead IP, which contained 34 interactors that were also found in the NSP2 IP previously carried out without infection. This might indicate that viral infection significantly altered the PPI landscape, and that the NSP2 protease catalytic activity might lead to a distinct interactome during infection. Specifically, although the previously identified NSP2 interactors DNA-directed RNA polymerase II subunit A (Rpb1) and transcription initiation factor IIE subunit 2 (TFIIE2) were detected in our experiment, they were not significantly enriched in the NSP2 IP over infected control. Notably, the interaction between NSP2 and the two host proteins was found in AP-MS without viral infection, and it was later characterized that MAYV NSP2 depletes Rpb1and TFIIE2 during viral infection. (234) Therefore, our experimental conditions with increased NSP2 expression under viral infection would further reduce the possibility of detecting Rpb1and TFIIE2, especially since our control IP was also performed in infected lysates. This could be a limitation of our study: viral infection can induce PPIs, but depending on the role of the proteins, their detection in AP-MS is not necessarily always enhanced.

4.5.3 Alternatives for AP-MS in infection

In this study, we performed AP-MS in infected cells to capture PPIs in viral infections. However, factors such as scalability and accessibility of the proper biosafety level facilities to work with different viruses can place limitations for adapting this approach. In addition, incorporating the infection step can be a labour-intensive and time-consuming bottleneck in a typically high-throughput AP-MS workflow, especially when studying viruses with a large number of viral proteins (e.g., poxviruses encoding > 200 proteins). Alternatively, cells can be treated with infection-like stimuli, such as cytokines. For instance, the Overall's group treated human lung epithelial cells with interferon α and β before lysis to examine antiviral response-

related substrates for SARS-CoV-2 Mpro *in vitro*. (63) This could serve as a cost- and resource-effective option for AP-MS or other *in vitro* studies.

4.5.4 Protein interactions shared by multiple viral proteins

Most AP-MS studies focused the analysis on host proteins that uniquely interact with each viral protein. Indeed, this is a good strategy to remove non-specific binders, false positives or artifacts in the dataset, as increased sensitivity in mass spectrometry has enabled thousands of protein identifications in a single co-immunoprecipitation run, making the selection process for follow-up candidates difficult. However, this filter also can be too harsh and can neglect that multiple viral proteins can interact with one key host factor, to regulate diverse functions or fine-tune a specific process. Such phenomenon was demonstrated in our SARS-CoV-2 study in Chapter 2, where BRD2 interacted with the viral Envelope protein to facilitate viral entry, (35) but was also cleaved by the viral protease Mpro. Investigating exclusive interactors alone will inevitably miss critical host factors and lead to misinterpretation of virus-host dynamics.

4.5.5 Further characterization of protein-protein interactions

Some of the ongoing and future work includes functional studies on the viral protein interactor candidates. For instance, interferon alpha and beta receptor subunit 1 (IFNAR1) is an antiviral factor that recruits STAT2 and activates interferon signaling induced by viral infection. (247, 248) In our study, the cytoplasmic tail of IFNAR1 uniquely interacted with MAYV NSP1. This was in line with our interactome analysis where NSP1 interactors were highly enriched in cell surface receptor signaling (**Fig. 4.10**), suggesting MAYV NSP1 may be highly involved in critical receptor interactions including IFNAR1.

Other interesting targets include those that are differentially interacting with NSP2 and NSP2-dead, which can be NSP2 protease substrates. In conjunction with the *in vitro* N-terminomics screening strategy presented in Chapter 2 and protease purification in Appendix D,

we can generate a list of viral protease substrates whose cleavage sites are labeled and interactions with the proteases are shown during infection. This will accelerate the identification of *bona fide* proteolysis targets that can potentially reveal novel therapeutic targets for intervention.

Chapter 5: Overall discussions and conclusions

5.1 Synopsis

Protein post-translational modifications (PTMs) and protein-protein interactions (PPIs) undergo significant changes during all stages of viral infection and replication. My thesis work aimed to unravel the intricate molecular mechanisms of viral infections by studying PTMs and PPIs using proteomics techniques. Our research has identified the proteolysis targets of SARS-CoV-2 proteases, profiled protein ubiquitination in VACV infection, and characterized the virus-host interaction networks in MAYV infection. Taken together, this thesis highlighted the various mass spectrometry approaches in illustrating the interplay in virus-host interactions, providing insights into the mechanisms of viral manipulation of host cellular processes and identifying putative targets for antiviral therapies. This work provided a basis for the potential repurposing of drugs or new therapeutics targeting either viral or host proteins.

5.2 Challenges in mapping PTMs in viral infection

PTM characterization is challenging. PTMs are highly dynamic and regulated, with many of them being reversible and low-abundance. It often requires the use of PTM-specific enrichment methods to effectively isolate the modifications for identification that would otherwise be undetectable by standard shotgun proteomic approaches. Moreover, the PTM location assignment is dependent on the peptide fragmentation, where a poor MS2 pattern could cause misassignment or failure in localization. (249) The peptide-spectrum assignment in LC-MS/MS analysis softwares for highly modified peptides is also more convoluted and can lead to identification of false positives, especially in data-independent acquisition that is gaining popularity in label-free protein quantifications (reviewed in (250)). Steps in sample processing can often produce artifacts that closely resemble PTM mass shifts. (249) For instance, overalkylation of lysines by iodoacetamide results in a covalent adduct that is identical to the

mass of diGlycyl ubiquitin remnant. (251) Meanwhile, true modifications, such as phosphate groups, are labile and can be easily lost in both sample preparation and phosphorylated peptide fragmentation in LC-MS/MS (Appendix C). (252)

Additionally, for most proteins, many sequences are simply inaccessible by bottom-up proteomics using common digestive enzymes (i.e. trypsin), or do not ionize/mobilize/fragment well, causing incomplete PTM mapping. (253) PTMs further amplify this inaccessibility, by potentially blocking cleavage by the digestive enzymes. Due to the complexity of modified proteoforms, multiple digestive enzymes are required for deep proteome profiling, (254) or additional strategies such as top- or middle-down proteomics need to be integrated to study all proteoforms of a target protein. (40, 255)

Viral infection adds another layer of complexity. Our studies in Chapters 2 and 3 exemplified that PTMs on a protein could vary as the viral infection progressed. BRD2 was required for SARS-CoV-2 entry, (171) but was cleaved by the viral protease during infection. In VACV infection, TRIM25 was highly ubiquitinated as early as 1 h.p.i., but its ubiquitinated forms were significantly reduced at 24 h.p.i. (**Fig. 3.6a**). Thus, infection time is another variable that affects the PTM landscape. Currently, we are uncertain whether the proteomic changes at an earlier infection time point fully represent the immediate antiviral responses, or the stable, persistent protein modifications fundamentally define the virus cycle. Since proteomic analysis provides only a snapshot of the dynamic interactions between cellular machinery and an invading pathogen, determining the optimal timing and frequency of sampling is crucial. Additionally, PTMs can vary depending on cell type, viral strain, and MOI. These factors should be carefully considered when investigating PTMs in virology studies.

Recent advances in single-cell transcriptomics and proteomics allow to capture proteomic changes by zooming into single virus-infected cells, eliminating the bulk analysis of a

heterogeneous cell population with various cellular and infection states. (256) It also permits the use of rare cell lines that correspond to viral tropism, such as the human primary cells, particularly in the cases where cellular responses to viral infections are dependent on cell types. However, while single-cell research may represent a new tool for virology studies, it will likely not enhance the identification of virus-induced PTMs at its current state, as PTMs are found in even more minute amounts when analyzed at the single-cell level. The development of more advanced methodologies are necessary to overcome these challenges and fully realize the potential of single-cell virology research.

5.3 Challenges in characterizing PPIs in viral infection

In Chapter 4, we proposed that performing AP-MS experiments in infected cells provided a better mimic of PPIs in viral infections. It also led to the identification of many viral protein common interactors. Current PPI scoring algorithms do not fully support the analysis of such interconnected PPI networks, and deciphering between common interactors or non-specific binders often entirely relies on the negative control. Furthermore, we employed anti-FLAG AP-MS for PPI identification, which has a smaller affinity tag (1 kDa) than popular alternatives such as using proximity labeling enzymes (Appendix F). However, anti-FLAG AP-MS can often miss transient interactions. On the other hand, while proximity labeling allows the use of harsh lysis conditions for membrane protein identification and virus inactivation, the larger fusion protein (BirA*) is more likely to interfere with PPIs. This can be particularly concerning if viral infection is implemented in the study, where viral proteins are processed from polyprotein precursor and BirA* could hinder the replication complex formation. That said, new smaller BirA* variants are being designed, such as MicroID2 with a size of 19 kDa and fast labeling time (1h). (257)

5.4 Crosstalk between PTMs and PPIs – what are we missing?

An interesting aspect that we have not fully explored in this thesis was the connection between PTMs and PPIs. We know that protein PTMs can alter PPIs, (258) and that is how they mediate protein signaling. However, we did not specifically quantify changes in the PTM of a viral protein binding prey, or characterize the interactome of a highly modified protein in viral infection, such as TRIM25. Mass spectrometry advances have enabled the sensitive, high-throughput and robust proteomic quantification of cells, tissues, bodily fluids, and now even in a single cell, but what we are potentially missing is the functional knowledge of all these protein levels. To what extent do the low-abundance protein decorations alter its interactions with other proteins? Moverover, PTMs also impact other PTMs. (259) As discussed in Chapter 3, VACV encodes SCF E3 ligase adaptors ANK-PRANC proteins, and we found these viral adaptor proteins to be ubiquitinated (**Table 3.2**). However, only phosphorylated proteins are recruited to the SCF ubiquitin-ligase complex. (77) Performing comparative analysis on phosphoproteomics and ubiquitomics in VACV infected cells may uncover unexpected findings such as phosphorylated and ubiquitinated viral ANK-PRANC substrates. Lastly, we assumed a single abundance per protein encoding gene in our AP-MS study, originating from the sum of the corresponding peptides abundances (MS2 peak area). This is a common practice in interactomics studies but perhaps neglects the variety of PTM-modified species, (260) which was estimated to be 100 proteoforms per protein encoding gene. (39) Each detected peptide may originate from different proteoforms, which may have varying abundances and carry diverse functions. Although current technology does not allow us to easily distinguish among the proteoforms and explore the proteome to such depths, considering all these factors is a step toward creating a more diverse network that better represents the virus-host interaction dynamics. Finally, while I have not address this in my thesis, I would expect that artificial intelligence tools will become more frequently used in the short future in proteomics research, to

study and analyze post-translational modifications and protein-protein interaction networks. This could be revolutionary, just like AlphaFold has recently transformed the field of structural biology. (261)

References

1. Hui, E. K. Reasons for the increase in emerging and re-emerging viral infectious diseases. *Microbes Infect* **2006**, *8* (3): 905-916. DOI: 10.1016/j.micinf.2005.06.032
2. Baker, R. E. et al. Infectious disease in an era of global change. *Nat Rev Microbiol* **2022**, *20* (4): 193-205. DOI: 10.1038/s41579-021-00639-z
3. Stanley, W. M.; Loring, H. S. The Isolation of Crystalline Tobacco Mosaic Virus Protein from Diseased Tomato Plants. *Science* **1936**, *83* (2143): 85. DOI: 10.1126/science.83.2143.85
4. Stanley, W. M.; Lauffer, M. A. DISINTEGRATION OF TOBACCO MOSAIC VIRUS IN UREA SOLUTIONS. *Science* **1939**, *89* (2311): 345-347. DOI: 10.1126/science.89.2311.345
5. Wang, D. et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* **2020**, *323* (11): 1061-1069. DOI: 10.1001/jama.2020.1585
6. Zhu, N. et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* **2020**, *382* (8): 727-733. DOI: 10.1056/NEJMoa2001017
7. Liu, J.; Xie, W.; Wang, Y.; Xiong, Y.; Chen, S.; Han, J.; Wu, Q. A comparative overview of COVID-19, MERS and SARS: Review article. *Int J Surg* **2020**, *81* 1-8. DOI: 10.1016/j.ijsu.2020.07.032
8. Hsieh, C. L. et al. Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. *Science* **2020**, *369* (6510): 1501-1505. DOI: 10.1126/science.abd0826
9. Baden, L. R. et al. Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl J Med* **2021**, *384* (5): 403-416. DOI: 10.1056/NEJMoa2035389
10. Polack, F. P. et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med* **2020**, *383* (27): 2603-2615. DOI: 10.1056/NEJMoa2034577
11. Sadoff, J. et al. Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19. *N Engl J Med* **2021**, *384* (23): 2187-2201. DOI: 10.1056/NEJMoa2101544
12. Heath, P. T. et al. Safety and Efficacy of NVX-CoV2373 Covid-19 Vaccine. *N Engl J Med* **2021**, *385* (13): 1172-1183. DOI: 10.1056/NEJMoa2107659
13. Tulman, E. R.; Delhon, G.; Afonso, C. L.; Lu, Z.; Zsak, L.; Sandybaev, N. T.; Kerembekova, U. Z.; Zaitsev, V. L.; Kutish, G. F.; Rock, D. L. Genome of horsepox virus. *J Virol* **2006**, *80* (18): 9244-9258. DOI: 10.1128/JVI.00945-06
14. Qin, L.; Favis, N.; Famulski, J.; Evans, D. H. Evolution of and evolutionary relationships between extant vaccinia virus strains. *J Virol* **2015**, *89* (3): 1809-1824. DOI: 10.1128/JVI.02797-14
15. Fenner, F.; D. A. Henderson; I. Arita; Z. Jezek; I. D. Ladnyi; H. O. World. 1988. *Smallpox and its eradication / F. Fenner. [et al.]*. World Health Organization.
16. Downie, A. W. The immunological relationship of the virus of spontaneous cowpox to vaccinia virus. *British Journal of Experimental Pathology* **1939**, *20* (2): 158. DOI:
17. G.L., S. 2008. Vaccinia Virus Encyclopedia of Virology (Third Edition). 243-250. Oxford: Academic Press.
18. Kim, J. H. et al. Systemic armed oncolytic and immunologic therapy for cancer with JX-594, a targeted poxvirus expressing GM-CSF. *Mol Ther* **2006**, *14* (3): 361-370. DOI: 10.1016/j.ymthe.2006.05.008
19. Guo, Z. S. et al. Vaccinia virus-mediated cancer immunotherapy: cancer vaccines and oncolytics. *J Immunother Cancer* **2019**, *7* (1): 6. DOI: 10.1186/s40425-018-0495-7
20. Moss, B. Regulation of vaccinia virus transcription. *Annu Rev Biochem* **1990**, *59* 661-688. DOI: 10.1146/annurev.bi.59.070190.003305

21. Hernandez-Gonzalez, M.; Calcraft, T.; Nans, A.; Rosenthal, P. B.; Way, M. A succession of two viral lattices drives vaccinia virus assembly. *PLoS Biol* **2023**, *21* (3): e3002005. DOI: 10.1371/journal.pbio.3002005
22. Shchelkunov, S. N.; Shchelkunova, G. A. Genes that Control Vaccinia Virus Immunogenicity. *Acta Naturae* **2020**, *12* (1): 33-41. DOI: 10.32607/actanaturae.10935
23. Smith, G. L.; Vanderplasschen, A.; Law, M. The formation and function of extracellular enveloped vaccinia virus. *J Gen Virol* **2002**, *83* (Pt 12): 2915-2931. DOI: 10.1099/0022-1317-83-12-2915
24. Caraballo, H.; King, K. Emergency department management of mosquito-borne illness: malaria, dengue, and West Nile virus. *Emerg Med Pract* **2014**, *16* (5): 1-23; quiz 23. DOI:
25. Mattar, S.; Morales, V.; Cassab, A.; Rodríguez-Morales, A. J. Effect of climate variables on dengue incidence in a tropical Caribbean municipality of Colombia, Cerete, 2003-2008. *Int J Infect Dis* **2013**, *17* (5): e358-9. DOI: 10.1016/j.ijid.2012.11.021
26. Rodriguez-Morales, A. J.; Patiño-Cadavid, L. J.; Lozada-Riascos, C. O.; Villamil-Gómez, W. E. Mapping Zika in municipalities of one coastal department of Colombia (Sucre) using geographic information systems during the 2015-2016 outbreak: implications for public health and travel advice. *Int J Infect Dis* **2016**, *48* 70-72. DOI: 10.1016/j.ijid.2016.05.012
27. Paniz-Mondolfi, A. E.; Rodriguez-Morales, A. J.; Blohm, G.; Marquez, M.; Villamil-Gomez, W. E. ChikDenMaZika Syndrome: the challenge of diagnosing arboviral infections in the midst of concurrent epidemics. *Ann Clin Microbiol Antimicrob* **2016**, *15* (1): 42. DOI: 10.1186/s12941-016-0157-x
28. Hoch, A. L.; Peterson, N. E.; LeDuc, J. W.; Pinheiro, F. P. An outbreak of Mayaro virus disease in Belterra, Brazil. III. Entomological and ecological studies. *Am J Trop Med Hyg* **1981**, *30* (3): 689-698. DOI: 10.4269/ajtmh.1981.30.689
29. Brustolin, M.; Bartholomeeusen, K.; Rezende, T.; Ariën, K. K.; Müller, R. Mayaro virus, a potential threat for Europe: vector competence of autochthonous vector species. *Parasit Vectors* **2024**, *17* (1): 200. DOI: 10.1186/s13071-024-06293-7
30. Nurk, S. et al. The complete sequence of a human genome. *Science* **2022**, *376* (6588): 44-53. DOI: 10.1126/science.abj6987
31. Vidal, C. J. 2010. *Post-Translational Modifications in Health and Disease*. Springer.
32. Gonzalez, M. W.; Kann, M. G. Chapter 4: Protein interactions and disease. *PLoS Comput Biol* **2012**, *8* (12): e1002819. DOI: 10.1371/journal.pcbi.1002819
33. Uetz, P. et al. Herpesviral protein networks and their interaction with the human proteome. *Science* **2006**, *311* (5758): 239-242. DOI: 10.1126/science.1116804
34. Jäger, S. et al. Global landscape of HIV-human protein complexes. *Nature* **2011**, *481* (7381): 365-370. DOI: 10.1038/nature10719
35. Gordon, D. E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020**, *583* (7816): 459-468. DOI: 10.1038/s41586-020-2286-9
36. Stukalov, A. et al. Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature* **2021**, *594* (7862): 246-252. DOI: 10.1038/s41586-021-03493-4
37. Prabakaran, S.; Lippens, G.; Steen, H.; Gunawardena, J. Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip Rev Syst Biol Med* **2012**, *4* (6): 565-583. DOI: 10.1002/wsbm.1185
38. Keenan, E. K.; Zachman, D. K.; Hirschey, M. D. Discovering the landscape of protein modifications. *Mol Cell* **2021**, *81* (9): 1868-1878. DOI: 10.1016/j.molcel.2021.03.015
39. Aebersold, R. et al. How many human proteoforms are there. *Nat Chem Biol* **2018**, *14* (3): 206-214. DOI: 10.1038/nchembio.2576
40. Sidoli, S.; Lin, S.; Karch, K. R.; Garcia, B. A. Bottom-up and middle-down proteomics have comparable accuracies in defining histone post-translational modification relative

abundance and stoichiometry. *Anal Chem* **2015**, *87* (6): 3129-3133. DOI: 10.1021/acs.analchem.5b00072
41. Radzicka, A.; Wolfenden, R. Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases. *J. Am. Chem. Soc.* **1996**, *118* (26): 6105-6109. DOI: 10.1021/ja954077c
42. Wang, L.; Main, K.; Wang, H.; Julien, O.; Dufour, A. Biochemical Tools for Tracking Proteolysis. *J Proteome Res* **2021**, *20* (12): 5264-5279. DOI: 10.1021/acs.jproteome.1c00289
43. Schechter, I.; Berger, A. On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications* **1967**, *27* (2): 157-162. DOI:
44. López-Otín, C.; Bond, J. S. Proteases: multifunctional enzymes in life and disease. *J Biol Chem* **2008**, *283* (45): 30433-30437. DOI: 10.1074/jbc.R800035200
45. Tözsér, J.; Tropea, J. E.; Cherry, S.; Bagossi, P.; Copeland, T. D.; Wlodawer, A.; Waugh, D. S. Comparison of the substrate specificity of two potyvirus proteases. *FEBS J* **2005**, *272* (2): 514-523. DOI: 10.1111/j.1742-4658.2004.04493.x
46. Harris, J. L.; Backes, B. J.; Leonetti, F.; Mahrus, S.; Ellman, J. A.; Craik, C. S. Rapid and general profiling of protease specificity by using combinatorial fluorogenic substrate libraries. *Proc Natl Acad Sci U S A* **2000**, *97* (14): 7754-7759. DOI: 10.1073/pnas.140132697
47. O'Donoghue, A. J. et al. Global identification of peptidase specificity by multiplex substrate profiling. *Nat Methods* **2012**, *9* (11): 1095-1100. DOI: 10.1038/nmeth.2182
48. Schilling, O.; Overall, C. M. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol* **2008**, *26* (6): 685-694. DOI: 10.1038/nbt1408
49. Varshavsky, A. The N-end rule pathway of protein degradation. *Genes Cells* **1997**, *2* (1): 13-28. DOI: 10.1046/j.1365-2443.1997.1020301.x
50. López-Otín, C.; Overall, C. M. Protease degradomics: a new challenge for proteomics. *Nat Rev Mol Cell Biol* **2002**, *3* (7): 509-519. DOI: 10.1038/nrm858
51. Klein, T.; Eckhard, U.; Dufour, A.; Solis, N.; Overall, C. M. Proteolytic Cleavage-Mechanisms, Function, and "Omic" Approaches for a Near-Ubiquitous Posttranslational Modification. *Chem Rev* **2018**, *118* (3): 1137-1168. DOI: 10.1021/acs.chemrev.7b00120
52. Osipiuk, J. et al. Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nat Commun* **2021**, *12* (1): 743. DOI: 10.1038/s41467-021-21060-3
53. Harcourt, B. H.; Jukneliene, D.; Kanjanahaluethai, A.; Bechill, J.; Severson, K. M.; Smith, C. M.; Rota, P. A.; Baker, S. C. Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. *J Virol* **2004**, *78* (24): 13600-13612. DOI: 10.1128/JVI.78.24.13600-13612.2004
54. Han, Y. S.; Chang, G. G.; Juo, C. G.; Lee, H. J.; Yeh, S. H.; Hsu, J. T.; Chen, X. Papain-like protease 2 (PLP2) from severe acute respiratory syndrome coronavirus (SARS-CoV): expression, purification, characterization, and inhibition. *Biochemistry* **2005**, *44* (30): 10349-10359. DOI: 10.1021/bi0504761
55. Shin, D. et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **2020**, *587* (7835): 657-662. DOI: 10.1038/s41586-020-2601-5
56. Ratia, K. et al. A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. *Proc Natl Acad Sci U S A* **2008**, *105* (42): 16119-16124. DOI: 10.1073/pnas.0805240105
57. Fu, Z. et al. The complex structure of GRL0617 and SARS-CoV-2 PLpro reveals a hot spot for antiviral drug discovery. *Nat Commun* **2021**, *12* (1): 488. DOI: 10.1038/s41467-020-20718-8

58. Zhang, S.; Wang, J.; Cheng, G. Protease cleavage of RNF20 facilitates coronavirus replication via stabilization of SREBP1. *Proc Natl Acad Sci U S A* **2021**, *118* (37): e2107108118. DOI: 10.1073/pnas.2107108118
59. Moustaqil, M. et al. SARS-CoV-2 proteases PLpro and 3CLpro cleave IRF3 and critical modulators of inflammatory pathways (NLRP12 and TAB1): implications for disease presentation across species. *Emerg Microbes Infect* **2021**, *10* (1): 178-195. DOI: 10.1080/22221751.2020.1870414
60. Tsu, B. V. et al. Host specific sensing of coronaviruses and picornaviruses by the CARD8 inflammasome. *bioRxiv* **2022**, 2022.09.21.508960. DOI: 10.1101/2022.09.21.508960
61. Wenzel, J. et al. The SARS-CoV-2 main protease M^{pro} causes microvascular brain pathology by cleaving NEMO in brain endothelial cells. *Nat Neurosci* **2021**, *24* (11): 1522-1533. DOI: 10.1038/s41593-021-00926-1
62. Koudelka, T.; Boger, J.; Henkel, A.; Schönherr, R.; Krantz, S.; Fuchs, S.; Rodríguez, E.; Redecke, L.; Tholey, A. N-Terminomics for the Identification of In Vitro Substrates and Cleavage Site Specificity of the SARS-CoV-2 Main Protease. *Proteomics* **2021**, *21* (2): e2000246. DOI: 10.1002/pmic.202000246
63. Pablos, I. et al. Mechanistic insights into COVID-19 by global analysis of the SARS-CoV-2 $3CL^{pro}$ substrate degradome. *Cell Rep* **2021**, *37* (4): 109892. DOI: 10.1016/j.celrep.2021.109892
64. Vuong, W. et al. Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat Commun* **2020**, *11* (1): 4282. DOI: 10.1038/s41467-020-18096-2
65. de Vries, M. et al. A comparative analysis of SARS-CoV-2 antivirals characterizes $3CL^{pro}$ inhibitor PF-00835231 as a potential new treatment for COVID-19. *J Virol* **2021**, JVI.01819-20. DOI: 10.1128/JVI.01819-20
66. Boras, B. et al. Preclinical characterization of an intravenous coronavirus 3CL protease inhibitor for the potential treatment of COVID19. *Nat Commun* **2021**, *12* (1): 6055. DOI: 10.1038/s41467-021-26239-2
67. Owen, D. R. et al. An oral SARS-CoV-2 M^{pro} inhibitor clinical candidate for the treatment of COVID-19. *Science* **2021**, *374* (6575): 1586-1593. DOI: 10.1126/science.abl4784
68. Abdelnabi, R. et al. The oral protease inhibitor (PF-07321332) protects Syrian hamsters against infection with SARS-CoV-2 variants of concern. *Nat Commun* **2022**, *13* (1): 719. DOI: 10.1038/s41467-022-28354-0
69. Allerton, C. M. N. et al. A Second-Generation Oral SARS-CoV-2 Main Protease Inhibitor Clinical Candidate for the Treatment of COVID-19. *J Med Chem* **2024**, DOI: 10.1021/acs.jmedchem.3c02469
70. Toma-Fukai, S.; Shimizu, T. Structural Diversity of Ubiquitin E3 Ligase. *Molecules* **2021**, *26* (21): 6682. DOI: 10.3390/molecules26216682
71. Humphreys, L. M.; Smith, P.; Chen, Z.; Fouad, S.; D'Angiolella, V. The role of E3 ubiquitin ligases in the development and progression of glioblastoma. *Cell Death Differ* **2021**, *28* (2): 522-537. DOI: 10.1038/s41418-020-00696-6
72. Cui, H.; Zhang, Y.; Zhang, L. Progress on Poxvirus E3 Ubiquitin Ligases and Adaptor Proteins. *Front Immunol* **2021**, *12* 740223. DOI: 10.3389/fimmu.2021.740223
73. Petroski, M. D.; Deshaies, R. J. Function and regulation of cullin-RING ubiquitin ligases. *Nat Rev Mol Cell Biol* **2005**, *6* (1): 9-20. DOI: 10.1038/nrm1547
74. Jeong, Y.; Oh, A. R.; Jung, Y. H.; Gi, H.; Kim, Y. U.; Kim, K. Targeting E3 ubiquitin ligases and their adaptors as a therapeutic strategy for metabolic diseases. *Exp Mol Med* **2023**, *55* (10): 2097-2104. DOI: 10.1038/s12276-023-01087-w
75. Bai, C.; Sen, P.; Hofmann, K.; Ma, L.; Goebl, M.; Harper, J. W.; Elledge, S. J. SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* **1996**, *86* (2): 263-274. DOI: 10.1016/s0092-8674(00)80098-7

76. Feldman, R. M.; Correll, C. C.; Kaplan, K. B.; Deshaies, R. J. A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. *Cell* **1997**, *91* (2): 221-230. DOI: 10.1016/s0092-8674(00)80404-3
77. Skowyra, D.; Craig, K. L.; Tyers, M.; Elledge, S. J.; Harper, J. W. F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* **1997**, *91* (2): 209-219. DOI: 10.1016/s0092-8674(00)80403-1
78. Zipper, L. M.; Mulcahy, R. T. The Keap1 BTB/POZ dimerization function is required to sequester Nrf2 in cytoplasm. *J Biol Chem* **2002**, *277* (39): 36544-36552. DOI: 10.1074/jbc.M206530200
79. Canning, P. et al. Structural basis for Cul3 protein assembly with the BTB-Kelch family of E3 ubiquitin ligases. *J Biol Chem* **2013**, *288* (11): 7803-7814. DOI: 10.1074/jbc.M112.437996
80. Ande, S. R.; Chen, J.; Maddika, S. The ubiquitin pathway: an emerging drug target in cancer therapy. *Eur J Pharmacol* **2009**, *625* (1-3): 199-205. DOI: 10.1016/j.ejphar.2009.08.042
81. Zhao, L.; Zhao, J.; Zhong, K.; Tong, A.; Jia, D. Targeted protein degradation: mechanisms, strategies and application. *Signal Transduct Target Ther* **2022**, *7* (1): 113. DOI: 10.1038/s41392-022-00966-4
82. Pohl, C.; Dikic, I. Cellular quality control by the ubiquitin-proteasome system and autophagy. *Science* **2019**, *366* (6467): 818-822. DOI: 10.1126/science.aax3769
83. Zientara-Rytter, K.; Subramani, S. The Roles of Ubiquitin-Binding Protein Shuttles in the Degradative Fate of Ubiquitinated Proteins in the Ubiquitin-Proteasome System and Autophagy. *Cells* **2019**, *8* (1): 40. DOI: 10.3390/cells8010040
84. Dong, Y.; Zhang, S.; Wu, Z.; Li, X.; Wang, W. L.; Zhu, Y.; Stoilova-McPhie, S.; Lu, Y.; Finley, D.; Mao, Y. Cryo-EM structures and dynamics of substrate-engaged human 26S proteasome. *Nature* **2019**, *565* (7737): 49-55. DOI: 10.1038/s41586-018-0736-4
85. Goh, L. K.; Sorkin, A. Endocytosis of receptor tyrosine kinases. *Cold Spring Harb Perspect Biol* **2013**, *5* a017459. DOI: 10.1101/cshperspect.a017459
86. Chandrasekharan, M. B.; Huang, F.; Sun, Z. W. Ubiquitination of histone H2B regulates chromatin dynamics by enhancing nucleosome stability. *Proc Natl Acad Sci U S A* **2009**, *106* (39): 16686-16691. DOI: 10.1073/pnas.0907862106
87. Braten, O. et al. Numerous proteins with unique characteristics are degraded by the 26S proteasome following monoubiquitination. *Proc Natl Acad Sci U S A* **2016**, *113* (32): E4639-47. DOI: 10.1073/pnas.1608644113
88. Hernaez, B.; Alcamí, A. Poxvirus Immune Evasion. *Annu Rev Immunol* **2024**, *42* (1): 551-584. DOI: 10.1146/annurev-immunol-090222-110227
89. Barry, M.; Van Buuren, N.; Burles, K.; Mottet, K.; Wang, Q.; Teale, A. Poxvirus exploitation of the ubiquitin-proteasome system. *Viruses* **2010**, *2* (10): 2356-2380. DOI: 10.3390/v2102356
90. Gustin, J. K.; Moses, A. V.; Früh, K.; Douglas, J. L. Viral takeover of the host ubiquitin system. *Front Microbiol* **2011**, *2* 161. DOI: 10.3389/fmicb.2011.00161
91. Randow, F.; Lehner, P. J. Viral avoidance and exploitation of the ubiquitin system. *Nat Cell Biol* **2009**, *11* (5): 527-534. DOI: 10.1038/ncb0509-527
92. Luo, H. Interplay between the virus and the ubiquitin-proteasome system: molecular mechanism of viral pathogenesis. *Curr Opin Virol* **2016**, *17* 1-10. DOI: 10.1016/j.coviro.2015.09.005
93. Shchelkunov, S. N. Interaction of orthopoxviruses with the cellular ubiquitin-ligase system. *Virus Genes* **2010**, *41* (3): 309-318. DOI: 10.1007/s11262-010-0519-y
94. Beard, P. M.; Froggatt, G. C.; Smith, G. L. Vaccinia virus kelch protein A55 is a 64 kDa intracellular factor that affects virus-induced cytopathic effect and the outcome of infection

in a murine intradermal model. *J Gen Virol* **2006**, *87* (Pt 6): 1521-1529. DOI: 10.1099/vir.0.81854-0
95. Pires de Miranda, M.; Reading, P. C.; Tscharke, D. C.; Murphy, B. J.; Smith, G. L. The vaccinia virus kelch-like protein C2L affects calcium-independent adhesion to the extracellular matrix and inflammation in a murine intradermal model. *J Gen Virol* **2003**, *84* (Pt 9): 2459-2471. DOI: 10.1099/vir.0.19292-0
96. Pallett, M. A.; Ren, H.; Zhang, R. Y.; Scutts, S. R.; Gonzalez, L.; Zhu, Z.; Maluquer de Motes, C.; Smith, G. L. Vaccinia Virus BBK E3 Ligase Adaptor A55 Targets Importin-Dependent NF-κB Activation and Inhibits $CD8^+$ T-Cell Memory. *J Virol* **2019**, *93* (10): e00051-19. DOI: 10.1128/JVI.00051-19
97. Froggatt, G. C.; Smith, G. L.; Beard, P. M. Vaccinia virus gene F3L encodes an intracellular protein that affects the innate immune response. *J Gen Virol* **2007**, *88* (Pt 7): 1917-1921. DOI: 10.1099/vir.0.82815-0
98. Shchelkunov, S. N.; Blinov, V. M.; Sandakhchiev, L. S. Ankyrin-like proteins of variola and vaccinia viruses. *FEBS Lett* **1993**, *319* (1-2): 163-165. DOI: 10.1016/0014-5793(93)80059-4
99. Senkevich, T. G.; Yutin, N.; Wolf, Y. I.; Koonin, E. V.; Moss, B. Ancient Gene Capture and Recent Gene Loss Shape the Evolution of Orthopoxvirus-Host Interaction Genes. *mBio* **2021**, *12* (4): e0149521. DOI: 10.1128/mBio.01495-21
100. Bravo Cruz, A. G.; Shisler, J. L. Vaccinia virus K1 ankyrin repeat protein inhibits NF-κB activation by preventing RelA acetylation. *J Gen Virol* **2016**, *97* (10): 2691-2702. DOI: 10.1099/jgv.0.000576
101. Waibler, Z.; Anzaghe, M.; Frenz, T.; Schwantes, A.; Pöhlmann, C.; Ludwig, H.; Palomo-Otero, M.; Alcamí, A.; Sutter, G.; Kalinke, U. Vaccinia virus-mediated inhibition of type I interferon responses is a multifactorial process involving the soluble type I interferon receptor B18 and intracellular components. *J Virol* **2009**, *83* (4): 1563-1571. DOI: 10.1128/JVI.01617-08
102. Liu, R.; Moss, B. Vaccinia Virus C9 Ankyrin Repeat/F-Box Protein Is a Newly Identified Antagonist of the Type I Interferon-Induced Antiviral State. *J Virol* **2018**, *92* (9): e00053-18. DOI: 10.1128/JVI.00053-18
103. Liu, R.; Olano, L. R.; Mirzakhanyan, Y.; Gershon, P. D.; Moss, B. Vaccinia Virus Ankyrin-Repeat/F-Box Protein Targets Interferon-Induced IFITs for Proteasomal Degradation. *Cell Rep* **2019**, *29* (4): 816-828.e6. DOI: 10.1016/j.celrep.2019.09.039
104. Burles, K.; Irwin, C. R.; Burton, R. L.; Schriewer, J.; Evans, D. H.; Buller, R. M.; Barry, M. Initial characterization of vaccinia virus B4 suggests a role in virus spread. *Virology* **2014**, *456-457* 108-120. DOI: 10.1016/j.virol.2014.03.019
105. Teale, A.; Campbell, S.; Van Buuren, N.; Magee, W. C.; Watmough, K.; Couturier, B.; Shipclark, R.; Barry, M. Orthopoxviruses require a functional ubiquitin-proteasome system for productive replication. *J Virol* **2009**, *83* (5): 2099-2108. DOI: 10.1128/JVI.01753-08
106. Satheshkumar, P. S.; Anton, L. C.; Sanz, P.; Moss, B. Inhibition of the ubiquitin-proteasome system prevents vaccinia virus DNA replication and expression of intermediate and late genes. *J Virol* **2009**, *83* (6): 2469-2479. DOI: 10.1128/JVI.01986-08
107. Mercer, J.; Snijder, B.; Sacher, R.; Burkard, C.; Bleck, C. K.; Stahlberg, H.; Pelkmans, L.; Helenius, A. RNAi screening reveals proteasome- and Cullin3-dependent stages in vaccinia virus infection. *Cell Rep* **2012**, *2* (4): 1036-1047. DOI: 10.1016/j.celrep.2012.09.003
108. Grossegesse, M.; Doellinger, J.; Fritsch, A.; Laue, M.; Piesker, J.; Schaade, L.; Nitsche, A. Global ubiquitination analysis reveals extensive modification and proteasomal degradation of cowpox virus proteins, but preservation of viral cores. *Sci Rep* **2018**, *8* (1): 1807. DOI: 10.1038/s41598-018-20130-9
109. Lodish, H. F. 2000. *Molecular Cell Biology*. Scientific American Library.

110. Kuzmanov, U.; Emili, A. Protein-protein interaction networks: probing disease mechanisms using model systems. *Genome Med* **2013**, *5* (4): 37. DOI: 10.1186/gm441
111. Mahalingam, S.; L. Herrero; B. Herring. 2016. *Alphaviruses : Current Biology*. Norfolk, UNITED KINGDOM: Caister Academic Press.
112. Schneider, M. et al. Safety and immunogenicity of a single-shot live-attenuated chikungunya vaccine: a double-blind, multicentre, randomised, placebo-controlled, phase 3 trial. *Lancet* **2023**, *401* (10394): 2138-2147. DOI: 10.1016/S0140-6736(23)00641-4
113. Mann, M.; Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat Biotechnol* **2003**, *21* (3): 255-261. DOI: 10.1038/nbt0303-255
114. Jiang, Y. et al. Comprehensive Overview of Bottom-Up Proteomics Using Mass Spectrometry. *ACS Measurement Science Au*
ACS Meas. Sci. Au **2024**, DOI: 10.1021/acsmeasuresciau.3c00068
115. Zhao, Y.; Jensen, O. N. Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques. *Proteomics* **2009**, *9* (20): 4632-4641. DOI: 10.1002/pmic.200900398
116. Richards, A. L.; Eckhardt, M.; Krogan, N. J. Mass spectrometry-based protein-protein interaction networks for the study of human diseases. *Mol Syst Biol* **2021**, *17* (1): e8792. DOI: 10.15252/msb.20188792
117. Gevaert, K.; Van Damme, J.; Goethals, M.; Thomas, G. R.; Hoorelbeke, B.; Demol, H.; Martens, L.; Puype, M.; Staes, A.; Vandekerckhove, J. Chromatographic isolation of methionine-containing peptides for gel-free proteome analysis: identification of more than 800 Escherichia coli proteins. *Mol Cell Proteomics* **2002**, *1* (11): 896-903. DOI: 10.1074/mcp.m200061-mcp200
118. Kleifeld, O.; Doucet, A.; Prudova, A.; auf dem Keller, U.; Gioia, M.; Kizhakkedathu, J. N.; Overall, C. M. Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat Protoc* **2011**, *6* (10): 1578-1611. DOI: 10.1038/nprot.2011.382
119. Wejda, M.; Impens, F.; Takahashi, N.; Van Damme, P.; Gevaert, K.; Vandenabeele, P. Degradomics reveals that cleavage specificity profiles of caspase-2 and effector caspases are alike. *J Biol Chem* **2012**, *287* (41): 33983-33995. DOI: 10.1074/jbc.M112.384552
120. Prudova, A.; auf dem Keller, U.; Butler, G. S.; Overall, C. M. Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. *Mol Cell Proteomics* **2010**, *9* (5): 894-911. DOI: 10.1074/mcp.M000050-MCP201
121. Walton, A. et al. It's Time for Some "Site"-Seeing: Novel Tools to Monitor the Ubiquitin Landscape in Arabidopsis thaliana. *Plant Cell* **2016**, *28* (1): 6-16. DOI: 10.1105/tpc.15.00878
122. Abrahmsén, L.; Tom, J.; Burnier, J.; Butcher, K. A.; Kossiakoff, A.; Wells, J. A. Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. *Biochemistry* **1991**, *30* (17): 4151-4159. DOI: 10.1021/bi00231a007
123. Weeks, A. M.; Wells, J. A. Engineering peptide ligase specificity by proteomic identification of ligation sites. *Nat Chem Biol* **2018**, *14* (1): 50-57. DOI: 10.1038/nchembio.2521
124. Weeks, A. M.; Byrnes, J. R.; Lui, I.; Wells, J. A. Mapping proteolytic neo-N termini at the surface of living cells. *Proc Natl Acad Sci U S A* **2021**, *118* (8): e2018809118. DOI: 10.1073/pnas.2018809118
125. Peng, J.; Schwartz, D.; Elias, J. E.; Thoreen, C. C.; Cheng, D.; Marsischky, G.; Roelofs, J.; Finley, D.; Gygi, S. P. A proteomics approach to understanding protein ubiquitination. *Nat Biotechnol* **2003**, *21* (8): 921-926. DOI: 10.1038/nbt849
126. Goldknopf, I. L.; Busch, H. Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate-protein A24. *Proc Natl Acad Sci U S A* **1977**, *74* (3): 864-868. DOI: 10.1073/pnas.74.3.864

127. Matsumoto, M.; Hatakeyama, S.; Oyamada, K.; Oda, Y.; Nishimura, T.; Nakayama, K. I. Large-scale analysis of the human ubiquitin-related proteome. *Proteomics* **2005**, *5* (16): 4145-4151. DOI: 10.1002/pmic.200401280
128. Tagwerker, C.; Flick, K.; Cui, M.; Guerrero, C.; Dou, Y.; Auer, B.; Baldi, P.; Huang, L.; Kaiser, P. A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivocross-linking. *Mol Cell Proteomics* **2006**, *5* (4): 737-748. DOI: 10.1074/mcp.M500368-MCP200
129. Meierhofer, D.; Wang, X.; Huang, L.; Kaiser, P. Quantitative analysis of global ubiquitination in HeLa cells by mass spectrometry. *J Proteome Res* **2008**, *7* (10): 4566-4576. DOI: 10.1021/pr800468j
130. Danielsen, J. M.; Sylvestersen, K. B.; Bekker-Jensen, S.; Szklarczyk, D.; Poulsen, J. W.; Horn, H.; Jensen, L. J.; Mailand, N.; Nielsen, M. L. Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. *Mol Cell Proteomics* **2011**, *10* (3): M110.003590. DOI: 10.1074/mcp.M110.003590
131. Xu, G.; Paige, J. S.; Jaffrey, S. R. Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat Biotechnol* **2010**, *28* (8): 868-873. DOI: 10.1038/nbt.1654
132. Heidelberger, J. B.; Wagner, S. A.; Beli, P. Mass Spectrometry-Based Proteomics for Investigating DNA Damage-Associated Protein Ubiquitylation. *Front Genet* **2016**, *7* 109. DOI: 10.3389/fgene.2016.00109
133. Swaney, D. L.; Beltrao, P.; Starita, L.; Guo, A.; Rush, J.; Fields, S.; Krogan, N. J.; Villén, J. Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat Methods* **2013**, *10* (7): 676-682. DOI: 10.1038/nmeth.2519
134. Zecha, J.; Gabriel, W.; Spallek, R.; Chang, Y. C.; Mergner, J.; Wilhelm, M.; Bassermann, F.; Kuster, B. Linking post-translational modifications and protein turnover by site-resolved protein turnover profiling. *Nat Commun* **2022**, *13* (1): 165. DOI: 10.1038/s41467-021-27639-0
135. Kim, W. et al. Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol Cell* **2011**, *44* (2): 325-340. DOI: 10.1016/j.molcel.2011.08.025
136. Akimov, V. et al. UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites. *Nat Struct Mol Biol* **2018**, *25* (7): 631-640. DOI: 10.1038/s41594-018-0084-y
137. Morris, J. H.; Knudsen, G. M.; Verschueren, E.; Johnson, J. R.; Cimermancic, P.; Greninger, A. L.; Pico, A. R. Affinity purification-mass spectrometry and network analysis to understand protein-protein interactions. *Nat Protoc* **2014**, *9* (11): 2539-2554. DOI: 10.1038/nprot.2014.164
138. Roux, K. J.; Kim, D. I.; Raida, M.; Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* **2012**, *196* (6): 801-810. DOI: 10.1083/jcb.201112098
139. Branon, T. C.; Bosch, J. A.; Sanchez, A. D.; Udeshi, N. D.; Svinkina, T.; Carr, S. A.; Feldman, J. L.; Perrimon, N.; Ting, A. Y. Efficient proximity labeling in living cells and organisms with TurboID. *Nat Biotechnol* **2018**, *36* (9): 880-887. DOI: 10.1038/nbt.4201
140. Kubitz, L.; Bitsch, S.; Zhao, X.; Schmitt, K.; Deweid, L.; Roehrig, A.; Barazzone, E. C.; Valerius, O.; Kolmar, H.; Béthune, J. Engineering of ultraID, a compact and hyperactive enzyme for proximity-dependent biotinylation in living cells. *Commun Biol* **2022**, *5* (1): 657. DOI: 10.1038/s42003-022-03604-5
141. Verschueren, E.; Von Dollen, J.; Cimermancic, P.; Gulbahce, N.; Sali, A.; Krogan, N. J. Scoring Large-Scale Affinity Purification Mass Spectrometry Datasets with MiST. *Curr Protoc Bioinformatics* **2015**, *49* 8.19.1-8.19.16. DOI: 10.1002/0471250953.bi0819s49

142. Breitkreutz, A. et al. A global protein kinase and phosphatase interaction network in yeast. *Science* **2010**, *328* (5981): 1043-1046. DOI: 10.1126/science.1176495
143. Jäger, S. et al. Purification and characterization of HIV-human protein complexes. *Methods* **2011**, *53* (1): 13-19. DOI: 10.1016/j.ymeth.2010.08.007
144. Sowa, M. E.; Bennett, E. J.; Gygi, S. P.; Harper, J. W. Defining the human deubiquitinating enzyme interaction landscape. *Cell* **2009**, *138* (2): 389-403. DOI: 10.1016/j.cell.2009.04.042
145. Choi, H.; Glatter, T.; Gstaiger, M.; Nesvizhskii, A. I. SAINT-MS1: protein-protein interaction scoring using label-free intensity data in affinity purification-mass spectrometry experiments. *J Proteome Res* **2012**, *11* (4): 2619-2624. DOI: 10.1021/pr201185r
146. Teo, G.; Liu, G.; Zhang, J.; Nesvizhskii, A. I.; Gingras, A. C.; Choi, H. SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. *J Proteomics* **2014**, *100* 37-43. DOI: 10.1016/j.jprot.2013.10.023
147. Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* **2012**, *11* (6): O111.016717. DOI: 10.1074/mcp.O111.016717
148. Weisbrod, C. R.; Eng, J. K.; Hoopmann, M. R.; Baker, T.; Bruce, J. E. Accurate peptide fragment mass analysis: multiplexed peptide identification and quantification. *J Proteome Res* **2012**, *11* (3): 1621-1632. DOI: 10.1021/pr2008175
149. Ullrich, S.; Nitsche, C. The SARS-CoV-2 main protease as drug target. *Bioorg Med Chem Lett* **2020**, *30* (17): 127377. DOI: 10.1016/j.bmcl.2020.127377
150. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020**, *368* (6489): 409-412. DOI: 10.1126/science.abb3405
151. Rut, W. et al. SARS-CoV-2 M^{pro} inhibitors and activity-based probes for patient-sample imaging. *Nat Chem Biol* **2021**, *17* (2): 222-228. DOI: 10.1038/s41589-020-00689-z
152. Chuck, C. P.; Chong, L. T.; Chen, C.; Chow, H. F.; Wan…, D. C. C. Profiling of Substrate Specificity of SARS-CoV 3CL<sup>pro</sup. *PloS one* **2010**, DOI:
153. Freitas, B. T.; Durie, I. A.; Murray, J.; Longo, J. E.; Miller, H. C.; Crich, D.; Hogan, R. J.; Tripp, R. A.; Pegan, S. D. Characterization and Noncovalent Inhibition of the Deubiquitinase and deISGylase Activity of SARS-CoV-2 Papain-Like Protease. *ACS Infect Dis* **2020**, *6* (8): 2099-2109. DOI: 10.1021/acsinfecdis.0c00168
154. Gao, X.; Qin, B.; Chen, P.; Zhu, K.; Hou, P.; Wojdyla, J. A.; Wang, M.; Cui, S. Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharm Sin B* **2021**, *11* (1): 237-245. DOI: 10.1016/j.apsb.2020.08.014
155. Rut, W. et al. Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti-COVID-19 drug design. *Sci Adv* **2020**, *6* (42): eabd4596. DOI: 10.1126/sciadv.abd4596
156. Hill, M. E.; Kumar, A.; Wells, J. A.; Hobman, T. C.; Julien, O.; Hardy, J. A. The Unique Cofactor Region of Zika Virus NS2B–NS3 Protease Facilitates Cleavage of Key Host Proteins. *ACS Chemical Biology* **2018**, *13* (9): 2398-2405. DOI: 10.1021/acschembio.8b00508
157. Jagdeo, J. M.; Dufour, A.; Klein, T.; Solis, N.; Kleifeld, O.; Kizhakkedathu, J.; Luo, H.; Overall, C. M.; Jan, E. N-Terminomics TAILS Identifies Host Cell Substrates of Poliovirus and Coxsackievirus B3 3C Proteinases That Modulate Virus Infection. *J Virol* **2018**, *92* (8): e02211-17. DOI: 10.1128/JVI.02211-17
158. May, D. G. et al. A BioID-Derived Proximity Interactome for SARS-CoV-2 Proteins. *Viruses* **2022**, *14* (3): 611. DOI: 10.3390/v14030611

159. Mohamud, Y.; Xue, Y. C.; Liu, H.; Ng, C. S.; Bahreyni, A.; Jan, E.; Luo, H. The papain-like protease of coronaviruses cleaves ULK1 to disrupt host autophagy. *Biochem Biophys Res Commun* **2021**, *540* 75-82. DOI: 10.1016/j.bbrc.2020.12.091
160. Reynolds, N. D. et al. The SARS-CoV-2 SSHHPS Recognized by the Papain-like Protease. *ACS Infect Dis* **2021**, *7* (6): 1483-1502. DOI: 10.1021/acsinfecdis.0c00866
161. Miczi, M.; Golda, M.; Kunkli, B.; Nagy, T.; Tőzsér, J.; Mótyán, J. A. Identification of Host Cellular Protein Substrates of SARS-COV-2 Main Protease. *Int J Mol Sci* **2020**, *21* (24): E9523. DOI: 10.3390/ijms21249523
162. Meyer, B. et al. Characterising proteolysis during SARS-CoV-2 infection identifies viral cleavage sites and cellular targets with therapeutic potential. *Nat Commun* **2021**, *12* (1): 5553. DOI: 10.1038/s41467-021-25796-w
163. Kleifeld, O.; Doucet, A.; auf dem Keller, U.; Prudova, A.; Schilling, O.; Kainthan, R. K.; Starr, A. E.; Foster, L. J.; Kizhakkedathu, J. N.; Overall, C. M. Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat Biotechnol* **2010**, *28* (3): 281-288. DOI: 10.1038/nbt.1611
164. Poreba, M.; Salvesen, G. S.; Drag, M. Synthesis of a HyCoSuL peptide substrate library to dissect protease substrate specificity. *Nat Protoc* **2017**, *12* (10): 2189-2214. DOI: 10.1038/nprot.2017.091
165. Kumar, A. et al. SARS-CoV-2 Nonstructural Protein 1 Inhibits the Interferon Response by Causing Depletion of Key Host Signaling Factors. *J Virol* **2021**, *95* (13): e0026621. DOI: 10.1128/JVI.00266-21
166. Luo, S. Y.; Araya, L. E.; Julien, O. Protease Substrate Identification Using N-terminomics. *ACS Chem Biol* **2019**, *14* (11): 2361-2371. DOI: 10.1021/acschembio.9b00398
167. Shen, X. R. et al. ACE2-independent infection of T lymphocytes by SARS-CoV-2. *Signal Transduct Target Ther* **2022**, *7* (1): 83. DOI: 10.1038/s41392-022-00919-x
168. Seaman, J. E.; Julien, O.; Lee, P. S.; Rettenmaier, T. J.; Thomsen, N. D.; Wells, J. A. Cacidases: caspases can cleave after aspartate, glutamate and phosphoserine residues. *Cell Death & Differentiation* **2016**, *23* (10): 1717-1726. DOI: 10.1038/cdd.2016.62
169. Crawford, E. D. et al. The DegraBase: a database of proteolysis in healthy and apoptotic human cells. *Mol Cell Proteomics* **2013**, *12* (3): 813-824. DOI: 10.1074/mcp.O112.024372
170. Taniguchi, Y. The Bromodomain and Extra-Terminal Domain (BET) Family: Functional Anatomy of BET Paralogous Proteins. *Int J Mol Sci* **2016**, *17* (11): E1849. DOI: 10.3390/ijms17111849
171. Samelson, A. J. et al. BRD2 inhibition blocks SARS-CoV-2 infection by reducing transcription of the host cell receptor ACE2. *Nat Cell Biol* **2022**, *24* (1): 24-34. DOI: 10.1038/s41556-021-00821-8
172. Chen, I. P. et al. Viral E protein neutralizes BET protein-mediated post-entry antagonism of SARS-CoV-2. *Cell Rep* **2022**, *40* (3): 111088. DOI: 10.1016/j.celrep.2022.111088
173. Landeras-Bueno, S.; Jorba, N.; Pérez-Cidoncha, M.; Ortín, J. The splicing factor proline-glutamine rich (SFPQ/PSF) is involved in influenza virus transcription. *PLoS Pathog* **2011**, *7* (11): e1002397. DOI: 10.1371/journal.ppat.1002397
174. Zhou, B.; Wu, F.; Han, J.; Qi, F.; Ni, T.; Qian, F. Exploitation of nuclear protein SFPQ by the encephalomyocarditis virus to facilitate its replication. *Biochem Biophys Res Commun* **2019**, *510* (1): 65-71. DOI: 10.1016/j.bbrc.2019.01.032
175. Greco-Stewart, V. S.; Thibault, C. S.; Pelchat, M. Binding of the polypyrimidine tract-binding protein-associated splicing factor (PSF) to the hepatitis delta virus RNA. *Virology* **2006**, *356* (1-2): 35-44. DOI: 10.1016/j.virol.2006.06.040
176. Flather, D.; Nguyen, J. H. C.; Semler, B. L.; Gershon, P. D. Exploitation of nuclear functions by human rhinovirus, a cytoplasmic RNA virus. *PLoS Pathog* **2018**, *14* (8): e1007277. DOI: 10.1371/journal.ppat.1007277

177. Park, N.; Katikaneni, P.; Skern, T.; Gustin, K. E. Differential targeting of nuclear pore complex proteins in poliovirus-infected cells. *J Virol* **2008**, *82* (4): 1647-1655. DOI: 10.1128/JVI.01670-07
178. De Jesús-González, L. A.; Palacios-Rápalo, S.; Reyes-Ruiz, J. M.; Osuna-Ramos, J. F.; Cordero-Rivera, C. D.; Farfan-Morales, C. N.; Gutiérrez-Escolano, A. L.; Del Ángel, R. M. The Nuclear Pore Complex Is a Key Target of Viral Proteases to Promote Viral Replication. *Viruses* **2021**, *13* (4): 706. DOI: 10.3390/v13040706
179. Bouhaddou, M. et al. The Global Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell* **2020**, *182* (3): 685-712.e19. DOI: 10.1016/j.cell.2020.06.034
180. Iftner, T.; Elbel, M.; Schopp, B.; Hiller, T.; Loizou, J. I.; Caldecott, K. W.; Stubenrauch, F. Interference of papillomavirus E6 protein with single-strand break repair by interaction with XRCC1. *EMBO J* **2002**, *21* (17): 4741-4748. DOI: 10.1093/emboj/cdf443
181. Almeida Pereira Leite, S. T.; Marques-Guimarães, N.; Silva-Oliveira, J. C.; Dutra-Souto, F. J.; Alves-dos-Santos, R.; Bassi-Branco, C. L. The X-ray repair cross complementing protein 1 (XRCC1) rs25487 polymorphism and susceptibility to cirrhosis in Brazilian patients with chronic viral hepatitis. *Ann Hepatol* **2013**, *12* (5): 733-739. DOI:
182. Li, Z.; Wu, J.; Chavez, L.; Hoh, R.; Deeks, S. G.; Pillai, S. K.; Zhou, Q. Reiterative Enrichment and Authentication of CRISPRi Targets (REACT) identifies the proteasome as a key contributor to HIV-1 latency. *PLoS Pathog* **2019**, *15* (1): e1007498. DOI: 10.1371/journal.ppat.1007498
183. Generous, A.; Thorson, M.; Barcus, J.; Jacher, J.; Busch, M.; Sleister, H. Identification of putative interactions between swine and human influenza A virus nucleoprotein and human host proteins. *Virol J* **2014**, *11* 228. DOI: 10.1186/s12985-014-0228-6
184. Li, J. et al. Antiviral activity of a purine synthesis enzyme reveals a key role of deamidation in regulating protein nuclear import. *Sci Adv* **2019**, *5* (10): eaaw7373. DOI: 10.1126/sciadv.aaw7373
185. Mazzarino, R. C. Targeting Future Pandemics, a Case for De Novo Purine Synthesis and Basic Research. *Front Immunol* **2021**, *12* 694300. DOI: 10.3389/fimmu.2021.694300
186. Wu, S. Y.; Lee, A. Y.; Hou, S. Y.; Kemper, J. K.; Erdjument-Bromage, H.; Tempst, P.; Chiang, C. M. Brd4 links chromatin targeting to HPV transcriptional silencing. *Genes Dev* **2006**, *20* (17): 2383-2396. DOI: 10.1101/gad.1448206
187. Viejo-Borbolla, A.; Ottinger, M.; Brüning, E.; Bürger, A.; König, R.; Kati, E.; Sheldon, J. A.; Schulz, T. F. Brd2/RING3 interacts with a chromatin-binding domain in the Kaposi's Sarcoma-associated herpesvirus latency-associated nuclear antigen 1 (LANA-1) that is required for multiple functions of LANA-1. *J Virol* **2005**, *79* (21): 13618-13629. DOI: 10.1128/JVI.79.21.13618-13629.2005
188. Filippakopoulos, P. et al. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **2012**, *149* (1): 214-231. DOI: 10.1016/j.cell.2012.02.013
189. Bisgrove, D. A.; Mahmoudi, T.; Henklein, P.; Verdin, E. Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription. *Proc Natl Acad Sci U S A* **2007**, *104* (34): 13690-13695. DOI: 10.1073/pnas.0705053104
190. Anand, K.; Ziebuhr, J.; Wadhwani, P.; Mesters, J. R.; Hilgenfeld, R. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **2003**, *300* (5626): 1763-1767. DOI: 10.1126/science.1085658
191. Zhou, Y.; Zhou, B.; Pache, L.; Chang, M.; Khodabakhshi, A. H.; Tanaseichuk, O.; Benner, C.; Chanda, S. K. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **2019**, *10* (1): 1523. DOI: 10.1038/s41467-019-09234-6
192. Fortelny, N.; Yang, S.; Pavlidis, P.; Lange, P. F.; Overall, C. M. Proteome TopFIND 3.0 with TopFINDer and PathFINDer: database and analysis tools for the association of

protein termini to pre- and post-translational events. *Nucleic Acids Res* **2015**, *43* (Database issue): D290-7. DOI: 10.1093/nar/gku1012
193. Mahrus, S.; Kisiel, W.; Craik, C. S. Granzyme M is a regulatory protease that inactivates proteinase inhibitor 9, an endogenous inhibitor of granzyme B. *J Biol Chem* **2004**, *279* (52): 54275-54282. DOI: 10.1074/jbc.M411482200
194. Bojkova, D.; Klann, K.; Koch, B.; Widera, M.; Krause, D.; Ciesek, S.; Cinatl, J.; Münch, C. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **2020**, *583* (7816): 469-472. DOI: 10.1038/s41586-020-2332-7
195. Rebendenne, A. et al. Bidirectional genome-wide CRISPR screens reveal host factors regulating SARS-CoV-2, MERS-CoV and seasonal HCoVs. *Nat Genet* **2022**, *54* (8): 1090-1102. DOI: 10.1038/s41588-022-01110-2
196. Mac Kain, A. et al. Identification of DAXX as a restriction factor of SARS-CoV-2 through a CRISPR/Cas9 screen. *Nat Commun* **2022**, *13* (1): 2442. DOI: 10.1038/s41467-022-30134-9
197. Wang, Y.; Fan, Y.; Huang, Y.; Du, T.; Liu, Z.; Huang, D.; Wang, Y.; Wang, N.; Zhang, P. TRIM28 regulates SARS-CoV-2 cell entry by targeting ACE2. *Cell Signal* **2021**, *85* 110064. DOI: 10.1016/j.cellsig.2021.110064
198. Araya, L. E.; Soni, I. V.; Hardy, J. A.; Julien, O. Deorphanizing Caspase-3 and Caspase-9 Substrates In and Out of Apoptosis with Deep Substrate Profiling. *ACS Chem Biol* **2021**, *16* (11): 2280-2296. DOI: 10.1021/acschembio.1c00456
199. Liu, Z.; Nailwal, H.; Rector, J.; Rahman, M. M.; Sam, R.; McFadden, G.; Chan, F. K. A class of viral inducer of degradation of the necroptosis adaptor RIPK3 regulates virus-induced inflammation. *Immunity* **2021**, *54* (2): 247-258.e7. DOI: 10.1016/j.immuni.2020.11.020
200. Chen, W. et al. Induction of Siglec-G by RNA viruses inhibits the innate immune response by promoting RIG-I degradation. *Cell* **2013**, *152* (3): 467-478. DOI: 10.1016/j.cell.2013.01.011
201. Manivannan, P.; Siddiqui, M. A.; Malathi, K. RNase L Amplifies Interferon Signaling by Inducing Protein Kinase R-Mediated Antiviral Stress Granules. *J Virol* **2020**, *94* (13): e00205-20. DOI: 10.1128/JVI.00205-20
202. Ivashkiv, L. B.; Donlin, L. T. Regulation of type I interferon responses. *Nat Rev Immunol* **2014**, *14* (1): 36-49. DOI: 10.1038/nri3581
203. Platanitis, E.; Decker, T. Regulatory Networks Involving STATs, IRFs, and NFκB in Inflammation. *Front Immunol* **2018**, *9* 2542. DOI: 10.3389/fimmu.2018.02542
204. Nathan, J. A.; Lehner, P. J. The trafficking and regulation of membrane receptors by the RING-CH ubiquitin E3 ligases. *Exp Cell Res* **2009**, *315* (9): 1593-1600. DOI: 10.1016/j.yexcr.2008.10.026
205. Sivan, G.; Ormanoglu, P.; Buehler, E. C.; Martin, S. E.; Moss, B. Identification of Restriction Factors by Human Genome-Wide RNA Interference Screening of Viral Host Range Mutants Exemplified by Discovery of SAMD9 and WDR6 as Inhibitors of the Vaccinia Virus K1L-C7L- Mutant. *mBio* **2015**, *6* (4): e01122. DOI: 10.1128/mBio.01122-15
206. Park, C.; Walsh, D. RACK1 Regulates Poxvirus Protein Synthesis Independently of Its Role in Ribosome-Based Stress Signaling. *J Virol* **2022**, *96* (18): e0109322. DOI: 10.1128/jvi.01093-22
207. Rahman, M. M.; Bagdassarian, E.; Ali, M. A. M.; McFadden, G. Identification of host DEAD-box RNA helicases that regulate cellular tropism of oncolytic Myxoma virus in human cancer cells. *Sci Rep* **2017**, *7* (1): 15710. DOI: 10.1038/s41598-017-15941-1
208. Beerli, C.; Yakimovich, A.; Kilcher, S.; Reynoso, G. V.; Fläschner, G.; Müller, D. J.; Hickman, H. D.; Mercer, J. Vaccinia virus hijacks EGFR signalling to enhance virus spread through rapid and directed infected cell motility. *Nat Microbiol* **2019**, *4* (2): 216-225. DOI: 10.1038/s41564-018-0288-2

209. Soday, L.; Lu, Y.; Albarnaz, J. D.; Davies, C. T. R.; Antrobus, R.; Smith, G. L.; Weekes, M. P. Quantitative Temporal Proteomic Analysis of Vaccinia Virus Infection Reveals Regulation of Histone Deacetylases by an Interferon Antagonist. *Cell Rep* **2019**, *27* (6): 1920-1933.e7. DOI: 10.1016/j.celrep.2019.04.042
210. Maluquer de Motes, C.; Schiffner, T.; Sumner, R. P.; Smith, G. L. Vaccinia virus virulence factor N1 can be ubiquitylated on multiple lysine residues. *J Gen Virol* **2014**, *95* (Pt 9): 2038-2049. DOI: 10.1099/vir.0.065664-0
211. Yang, N. et al. Vaccinia E5 is a major inhibitor of the DNA sensor cGAS. *Nat Commun* **2023**, *14* (1): 2898. DOI: 10.1038/s41467-023-38514-5
212. MAGEE, W. E.; SAGIK, B. P. The synthesis of deoxyribonucleic acid by HeLa cells infected with vaccinia virus. *Virology* **1959**, *8* (1): 134-137. DOI: 10.1016/0042-6822(59)90029-7
213. MAGEE, W. E.; SHEEK, M. R.; BURROUS, M. J. The synthesis of vaccinial deoxyribonucleic acid. *Virology* **1960**, *11* 296-299. DOI: 10.1016/0042-6822(60)90070-2
214. KIT, S.; DUBBS, D. R. Biochemistry of vaccinia-infected mouse fibroblasts (strain L-M). I. Effects on nucleic acid and protein synthesis. *Virology* **1962**, *18* 274-285. DOI: 10.1016/0042-6822(62)90014-4
215. Jungwirth, C.; Launer, J. Effect of poxvirus infection on host cell deoxyribonucleic acid synthesis. *J Virol* **1968**, *2* (5): 401-408. DOI: 10.1128/JVI.2.5.401-408.1968
216. Martin, C. K.; Samolej, J.; Olson, A. T.; Bertoli, C.; Wiebe, M. S.; de Bruin, R. A. M.; Mercer, J. Vaccinia Virus Arrests and Shifts the Cell Cycle. *Viruses* **2022**, *14* (2): 431. DOI: 10.3390/v14020431
217. Postigo, A.; Ramsden, A. E.; Howell, M.; Way, M. Cytoplasmic ATR Activation Promotes Vaccinia Virus Genome Replication. *Cell Rep* **2017**, *19* (5): 1022-1032. DOI: 10.1016/j.celrep.2017.04.025
218. Zou, W.; Zhang, D. E. The interferon-inducible ubiquitin-protein isopeptide ligase (E3) EFP also functions as an ISG15 E3 ligase. *J Biol Chem* **2006**, *281* (7): 3989-3994. DOI: 10.1074/jbc.M510787200
219. Martín-Vicente, M.; Medrano, L. M.; Resino, S.; García-Sastre, A.; Martínez, I. TRIM25 in the Regulation of the Antiviral Innate Immunity. *Front Immunol* **2017**, *8* 1187. DOI: 10.3389/fimmu.2017.01187
220. Gack, M. U. et al. TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* **2007**, *446* (7138): 916-920. DOI: 10.1038/nature05732
221. Gack, M. U.; Kirchhofer, A.; Shin, Y. C.; Inn, K. S.; Liang, C.; Cui, S.; Myong, S.; Ha, T.; Hopfner, K. P.; Jung, J. U. Roles of RIG-I N-terminal tandem CARD and splice variant in TRIM25-mediated antiviral signal transduction. *Proc Natl Acad Sci U S A* **2008**, *105* (43): 16743-16748. DOI: 10.1073/pnas.0804947105
222. Thoresen, D.; Wang, W.; Galls, D.; Guo, R.; Xu, L.; Pyle, A. M. The molecular mechanism of RIG-I activation and signaling. *Immunol Rev* **2021**, *304* (1): 154-168. DOI: 10.1111/imr.13022
223. Gack, M. U.; Albrecht, R. A.; Urano, T.; Inn, K. S.; Huang, I. C.; Carnero, E.; Farzan, M.; Inoue, S.; Jung, J. U.; García-Sastre, A. Influenza A virus NS1 targets the ubiquitin ligase TRIM25 to evade recognition by the host viral RNA sensor RIG-I. *Cell Host Microbe* **2009**, *5* 439-449. DOI: 10.1016/j.chom.2009.04.006
224. Gupta, S.; Ylä-Anttila, P.; Sandalova, T.; Sun, R.; Achour, A.; Masucci, M. G. 14-3-3 scaffold proteins mediate the inactivation of trim25 and inhibition of the type I interferon response by herpesvirus deconjugases. *PLoS Pathog* **2019**, *15* (11): e1008146. DOI: 10.1371/journal.ppat.1008146
225. Hu, Y.; Li, W.; Gao, T.; Cui, Y.; Jin, Y.; Li, P.; Ma, Q.; Liu, X.; Cao, C. The Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Inhibits Type I Interferon Production by

Interfering with TRIM25-Mediated RIG-I Ubiquitination. *J Virol* **2017**, *91* (8): e02143-16. DOI: 10.1128/JVI.02143-16
226. Hyer, M. L. et al. A small-molecule inhibitor of the ubiquitin activating enzyme for cancer treatment. *Nat Med* **2018**, *24* (2): 186-193. DOI: 10.1038/nm.4474
227. Soucy, T. A. et al. An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer. *Nature* **2009**, *458* (7239): 732-736. DOI: 10.1038/nature07884
228. Morimoto, M.; Nishida, T.; Nagayama, Y.; Yasuda, H. Nedd8-modification of Cul1 is promoted by Roc1 as a Nedd8-E3 ligase and regulates its stability. *Biochem Biophys Res Commun* **2003**, *301* (2): 392-398. DOI: 10.1016/s0006-291x(02)03051-6
229. Kayagaki, N. et al. DUBA: a deubiquitinase that regulates type I interferon production. *Science* **2007**, *318* (5856): 1628-1632. DOI: 10.1126/science.1145918
230. Huang, O. W. et al. Phosphorylation-dependent activity of the deubiquitinase DUBA. *Nat Struct Mol Biol* **2012**, *19* (2): 171-175. DOI: 10.1038/nsmb.2206
231. Tesh, R. B. et al. Mayaro virus disease: an emerging mosquito-borne zoonosis in tropical South America. *Clin Infect Dis* **1999**, *28* (1): 67-73. DOI: 10.1086/515070
232. Long, K. C.; Ziegler, S. A.; Thangamani, S.; Hausser, N. L.; Kochel, T. J.; Higgs, S.; Tesh, R. B. Experimental transmission of Mayaro virus by Aedes aegypti. *Am J Trop Med Hyg* **2011**, *85* (4): 750-757. DOI: 10.4269/ajtmh.2011.11-0359
233. Smith, G. C.; Francy, D. B. Laboratory studies of a Brazilian strain of Aedes albopictus as a potential vector of Mayaro and Oropouche viruses. *J Am Mosq Control Assoc* **1991**, *7* (1): 89-93. DOI:
234. Ishida, R.; Cole, J.; Lopez-Orozco, J.; Fayad, N.; Felix-Lopez, A.; Elaish, M.; Luo, S. Y.; Julien, O.; Kumar, A.; Hobman, T. C. Mayaro Virus Non-Structural Protein 2 Circumvents the Induction of Interferon in Part by Depleting Host Transcription Initiation Factor IIE Subunit 2. *Cells* **2021**, *10* (12): 3510. DOI: 10.3390/cells10123510
235. Zhou, Y. et al. A comprehensive SARS-CoV-2-human protein-protein interactome reveals COVID-19 pathobiology and potential host therapeutic targets. *Nat Biotechnol* **2023**, *41* (1): 128-139. DOI: 10.1038/s41587-022-01474-0
236. Kumar, A.; Hou, S.; Airo, A. M.; Limonta, D.; Mancinelli, V.; Branton, W.; Power, C.; Hobman, T. C. Zika virus inhibits type-I interferon production and downstream signaling. *EMBO Rep* **2016**, *17* (12): 1766-1775. DOI: 10.15252/embr.201642627
237. Auguste, A. J. et al. Evolutionary and Ecological Characterization of Mayaro Virus Strains Isolated during an Outbreak, Venezuela, 2010. *Emerg Infect Dis* **2015**, *21* (10): 1742-1750. DOI: 10.3201/eid2110.141660
238. Knight, J. D. R.; Choi, H.; Gupta, G. D.; Pelletier, L.; Raught, B.; Nesvizhskii, A. I.; Gingras, A. C. ProHits-viz: a suite of web tools for visualizing interaction proteomics data. *Nat Methods* **2017**, *14* (7): 645-646. DOI: 10.1038/nmeth.4330
239. Akhrymuk, I.; Lukash, T.; Frolov, I.; Frolova, E. I. Novel Mutations in nsP2 Abolish Chikungunya Virus-Induced Transcriptional Shutoff and Make the Virus Less Cytopathic without Affecting Its Replication Rates. *J Virol* **2019**, *93* (4): e02062-18. DOI: 10.1128/JVI.02062-18
240. Salonen, A.; Ahola, T.; Kääriäinen, L. Viral RNA replication in association with cellular membranes. *Curr Top Microbiol Immunol* **2005**, *285* 139-173. DOI: 10.1007/3-540-26764-6_5
241. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **2003**, *13* (11): 2498-2504. DOI: 10.1101/gr.1239303
242. Luo, S. Y. et al. Identification of Human Host Substrates of the SARS-CoV-2 M^{pro} and PL^{pro} Using Subtiligase N-Terminomics. *ACS Infect Dis* **2023**, *9* (4): 749-761. DOI: 10.1021/acsinfecdis.2c00458

243. Hsieh, J. J.; Nofziger, D. E.; Weinmaster, G.; Hayward, S. D. Epstein-Barr virus immortalization: Notch2 interacts with CBF1 and blocks differentiation. *J Virol* **1997**, *71* (3): 1938-1945. DOI: 10.1128/JVI.71.3.1938-1945.1997
244. Kohlhof, H.; Hampel, F.; Hoffmann, R.; Burtscher, H.; Weidle, U. H.; Hölzel, M.; Eick, D.; Zimber-Strobl, U.; Strobl, L. J. Notch1, Notch2, and Epstein-Barr virus-encoded nuclear antigen 2 signaling differentially affects proliferation and survival of Epstein-Barr virus-infected B cells. *Blood* **2009**, *113* (22): 5506-5515. DOI: 10.1182/blood-2008-11-190090
245. Choi, H.; Larsen, B.; Lin, Z. Y.; Breitkreutz, A.; Mellacheruvu, D.; Fermin, D.; Qin, Z. S.; Tyers, M.; Gingras, A. C.; Nesvizhskii, A. I. SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat Methods* **2011**, *8* (1): 70-73. DOI: 10.1038/nmeth.1541
246. da Silva Caetano, C. C. et al. Mayaro Virus Induction of Oxidative Stress is Associated With Liver Pathology in a Non-Lethal Mouse Model. *Sci Rep* **2019**, *9* (1): 15289. DOI: 10.1038/s41598-019-51713-9
247. Yan, H.; Krishnan, K.; Greenlund, A. C.; Gupta…, S. Phosphorylated interferon-alpha receptor 1 subunit (IFNaR1) acts as a docking site for the latent form of the 113 kDa STAT2 protein. *The EMBO …* **1996**, DOI:
248. Wang, Y. et al. A virus-induced conformational switch of STAT1-STAT2 dimers boosts antiviral defenses. *Cell Res* **2021**, *31* (2): 206-218. DOI: 10.1038/s41422-020-0386-6
249. Kim, M. S.; Zhong, J.; Pandey, A. Common errors in mass spectrometry-based analysis of post-translational modifications. *Proteomics* **2016**, *16* (5): 700-714. DOI: 10.1002/pmic.201500355
250. Yang, Y.; Qiao, L. Data-independent acquisition proteomics methods for analyzing post-translational modifications. *Proteomics* **2023**, *23* (7-8): e2200046. DOI: 10.1002/pmic.202200046
251. Nielsen, M. L.; Vermeulen, M.; Bonaldi, T.; Cox, J.; Moroder, L.; Mann, M. Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nat Methods* **2008**, *5* (6): 459-460. DOI: 10.1038/nmeth0608-459
252. Potel, C. M.; Lemeer, S.; Heck, A. J. R. Phosphopeptide Fragmentation and Site Localization by Mass Spectrometry: An Update. *Anal Chem* **2019**, *91* (1): 126-141. DOI: 10.1021/acs.analchem.8b04746
253. Farley, A. R.; Link, A. J. Identification and quantification of protein posttranslational modifications. *Methods Enzymol* **2009**, *463* 725-763. DOI: 10.1016/S0076-6879(09)63040-8
254. Sinitcyn, P. et al. Global detection of human variants and isoforms by deep proteome sequencing. *Nat Biotechnol* **2023**, *41* (12): 1776-1786. DOI: 10.1038/s41587-023-01714-x
255. Siuti, N.; Kelleher, N. L. Decoding protein modifications using top-down mass spectrometry. *Nat Methods* **2007**, *4* (10): 817-821. DOI: 10.1038/nmeth1097
256. Matía, A. et al. Spatio-temporal analysis of Vaccinia virus infection and host response dynamics using single-cell transcriptomics and proteomics. *bioRxiv* **2024**, 2024.01.13.575413. DOI: 10.1101/2024.01.13.575413
257. Johnson, B. S.; Chafin, L.; Farkas, D.; Adair, J.; Elhance, A.; Farkas, L.; Bednash, J. S.; Londino, J. D. MicroID2: A Novel Biotin Ligase Enables Rapid Proximity-Dependent Proteomics. *Mol Cell Proteomics* **2022**, *21* (7): 100256. DOI: 10.1016/j.mcpro.2022.100256
258. Duan, G.; Walther, D. The roles of post-translational modifications in the context of protein interaction networks. *PLoS Comput Biol* **2015**, *11* (2): e1004049. DOI: 10.1371/journal.pcbi.1004049
259. Minguez, P.; Parca, L.; Diella, F.; Mende, D. R.; Kumar, R.; Helmer-Citterich, M.; Gavin, A. C.; van Noort, V.; Bork, P. Deciphering a global network of functionally associated post-translational modifications. *Mol Syst Biol* **2012**, *8* 599. DOI: 10.1038/msb.2012.31

260. Plubell, D. L.; Käll, L.; Webb-Robertson, B. J.; Bramer, L. M.; Ives, A.; Kelleher, N. L.; Smith, L. M.; Montine, T. J.; Wu, C. C.; MacCoss, M. J. Putting Humpty Dumpty Back Together Again: What Does Protein Quantification Mean in Bottom-Up Proteomics. *J Proteome Res* **2022**, *21* (4): 891-898. DOI: 10.1021/acs.jproteome.1c00894
261. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596* (7873): 583-589. DOI: 10.1038/s41586-021-03819-2

Appendix A – Supporting information for Chapter 2: The role of viral protease-induced proteolysis in SARS-CoV-2

Preface

Appendix A includes supporting information for Chapter 2: The role of viral protease-induced proteolysis in SARS-CoV-2, and has been published as part of the manuscript:

> Luo, S. Y.; Moussa, E. W.; Lopez-Orozco, J.; Felix-Lopez, A.; Ishida, R.; Fayad, N.; Gomez-Cardona, E.; Wang, H.; Wilson, J. A.; Kumar, A.; Hobman, T. C.; Julien, O. Identification of Human Host Substrates of the SARS-CoV-2 Mpro and PLpro Using Subtiligase N-Terminomics. ACS Infect Dis 2023, 9 (4): 749-761.

List of the supporting figures:

A.1 Supporting information for Chapter 2

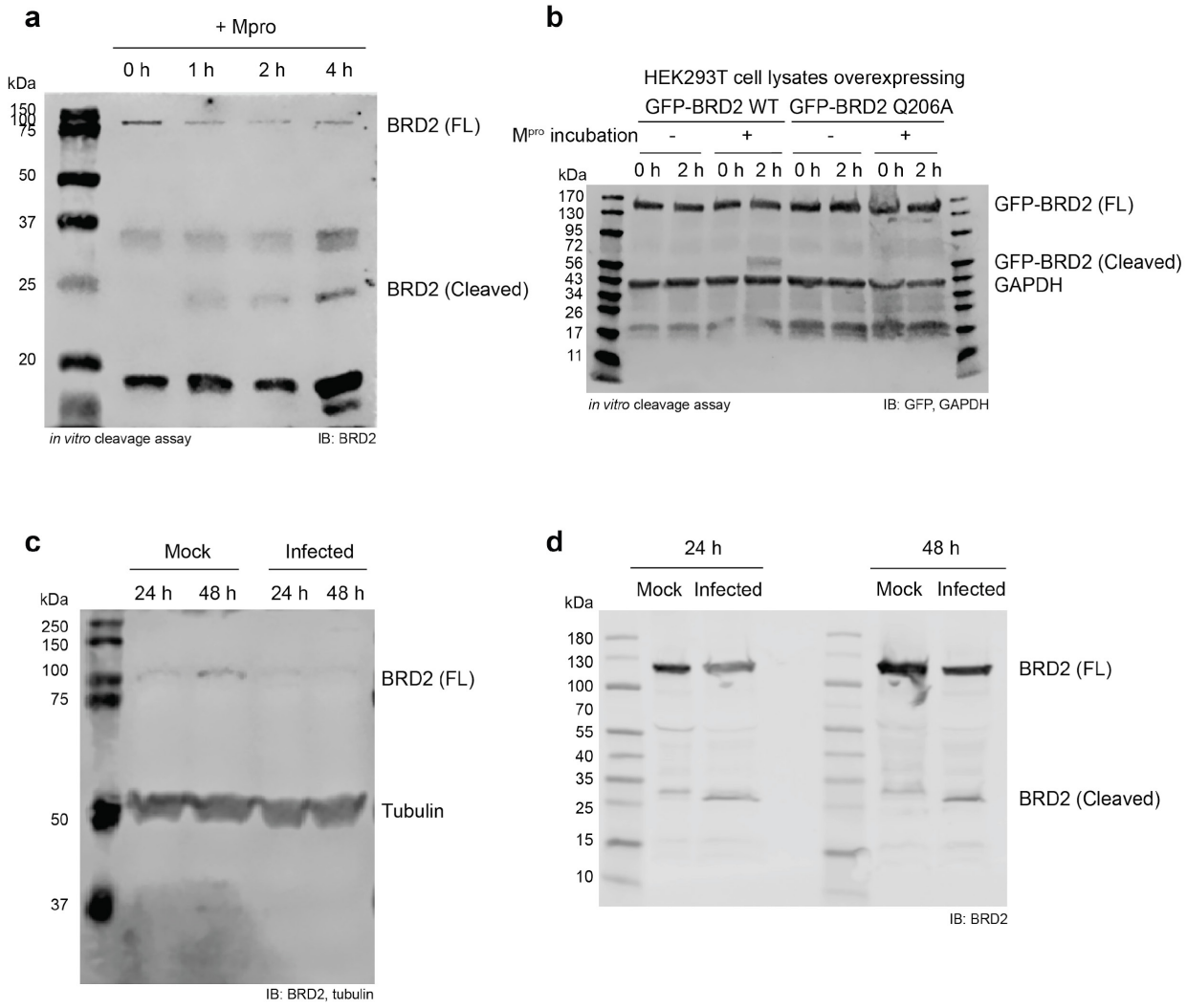


Figure A.1. **Full immunoblot images of Fig. 2.7. Proteolysis of BRD2 by M^{pro} *in vitro* and in SARS-CoV-2 infected cells. a**) BRD2 was cleaved by recombinant M^{pro} in Jurkat cell lysates. Jurkat cell lysates were incubated with recombinant M^{pro} for 0-4 hours, and immunoblotted against BRD2. A cleavage product at 23 kDa appeared with incubation time as the full length BRD2 level decreased. **b**) GFP-BRD2 WT and mutant Q206A overexpression in HEK293T-ACE2 and *in vitro* cleavage by recombinant SARS-CoV-2 M^{pro}. HEK293T-ACE2 cells overexpressing GFP-BRD2 were lysed, and the cell lysates were incubated with M^{pro} for 2 hours and immunoblotted against GFP. Cleavage was only observed with GFP-BRD2 WT. Depletion of full-length BRD2 was also observed in SARS-CoV-2 infected **c**) A549-ACE2 and **d**) HEK293T-ACE2 at 24 and 48 h.p.i.

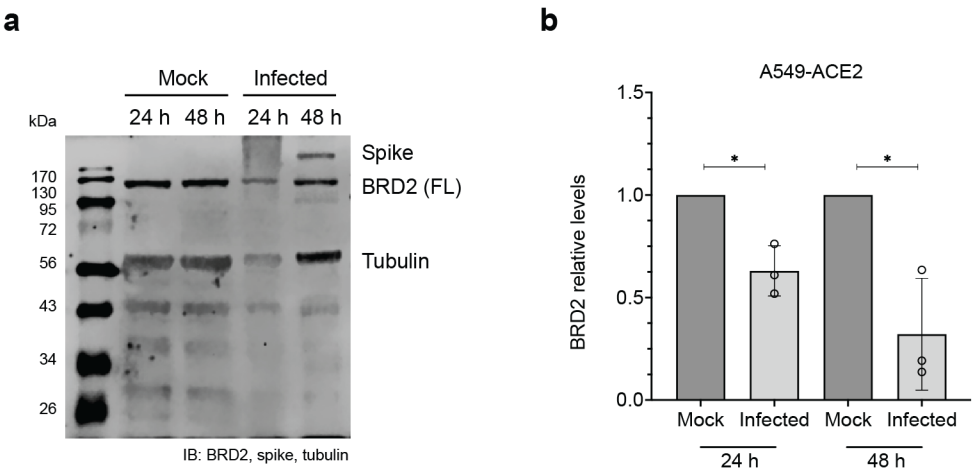


Figure A.2. Additional immunoblot of SARS-CoV-2 infected A549-ACE2 lysates and BRD2 quantification. a) Another representative replicate was performed in addition to the blot presented in Figure 3 of the main manuscript, for a total of n=3, biological replicates. **b**) Quantification of BRD2 levels at 24 h and 48 h after infection with SARS-CoV-2 compared to mock. BRD2 levels measured were 0.6 ± 0.1 (24 h) and 0.3 ± 0.3 (48 h), with a *$p < 0.05$ using Student's t-test.

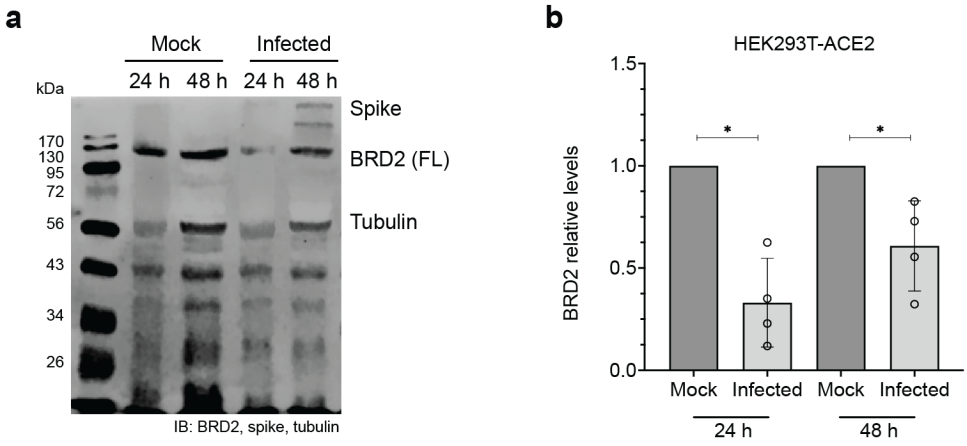


Figure A.3. **Additional immunoblot of SARS-CoV-2 infected HEK293T-ACE2 lysates and BRD2 quantification. a**) Another representative replicate was performed in addition to the blot presented in Figure 3 of the main manuscript, for a total of n=4, biological replicates. **b**) Quantification of BRD2 levels at 24 h and 48 h after infection with SARS-CoV-2 compared to mock. BRD2 levels measured were 0.3 ± 0.2 (24 h) and 0.6 ± 0.2 (48 h), with a *$p < 0.05$ using Student's t-test.

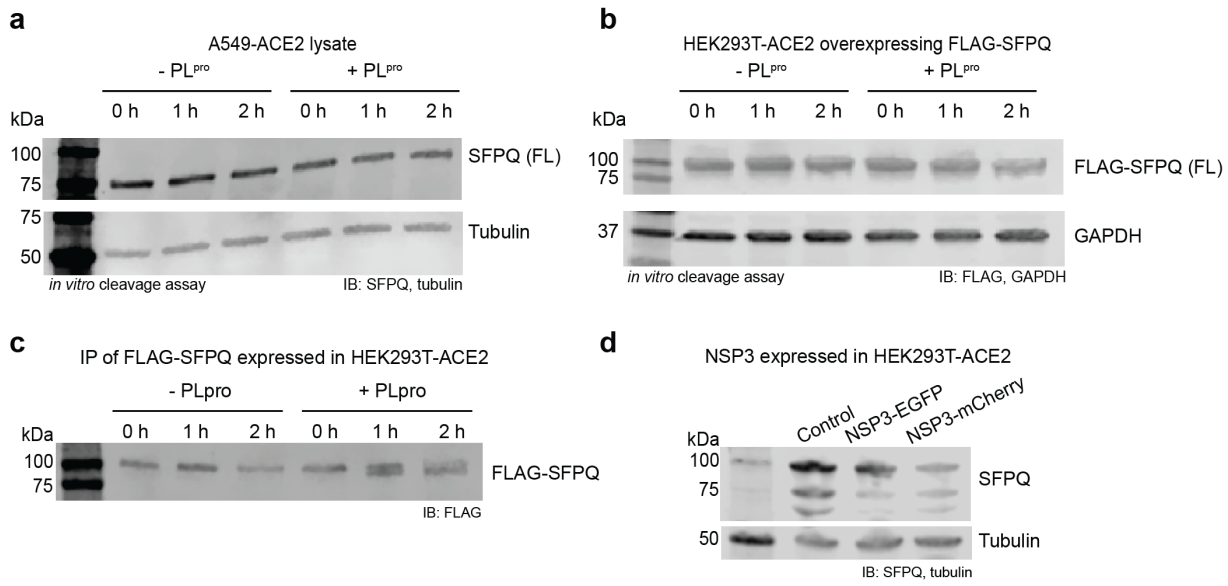


Figure A.4. Additional investigation of cleavage of SFPQ by PL^{pro} *in vitro* (n=2, biological replicates)**.** **a**) Uninfected A549-ACE2 cell lysates were incubated with PL^{pro} and SFPQ cleavage by PL^{pro} could not be detected using immunoblotting. **b**) Cleavage of overexpressed FLAG-tagged SFPQ in HEK293T-ACE2 cells by PL^{pro} was also not detected on immunoblot. **c**) A potential cleavage product was observed when incubating immunoprecipitated FLAG-SFPQ with SARS-CoV-2 PL^{pro}. **d**) Expression of full length NSP3 in HEK293T-ACE2 did not show distinct cleavage of SFPQ compared to the control.

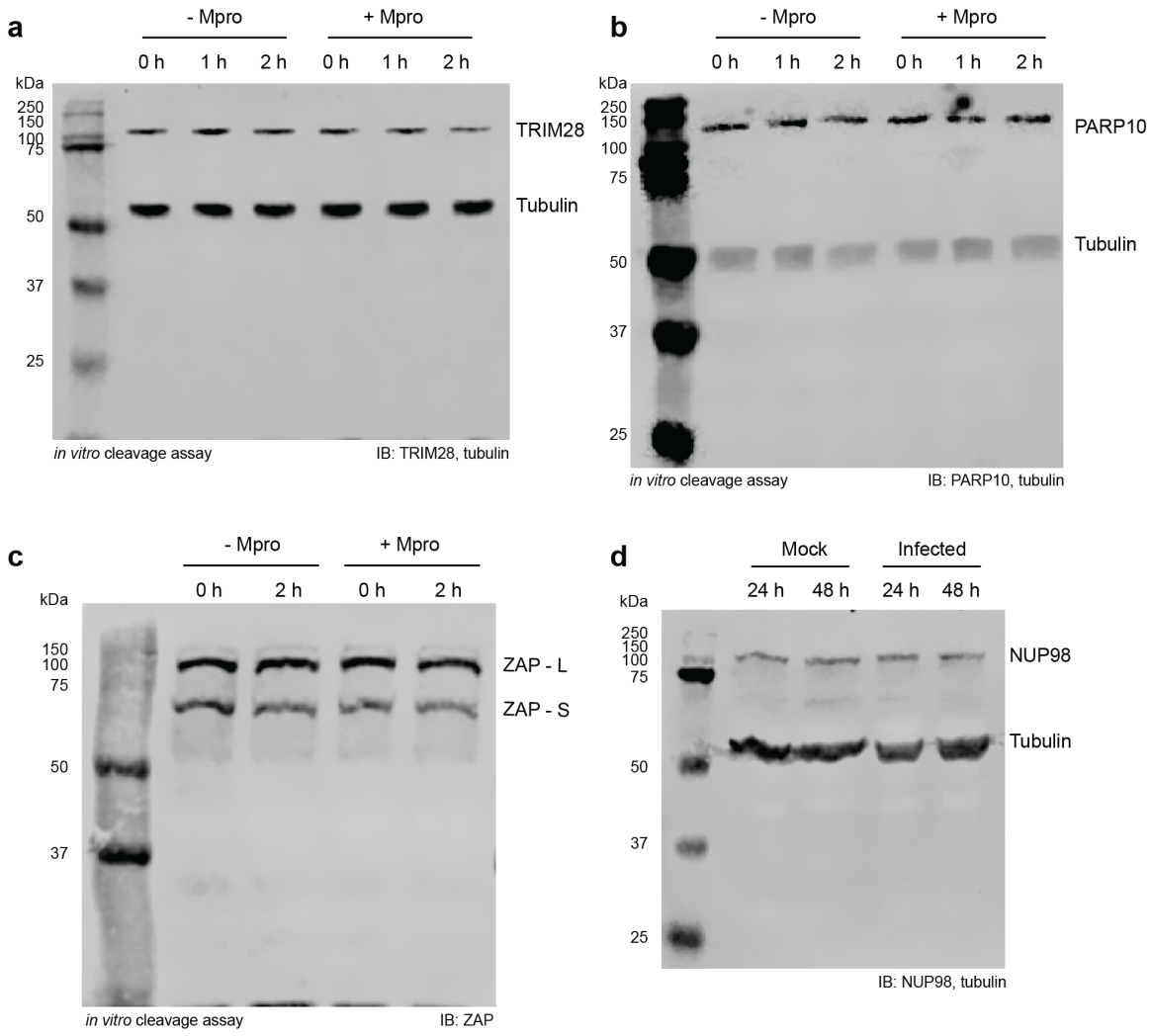


Figure A.5. Substrate proteolysis by M^{pro} was not detectable by immunoblot for **a**) *in vitro* cleavage assays of TRIM28 in Jurkat lysates; **b**) *in vitro* cleavage assays of PARP10 in Jurkat lysates; **c**) *in vitro* cleavage assays of ZAP in A549 lysates; and **d**) endogenous NUP98 level in infected A549-ACE2 cells. This suggests that these targets can be cleaved by M^{pro}, but only at a low level detectable only by mass spectrometry.

Appendix B – Protein-level enrichment for PTM detection

Preface

Appendix B includes method optimizations for protein-level enrichment to detect diGlycyl sites on a target protein, and contains preliminary unpublished results for the project in the manuscript:

> Delyea C., Forster M., Luo S., Dubrule B.E., Julien O., and Bhavsar A.P. The Salmonella effector SspH2 facilitates spatially selective ubiquitination of NOD1 to enhance inflammatory signaling. *Biochemistry*. (accepted)

B.1 Summary

Salmonella enterica serovar Typhimurium (*S*. Typhimurium) is a Gram-negative facultative intracellular pathogen. It is a major cause of diarrhoeal disease worldwide, and results in 33 million healthy life years being lost yearly. (World Health Organization, 2018) As part of its pathogenesis, *S*. Typhimurium delivers effector proteins into host cells. One effector is SspH2, a member of the novel E3 ubiquitin ligase family, interacts with, and enhances, NOD1 pro-inflammatory signaling, though the underlying mechanisms are unclear. We aimed to use mass spectrometry analysis to uncover the ubiquitination sites on host NOD1 by *Salmonella* SspH2. We overexpressed FLAG-tagged NOD1 and HA-tagged SspH2 in HEK293T cells, and tested various methods to optimize the detection of ubiquitination sites on NOD1.

B.2 Evaluation of ubiquitin overexpression, protein-level enrichment, and bead selection

We first would like to evaluate whether the co-overexpression of ubiquitin and the protein-level enrichment steps were necessary to detect diGlycyl sites on NOD1. Additionally, we compared the target protein coverage and the number of associating diGlycyl sites identified using two immunoprecipitation (IP) methods. We tested pre-conjugated commercial anti-FLAG magnetic beads (Pierce), and alternatively, we pre-incubated a mouse anti-FLAG antibody with Protein G magnetic beads, and proceeded to binding in lysates (direct IP).

Table B.1 provided the summary of results. We confirmed that we were able to detect numbers of proteins and peptides corresponding to the sample complexities of IP and lysates, and we detected the overexpressed *Salmonella* E3 ligase SspH2. Next, we performed more targeted analysis on the NOD1 sequence coverage and its diGlycyl sites (**Table B.2**). We found that the overexpression of ubiquitin did not affect the total number of IPed proteins or peptides (**Table B.1**), but drastically increased the number of diGlycyl sites on NOD1 (**Table B.2**).

Moreover, the anti-FLAG enrichment significantly enhanced the identification of diGlycyl sites on NOD1 (**Table B.2**), even when its coverage was about the same in the IP and lysates. Lastly, although we were able to detect more proteins and peptides using Pierce anti-FLAG in the IP samples (**Table B.1**), direct IP using anti-FLAG antibody and Protein G beads in ubiquitin-overexpressing lysates resulted in the highest number of diGlycyl sites identified on NOD1 (**Table B.2**). Therefore, we showed that protein-level enrichment was an effective alternative in characterizing PTMs on a target protein with these optimizations.

It was however important to note that, since the enrichment was protein-level, both diGlycyl and unmodified NOD1 peptides were pulled down (**Fig. B.1**). Given that the identification of low-abundance modified peptides could be obscured by their unmodified counterparts, this might partially elucidate why the identification of diGlycyl sites on the target protein required specific conditions and approaches.

Overexpressed protein	Sample type	Beads	# Proteins	# Peptides	% Coverage of SspH2 (# peptides)
NOD1+SspH2	Manual IP	Pierce	1614	6864	68.9% (45)
	Manual IP	Protein G + antibody	954	3169	79.9% (48)
	Lysates	-	2234	19437	79.8% (49)
NOD1+SspH2+Ub	Manual IP	Pierce	1662	6656	62.2% (44)
	Manual IP	Protein G + antibody	1144	3938	80.1% (49)
	Lysates	-	2014	18074	75.6% (44)

Table B.1. Summary of results comparing 1) effects of overexpressing ubiquitin; 2) effects of protein-level enrichment; 3) effects of using different beads in identifying proteins and peptides. Pierce: anti-FLAG (Anti-DYKDDDDK) magnetic agarose (Pierce, A36797); Protein G + antibody: Dynabeads Protein G beads (Invitrogen) and mouse α-FLAG (Sigma; M). Samples were run on the Orbitrap Fusion Lumos.

Overexpressed protein	Sample type	Beads	# NOD1 peptides	NOD1 coverage	# diGlycyl sites on NOD1	Lys position
NOD1+SspH2	Manual IP	Pierce	56	65.8%	1	K473
	Manual IP	Protein G + antibody	52	59.2%	2	K473 K582
	Lysates	-	43	58.6%	0	-
NOD1+SspH2+Ub	Manual IP	Pierce	53	59.1%	3	K473 K746 K666
	Manual IP	Protein G + antibody	66	67.9%	12	K473 K858 K746 K812 K756 K778 K754 K142 K598 K809 K666 K70
	Lysates	-	44	58.7%	1	K746

Table B.2. Effects of experimental conditions on profiling diGlycyl sites on NOD1. Pierce: anti-FLAG (Anti-DYKDDDDK) magnetic agarose (Pierce, A36797); Protein G + antibody: Dynabeads Protein G beads (Invitrogen) and mouse α-FLAG (Sigma; M). Samples were run on the Orbitrap Fusion Lumos.

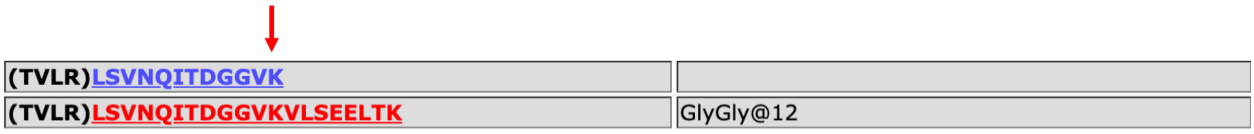


Figure B.1. Protein-level enrichment yielded both modified and unmodified peptides from the same protein. The diGly- state of K co-existed with unmodified K; for instance, the peptides above were detected in the same sample (Nod1+SspH2+Ub, Protein G + anti-FLAG antibody).

B.3 Effects of automated enrichment

We have also tested manual and automated anti-FLAG protein enrichment using KingFisher Duo (ThermoFisher) with anti-FLAG antibody and Dynabeads Protein G, under conditions optimized above. C.D. generated a catalytically dead mutant SspH2 C580A as a negative control. The automated anti-FLAG IP greatly enhanced the capture of FLAG-NOD1 (**Fig. B.2**). Data analysis also revealed the automated IP samples contained overall more

proteins, and higher NOD1 coverage with increased diGlycyl site identification (**Table B.3**). All samples were searched against the SspH2 WT/C580A sequence and the active site mutation was confirmed (**Fig. B.3**).

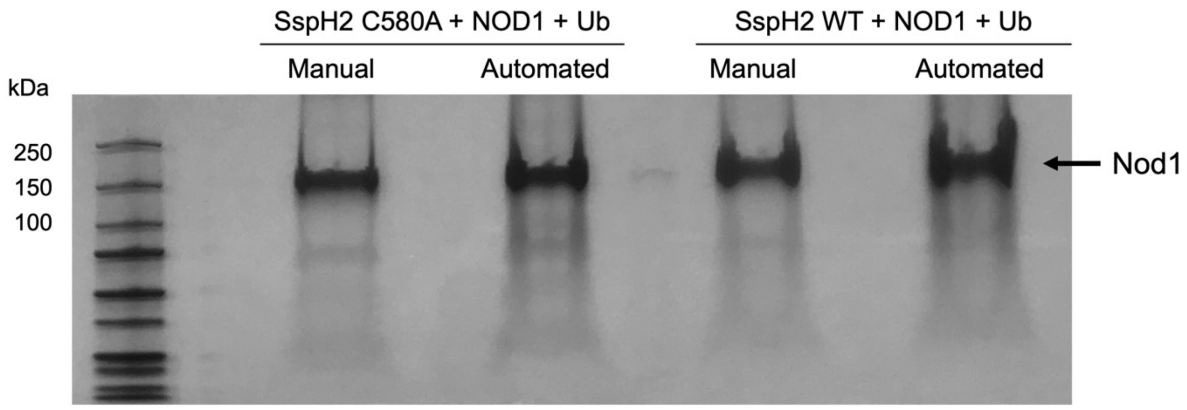


Figure B.2. Enhanced FLAG-NOD1 capture using automated IP. Coomassie-stained SDS-PAGE of eluted anti-FLAG IP in HEK293T lysates overexpressing the indicated constructs.

Sample		# proteins	# peptides	% coverage on NOD1	# NOD1 peptides	# glygly sites on NOD1	# glygly sites in the pulldown
NOD1+Ub+ SspH2 C580A	Manual	637	3851	76.50%	99	14	83
	Auto	905	5411	78.50%	118	27	131
NOD1+Ub+ SspH2 WT	Manual	602	3429	73.30%	93	12	78
	Auto	933	5420	76.00%	111	22	104

Table B.3. Enhanced identification on overall IPed proteins/peptides, and increased NOD1 coverage and diGlycyl sites using automated direct IP. All IPs were performed with anti-FLAG antibody and Dynabeads Protein G.

Figure B.3. Detection of SspH2 sequences corresponding to WT and C580A. Red: SspH2 peptides found in the IP. The residue at 580 was indicated by the green arrow.

Appendix C – Optimizations of phosphoproteomics workflow

Preface

Appendix C includes method optimizations for the phosphoproteomics workflow used in the unpublished projects, and contains some collaborative work with Jason Lane and Dr. Amit Bhavsar (University of Alberta).

C.1 Summary

Phosphorylation is a critical PTM that plays a vital role in regulating numerous cellular processes, including signal transduction, cell division, and metabolic control. Phosphoproteomics coupled with LC-MS/MS analysis enables the characterization and quantification of the phosphoproteome in a biological sample, elucidating the dynamic changes in protein phosphorylation states and sites in response to various stimuli, and providing insights into cellular pathways and disease mechanisms. Enrichment strategies such as immobilized metal affinity chromatography (IMAC) are employed to selectively isolate phosphopeptides from the proteome. Here, I investigated variations in each step of the phosphoproteomics workflow to improve enrichment and detection of phosphopeptides.

C.2 Effects of sample lysis and desalting methods on identifications

We tested two alternative approaches for sample preparation in phosphoproteomics workflow based on Jersie-Christensen *et al.* with some modifications (**Fig. C.1**). (1) Cells were lysed either on ice in classic RIPA buffer [50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS and protease and phosphatase inhibitors (ThermoFisher, Halt)] or by boiling in GdnHCl (6 M guanidine hydrochloride and 100 mM Tris-HCl, pH 8.5). The RIPA sample underwent acetone precipitation, reduction and alkylation, LysC and trypsin digestion, and peptide desalting using ProTrap XG (Allumiqs). The GdnHCl protein extract was directly reduced, alkylated and digested by LysC and trypsin in solution. The peptides were desalted using SOLA HRP SPE cartridges (ThermoFisher, #03-150-391). Desalted peptides from both samples proceeded to the same Ti-IMAC HP (MagReSyn) phosphopeptide enrichment automated using KingFisher Duo Prime. Both the flow-through (total peptides; TP) and the eluted fractions (phosphopeptides; PP) were desalted again and analyzed on LC-MS/MS.

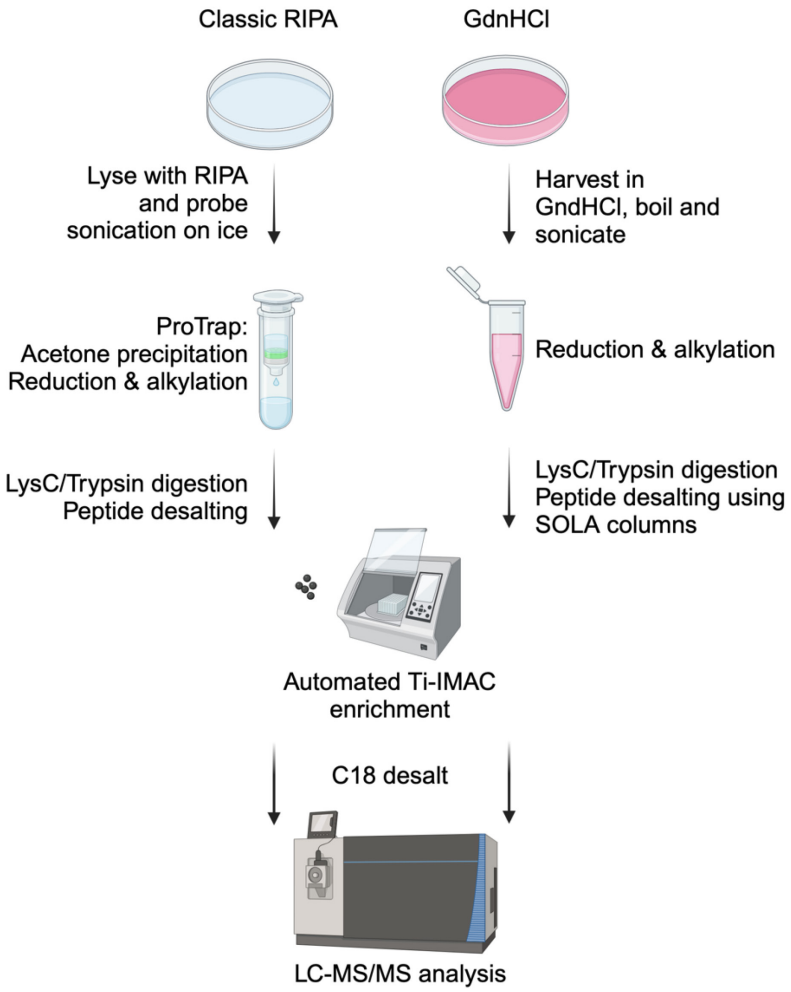


Figure C.1. Two alternative phosphoproteomics workflows featuring different lysis buffers and desalting methods.

We assessed the number of peptides and phosphopeptides in the samples processed using the two workflows. We found that the use of ProTrap resulted in no peptide identification after phosphopeptide enrichment (**Fig. C.2**; left), although other non-phosphorylated peptides were identified in the unbound fractions (**Fig. C.3**; left). On the other hand, the sample with heated GndHCl lysis and SOLA-column desalting yielded sufficient numbers of phosphopeptides in the PP fraction (**Fig. C.2**; right), and higher numbers of unmodified peptides in both PP and TP (**Fig. C.2**; right). Therefore, all subsequent experiments were performed with the GdnHCl workflow indicated in **Fig. C.1**.

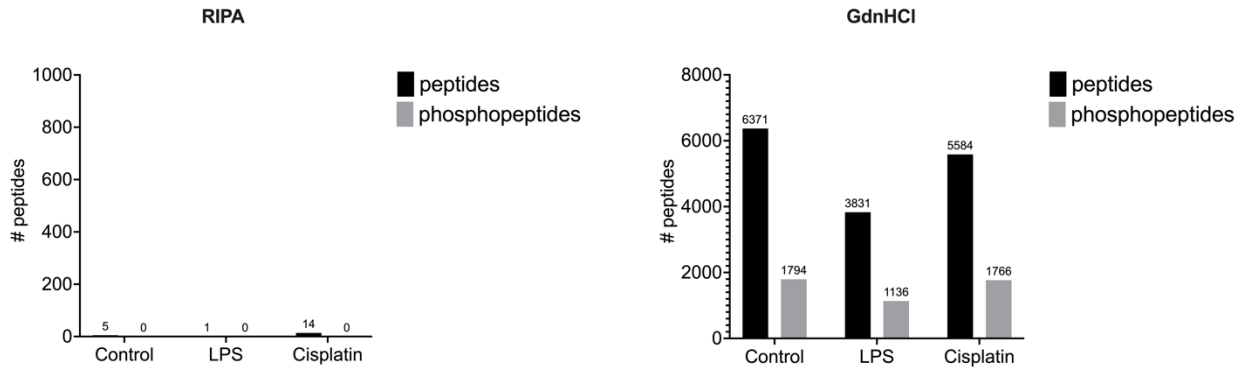


Figure C.2. Number of identified peptides and phosphopeptides in elution after Ti-IMAC enrichment. Cells were cultured and treated with the PBS (control) or the indicated molecule by J.L.

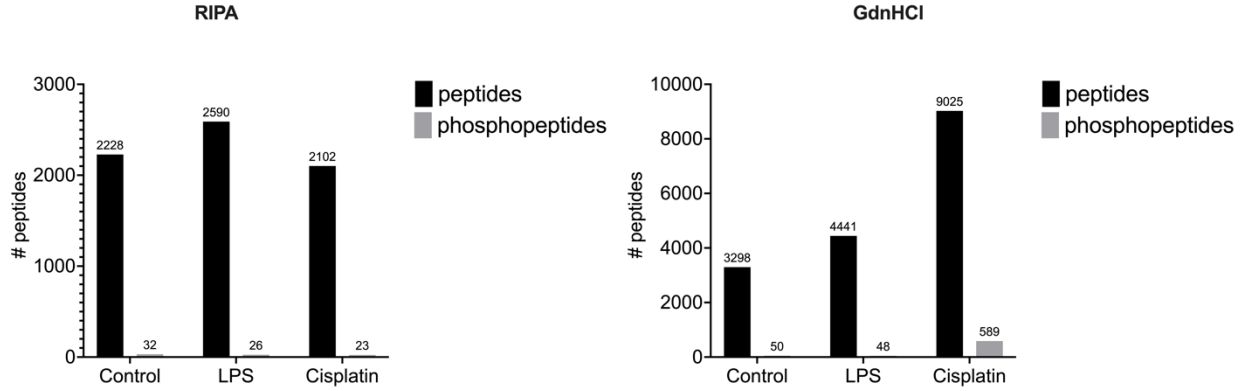


Figure C.3. Number of identified peptides and phosphopeptides in the flow-through after Ti-IMAC enrichment. Cells were cultured and treated with the PBS (control) or the indicated molecule by J.L.

C.3 Types of phospho beads and enrichment buffer

We further investigated the use of varying phosphopeptide enrichment beads and conditions with 270 µg Jurkat peptides per sample. For Ti-IMAC beads, we used 25 µL of 20 µg/µL beads per 500 µg peptides, pre-washed beads 3X with binding buffer [80% acetonitrile (ACN), 0.1% trifluoroacetic acid (TFA)], and proceeded to peptide binding at RT for 30 min. The beads were washed 3X with the binding buffer, eluted 2X with 50% ACN, 2.5% NH_4OH, and neutralized with 75% ACN, 10% FA. We also tested another enrichment protocol with increased TFA concentration in the binding buffer to 6% and more thorough washing. (2) The beads were pre-washed 3X with the binding buffer (80% ACN, 6% TFA) and were incubated with peptides in

the binding buffer with mixing at RT for 20 min. Beads were washed 3X with the binding buffer, 1X with 80% ACN, 1X with 80% ACN, 0.5 M glycolic acid, and 1X with 80% ACN. The peptides were eluted 2X by 50% ACN, 1% NH_4OH.

We additionally evaluated TiO_2 beads (Titansphere, GL Sciences) and followed the enrichment protocol by Jersie Christensen *et al.*, (1) where the enrichment was carried out in 80% ACN, 12% TFA, and beads were washed with increasing concentrations of ACN (10%, 40%, and 60%) and 6% TFA, and finally eluted first with 5% NH_4OH, and then with 25% ACN, 10% NH_4OH.

We found that enrichment with Ti-IMAC beads and 6% TFA yielded the highest numbers of phosphopeptides and phosphoproteins (**Fig. C.4**). Notably, there was high overlap among the peptides identified in the PP fractions among the three conditions (**Fig. C.5a**), especially in phosphopeptides (**Fig. C.5b**), with exclusive gain in Ti-IMAC 6% TFA.

Lastly, we explored the phosphopeptide enrichment preference of Ti-IMAC beads between 0.1% and 6% TFA. We found that the phosphopeptide abundances in the two samples correlated (**Fig. C.6**; r = 0.80). It is noteworthy that the unique phosphopeptides in Ti-IMAC 6% TFA also did not exclusively contain multi-phosphorylated peptides.

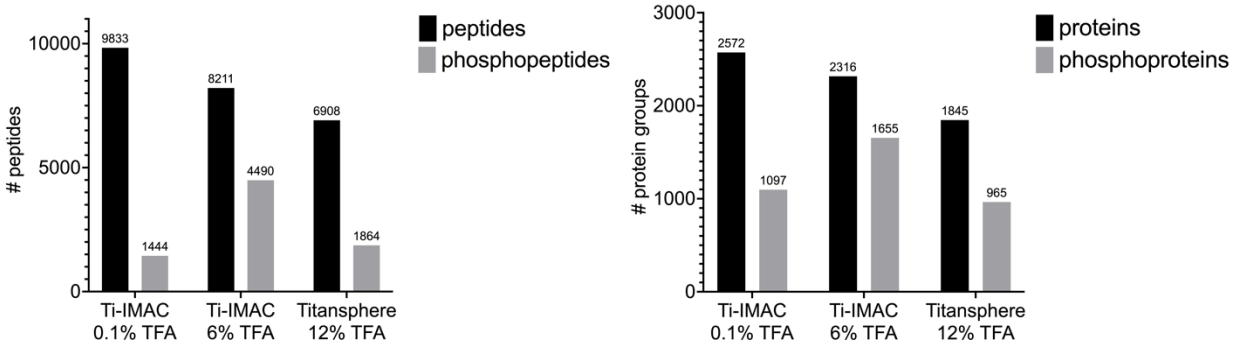


Figure C.4. Number of peptides and phosphopeptides and the corresponding proteins enriched with varying beads and binding conditions. All enrichment steps were automated by KingFisher Duo Prime, with 270 µg starting peptides in Jurkat (peptide concentration measured by Nanodrop).

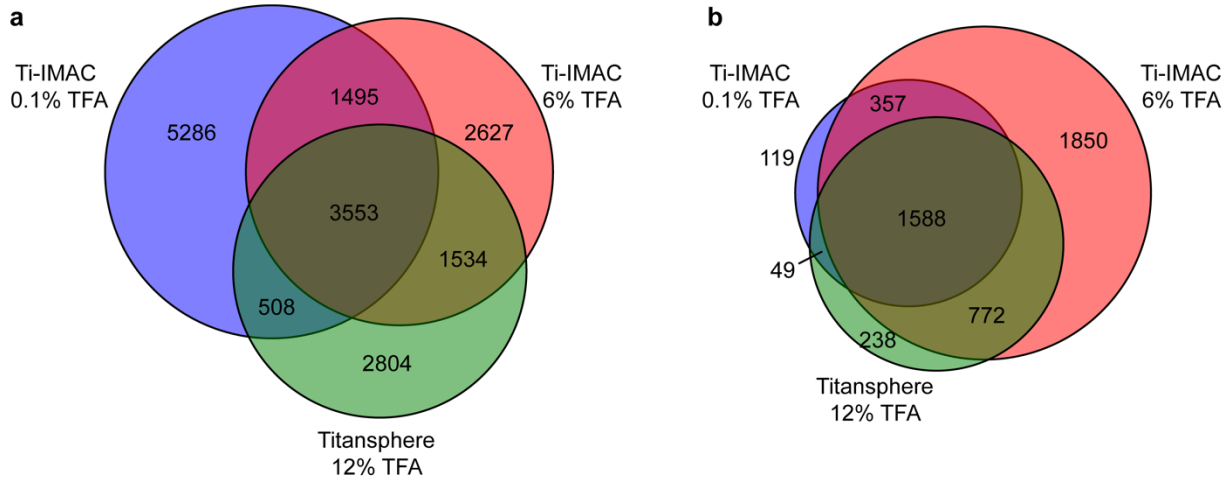


Figure C.5. Overlap among the enrichment conditions in the identification of **a**) peptides and **b**) phosphopeptides. All enrichment steps were automated by KingFisher Duo Prime, with 270 µg starting peptides in Jurkat (peptide concentration measured by Nanodrop).

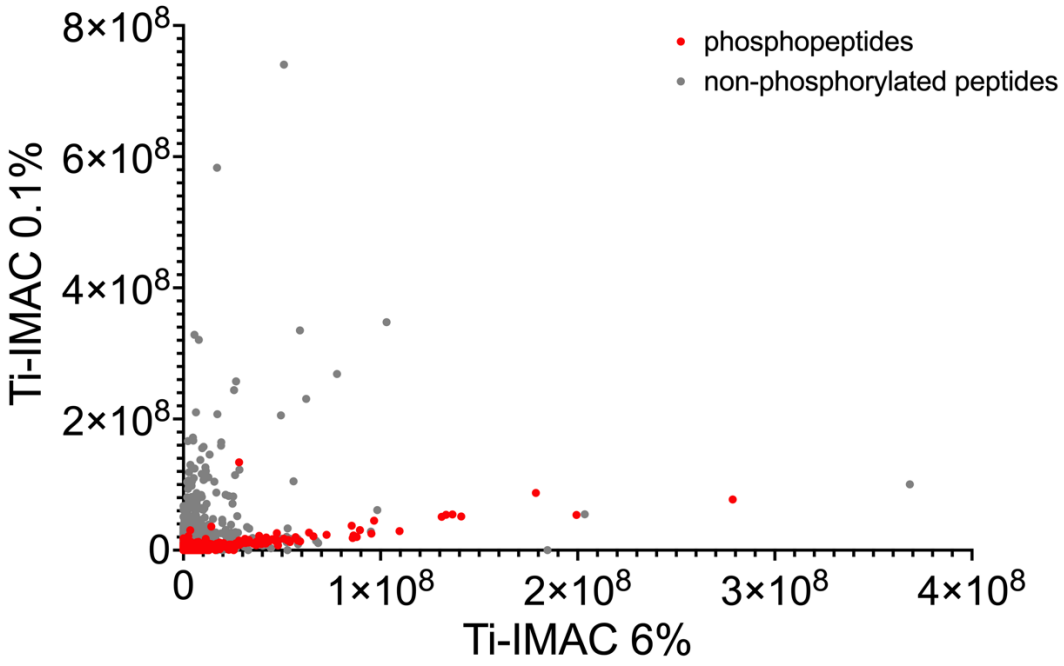


Figure C.6. Correlation between phosphopeptides identified in Ti-IMAC 6% TFA (x-axis) and Ti-IMAC 0.1% TFA (y-axis). The abundances of quantified peptides enriched by Ti-IMAC beads were plotted using a scatter plot. The Pearson correlation coefficient (r) showed very strongly positive correlation between the two conditions for phosphopeptides (r = 0.80); r = 0.33 for non-phosphorylated peptides. Grey dots: non-phosphorylated peptides; red dots: phosphorylated peptides.

C.3 Effects of acquisition methods on phosphopeptide IDs

The parameter optimization on LC-MS/MS for phosphopeptide analysis is critical for their fragmentation and assignment. We found that in a tribrid mass spectrometer, using orbitrap for both MS1 and MS2 (OT-OT) ion measurements yielded better results than conventional methods of measuring MS1 ions in orbitrap and MS2 by ion trap (OT-IT) (**Fig. C.7**).

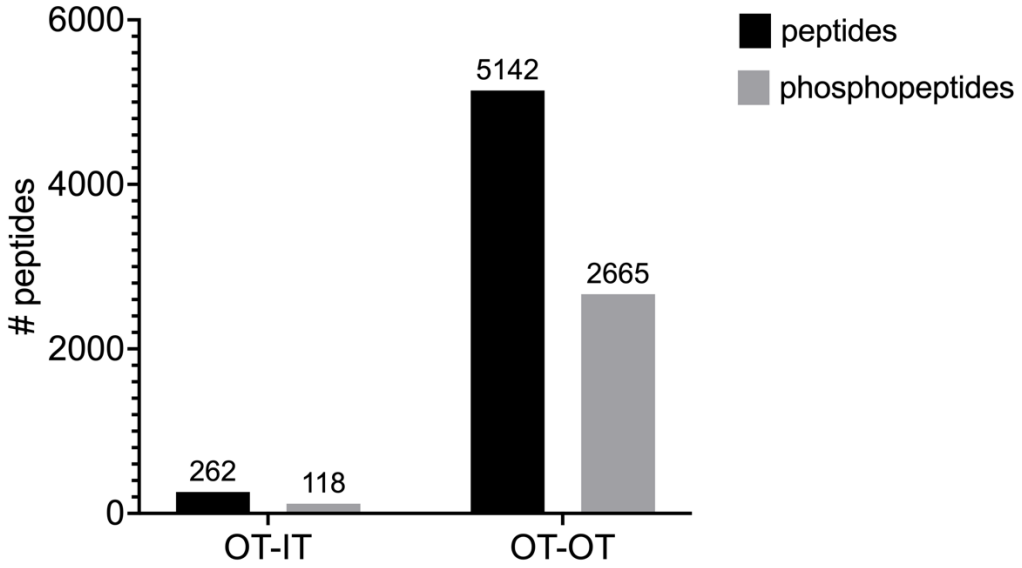


Figure C.7. Numbers of peptides and phosphopeptides identified using the above LC-MS/MS methods. Jurkat peptides were enriched with Ti-IMAC beads, and the same sample was injected twice, analyzed either with OT-IT (MS1 by orbitrap, MS2 by ion trap), or OT-OT (MS1 and MS2 both by orbitrap).

C.4 References

1. Jersie-Christensen, R. R.; Sultan, A.; Olsen, J. V. Simple and Reproducible Sample Preparation for Single-Shot Phosphoproteomics with High Sensitivity. *Methods Mol Biol* **2016**, *1355* 251-260. DOI: 10.1007/978-1-4939-3049-4_17

2. Muehlbauer, L. K.; Hebert, A. S.; Westphall, M. S.; Shishkova, E.; Coon, J. J. Global Phosphoproteome Analysis Using High-Field Asymmetric Waveform Ion Mobility Spectrometry on a Hybrid Orbitrap Mass Spectrometer. *Analytical Chemistry Anal. Chem.* **2020**, *92* (24): 15959-15967. DOI: 10.1021/acs.analchem.0c03415

Appendix D – Protein expression, purification, and activity assays of Chikungunya and Mayaro viral proteases

Preface

Appendix D includes collaborative work with Drs. Tom Hobman (University of Alberta), Jeanne Hardy (University of Massachusetts Amherst) and Anil Kumar (University of Saskatchewan). I modified some purification and assay conditions from Kristalle G. Cruz (University of Massachusetts Amherst).

D.1 Summary

Outbreaks of Zika, Dengue and West Nile virus have drawn the world's attention to the mosquito-borne viruses, and the scarcity of corresponding vaccines accentuated the significance of virus-host interactions research at the molecular level. The successful development of SARS-CoV-2 main protease inhibitor Paxlovid (Pfizer) as antiviral treatment has especially underscored the importance of viral protease research. In alphavirus, preliminary studies have shown that the viral protease plays a key role in suppressing cellular transcription and blocking interferon signaling in immune response during infection. (1,2) However, the specific cellular targets of these proteases remain uncharacterized. To identify host substrates cleaved by viral proteases in alphaviruses, we aimed to express and purify active Chikungunya virus (CHIKV) and Mayaro virus (MAYV) proteases. The viral proteases can be added to human cell lysates for N-terminomics labeling experiments and substrate characterization.

D.2 Expression and purification of CHIKVP

E. coli codon optimized sequence encoding the CHIKV NSP2 protease domain (CHIKVP) was cloned in the vector pET-15b(-) with N-terminal 6X His tag and cleavable TEV protease recognition sequence (His6-TEV-CHIKVP) by A.K. The plasmid was transformed in BL21(DE3)pLysS cells and cultured in 2X YT induced with 1 mM IPTG at OD_{600} = 0.7 incubated for 6 hours at 170 rpm and 15˚C. Cells were lysed in lysis buffer (50 mM Tris, pH 8.0, 300 mM NaCl, 5% glycerol) by Emulsiflex and the lysates were clarified by centrifugation at 40,000 x g for 4˚C at 30 min. The soluble protein extract then proceeded to Ni column purification using HisTrap (Cytiva), with wash buffer (50 mM Tris, pH 8.0, 300 mM NaCl, 5% glycerol, 50 mM imidazole) and elution buffer (50 mM Tris, pH 8.0, 300 mM NaCl, 5% glycerol, 300 mM imidazole). The eluted fractions containing His-TEV-CHIKVP were pooled and dialyzed overnight with 1:50 TEV protease:protein ratio in 2 L of dialysis buffer (50 mM Tris, pH 8.0, 5% glycerol, 1 mM TCEP) with a stirrer at 4˚C. The untagged CHIKVP was purified by a second Ni

column purification using the previous wash and elution buffers, and the flow-through and wash fractions containing the protein was collected.

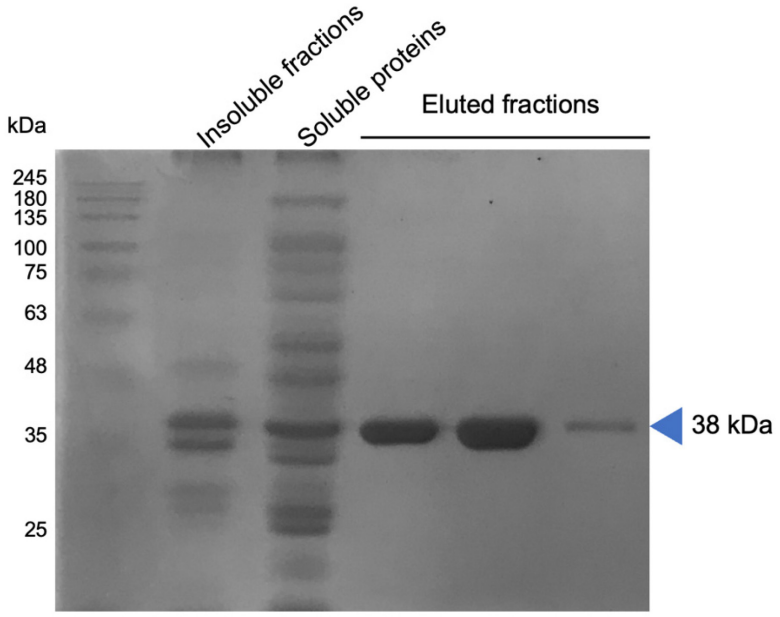


Figure D.1. Expression and purification of His-TEV-CHIKVP. His-TEV-CHIKVP was expressed in BL21(DE3)pLysS and purified by Ni column purification.

D.3 Activity assay of CHIKVP

Recombinant CHIKVP activity was monitored using fluorescence resonance energy transfer (FRET) probe substrate DABCYL-RAGGYIFS-Glu(EDANS)-NH2 (Biomatik) with λ_{ex} = 340 nm and λ_{em} = 490 nm. Under cell-free assay conditions, 2 μM of untagged CHIKVP was incubated with 30 μM substrate in assay buffer (50mM HEPES, pH 7.0, 1mM CHAPS) in 96-well opaque plate (Corning) (**Fig. D.2**). To perform activity assays in lysates, 5-10 million A549 cells were lysed in 1 mL assay buffer supplemented with protease inhibitors (5 mM EDTA, 1 mM AEBSF, 1 mM PMSF, 4 mM IAM) using probe sonication, and lysates were clarified by centrifugation at 13,000 rpm for 10 min. 20 mM of DTT was added to quench excess IAM.

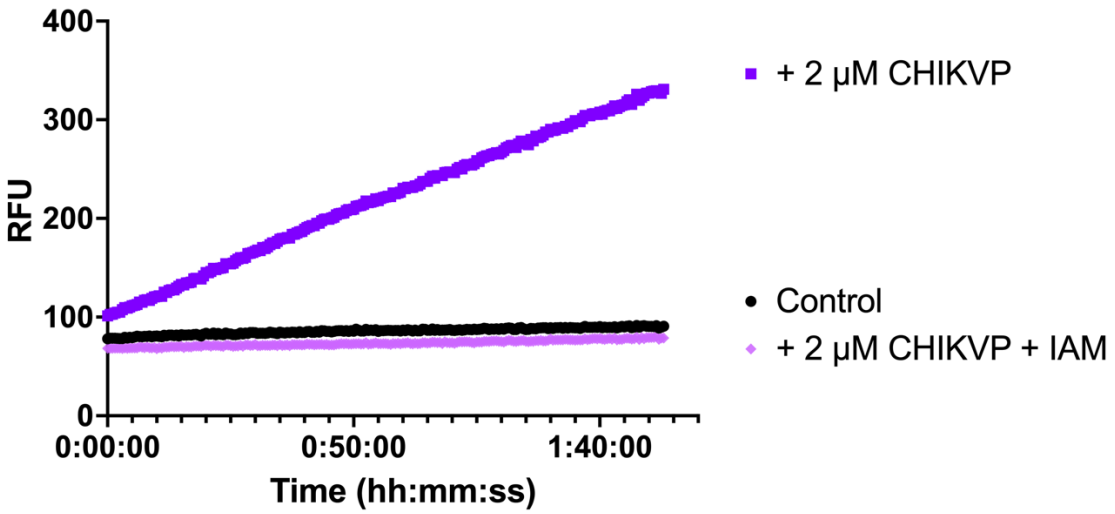


Figure D.2. Recombinant CHIKVP activity monitored by FRET assays. 2 μM of CHIKVP was incubated with 30 μM probe DABCYL-RAGGYIFS-Glu(EDANS)-NH2 in the assay buffer. 10 mM of IAM completely inhibited CHIKVP activity.

D.4 Expression and purification of MAYVP

Despite the sequence homology between CHIKVP and MAYVP (**Fig. D.3**), MAYVP exhibited very different biochemical properties, where the MAYVP expression in various constructs in bacterial systems formed inclusion bodies (**Fig. D.4**). Moreover, its expression in mammalian cells was also highly cytotoxic, imposing many challenges for performing *in vitro* studies.

Score	Expect	Method	Identities	Positives	Gaps
462 bits(1189)	6e-170	Compositional matrix adjust.	214/321(67%)	263/321(81%)	1/321(0%)

```
CHIKVP 1    FQNKANVCWAKSLVPILETAGIKLNDRQWSQIIQAFKEDKAYSPEVALNEICTRMYGVDL  60
MAYVP 1     .....K.....C...V.........SAAD..S..L.....R.............KI.....  60

CHIKVP 61   DSGLFSKPLVSVYYADNHWDNRPGGKMFGFNPEAASILERKYPFTKGKWNINKQICVTTR  120
MAYVP 61    ......A.R..LH.TT.....S...R.Y..SV...NR..QRH..YR.R.-ASG.VL.AE.  119

CHIKVP 121  RIEDFNPTTNIIPANRRLPHSLVAEHRPVKGERMEWLVNKINGHHVLLVSGCSLALPTKR  180
MAYVP 120   .TQPIDI.C.L..F......A..T.YH......V.......P.Y.L....EYN.I..RRK  179

CHIKVP 181  VTWVAPLGVRGADYTYNLELGLPATLGRYDLVVINIHTPFRIHHYQQCVDHAMKLQMLGG  240
MAYVP 180   ...I..PT.T...L.HD.D....PNA......FV.M...Y.L..................  239

CHIKVP 241  DSLRLLKPGGSLLIRAYGYADRTSERVICVLGRKFRSSRALKPPCVTSNTEMFFLFSNFD  300
MAYVP 240   .A.Y.........L...........A.VTA.A.R.S.F..VR.........V.L..T...  299

CHIKVP 301  NGRRNFTTHVMNNQLNAAFVG  321
MAYVP 300   ....TV.L.PT.GK.SSIYA.  320
```

Figure D.3. Sequence alignment between CHIKVP and MAYVP showed 67% identical and 81% similar homology.

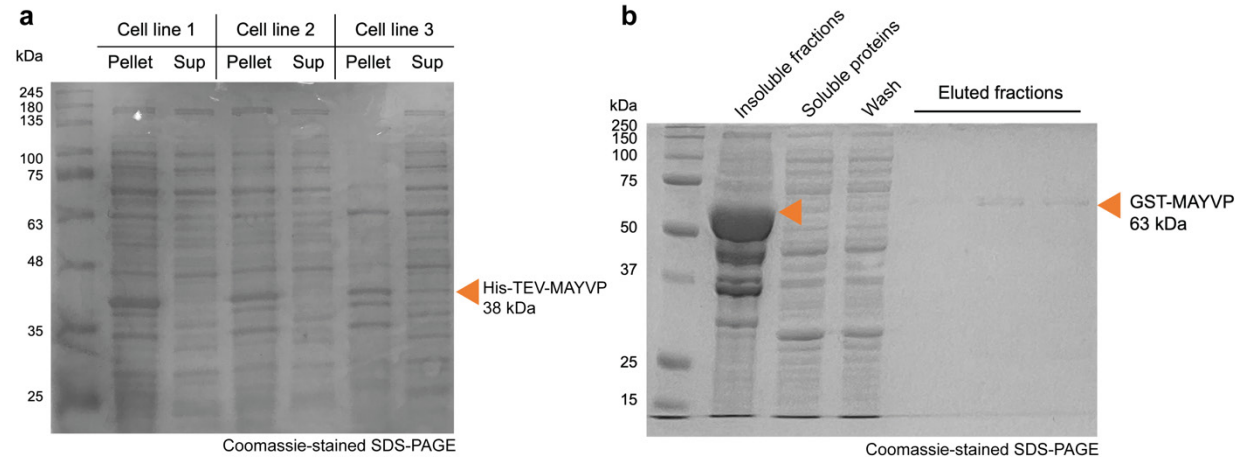


Figure D.4. Unsuccessful expression of recombinant MAYVP in bacterial cells. a) Expression of His-TEV-MAYVP in cell lines 1 (BL21-CodonPlus-RIL), 2 [BL21-CodonPlus(DE3)-RIPL] and 3 (Arctic express(DE3)RIL) was mostly in the insoluble cell pellets. The bacterial cell lines were kindly gifted by Dr. R. Glen Uhrig (University of Alberta). **b**) Expression of GST-MAYVP did not improve its solubility.

Finally, we found SUMO tag fusion significantly improved MAYVP solubility. *E. coli* codon optimized sequence encoding the MAYV NSP2 protease domain (MAYVP) was cloned in the pET-28a backbone vector fused with His-SUMO tag cleavable by His-Ulp1 (SUMO protease) to maintain MAYVP native N- and C-termini. The pET-28a vector plasmid and the plasmid encoding His-Ulp1 were kindly gifted by Dr. Joanne Lemieux. His-SUMO-MAYVP was transformed in BL21(DE3) and the bacteria were cultured in 2X YT + 1% glycerol. Protein expression was induced with 0.5 mM IPTG at OD_{600} = 0.65 at 170 rpm, 15˚C overnight. Cells were lysed in binding buffer (50 mM Tris, pH 7.5, 300 mM NaCl, 5% glycerol and 25 mM imidazole) by Emulsiflex and the lysates were clarified by centrifugation at 40,000 x g for 4˚C at 30 min. The soluble protein extract then proceeded to Ni column purification using HisTrap (Cytiva), with the above binding buffer and elution buffer (50 mM Tris, pH 7.5, 300 mM NaCl, 5% glycerol and 500 mM imidazole). Notably, all of MAYVP was precipitated after SUMO tag removal during dialysis with His-Ulp1. Therefore, we kept the SUMO tag for subsequent studies.

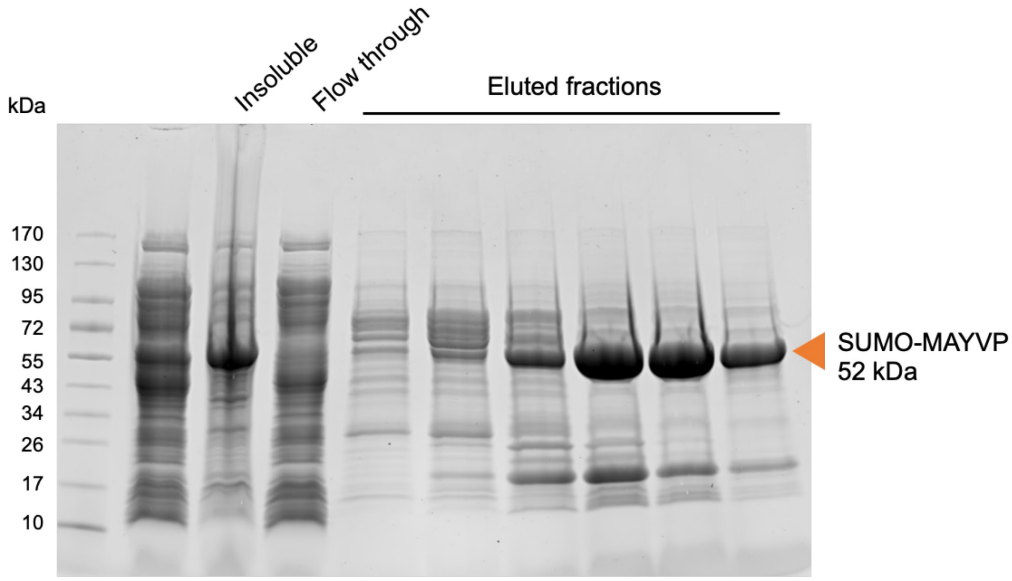


Figure D.5. Expression and purification of His-SUMO-MAYVP. His-SUMO tagged MAYVP was expressed in BL21(DE3) and purified with Ni column purification.

D.5 References

1. Fros, J. J.; Liu, W. J.; Prow, N. A.; Geertsema, C.; Ligtenberg, M.; Vanlandingham, D. L.; Schnettler, E.; Vlak, J. M.; Suhrbier, A.; Khromykh, A. A.; Pijlman, G. P. Chikungunya virus nonstructural protein 2 inhibits type I/II interferon-stimulated JAK-STAT signaling. *J Virol* **2010**, *84* (20): 10877-10887. DOI: 10.1128/JVI.00949-10

2. Ishida, R.; Cole, J.; Lopez-Orozco, J.; Fayad, N.; Felix-Lopez, A.; Elaish, M.; Luo, S. Y.; Julien, O.; Kumar, A.; Hobman, T. C. Mayaro Virus Non-Structural Protein 2 Circumvents the Induction of Interferon in Part by Depleting Host Transcription Initiation Factor IIE Subunit 2. *Cells* **2021**, *10* (12): 3510. DOI: 10.3390/cells10123510

Appendix E – Protein expression, purification, and activity assays of Norwalk viral protease

Preface

Appendix D includes collaborative work with Dr. John (Lok Man) Law (Memorial University of Newfoundland).

E.1 Summary

Noroviruses (also known as Norwalk viruses; NV) are a leading cause of acute gastroenteritis globally, resulting in widespread outbreaks in various settings such as hospitals, schools, and cruise ships. Currently, there are no vaccines or therapies for NV infections. NV encodes a viral protease (NVP) in non-structural protein 6 (NS6), that is essential for viral polyprotein processing and viral replication. Research on NV and its protease is fundamental for the development of viral protease inhibitors and other targeted antiviral therapies. In this study, we aimed to express and purify active recombinant NVP for future experiments such as identification of the viral protease substrates in host cells.

E.2 Expression and purification of NVP

E. coli codon optimized sequence encoding the NVP protease domain (Norovirus GI, obtained from NCBI NC-001959, nucleotide 3305-3847) was cloned into the pET-28a backbone vector fused with His-SUMO tag cleavable by His-Ulp1 (SUMO protease) to maintain NVP native N- and C-termini. The pET-28a vector plasmid and the plasmid encoding His-Ulp1 were kindly gifted by Dr. Joanne Lemieux. The His-SUMO-NVP construct was transformed in BL21(DE3), and bacteria were cultured in 2X YT. Protein expression was induced with 0.5 mM IPTG at OD_{600} = 0.65, incubated overnight at 22˚C at 200 rpm. Cells were lysed in binding buffer

(50 mM Tris, pH 7.5, 300 mM NaCl, 25 mM imidazole) by Emulsiflex and the lysates were clarified by centrifugation at 40,000 x g for 4˚C at 30 min. The soluble protein extract then proceeded to the first Ni column purification using HisTrap (Cytiva), with binding buffer and elution buffer (50 mM Tris, pH 7.5, 300 mM NaCl, 500 mM imidazole) (**Fig. E.1a**). The eluted fractions containing His-SUMO-NVP (B3-B12) were pooled and dialyzed with SUMO protease in 1:500 SUMO:protein ratio in 2 L dialysis buffer (50 mM Tris, pH 7.5, 300 mM NaCl, 1 mM TCEP) overnight at 4˚C. The cleaved protein was then subjected to a second Ni column purification using the same buffers, where the fractions containing untagged NVP in flow-through and wash were collected (**Fig. E.1b**).

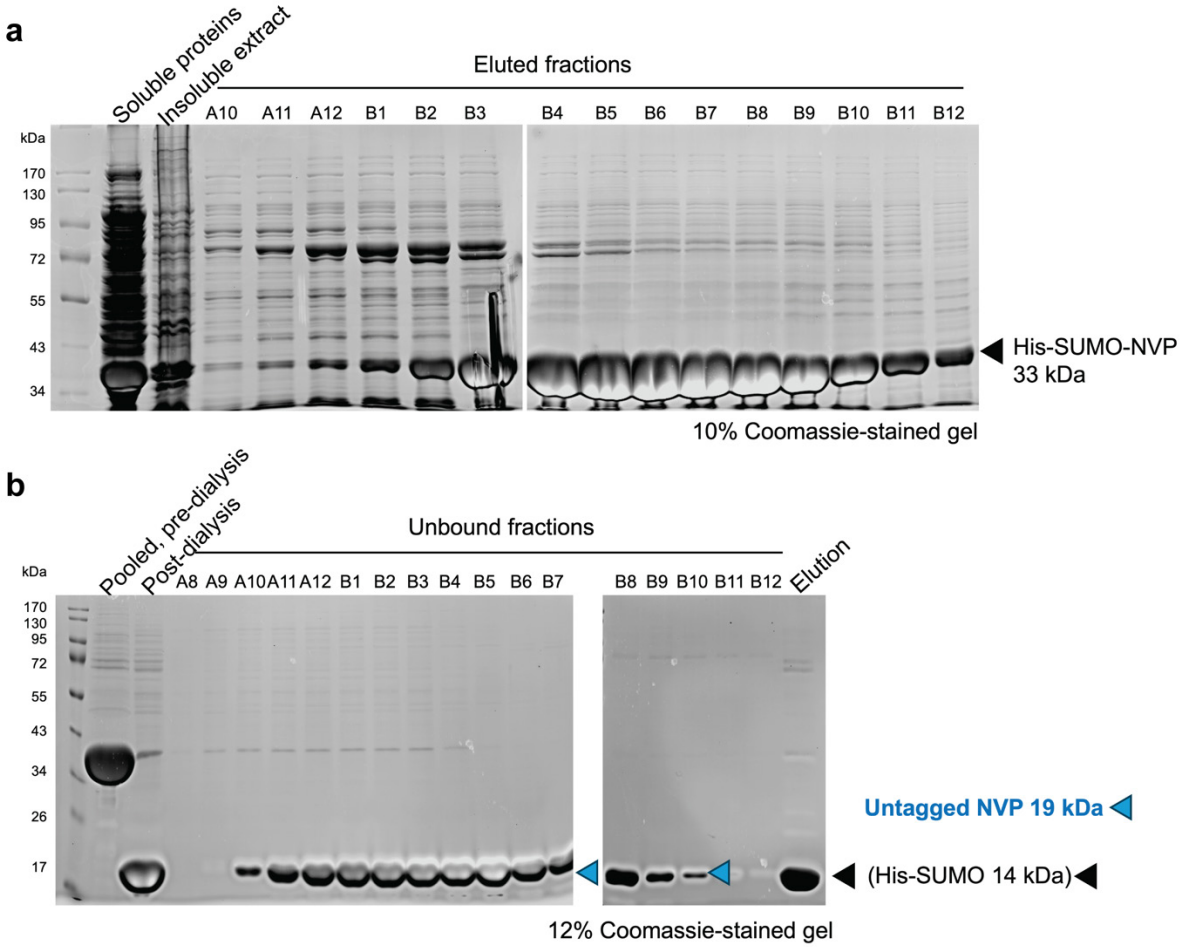


Figure. E.1. Expression and purification of recombinant NVP. His-SUMO-NVP was expressed in BL21(DE3), purified by **a**) first Ni column purification, cleaved and dialyzed with SUMO protease and further subjected to **b**) a second Ni column purification. The collected untagged NVP contained native N- and C-terminal sequences without additional residues.

E.3 Activity assay of NVP

Both SARS-CoV-2 Mpro and NVP are 3C-like cysteine proteases and exhibit similar sequence preferences at non-prime sites (P2P1 = LQ). However, we observed little to no NVP proteolytic activity with Mpro probes (**Fig. E.2**). When monitoring its activity using FRET probe Dabcyl-FHLQ|GPED-EDANS (Biomatik) in the assay buffer (20 mM Tris, pH 7.5, 1 mM EDTA, 10 mM DTT), NVP activity levels were substantially higher (**Fig. E.3**). We speculated that prime site sequence was required for NVP activity, and P1′P2′ = GP were critical sites highly preferred for GI NVP. This was consistent with previous kinetics studies performed on GII NVP using 14-amino-acid activity probes. (1)

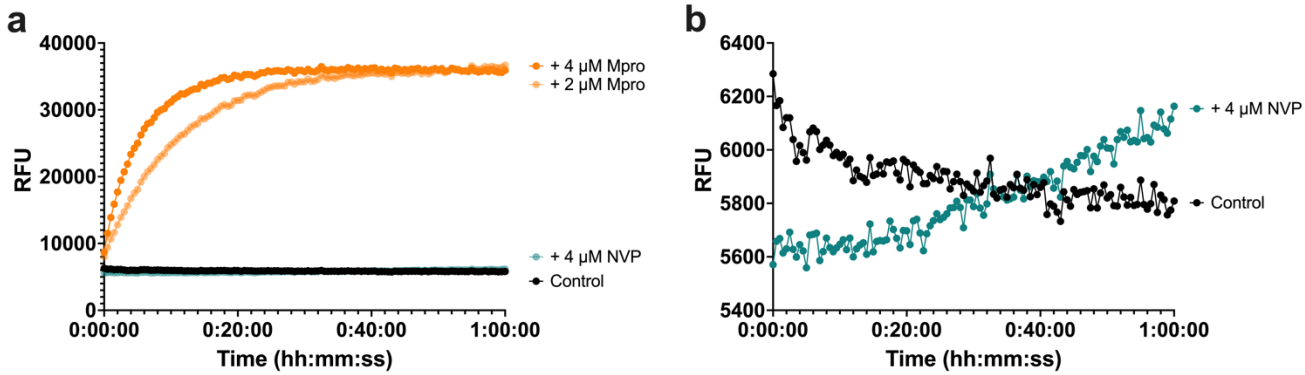


Figure E.2. Activity of 4 μM NVP against the Mpro coumarin probe Ac-Abu-Tle-Leu-Gln-ACC. a) NVP activity was close to the control baseline (no protease) relative to Mpro. **b**) Zoomed in view of the NVP activity showed limited increase in RFU over one hour.

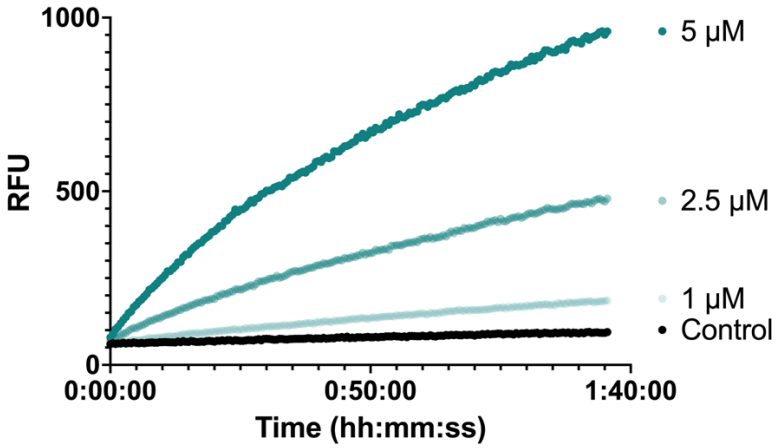


Figure E.3. Activity of 1, 2.5 and 5 μM of NVP. Activity monitored using 15 μM assay probe Dabcyl-FHLQ|GPED-EDANS.

E.4 References

1. May, J.; Korba, B.; Medvedev, A.; Viswanathan, P. Enzyme kinetics of the human norovirus protease control virus polyprotein processing order. *Virology* **2013**, *444* (1-2): 218-224. DOI: 10.1016/j.virol.2013.06.013

Appendix F – Comparing the BAD Protein Interactomes in 2D and 3D Cell Culture Using Proximity Labeling

Preface

Appendix F has been published on the manuscript:

> Pirayeshfard, L.,* Luo, S.,* Githaka, J. M.,* Saini, A., Touret, N., Goping, I. S. and Julien, O. (2024) Comparing the BAD Protein Interactomes in 2D and 3D Cell Culture Using Proximity Labeling. *J. Proteome Res.* 23, 3433-3443. DOI: 10.1021/acs.jproteome.4c00111
>
> * indicate co-authorship

F.1 Abstract

Protein-protein interaction studies using proximity labeling techniques, such as biotin ligase-based BioID, have become integral in understanding cellular processes. Most studies utilize conventional 2D cell culture systems, potentially missing important differences in protein behavior found in 3D tissues. In this study, we investigated the protein-protein interactions of a protein, Bcl-2 Agonist of cell death (BAD) and compared conventional 2D culture conditions to a 3D system wherein cells were embedded within a 3D extracellular matrix mimic. Using BAD fused to the engineered biotin ligase miniTurbo (BirA*), we identified both overlapping and distinct BAD interactomes in 2D and 3D conditions. The known BAD binding proteins 14-3-3 isoforms and Bcl-XL interacted with BAD in both 2D and 3D. Of the 131 BAD interactors identified, 56% were specific to 2D, 14% were specific to 3D, and 30% were common to both conditions. Interaction network analysis demonstrated differential associations between 2D and 3D interactomes, emphasizing the impact of culture conditions on protein interactions. The 2D-3D overlap interactome encapsulated the apoptotic program, which is a well-known role of BAD. The 3D unique pathways were enriched in extracellular matrix signaling, suggestive of hitherto unknown functions for BAD. Thus, exploring protein-protein interactions in 3D provides novel clues into cell behavior. This exciting approach has the potential to bridge the knowledge gap between tractable 2D cell culture and organoid-like 3D systems.

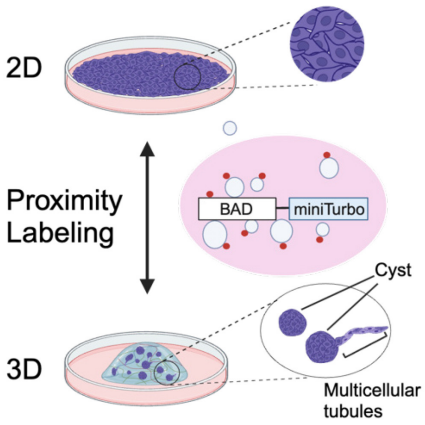


Figure F.1. Graphical abstract of Appendix F.

F.2 Introduction

Characterizing protein-protein interactions and the resultant protein interactomes is an expanding field in proteomics[1–3]. In particular, advances in proximity labeling techniques have enabled unbiased identification of both stable and transient protein interactors in living cells[4]. While these methods have been successful at mapping protein-protein interactions in most regions of the human cell[5], most BioID experiments are typically carried out in cells cultured on a plastic surface in a 2-dimensional (2D) monolayer for experimental simplicity. However, these conditions do not take into account the possibilities of proteins behaving differently in cells embedded in the extracellular matrix (ECM) in 3-dimensional (3D) tissues or organs. Differences in 2D and 3D cell morphology indicate that inter- and intra-cellular signaling in 2D versus 3D cell culture are inherently different. There are profound variations in protein-protein interactions involved in a variety of molecular pathways, such as differential molecular machineries that drive cell mobility and migration in 2D or 3D model systems[6–8].

The protein Bcl-2 Agonist of cell death (BAD) was first identified as a protein that binds the anti-apoptotic proteins BCL-2 and BCL-XL[9]. Although BAD sensitizes cells to apoptosis in cell culture, it has complex roles *in vivo*[10–12], including altering glycolysis via binding to glucokinase[13,14] and modulating mammary epithelial cell mobility[15]. To examine the extent to which 2D and 3D cellular environments affect protein interactions, we queried the protein-protein interactions in 2D and 3D culture using BAD as a bait within human mammary epithelial cells (MCF10A). While some of the proteins identified overlapped, we find significant differences in BAD interactomes between 2D and 3D conditions, highlighting the importance of incorporating the effect of cellular context to define protein-protein interaction networks.

F.3 Results

F.3.1 Establishing experimental model systems

To identify the differential protein interactome in 2D versus 3D systems, we employed proximity labeling[1,2]. Given the size of BAD (23 kDa), we decided to use the smaller biotin ligase miniTurbo, which promiscuously biotinylates interacting proteins effectively and has a higher biotin ligation efficiency with a shorter incubation time compared to other biotin ligase variants[1,2]. We fused miniTurbo to the C-terminus of our bait protein BAD where it showed similar subcellular localization to untagged BAD (**Fig. F.2**), and generated an MCF10A cell line stably expressing the fusion protein where endogenous BAD expression is knocked-out[16]. As a control, we generated MCF10A cell lines stably expressing miniTurbo tagged to the monomeric green fluorescent protein, mEmerald. For 2D culture, we grew cells as monolayers on plastic dishes under conventional conditions, where cells grow and migrate on the surface and eventually form a confluent monolayer. For 3D cultures, single cells were embedded in an ECM mimic of basement membrane mixed with collagen-I[17] (**Fig. F.3A, B**). In 3D culture, cells form multicellular clusters known as cysts. Cells on the exterior surface of the cyst interact with the surrounding ECM with some cells initiating collective cell migration to form multicellular tubules. Confocal microscopy showed this differential morphology of cells cultured in 2D vs 3D (**Fig. F.3B**). Intracellular localization of BAD was similar in both conditions. We also assessed the heterogeneity of the 3D culture conditions. As expected, the 3D cysts were filled and lacked a central lumen as described by others[18].

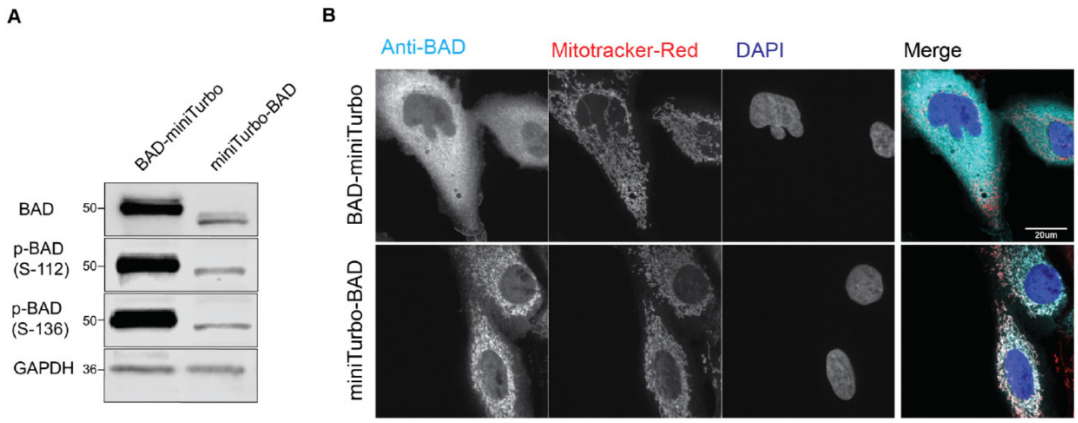


Figure F.2. Expression level of N- and C-terminal fusion of miniTurbo to BAD. (**A**) Immunoblot shows expression and phosphorylation status in MCF10A BAD-miniTurbo (C-terminal fusion) and miniTurbo-BAD (N-terminal fusion). The expression of BAD is lower in the C-terminal fusion versus N-terminal fusion. (**B**) Immunofluorescence images show localization of BAD-miniTurbo (C-terminal fusion) both in the cytoplasm and mitochondria, similar to untagged-BAD, while miniTurbo-BAD (N-terminal fusion) was largely found in the mitochondria. Consequently, the BAD-miniTurbo (C-terminal fusion) was chosen for proximity labeling experiments.

We next verified that the miniTurbo moiety retained biotin ligase activity. Time course labeling in the presence and absence of exogenously added biotin showed a time-dependent and biotin-dependent increase of protein biotinylation. Biotinylated proteins could be observed after 30 min in 2D and 1 hour in 3D (**Fig. F.3C**) – and these time points were chosen for further experiments to minimize non-specific protein-protein interactions. Biotinylation was monitored using immunofluorescence microscopy and showed expected localization in both 2D and 3D conditions (**Fig. F.3D**). We also observed consistent biotinylation signals distributed throughout the cyst, except in the nucleus as expected, demonstrating adequate biotin diffusion throughout the 3D multicellular cysts. We next assessed the ability to capture biotinylated proteins by streptavidin pull-down. Biotinylated proteins were efficiently depleted from the streptavidin flow-through (FT) fraction in both 2D and 3D. These findings demonstrate that miniTurbo expressing cells cultured in 2D and 3D successfully produced biotinylated proteins that could be captured by streptavidin.

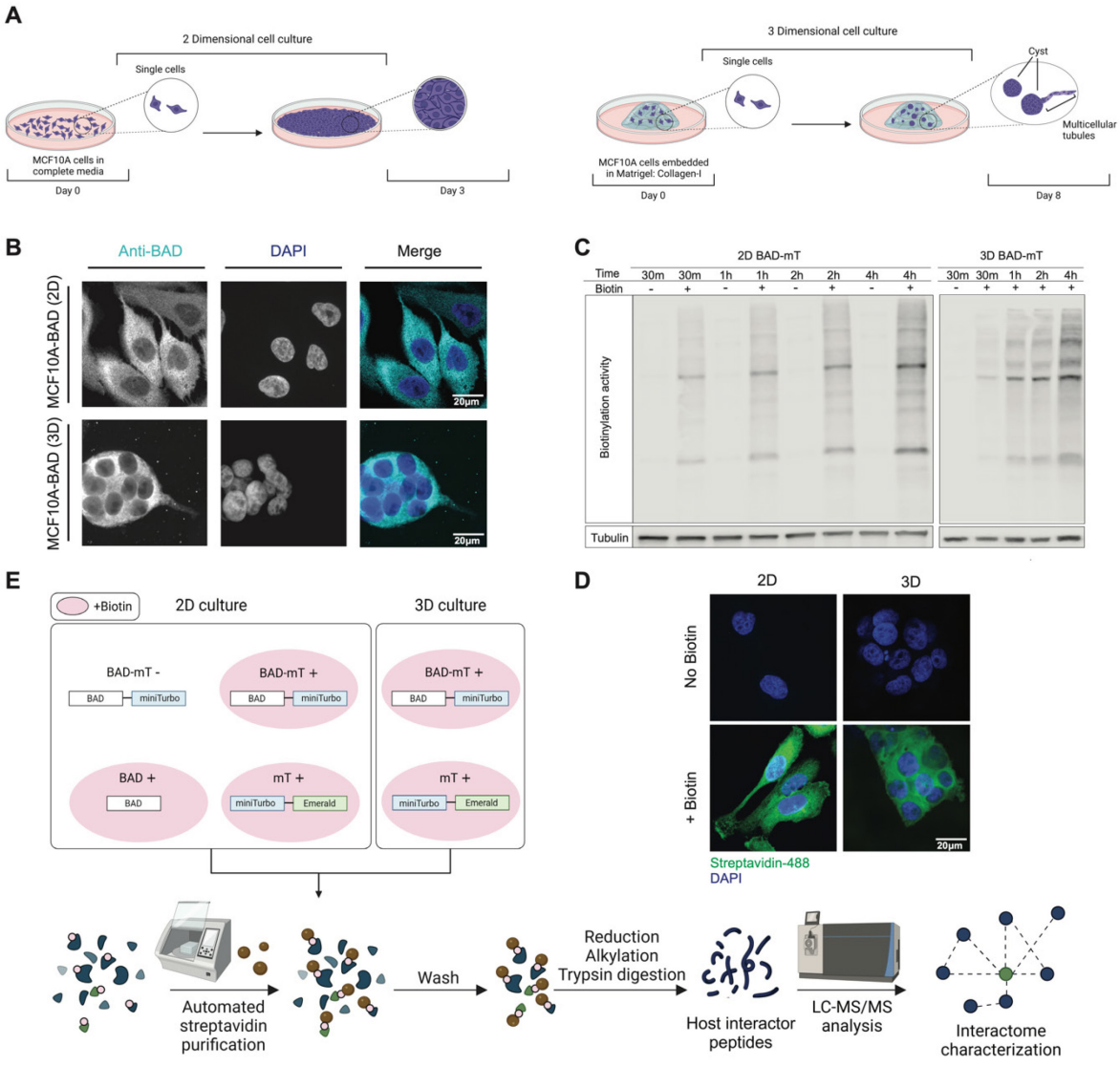


Figure F.3. Biotin ligase methodology in 2D and 3D cell culture. (**A**) Schematic of MCF10A human mammary epithelial cell model systems in 2 dimension (2D) or 3 dimensions (3D). (**B**) Immunofluorescence of BAD subcellular localization in MCF10A cells grown in 2D or 3D conditions. BAD (cyan); DAPI (blue); scalebar = 20 μm. (**C**) Streptavidin blot confirming biotin ligase activity in 2D and 3D conditions. Incubation times with biotin and the presence or absence of biotin are indicated. Biotinylated proteins were visualized using streptavidin-680. (**D**) Biotinylation was also monitored using immunofluorescence microscopy and showed expected localization in both 2D and 3D conditions. (**E**) Schematic of the proximity labeling experimental design and controls. Addition of biotin is indicated with pink shading. (mT= miniTurbo; + indicates biotin treatment).

F.3.2 Identification of potential BAD interacting and proximal proteins

Purified biotinylated proteins were reduced, alkylated, and subjected to trypsin digestion, using automated magnetic bead capture to minimize variability (**Fig. F.3E**). The resulting peptides were analyzed by liquid chromatography followed by tandem mass spectrometry (LC-MS/MS) using an Orbitrap Fusion Lumos (see methods). We identified a combined 4056 total proteins in the 2D culture samples, and 2531 total proteins in the 3D culture samples (2D: n=3; 3D: n=4). Next, we identified the BAD proximal proteins by comparing the abundance of proteins in the BAD-miniTurbo sample with biotin (BAD-mT+) to those in the mEmerald-miniTurbo control samples (mT+) (**Fig. F.4A, B**). Volcano plot analysis of this comparison indicated that 124 and 66 protein hits were significantly enriched in 2D and 3D culture conditions, respectively (p-value < 0.05 and fold change > 3) (**Fig. F.4A, B**). We generated a list of known interacting proteins obtained by querying the BioGRID interaction repository of published physical and genetic interactions[19]. From the 121 BAD-interacting proteins that were curated in the database, we filtered for the ones derived in human cells that had evidence of physical interaction and were represented by at least 2 independent publications. This generated a list of 18 proteins with high-confidence as *known* BAD-interactors. Of these known BAD-interactors, 6 (33%) proteins were identified in our 2D and 3D culture interactomes (**Fig. F.4C**). Importantly, both the known BAD-interactors and our potential BAD-interactome contained 14-3-3 isoforms (four of seven in the 2D and 3D samples), and BCL-XL (in 3D samples), which constitute the best documented canonical interactors of BAD (**Fig. F.4A, B**)[20]. Thus, our screen captured both well-documented BAD interactors and identified potential novel BAD interacting proteins.

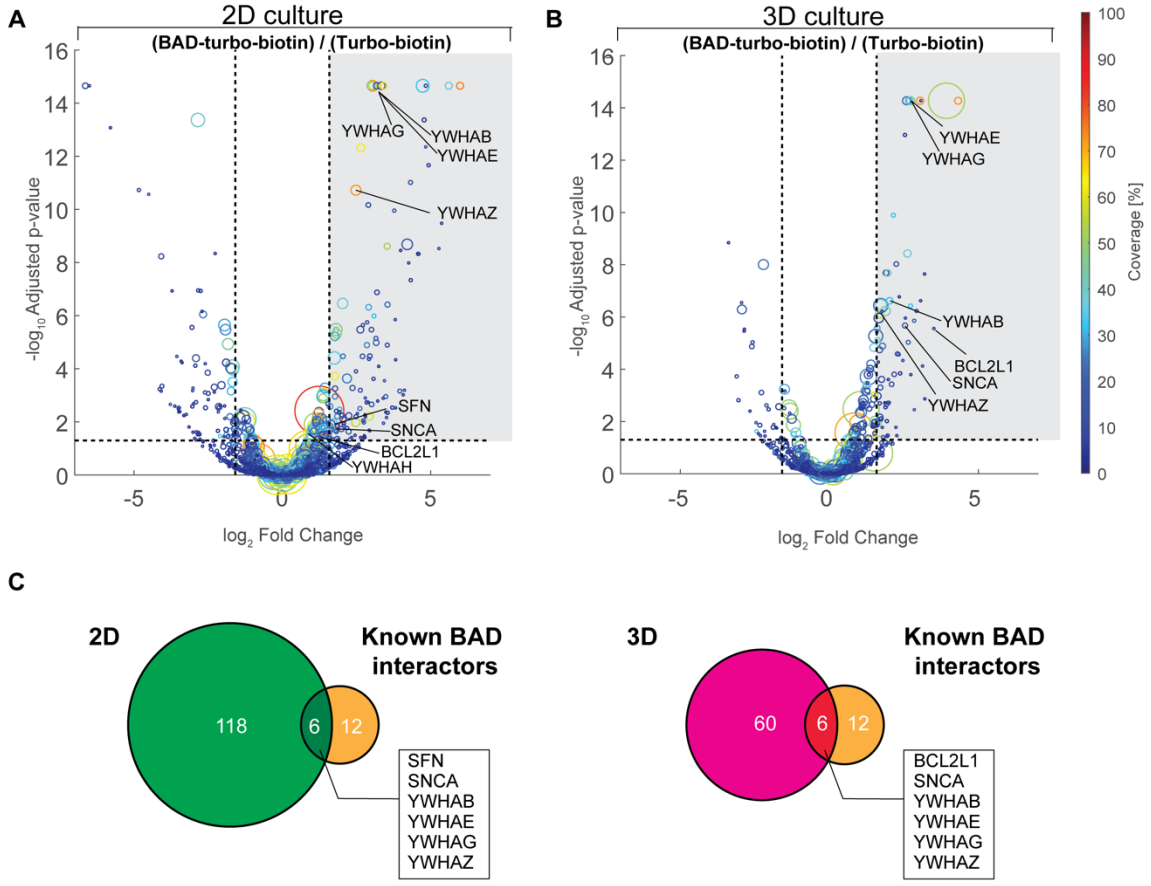


Figure F.4. Protein interactors identified by mass spectrometry. Volcano plots of all biotinylated proteins identified in 2D (**A**; n=3) and 3D (**B**; n=4). The comparison of BAD-miniTurbo+ to miniTurbo+ control is shown. Dashed lines demarcate adjusted p-value of < 0.05 and a fold change greater than 3 with proteins inside this area shaded in grey. The circle size and line color indicate the number of peptides and percentage of the protein coverage, respectively. Previously known BAD interactors from literature are labeled. (**C**) Venn diagram comparing overlap between our identified potential BAD interactors with previously known BAD interactors from BioGrid[19].

F.3.2 Screening for BAD specific interacting proteins in 2D and 3D

To further filter out non-specific interactors, we incorporated additional negative controls and subjected the data to unsupervised agglomerative hierarchical clustering. Specifically, in addition to the aforementioned mEmerald-miniTurbo control (mT+), we included untagged BAD incubated with biotin (BAD+) and BAD-miniTurbo without biotin (BAD-mT-) in the 2D experimental setup (**Fig. F.3E**; n=3). These controls allowed for two more quality control

comparisons to ensure minimizing non-specific identification of BAD interactors. However, given the experimental constraints of the 3D cultures (e.g. time and cost), only BAD-mT+ compared to mT+ were included in the 3D conditions. For analysis, we considered only proteins with a p-value < 0.05 and ≥ 3-fold change in BAD-mT+ compared to mT+ in both 2D and 3D cultures, which pared our list down to 141 proteins. Unsupervised hierarchical clustering heatmap of these lists of proteins ultimately identified three main general clusters containing 2D unique, 3D unique, or 2D/3D overlapping proteins (n=131 proteins, in **Fig. F.5A**).

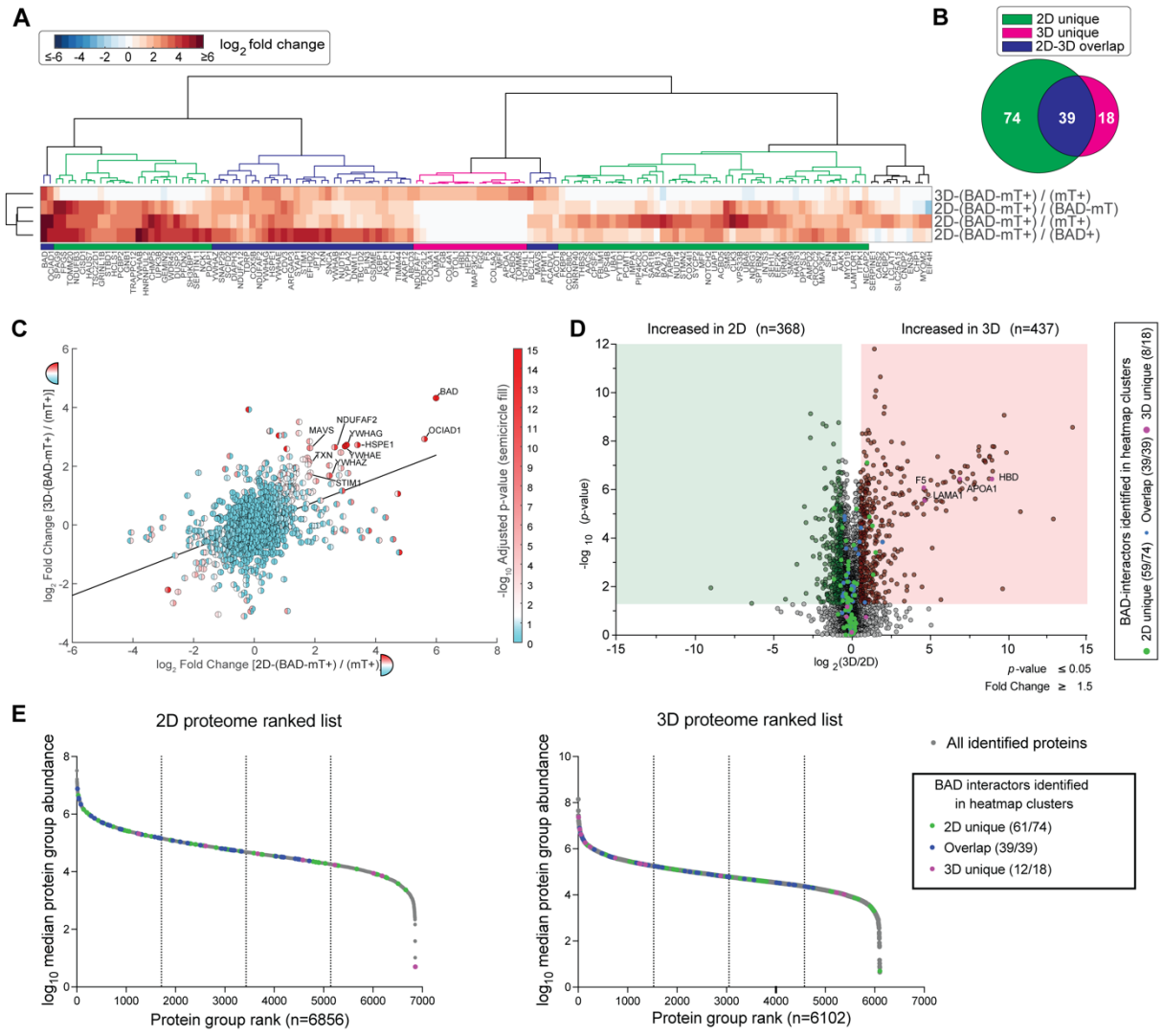


Figure F.5. Unsupervised hierarchical clustering identifies differential BAD-interacting proteins in 2D and 3D. (**A**) Possible BAD interactors (p-value < 0.05 and fold change ≥ 3) were derived from the ratios indicated on the y-axes and subjected to unsupervised hierarchical clustering

heatmaps to agglomerate BAD-specific interactors in 2D (n=3) and 3D (n=4). BAD-miniTurbo with biotin (BAD-mT+) was compared to four controls (+ indicates biotin treatment): miniTurbo without BAD in 2D and 3D (mT+), BAD-miniTurbo without biotin (BAD-mT-), BAD without miniTurbo (BAD+) in 2D. 131 interactors were identified for BAD which are highlighted as three main clusters; 2D-unique (green), 3D-unique (magenta) and 2D-3D overlap interactors' clusters (blue). (**B**) Venn diagram comparing overlap between potential BAD interactors in 2D and 3D conditions. (**C**) Correlations of the log2 fold change from all common proteins between 2D (BAD-mT+/mT+) & 3D (BAD-mT+/mT+) (Rho=0.45259). The color of the circles indicate the -log10 adjusted p-value for 2D (right semicircle) and 3D (left semicircle), with red being the most significant. The alpha threshold value (adjusted p-value = 0.05) is shown in white. The top 10 proteins with the lowest p-values are labeled. (**D**) Volcano plot comparing proteome changes of MCF10A in 2D and 3D culture stably expressing miniTurbo-BAD with biotin treatment (2D: n=3; 3D: n=4), before biotin capture. Proteins with significantly increased abundance in 3D were shaded in the red box (fold change ≥ 1.5, p-value ≤ 0.05), and proteins with significantly increased abundance in 2D were shaded in the green box. The enriched BAD interactors identified from the subsequent proximity labeling were highlighted according to the heatmap clusters. (**E**) Ranked protein abundances from the volcano plot. The BAD-interactors identified from the proximity labeling heatmap clusters are highlighted, showing a dispersed distribution from high to low abundant proteins. The dashed lines correspond to the quartiles of the total protein abundances distribution prior to the enrichment step.

F.3.2 BAD interactome varies in 2D vs 3D culture conditions

From these 131 potential BAD interactors, 2D unique and 3D unique clusters consisted of 74 hits and 18 hits, exclusively identified in 2D and 3D conditions respectively. The overlap cluster consisted of the remaining 39 hits found in both 2D and 3D conditions (**Fig. F.5B**). Correlation of all 131 potential BAD interactors in 2D vs 3D gave a negative correlation (Pearson correlation coefficient = -0.41), as would be expected since the hits in the 2D and 3D unique clusters were more than the overlapping hits. However, the enrichment of all overlapping hits was positively correlated, as shown in **Fig. F.5C**. Interestingly, the most enriched common interactors in 2D and 3D culture were also the most significant hits.

We also assessed the global proteome changes in response to 2D and 3D culture. Specifically, an aliquot of the input cell lysates from MCF10A cells expressing miniTurbo-BAD in 2D and 3D with biotin treatment before the enrichment step of our proximity labeling experiments were subjected to total protein abundance quantitation using data independent

acquisition (DIA) (**Fig. F.5D**). Culture conditions altered the abundance of a subset of proteins. We ranked the 6,000+ proteins quantified in the lysates based on their abundances (**Fig. F.5E**), and indicated the BAD interactors we had identified from the proximity labeling experiments (2D unique, 3D unique, and 2D-3D overlap). This data showed that protein levels of some of the 3D-specific interactors were elevated in response to 3D culture. The majority of interactors, however, were similarly expressed in either culture condition, and of these, a subset was differentially biotinylated in 2D vs 3D, indicating that protein interactome changes were likely specific and not simply a function of protein abundance. Finally, the highest-fold enriched hits from each cluster are shown (**Fig. F.6**). Of note, all the previously known BAD-interacting proteins were found in the 2D-unique or 2D-3D overlapping clusters. In summary, our analyses identified 131 potential BAD interacting proteins; 56% are unique to 2D culture, 14% are unique to 3D culture and 30% were identified to interact with BAD in both 2D and 3D culture conditions. Thus, the BAD-interactome differs depending on culture conditions, with nearly two thirds of BAD-proximal proteins showing exclusive interactions in either 2D or 3D.

	Description	Gene Symbol	Uniprot Accession	2D- [%] Coverage	3D- [%] Coverage	2D-Fold Change	3D-Fold Change	2D- Adj p=Value	3D- Adj p=Value	Cellular Location
2D unique	Vacuolar protein sorting-associated protein 33B	VPS33B	Q9H267	3		30.5		2.2E-12		C, E
	Folylpolyglutamate synthase, mitochondrial	FPGS	Q05932	6		28.6		2.2E-15		M, C
	Bromodomain adjacent to zinc finger domain protein 1A	BAZ1A	Q9NRL2	3		27.5		4.3E-14		N
	Ubiquitin-like modifier-activating enzyme 6	UBA6	A0AVT1	4		23.9		4.7E-09		C
	Heterogeneous nuclear ribonucleoprotein U-like protein 2	HNRNPUL2	Q1KMD3	3		20.1		4.6E-08		N
	Prefoldin subunit 2	PFDN2	Q9UHV9	17		13.7		1.1E-10		C, M, N
	Integrator complex subunit 3	INTS3	Q68E01	3		13.1		2.9E-03		N, C
	Protein NDRG1	NDRG1	Q92597	22		11.7		3.8E-07		N, C, CM
	UDP-N-acetylhexosamine pyrophosphorylase	UAP1	Q16222	8		11.6		1.3E-05		C
	CDGSH iron-sulfur domain-containing protein 1	CISD1	Q9NZ45	53		11.6		2.5E-09		M
3D unique	Mitochondrial fission factor	MFF	Q9GZY8		17		5.8		2.1E-06	M, P
	Collagen alpha-2(V) chain	COL5A2	P05997		3		5.8		1.1E-13	S
	Coagulation factor V	F5	P12259		1		5.2		1.4E-04	S
	Fibrinogen gamma chain	FGG	P02679		11		4.7		9.5E-09	S
	Protein arginine methyltransferase NDUFAF7, mitochondrial	NDUFAF7	Q7L592		9		4.3		3.8E-04	M
	Tumor protein D54	TPD52L2	O43399		21		3.8		2.2E-03	C
	Hemoglobin subunit delta	HBD	P02042		15		3.5		3.9E-02	C
	Collagen alpha-1(IV) chain	COL4A1	P02462		2		3.3		8.0E-03	S
	Fibrinogen beta chain	FGB	P02675		10		3.3		5.0E-04	S
	Laminin subunit alpha-1	LAMA1	P25391		5		3.2		1.1E-06	S
2D-3D overlap	OCIA domain-containing protein 1	OCIAD1	Q9NX40	38	11	48.8	7.6	2.2E-15	5.9E-07	E
	10 kDa heat shock protein, mitochondrial	HSPE1	P61604	67	32	10.6	6.6	2.2E-15	3.8E-07	M
	14-3-3 protein epsilon	YWHAE	P62258	73	46	8.2	6.6	2.2E-15	5.4E-15	N, C
	14-3-3 protein gamma	YWHAG	P61981	47	32	8.0	6.4	2.2E-15	5.4E-15	C
	Bcl-2-like protein 1	BCL2L1	Q07817	14	4	2.8	11.4	2.8E-02	2.8E-06	M, C
	Cytochrome c oxidase subunit 5B, mitochondrial	COX5B	P10606	26	6	7.2	5.5	1.3E-05	4.6E-05	M
	NADH dehydrogenase 1 alpha subcomplex assembly factor 2	NDUFAF2	Q8N183	63	21	6.3	6.3	4.8E-13	9.3E-06	M
	Protein CDV3 homolog	CDV3	Q9UKY7	32	22	8.5	3.4	1.0E-06	9.1E-02	C
	Testis development-related protein	TDRP	Q86YL5	35	6	2.8	8.7	5.4E-02	7.6E-05	C, N
	Alpha-synuclein	SNCA	P37840	39	19	3.5	7.2	1.4E-02	1.4E-06	N, C, CM, S

C=Cytoplasm, CM= Membrane, E=Endosome, M=Mitochondria, N= Nucleus, P= Peroxisome, S= Secreted

Figure F.6. List of top hits for BAD interactomes in 2D and 3D. The table shows the top 10 hits for each cluster. The proteins were ranked based on the fold change in the 2D and 3D groups. For the 2D and 3D overlap group, a combined fold-change score was calculated by the sum of the combined 2D and 3D fold changes. Only proteins with ≥ 2 unique peptides in either 2D or 3D and presence in ≥ 2 replicates were selected for the final list.

F.3.2 Association of 2D and 3D interactome with each other and known interactors

To further characterize the proteins within each cluster, we analyzed their subcellular localizations (**Fig. F.7A**). The subcellular protein distributions for the 2D-3D overlap group were fairly similar to the 2D unique hits, with the majority of proteins found in the cytoplasm, nucleus and mitochondrion. On the other hand, the subcellular distributions of the 3D unique hits were markedly different with greater than 60% representation of secreted proteins. This strongly suggested distinct biological consequences of these differential BAD interactomes in 2D vs 3D conditions. To further explore this, we leveraged interactomics STRING database analysis[21].

Using BAD as our bait, we queried our interactome clusters using all curated interactions (known and predicted interactions) in the STRING database. 2D-3D overlap hits gave a robust interaction network, with BAD and its canonical interactors such as BCL2L1 (BCL-XL) and 14-3-3 isoforms forming part of the core network (**Fig. F.7B**). 2D-unique interactors showed weaker association with each other. Interestingly, similar to the 2D-3D overlap hits, 3D-unique hits showed a robust interaction network that has not yet been linked to BAD (**Fig. F.7B**). This suggests that BAD is a novel interactor of an established interaction network in 3D culture, further emphasizing the value of incorporating 3D models to capture molecular networks that might be otherwise missed in 2D models.

F.3.2 The 2D- vs 3D-specific BAD interactomes differ in cellular pathways

We then queried the Reactome database to identify significant pathways from 2D unique, 3D unique and 2D-3D overlap interactomes (**Fig. F.7C**). The most significant and represented pathway in the 2D-3D overlap BAD-interactome was apoptosis. This was not surprising, as BAD is a well-known pro-apoptotic protein[20]. The 3D unique interactome pathways included ECM components, including collagen. This was consistent with the identified prevalence of secreted proteins, and STRING analysis associations with multiple ECM nodes. Altogether our data identified clear differences in BAD-proximal proteins in 2D vs 3D conditions, and these differential interactomes were functionally distinct. Apoptotic signaling was associated with 2D or 3D, whereas ECM-related pathways were exclusively in 3D. Thus, BAD likely has functionally different roles in 2D vs 3D conditions. Importantly, our work highlights the value of incorporating 3D culture analysis to study protein-protein interactions.

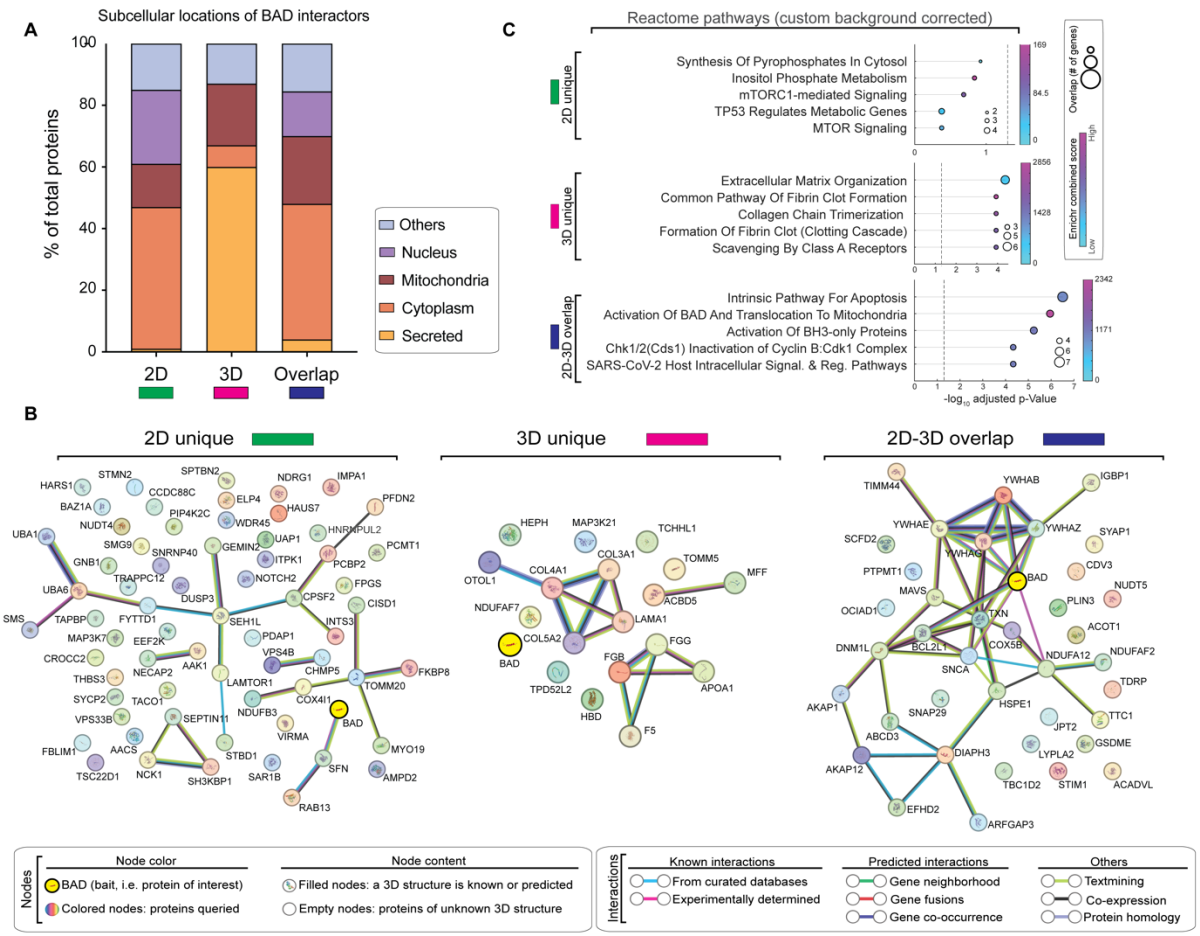


Figure F.7. Characterization of the differential BAD interactomes in 2D and 3D. (**A**) The subcellular localization of the proteins found in the 2D unique, 3D unique and 2D-3D overlap groups. Localization data was extracted from the UniProt. (**B**) STRING interaction connections between proteins from the 2D-unique, 3D-unique and 2D-3D overlap clusters. Note, the bait protein (BAD) was included in each analysis. The network nodes are proteins, filled with known/predicted protein structure, if available. Differently colored lines represent different types of evidence used in predicting the associations. Types of evidence include presence of fusion, neighborhood, co-occurrence, experimental, text-mining, database, co-expression & homology evidence. (**C**) Reactome pathway analysis of 2D unique, 3D unique and 2D-3D overlap clusters with custom background genes correction which are all the proteins that are identified as biotinylated in our 2D and 3D experiments. Circle size and color indicate the number of proteins and Enrichr combined score, respectively. Dashed line demarcates adjusted p-value < 0.05.

F.4 Discussion

Though *in vivo* tissues and organs are organized in a 3D space, most *in vitro* experimental models employ traditional 2D cell cultures which lack the rich 3D informative environment important in tissue physiology[7]. 3D cultures aim to narrow the gap between *in vitro*

models and the *in vivo* systems that are the study of interest. Some of the essential aspects that are affected by different culturing systems are cell morphology and cell-cell interactions. For instance, when cultured in 2D, non-transformed mammary epithelial cells, such as MCF10A, form monolayer sheets of cells that lack polarization. In contrast, when embedded within an ECM mimic of basement membrane as 3D cultures, those same MCF10A cells form multicellular cysts of polarized cells[18]. The addition of extracellular collagen-I to the ECM stimulates collective cell invasion and produces multicellular tubules similar to the process of mammary gland morphogenesis[17] (**Fig. F.3A**). This demonstrates that cell morphology and migration are affected by the environment[22]. At the molecular level, gene expression and protein abundance changes and consequent signaling pathways are profoundly different between 2D and 3D cultures[7,23]. For example, in a breast cancer model, 3D cultures showed an increase in expression of genes that are involved in initiation, progression and metastasis compared to conventional 2D cell culture[24]. Further functional differences include cellular drug response, wherein 3D systems showed different cell viability and molecular mechanisms after treatment[25,26]. Although the 3D culture models are hampered by higher costs, time constraints, and requirement for method expertise, studies in 3D provide an enhanced and likely more physiologically relevant understanding of *in vivo* biological processes compared to conventional 2D models.

To investigate molecular mechanisms of cellular processes, identifying protein-protein interactions is a powerful approach. We employed the proximity labelling methodology wherein our bait protein-of-interest was fused to a biotin ligase to label and purify proximal/interacting proteins. We used the miniTurbo biotin ligase, which requires short labeling times, low temperature and produces low background[1–4]. Our model protein-of-interest was the Bcl-2 agonist of cell death (BAD), which we used as our bait. BAD is a pro-apoptotic protein that has various roles in cell cycle, glucose metabolism, immune system development, autophagy, breast cancer biomarker and mammary gland development[15,20,27]. We identified 74, 18 and 39

BAD-interacting proteins in 2D unique, 3D unique and 2D-3D overlap, respectively. This shows that our protein interacts with both common and unique proteins in 2D and 3D. To our knowledge very few groups have explored differential protein interactomes imparted by 3D culture model systems. Perez-White *et al.*[28] compared the 2D and 3D interactome of EphA2-proximal proteins in keratinocytes and identified 3D-specific interactors associated with different biological pathways[28,29]. Wang *et al.* explored 2D vs 3D unique and shared PAR6B-proximal proteins in colon carcinoma cells and identified PARD3B as a novel interactor required for lumen formation in 3D multicellular cysts[29].

In our study, we aimed to elucidate the impact of 2D and 3D cultures by identifying interactors to a specific protein (BAD) in mammary epithelial cells. BAD regulates collective cell invasion during pubertal mammary gland morphogenesis[15] and thus identifying BAD interactors in 3D may be physiologically relevant. We found that 14% of interacting proteins were specific to 3D conditions and 56% were specific to 2D conditions. Pathway analysis of 2D-3D overlap proteins showed apoptosis as the most significant pathway, which is the canonical function for BAD[30]. This observation not only highlights the impact of culture methods on the protein interactomes but also prompts consideration that BAD can execute its apoptotic function in both 2D and 3D environments. The 3D unique proteins suggest an involvement of BAD in ECM organization. This observation points to a novel role for this protein; an exciting avenue we are currently pursuing in our studies. This role is likely related to the morphology of mammary epithelial cells that are unique to 3D cultures. MCF10A cells form multicellular cysts with tubules when they are embedded in basement membrane and collagen-I. Basement membrane driven stiffness provides an environment conducive for cyst formation. The reorganization of collagen fiber networks facilitates the multicellular process of tubule formation[17]. Notably, the lack of a 3D ECM in 2D culture, leads to the loss of BAD interactions with proteins involved in ECM reorganization.

Overall, this study signifies the importance of different culturing systems in affecting protein-protein interactions. As 3D cultures are designed to mimic tissue, proximity labeling in 3D systems has excellent potential to provide further insight into physiologically relevant molecular processes.

F.5 Materials and Methods

F.5.1 Generating expression constructs and stable cell lines

The BAD-miniTurbo and miniTurbo-mEmerald constructs were generated in the pWPI-IRES-Bla-AK (Addgene Cat#154980) lentivirus vector backbone. Lentivirus packaging particles were generated in HEK293T cells that were transfected by Lipofectamine 2000 (Invitrogen Cat#11668-027) with a plasmid mixture of BAD-miniTurbo or miniTurbo-mEmerald, pVSV-G and PCMV Gag Pol (10.5:3:10.5 µg of plasmid ratio, respectively). Stable ectopic expression of constructs were generated in MCF10A $BAD^{-/-}$ (BAD knockout) cells we previously described[16]. The MCF10A $BAD^{-/-}$ were transduced with lentiviral particles and were selected using Blasticidine S hydrochloride at 10 µg/mL (Sigma Cat#15205).

F.5.2 Cell line culture

Cell lines were maintained in growth medium[18]; DMEM/F12 (Sigma Cat#D6421) supplemented with 5% horse serum (Invitrogen Cat#16050-122), 200 nM L-Glutamine, 20ng/mL EGF (Peprotech Cat#AF100-15), 0.5µg/mL Hydrocortisone (Sigma Cat# H-0888),100 ng/mL cholera toxin (Sigma Cat#C-8052) and10 µg/mL insulin (Sigma Cat# I-1882) in a humidified incubator at 37°C with a 5% CO_2 atmosphere. The same media were used for both 2D and 3D cultures. 2D cell culture was using conventional methods, and grown to 100% confluency to maintain a similar growth rate as found in 3D cultures. For 3D culture, cells were embedded in a 3D matrix. The 3D collagen-I component was prepared by neutralization of bovine collagen-I (Corning Cat#354231) with 1 N NaOH and 10x DMEM (217:8:25 volume ratio, respectively) and

incubated for 1 h on ice. Single cells were then embedded (50 cells per µL) in pre-thawed 1 parts volume growth factor reduced basement membrane (Matrigel, Corning Cat#354230) and 9 parts volume collagen-I solution (1 mg/mL final concentration). The cell-gel suspension was gently plated in 12-well plates (60 µL per well). The plates were incubated in 37°C for 1h. Pre-warmed growth medium was then added and replenished every 3 days[15].

F.5.2 Immunoblotting

Cell lysates were generated using RIPA lysis buffer (50mM Tris, 150mM NaCl, 0.1%SDS, 0.5% Sodium deoxycholate, 1% Triton X-100, Protease and Phosphatase inhibitor, PH=7.5). 3D culture lysates were made by shearing the gels 5 times in a 27-gauge needle in RIPA lysis buffer. The protein lysates were collected from the supernatant after 14,000g, 20 min centrifugation at 4°C. Lysates were resolved on SDS-PAGE gels and transferred to nitrocellulose membranes. For biotinylation detection, membranes were incubated with Streptavidin-680 (Licor Cat#D2040-35). For BAD expression and phosphorylation status, membranes were incubated with anti-BAD (Sigma#B0684), anti-phospho-BAD(Ser112) (Cell Signaling Technology# 9291) and anti-phospho-BAD(Ser136) (Cell Signaling Technology# 5286). Anti-Tubulin (Sigma Cat#T5168) or anti-GAPDH (Sigma Cat#G8795) were used as controls. HRP and IRDye-coupled secondary antibodies were used for detection and blots were scanned using Odyssey LI-COR Fc imager (LI-COR Biosciences) in the chemiluminescence and 700 nm channels.

F.5.3 Immunofluorescence

Cells (1x 10^5 cells) were seeded in a 12-well plate on coverslips. After 2 days cells were pre-incubated with 200nM MitoTracker-Red (Invitrogen#479525) (15min), then were fixed in 4% PFA (15 min) and permeabilized with 0.5% Triton X-100 (20 min). For staining, coverslips were blocked (3% BSA in PBS, 30 min) and incubated with anti-BAD (Sigma Cat#B0684) at 4°C overnight. Anti-rabbit (Alexa-Fluor 647) was used as a secondary antibody. Biotin signal was

detected by Streptavidin AF488 (Jackson ImmunoRsearch Cat# 016-540-084). Nuclei were stained with DAPI. Methodology was the same for 2D and 3D with the exception that 3D cultures were quenched with 50 mM NH_4Cl (30 min) after fixation.

F.5.4 Imaging and analysis

Confocal imaging was acquired with Volocity software (PerkinElmer, USA) using a ×60 oil immersion objective on WaveFx spinning-disk microscope (Quorum Technologies,ON, Canada). The microscope was set up on an Olympus IX-81 inverted stand (Olympus, Japan), equipped with an EM-CCD camera (Hamamatsu, Japan). Brightfield imaging was done on Zeiss AxioObserver.Z1 Microscope, ×2.5 air objective with an extra x1.6 magnifier. Image analyses were performed in MATLAB (MathWorks) and ImageJ. To extract 3D structure properties in brightfield images, images were corrected for non-uniform background as we earlier detailed[31], using a wide Gaussian (standard deviation 10 pixels) to estimate the background image. Next, edge and Sobel operator was applied to the background corrected image to get a binary gradient image mask of the 3D structures as we earlier detailed[15,32]. 3D structures' circularity and areas were computed using MATLAB's '*regionprops*' function on the binary image. To characterize the biotin diffusion in the cyst, the fluorescence intensities for each channel were measured in the middle z-stack using line plot profile in ImageJ. Mean and standard deviation of the fluorescence intensity for each pixel measured from the center of the cysts were calculated and plotted using Prism (Graphpad version 9.0).

F.5.5 Proximity labeling

Bio-ID was done based on the protocol that was previously described[1]. MCF10A BAD-miniTurbo and miniTurbo-Emerald were incubated in complete media supplemented with 50 μM biotin for 30 min and 1 hour in 2D and 3D culture, respectively. After PBS wash, cell lysates were prepared for purification as described above.

F.5.6 Automated purification

Streptavidin purification of biotinylated proteins was carried out using KingFisher Prime Duo automation (Thermo inc.). For each sample, 200 µL of streptavidin magnetic beads (Thermo Fisher Cat#88817) were pre-washed with 500 µL lysis buffer and protease inhibitor cocktail (described above). Beads were then incubated with clarified cell lysates at 4˚C for 18 h (with gentle mixing on the KingFisher), washed 2x with lysis buffer with protease inhibitors, and 1x with PBS. Protein-bound beads underwent reduction with 300 µL of 10 mM dithiothreitol for 45 min and alkylation with 300 µL of 50 mM iodoacetamide for 45 min and washed with 100 mM ammonium bicarbonate. On-bead trypsin digestion was performed at 37˚C for 5 h on the KingFisher.

F.5.7 Mass spectrometry analysis

For input lysates prior to streptavidin enrichment (MCF10A cells stably expressing miniTurbo-BAD in 2D and 3D culture with biotin treatment), sample cleanup was performed using S-trap micros (ProtiFi). Peptides were desalted with C18 ziptips (Millipore Cat#ZTC18S960) and recovered in buffer A [0.1% formic acid (FA)] prior to mass spectrometry analysis. The samples were analyzed using a nanoflow-HPLC (Thermo Scientific EASY-nLC 1200 System) coupled to an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific). Reverse phase separation of the peptides was done with an Aurora Ultimate analytical column (25 cm x 75 µm ID with 1.7 µm media, IonOpticks). Peptides were eluted with a solvent B gradient (80% ACN, 0.1% FA) for 90 min. The gradient was run at 400 nL/min with analytical column temperature set at 45°C. For input lysates, DIA analysis was used using 38.5 m/z isolation window widths, and data were searched using Spectronaut with directDIA library-free workflow and factory settings. Enriched data from proximity labeling were acquired using DDA, and data were analyzed using Proteome Discoverer (v2.4.1.15) against the concatenated database of the human proteome (UP000005640, v2021-08-30), with false discovery of

peptides restricted at 1%. Search parameters included a maximum of two missed trypsin cleavages, a precursor mass tolerance of 15 ppm, a fragment mass tolerance of 0.8 Da, with the constant modification carbamidomethylation (C), and variable modifications of acetyl (protein N-term), deamidated (N/Q), and oxidation (M). The maximum number of variable modifications was set to three. Only tryptic peptides were searched, and MS1 intensities were normalized using total peptide amount. Statistical analysis was performed using Proteome Discoverer (v2.4). Protein abundances were calculated using the sum of peptide peak MS1 intensities. Protein abundance ratios were calculated using the median of all peptides pairwise ratios, and hypothesis testing was performed using background-based t-test. Based on the number of measurements, the *p*-values were adjusted using the Benjamini-Hochberg method by Proteome Discoverer. Missing values were imputed with low abundance resampling using bottom 5% of all detected values, for all protein abundance ratio calculations.

F.5.8 BAD interactome identification

Data analysis and plotting was done on MATLAB and R. Briefly, we first computed Volcano plots in 2D and 3D datasets and verified enrichment of canonical BAD interactors as expected (**Fig. F.4**). We considered significant proteins (adjusted p-value < 0.05) whose fold change > 3. Using unsupervised hierarchical clustering heatmaps (Euclidean distance metric and Ward's linkage algorithm[33]) as we earlier detailed in code linked in “Availability of data and materials” section in Zare *et al*[34], these candidates clustered into 3 groups of possible BAD interactors (n=131) in 2D unique (n=74), 3D unique (n=18) and 2D-3D-overlap (n=39) clusters (**Fig. F.5A**). We tested the strength of these patterns using the function ‘*clusterboot*’ in R’s ‘*fpc*’ package. For each cluster, cluster stability was the resultant mean of Jaccard coefficients obtained from 3000 bootstrap resampling iterations. The 2D-3D-overlap cluster was enriched with canonical BAD interactors, validating our analysis approach. The online STRING database (version 12) was used on these clusters for generation of protein interactions networks[21] with

default confidence score cutoff of 0.4 to determine whether two nodes were functionally related. The bait protein, BAD, was included in each analysis. The subcellular locations of interactors were extracted from UniProt/Swiss-Prot human proteome (2023-11-08) using Biopython. The calculated percentages were proportional to all proteins containing subcellular location annotations under the comments section. Pathways (Reactome) analysis was computed using Enrichr for each cluster, with custom background genes correction[35]. All proteins identified as biotinylated in our 2D and 3D experiments were used as background in our 'custom background corrected' approach[35].

F.5.9 Data availability

Mass spectrometry spectrum files are available on MassIVE: MSV000093357 (2D miniTurbo), MSV000093358 (3D miniTurbo); MSV000094824 (2D shotgun), MSV000094825 (3D shotgun).

F.5.10 Funding

The authors declare no competing financial interest. This work was supported by an infrastructure CFI-JELF award (O.J. 37833 and 39051), and operating grants from the Natural Sciences and Engineering Research Council of Canada (O.J. RGPIN-2018-05881), the Canadian Institutes of Health Research (I.S.G. 419793), and the Lilian McCullough Breast Cancer Research Endowed Chair (I.S.G). LP gratefully acknowledges partial scholarship support from the Cancer Research Institute of Northern Alberta (CRINA), John & Rose McAllister scholarship, Delnor scholarship and FoMD 75th Anniversary Award.

F.5.10 Acknowledgements

We would like to thank the Goping and Julien lab members for the helpful discussions. We acknowledge the Faculty of Medicine and Dentistry (FoMD) Cell Imaging Core and are grateful to Jack Moore and the Alberta Proteomics and Mass Spectrometry facility for their help.

The miniTurbo was a generous gift from Dr. Alice Ting, Stanford University, Stanford, CA.

Figures F.1 and F.3 were created in part with BioRender.

F.5.11 References

(1) Roux, K. J.; Kim, D. I.; Raida, M.; Burke, B. A Promiscuous Biotin Ligase Fusion Protein Identifies Proximal and Interacting Proteins in Mammalian Cells. *Journal of Cell Biology* **2012**, *196* (6), 801–810. https://doi.org/10.1083/jcb.201112098.
(2) Branon, T. C.; Bosch, J. A.; Sanchez, A. D.; Udeshi, N. D.; Svinkina, T.; Carr, S. A.; Feldman, J. L.; Perrimon, N.; Ting, A. Y. Efficient Proximity Labeling in Living Cells and Organisms with TurboID. *Nat Biotechnol* **2018**, *36* (9), 880–898. https://doi.org/10.1038/nbt.4201.
(3) Samavarchi-Tehrani, P.; Samson, R.; Gingras, A. C. Proximity Dependent Biotinylation: Key Enzymes and Adaptation to Proteomics Approaches. *Molecular and Cellular Proteomics*. American Society for Biochemistry and Molecular Biology Inc. May 1, 2020, pp 757–773. https://doi.org/10.1074/mcp.R120.001941.
(4) Cho, K. F.; Branon, T. C.; Udeshi, N. D.; Myers, S. A.; Carr, S. A.; Ting, A. Y. Proximity Labeling in Mammalian Cells with TurboID and Split-TurboID. *Nat Protoc* **2020**, *15* (12), 3971–3999. https://doi.org/10.1038/s41596-020-0399-0.
(5) Go, C. D.; Knight, J. D. R.; Rajasekharan, A.; Rathod, B.; Hesketh, G. G.; Abe, K. T.; Youn, J. Y.; Samavarchi-Tehrani, P.; Zhang, H.; Zhu, L. Y.; Popiel, E.; Lambert, J. P.; Coyaud, É.; Cheung, S. W. T.; Rajendran, D.; Wong, C. J.; Antonicka, H.; Pelletier, L.; Palazzo, A. F.; Shoubridge, E. A.; Raught, B.; Gingras, A. C. A Proximity-Dependent Biotinylation Map of a Human Cell. *Nature* **2021**, *595* (7865), 120–124. https://doi.org/10.1038/s41586-021-03592-2.
(6) Kimlin, L.; Kassis, J.; Virador, V. 3D in Vitro Tissue Models and Their Potential for Drug Screening. *Expert Opinion on Drug Discovery*. 2013, pp 1455–1466. https://doi.org/10.1517/17460441.2013.852181.
(7) Duval, K.; Grover, H.; Han, L. H.; Mou, Y.; Pegoraro, A. F.; Fredberg, J.; Chen, Z. Modeling Physiological Events in 2D vs. 3D Cell Culture. *Physiology*. American Physiological Society June 14, 2017, pp 266–277. https://doi.org/10.1152/physiol.00036.2016.
(8) Caswell, P. T.; Zech, T. Actin-Based Cell Protrusion in a 3D Matrix. *Trends in Cell Biology*. Elsevier Ltd October 1, 2018, pp 823–834. https://doi.org/10.1016/j.tcb.2018.06.003.
(9) Yang, E.; Zha, J.; Jockel, J.; Boise, L. H.; Thompson, C. B.; Korsmeyer, S. J. Bad, a Heterodimeric Partner for Bcl-XL and Bcl-2, Displaces Bax and Promotes Cell Death. *Cell* **1995**, *80* (2), 285–291. https://doi.org/10.1016/0092-8674(95)90411-5.
(10) Kitada, S.; Krajewska, M.; Zhang, X.; Scudiero, D.; Zapata, J. M.; Wang, H.-G.; Shabaik, A.; Tudor, G.; Krajewski, S.; Myers, T. G.; Johnson, G. S.; Sausville, E. A.; Reed, J. C. *Expression and Location of Pro-Apoptotic Bcl-2 Family Protein BAD in Normal Human Tissues and Tumor Cell Lines*; 1998; Vol. 152.
(11) Zha, J.; Harada, H.; Osipov, K.; Jockel, J.; Waksman, G.; Korsmeyer, S. J. *The Bad DNA Fragments*; 1997. http://www.jbc.org.
(12) Gilmore, A. P.; Valentijn, A. J.; Wang, P.; Ranger, A. M.; Bundred, N.; O'Hare, M. J.; Wakeling, A.; Korsmeyer, S. J.; Streuli, C. H. Activation of BAD by Therapeutic Inhibition of

Epidermal Growth Factor Receptor and Transactivation by Insulin-like Growth Factor Receptor. *Journal of Biological Chemistry* **2002**, *277* (31), 27643–27650. https://doi.org/10.1074/jbc.M108863200.
(13) Danial, N. N.; Gramm, C. F.; Scorrano, L.; Zhang, C.-Y.; Krauss, S.; Ranger, A. M.; Datta, S. R.; Greenberg, M. E.; Licklider, L. J.; Lowell, B. B.; Gygi, S. P.; Korsmeyer, S. J. BAD and Glucokinase Reside in a Mitochondrial Complex That Integrates Glycolysis and Apoptosis. **2003**.
(14) Danial, N. N.; Walensky, L. D.; Zhang, C. Y.; Choi, C. S.; Fisher, J. K.; Molina, A. J. A.; Datta, S. R.; Pitter, K. L.; Bird, G. H.; Wikstrom, J. D.; Deeney, J. T.; Robertson, K.; Morash, J.; Kulkarni, A.; Neschen, S.; Kim, S.; Greenberg, M. E.; Corkey, B. E.; Shirihai, O. S.; Shulman, G. I.; Lowell, B. B.; Korsmeyer, S. J. Dual Role of Proapoptotic BAD in Insulin Secretion and Beta Cell Survival. *Nat Med* **2008**, *14* (2), 144–153. https://doi.org/10.1038/nm1717.
(15) Githaka, J. M.; Tripathi, N.; Kirschenman, R.; Patel, N.; Pandya, V.; Kramer, D. A.; Montpetit, R.; Zhu, L. F.; Sonenberg, N.; Fahlman, R. P.; Danial, N. N.; Underhill, D. A.; Goping, I. S. BAD Regulates Mammary Gland Morphogenesis by 4E-BP1-Mediated Control of Localized Translation in Mouse and Human Models. *Nat Commun* **2021**, *12* (1). https://doi.org/10.1038/s41467-021-23269-8.
(16) Mann, J.; Githaka, J. M.; Buckland, T. W.; Yang, N.; Montpetit, R.; Patel, N.; Li, L.; Baksh, S.; Godbout, R.; Lemieux, H.; Goping, I. S. Non-Canonical BAD Activity Regulates Breast Cancer Cell and Tumor Growth via 14-3-3 Binding and Mitochondrial Metabolism. *Oncogene* **2019**, *38* (18), 3325–3339. https://doi.org/10.1038/s41388-018-0673-6.
(17) Barnes, C.; Speroni, L.; Quinn, K. P.; Montevil, M.; Saetzler, K.; Bode-Animashaun, G.; McKerr, G.; Georgakoudi, I.; Downes, C. S.; Sonnenschein, C.; Howard, C. V.; Soto, A. M. From Single Cells to Tissues: Interactions between the Matrix and Human Breast Cells in Real Time. *PLoS One* **2014**, *9* (4). https://doi.org/10.1371/journal.pone.0093325.
(18) Debnath, J.; Muthuswamy, S. K.; Brugge, J. S. Morphogenesis and Oncogenesis of MCF-10A Mammary Epithelial Acini Grown in Three-Dimensional Basement Membrane Cultures. *Methods*. Academic Press Inc. July 1, 2003, pp 256–268. https://doi.org/10.1016/S1046-2023(03)00032-X.
(19) Chatr-Aryamontri, A.; Breitkreutz, B. J.; Oughtred, R.; Boucher, L.; Heinicke, S.; Chen, D.; Stark, C.; Breitkreutz, A.; Kolas, N.; O'Donnell, L.; Reguly, T.; Nixon, J.; Ramage, L.; Winter, A.; Sellam, A.; Chang, C.; Hirschman, J.; Theesfeld, C.; Rust, J.; Livstone, M. S.; Dolinski, K.; Tyers, M. The BioGRID Interaction Database: 2015 Update. *Nucleic Acids Res* **2015**, *43* (D1), D470–D478. https://doi.org/10.1093/nar/gku1204.
(20) Danial, N. N. Bad: Undertaker by Night, Candyman by Day. *Oncogene* **2008**, *27* (SUPPL. 1), 53–70. https://doi.org/10.1038/onc.2009.44.
(21) Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A. L.; Fang, T.; Doncheva, N. T.; Pyysalo, S.; Bork, P.; Jensen, L. J.; Von Mering, C. The STRING Database in 2023: Protein-Protein Association Networks and Functional Enrichment Analyses for Any Sequenced Genome of Interest. *Nucleic Acids Res* **2023**, *51* (1 D), D638–D646. https://doi.org/10.1093/nar/gkac1000.
(22) Ewald, A. J.; Brenot, A.; Duong, M.; Chan, B. S.; Werb, Z. Collective Epithelial Migration and Cell Rearrangements Drive Mammary Branching Morphogenesis. *Dev Cell* **2008**, *14* (4), 570–581. https://doi.org/10.1016/J.DEVCEL.2008.03.003.

(23) He, W.; He, W.; Kuang, Y.; Xing, X.; Simpson, R. J.; Huang, H.; Yang, T.; Chen, J.; Yang, L.; Liu, E.; Gu, J. Proteomic Comparison of 3D and 2D Glioma Models Reveals Increased HLA-E Expression in 3D Models Is Associated with Resistance to NK Cell-Mediated Cytotoxicity. *J Proteome Res* **2014**, *13* (5), 2272–2281. https://doi.org/10.1021/pr500064m.
(24) Balachander, G. M.; Balaji, S. A.; Rangarajan, A.; Chatterjee, K. Enhanced Metastatic Potential in a 3D Tissue Scaffold toward a Comprehensive in Vitro Model for Breast Cancer Metastasis. *ACS Appl Mater Interfaces* **2015**, *7* (50), 27810–27822. https://doi.org/10.1021/acsami.5b09064.
(25) Vinci, M.; Gowan, S.; Boxall, F.; Patterson, L.; Zimmermann, M.; Court, W.; Lomas, C.; Mendiola, M.; Hardisson, D.; Eccles, S. A. Advances in Establishment and Analysis of Three-Dimensional Tumor Spheroid-Based Functional Assays for Target Validation and Drug Evaluation. *BMC Biol* **2012**, *10*. https://doi.org/10.1186/1741-7007-10-29.
(26) Kim, Y. E.; Jeon, H. J.; Kim, D.; Lee, S. Y.; Kim, K. Y.; Hong, J.; Maeng, P. J.; Kim, K. R.; Kang, D. Quantitative Proteomic Analysis of 2D and 3D Cultured Colorectal Cancer Cells: Profiling of Tankyrase Inhibitor XAV939-Induced Proteome. *Sci Rep* **2018**, *8* (1). https://doi.org/10.1038/s41598-018-31564-6.
(27) Craik, A. C.; Veldhoen, R. A.; Czernick, M.; Buckland, T. W.; Kyselytzia, K.; Ghosh, S.; Lai, R.; Damaraju, S.; Underhill, D. A.; Mackey, J. R.; Goping, I. S. The BH3-Only Protein Bad Confers Breast Cancer Taxane Sensitivity through a Nonapoptotic Mechanism. *Oncogene* **2010**, *29*, 5381–5391. https://doi.org/10.1038/onc.2010.272.
(28) Perez White, B. E.; Ventrella, R.; Kaplan, N.; Cable, C. J.; Thomas, P. M.; Getsios, S. EphA2 Proteomics in Human Keratinocytes Reveals a Novel Association with Afadin and Epidermal Tight Junctions. *J Cell Sci* **2017**, *130* (1), 111–118. https://doi.org/10.1242/jcs.188169.
(29) Wang, L. T.; Proulx, M. È.; Kim, A. D.; Lelarge, V.; McCaffrey, L. A Proximity Proteomics Screen in Three-Dimensional Spheroid Cultures Identifies Novel Regulators of Lumen Formation. *Sci Rep* **2021**, *11* (1). https://doi.org/10.1038/s41598-021-02178-2.
(30) Datta, S. R.; Katsov, A.; Hu, L.; Petros, A.; Fesik, S. W.; Yaffe, M. B.; Greenberg, M. E. 14-3-3 Proteins and Survival Kinases Cooperate to Inactivate BAD by BH3 Domain Phosphorylation. *Mol Cell* **2000**, *6* (1), 41–51. https://doi.org/10.1016/S1097-2765(05)00012-2.
(31) Githaka, J. M.; Vega, A. R.; Baird, M. A.; Davidson, M. W.; Jaqaman, K.; Touret, N. Ligand-Induced Growth and Compaction of CD36 Nanoclusters Enriched in Fyn Induces Fyn Signaling. *J Cell Sci* **2016**, *129* (22), 4175–4189. https://doi.org/10.1242/jcs.188946.
(32) Pandya, V.; Githaka, J. M.; Patel, N.; Veldhoen, R.; Hugh, J.; Damaraju, S.; McMullen, T.; Mackey, J.; Goping, I. S. BIK Drives an Aggressive Breast Cancer Phenotype through Sublethal Apoptosis and Predicts Poor Prognosis of ER-Positive Breast Cancer. *Cell Death & Disease 2020 11:6* **2020**, *11* (6), 1–19. https://doi.org/10.1038/s41419-020-2654-2.
(33) Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *J Am Stat Assoc* **1963**, *58* (301), 236–244. https://doi.org/10.1080/01621459.1963.10500845.
(34) Zare, A.; Postovit, L. M.; Githaka, J. M. Robust Inflammatory Breast Cancer Gene Signature Using Nonparametric Random Forest Analysis. *Breast Cancer Research* **2021**, *23* (1). https://doi.org/10.1186/s13058-021-01467-y.
(35) Xie, Z.; Bailey, A.; Kuleshov, M. V.; Clarke, D. J. B.; Evangelista, J. E.; Jenkins, S. L.; Lachmann, A.; Wojciechowicz, M. L.; Kropiwnicki, E.; Jagodnik, K. M.; Jeon, M.; Ma'ayan, A.

Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* **2021**, *1* (3). https://doi.org/10.1002/cpz1.90.