Measuring proteomic changes associated with viral infections and neurodegenerative diseases using mass spectrometry

By

Eman Walied Moussa

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

Master of Science

Department of Biochemistry

University of Alberta

Abstract

Viral infection and the resultant cellular stress can trigger significant changes to the cellular host proteome. Viral proteases essential to viral replication have been shown to cleave human host protein substrates as a mechanism of immune response evasion. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of COVID-19 and the resulting COVID-19 pandemic. The widespread impact of this virus illustrates the need to better understand its biology and the role of its viral proteases. We used mass spectrometry to characterize the host protein substrates of the SARS-CoV-2 viral proteases that play a role in its evasion of the immune response. We show that bromodomain-containing protein 2 (BRD2) and splicing factor proline- and glutamine-rich (SFPQ) are potential substrates of the SARS-CoV-2 main (M^{pro}) and papain-like viral proteases (PL^{pro}), respectively. The identification of these human host protein substrates aids in our understanding of the biology of viral infection and how the SARS-CoV-2 virus may evade the host's immune response. Our results may also provide a foundation of knowledge that can be used to guide studies of potential future coronaviruses, similar to how SARS-CoV-1 studies have guided SARS-CoV-2 studies.

Additionally, we highlight the use of mass spectrometry as a powerful tool to study proteomic changes in neurodegenerative diseases, viral infections in the cells of the central nervous system, and protein mistranslation. Our investigation into the role of gasdermin-D in multiple sclerosis, using a cuprizone mouse model, demonstrates changes in protein abundance with disease suggesting increased astrocyte and microglia activity. Furthermore, infection of human astrocytes with monkeypox virus (MPXV) and vaccinia virus (VACV) demonstrates the greater impact of MPXV infection on mitochondrial protein functions, compared to VACV infection. Lastly, we present alternative approaches for the identification of protein mistranslation due to oxidative stress and the disrupted function of the methionyl-tRNA synthetase (MARS). These

results highlight the need for bioinformatic and database-driven analysis, in addition to traditional mass spectrometry, for the identification of low abundance mistranslational events.

Preface

Sections of the material in Chapter two were initially published in *ACS infectious diseases* journal. SARS-CoV-2 PL^{pro} experiments were conducted by Eman Moussa. SARS-CoV-2 M^{pro} experiments were conducted by Shu Luo. Reprinted (adapted) with permission from Luo *et al*., 2023 (complete reference citation below). Copyright 2023 American Chemical Society.

> 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. (2023). Identification of Human Host Substrates of the SARS-CoV-2 Mpro and PLpro Using Subtiligase N-Terminomics. *ACS Infectious Diseases*, *9*(4), 749–761. https://doi.org/10.1021/acsinfecdis.2c00458.

Sections of the material in Chapter two were initially published in *ACS central science* journal. Mass spectrometry experiments were conducted by Eman Moussa. Galectin-8 cleavage assay samples were provided by Sizhu Amelia Chen and Elena Arutyunova. Reprinted (adapted) from Chen *et al*., 2023 (complete reference citation below). Copyright 2023 American Chemical Society. Published article is licensed under a Creative Commons Attribution 4.0 International license.

> Chen, S. A., Arutyunova, E., Lu, J., Khan, M. B., Rut, W., Zmudzinski, M., Shahbaz, S., Iyyathurai, J., **Moussa, E. W.**, Turner, Z., Bai, B., Lamer, T., Nieman, J. A., Vederas, J. C., Julien, O., Drag, M., Elahi, S., Young, H. S., & Lemieux, M. J. (2023). SARS-CoV-2 Mpro Protease Variants of Concern Display Altered Viral Substrate and Cell Host Target Galectin-8 Processing but Retain Sensitivity toward Antivirals. *ACS Central Science*, *9*(4), 696–708. https://doi.org/10.1021/acscentsci.3c00054

Sections of the materials in Chapter three have been submitted for publication and are currently under review. Mass spectrometry experiments were conducted by Eman Moussa.

> Pollock, N.M., Fernandes, J.P, **Moussa, E.W.**, Hlavay, B., Branton, W.G., Wuest, M., Mohammadzadeh, N., Laura Schmitt, L., Plemel, J.R., Julien, O., Wuest, F., & Power, C. (2023). Gasdermin D activation in CNS macrophages regulates inflammatory demyelination and axonal injury in progressive multiple sclerosis. (Under review)

The materials in Appendix D have been submitted for publication and are currently under review. Mass spectrometry experiments were conducted by Eman Moussa.

> Beauchamp, E., Gamma, J., Cromwell, C., **Moussa, E.W.**, Iyer, A., Yap, M., Pain, R., Kostiuk, M., Julien, O., Hubbard, B., Mackey, J.R., & Berthiaume, L.G. (2023). Myristoylation inhibition impacts oxidative phosphorylation and defines a 91 gene sensitivity signature, MISS-91, that identifies new cancer indications for NMT inhibitor PCLX-001. (Under review)

The mass spectrometry experiments and analysis in section 3.1 were conducted by Eman Moussa, excluding cyanogen bromide-based experiments conducted by Angelle Britton (Dr. Sue-Ann Mok lab) with assistance from Jack Moore at the Alberta Proteomics and Mass Spectrometry core. Immunoblot experiments were performed by Angelle Britton and analyzed by Eman Moussa. Samples for proteomic analysis were provided by Angelle Britton.

Human cell samples in sections 3.2.2 and 3.2.3 as well as monkeypox and vaccinia virus-infected samples in section 3.3 were provided by Dr. Hajar Miranzadeh Mahabadi, of the Dr. Christopher Power lab. The associated mass spectrometry experiments and analysis were conducted by Eman Moussa.

Proteomics in section 3.2.2 was conducted with Henry Wang and in Appendix G was conducted with Kolden Van Baar. Tau protein samples in Appendix E were provided by Angelle Britton (Dr. Sue-Ann Mok lab). Cell samples in Appendix D were provided by Dr. Erwan Beauchamp (Dr. Luc Berthiaume lab), in Appendix F were provided by Emily McNamara (Dr. Sue-Ann Mok lab), and in Appendix G were provided by Angelle Britton.

Acknowledgments

I would like to thank my supervisor Dr. Oliver Julien for his guidance and support throughout my time in the lab. I value the opportunities and experiences that my time in the lab has afforded me. I will take this knowledge with me on my journey forward to great places.

Thank you to my committee members Dr. Richard Fahlman and Dr. Holger Wille for their helpful advice and words of wisdom. Thank you Dr. Fahlman for your conversations in passing, from which I learned many interesting facts that have guided my curiosity and inquisition. An additional thank you to our many collaborators without whom I could not have accomplished all this work.

Thank you, Dr. Adrienne Wright and Dr. Jonathan Parrish, for educating me with your knowledge and improving my skills as an educator.

I'd like to thank my senior graduate students Erik Gomez-Cardona, Bridgette Hartley, Shu "Lucy" Luo, and past lab mate Luam Araya for their help and assistance in the lab. I'd also like to thank Kolden Van Baar who has been a welcome addition to our lab, and our long-term undergraduate student Henry Wang. Thank you for game nights, birthday cakes, lunches, road trips, and the rest of our misadventures.

I thank my friends in the department and the Biochemistry Graduate Students Association for making these years, despite the pandemic, an enjoyable time. Thank you, Angelle Britton, for your constant support and comradery. I would also like to thank my friends who I met at this university and have now gone far and wide in pursuit of their dreams. Thank you for your friendship, virtual video game nights, gift exchanges, and much more.

I would especially like to thank my family and lifelong friends for their support, love, shoulder to lean on, and listening ear, without whom I would not have gotten through these last years and our loss of my father. Mom and Yousef, you have been a constant shining north star in an endless sky. Thank you, Tante Soha, for being a second mother to me. I am forever thankful for your constant encouragement, words of wisdom, willingness to listen, late-night calls, and overall caring nature that means more to me than words can describe.

عَنْ أَبِي هُرَيْرَةَ رَضِيَ اللَّهُ عَنْهُ :قَالَ قَالَ رَسُولُ اللَّهِ ,صَلَّى اللَّهُ عَلَيْهِ وَسَلَّم **اللَّهُمَّ انْفَعْنِي بِمَا عَلَّمْتَنِي وَعَلِّمْنِي مَا يَنْفَعُنِي وَزِدْنِي عِلْمًا**

Abu Huraira reported: The Messenger of Allah, peace and blessings be upon him, said, “O Allah, benefit me with what you have taught me, teach me what will benefit me, and increase my knowledge.”

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Terminology

A549: Immortalized human lung carcinoma cells

A549-ACE2: Human lung carcinoma cells overexpressing the ACE2 receptor

Abu: Aminobutyric acid

ACAM2000: A strain of the vaccinia virus

AI: Artificial intelligence

CID: Collision induced dissociation

CNS: Central nervous system

COVID-19: Coronavirus disease of 2019

DNA: Deoxyribonucleic acid

dotp: Dot product

E. coli: *Escherichia coli*

EBV: Epstein-Barr virus

EC: Effective concentration

ER: Endoplasmic reticulum

ESI: Electrospray ionization

HEK293 P301L: Immortalized human embryonic kidney cells that express inducible tau P301L

HEK293: Immortalized human embryonic kidney cells

HeLa: Immortalized human epithelial cells

h.p.i: Hours post-infection

HPV: Human papillomavirus

iPSCs: Induced pluripotent stem cells

iTRAQ: Isobaric tag for relative and absolute quantification

Jurkat: Immortalized human T lymphocyte cells

KO: Knockout

LC: Liquid chromatography

LC-MS/MS: Liquid chromatography tandem mass spectrometry

m/z: Mass-to-charge ratio

MALDI: Matrix-assisted laser desorption ionization

MALDI-TOF: Matrix-assisted laser desorption ionization-time of flight

MERS: Middle east respiratory syndrome

MO3.13: Human glial oligodendrocytic hybrid cells

MOI: Multiplicity of infection

Mpox: Monkeypox disease

MPXV: Monkeypox virus

mRNA: Messenger ribonucleic acid

MS/MS (or MS^2): Tandem mass spectrometry

MS: Mass spectrometry

PCR: Polymerase chain reaction

PMT: Photomultiplier tube

PRM: Parallel reaction monitoring

PTMs: Post-translational modifications

RFLC: Reverse phase liquid chromatography

RNA: Ribonucleic acid

SARS-CoV-1: Severe acute respiratory syndrome coronavirus 1

SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2

SDS-PAGE: Sodium dodecyl-sulfate polyacrylamide gel electrophoresis

SILAC: Stable isotope labeling by amino acids in cell culture

SPE: Solid phase extraction

SRM/MRM: Selective reaction monitoring or multiple reaction monitoring

TAILS: Terminal amine isotopic labeling of substrates

tRNA: Transfer ribonucleic acid

VACV: Vaccinia virus

Vero E6: African green monkey kidney cells

WT: Wildtype

Proteins

ACE2: Angiotensin-converting enzyme 2

ATF-1: Cyclic AMP-dependent transcription factor ATF-1 (activating transcription factor 1)

ATG16L1: Autophagy-related protein 16-1

BDI: Bromodomain I

BET: Bromodomain and extra-terminal domain

BRD2: Bromodomain-containing protein 2

BRD3: Bromodomain-containing protein 3

BRD4: Bromodomain-containing protein 4

BRE1A: E3 ubiquitin-protein ligase BRE1A (RNF20 gene)

CARD8: Caspase recruitment domain-containing protein 8

cGAS-STING-IRF3: cyclic GMP-AMP synthase-stimulator of interferon genes- interferon regulatory transcription factor 3

CREB1: Cyclic AMP-responsive element-binding protein 1

DNAJA2: DnaJ homolog subfamily A member 2

eIF4G1: Eukaryotic translation initiation factor 4 gamma 1

ERK 1/2: extracellular signal-related kinase ½

FOXP3: Forkhead box P3 protein

FYCO1: FYVE and coiled-coil domain-containing protein 1

Gal8: Galectin-8

GFAP: Glial fibrillary acidic protein

GFP: Green fluorescent protein

GSDMD: Gasdermin-D

HDP-RNP: HEXIM1-DNA-PK-paraspeckle components-ribonucleoprotein complex

HER4: Receptor tyrosine-protein kinase erbB-4

IFN: Interferon

IRF3: Interferon regulatory transcription factor 3

ISG15: Interferon-stimulated gene 15

JIP4: c-Jun amino-terminal kinase-interacting protein 4

LANA1: latency-associated nuclear antigen 1

MAP4K5: Mitogen-activated protein kinase kinase kinase kinase 5

MARS: Methionyl-tRNA synthetase

MARS-SA: Methionyl-tRNA synthetase non-phosphomimetic (dual serine to alanine mutant)

MARS-SD: Methionyl-tRNA synthetase phosphomimetic (dual serine to aspartate mutant)

M^{pro}: Main protease

MYH6: Myosin heavy chain 6

MYH7: Myosin heavy chain 7

NEMO: NF-kappa-B essential modulator

NF-kappa-B: Nuclear factor kappa B

NLRP12: NACHT, LRR and PYD domains-containing protein 12

NMT1: N-myristoyltransferase 1

NMT2: N-myristoyltransferase 2

NSP: Non-structural protein

Nup98-Nup96: Nuclear pore complex protein

PARP10: Poly (ADP-ribose) polymerase 10

PL^{pro}: Papain-like protease

PROS1: Vitamin K-dependent protein S

PTBP1: Polypyrimidine tract binding protein 1

P-TEFb: Positive transcription elongation factor b

RAPGEF1: Rap guanine nucleotide exchange factor 1

RPAP1: RNA polymerase II-associated protein 1

SFPQ: Splicing factor proline- and glutamine-rich

TAB1: TGF-beta-activated kinase 1 and MAP3K7-binding protein 1

TEV: Tobacco etch viral

TLRs: Toll-like receptors

TRIM28: Transcription intermediary factor 1-beta

UAE: Ubiquitin-activating enzyme

ULK1: Unc51-like kinase 1

YAP1: Yes-associated protein 1

ZAP (or ZC3HAV1): The zinc finger antiviral protein

Reagents

ACC: 7-amino-4-carbamoylmethylcoumarin

ACN: Acetonitrile

AEBSF: 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride

BME: β-mercaptoethanol, also written as 2-mercaptoethanol

CNBr: Cyanogen bromide

CPZ: Cuprizone

DCM: Dichloromethane

ddH_2O: Double distilled (or MilliQ system purified by reverse osmosis) water

DMEM: Dulbecco's Modified Eagle medium

DMF: Dimethylformamide

DTT: Dithiothreitol

EDTA: Ethylenediaminetetraacetic acid

FBS: Fetal bovine serum

Fmoc: Fluorenylmethoxycarbonyl protecting group

Gdn-HCl: Guanidine hydrochloride

GST: Glutathione S-transferases

H_2O_2: Hydrogen peroxide

HEPES: N-2-hydroxyethylpiperazine-N-2-ethane sulfonic acid

IAM: Iodoacetamide

IPTG: Isopropyl β-d-1-thiogalactopyranoside

KCN: Potassium cyanide

LB: Lysogeny Broth

MEM: Minimum Essential Medium

NaCl: Sodium chloride

PBS: Phosphate-buffered saline

PCLX-001: A pan-N-myristoyltransferase (NMT) inhibitor

PMA: Phorbol-12-myristate-13-acetate

PMSF: Phenylmethylsulfonyl fluoride

RIPA: Radioimmunoprecipitation assay

SDS: Sodium dodecyl-sulfate

TAK-243: An inhibitor of the Ubiquitin-activating, or E1, enzyme

TBS: Tris-buffered saline

TBST: Tris-buffered saline with Tween® 20 detergent

TCEP: Tris(2-carboxyethyl)phosphine

TFA: Trifluoroacetic acid

z-VAD-fmk: A pan-caspase inhibitor

Glossary of Terms

CRISPR screening: Clustered regularly interspaced short palindromic repeats. Refers to the use of large-scale genomic screening with the incorporation of mutations to determine gene function and phenotype.

C-terminus (plural. C-termini): The free carboxyl group bearing end of a protein or polypeptide, referred to as the end of the protein or polypeptide.

DegraBase: A database of previously observed cleavage sites in healthy and apoptotic cells from subtiligase-based N-terminomics experiments (Crawford et al., 2013).

IceLogo: A tool for visualization of conserved patterns in protein and nucleotide sequences that incorporates probability theory (Colaert et al., 2009). This tool is utilized in protease studies to visualize the pattern observed for the cleavage site sequences of identified substrates.

MassIVE: Mass Spectrometry Interactive Virtual Environment. Refers a data repository for the storage and public assess of raw mass spectrometry data files. This repository can be accessed at https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp.

Metascape: A web resource for gene ontology analysis and annotation (Y. Zhou et al., 2019).

N-terminomics: The study of peptide or protein N-termini, typically using N-terminal labeling.

N-terminus (plural. N-termini): The free amine group bearing end of a protein or polypeptide, referred to as the start of the protein or polypeptide.

ProteinProspector: An online proteomics tool for mass spectrometry data analysis by mining proteomic sequence databases.

Proteome Discoverer: A Thermo Fisher Scientific software for the analysis of mass spectrometry experimental data, utilized in the following studies for label-free quantification by data-dependent acquisition (DDA).

Reverse N-terminomics: The utilization of N-terminal labeling to identify cleavage events by mass spectrometry. The cleavage events are produced by the addition of a purified known protease of interest.

Schechter and Berger nomenclature: Refers to the nomenclature utilize to describe the residue positions within the site cleaved by a protease as well as the substrate pocket within the protease itself (Schechter & Berger, 1967). Residues N-terminal and C-terminal to the cleavage site are denoted as P# and P#′, respectively, where the numbering begins at 1 and increases in ascending order for residues further from the cut site.

Spectronaut®: A Biognosys software designed and optimized for the analysis of mass spectrometry experimental data acquired by data-independent acquisition (DIA).

Subtiligase-based N-terminomics: The utilization of the enzyme subtiligase to label new N-termini with a biotinylated peptide ester tag for enrichment and identification of cleavage events by mass spectrometry.

TopFind 4.1: A publicly available knowledgebase for proteolytic processing that ca be utilized to assign protein fragments and their N-termini with possible protease cleavage events (Fortelny et al., 2015).

Chapter One: Introduction

1.1 Proteomics

1.1.1 Proteomes can be modified in disease

Proteins are essential for cellular function and life. The central dogma defines the process by which deoxyribonucleic acid (DNA) is transcribed into ribonucleic acid (RNA), RNA is processed by splicing into messenger RNA (mRNA), and mRNA is translated into proteins made up of amino acids. These proteins play a fundamental role in the physiology of cellular organisms. The profile of all proteins expressed by a species or cell line is known as its proteome. Different species and cell types express differing proteomes unique to their needs.

Under healthy conditions, these proteomes operate to maintain normal cell function and homeostasis but during disease, these proteomes can change. The study of these proteomes and their changes in health and diseases is known as proteomics. These changes can vary from post-translational modifications (PTM) to decreased or increased protein expression and abundance. One such PTM is proteolysis.

1.1.2 Proteolysis in viral infection

Proteolysis is the hydrolysis of peptide bonds by enzymatic proteins called proteases. Proteases are present in all kingdoms of life as well as viruses. These proteins can be endogenous to cells, with their own set of functions, or exogenous and introduced to cells, for example during viral infection. Endogenous proteases play an essential role in maintaining cellular homeostasis and regulation through processes, such as protein processing and function, apoptosis, and digestion. In contrast, exogenous proteases serve an essential role for their source organism or virus and can serve a disruptive role for the host organism.

As a case example, viruses can possess virally encoded protease(s) that are responsible for the cleavage of translated viral polyproteins into their functional protein units. But in addition to this role, there have been many studies conducted that describe the role of viral proteases in the cleavage of host substrates. The viral proteases of several RNA viruses including coronaviruses and flaviviruses such as the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and Zika virus, respectively, have been widely studied in their role of host substrate targeting and immune response evasion (Hill et al., 2018; Miczi et al., 2020; Mohamud et al., 2021; Moustaqil et al., 2021; Pablos et al., 2021; Reynolds et al., 2021; Tsu et al., 2022; Wenzel et al., 2021; S. Zhang et al., 2021). The host substrates of the Zika virus protease NS2B–NS3 have been successfully studied utilizing a protein cleavage site labeling technique called subtiligase-based

N-terminomics coupled with mass spectrometry (Hill et al., 2018). Hill et al. identified 31 putative Zika virus protease substrates, including proteins that play a role in or are related to the antiviral immune response, such as c-Jun amino-terminal kinase-interacting protein 4 (JIP4), autophagy-related protein 16-1 (ATG16L1), and eukaryotic translation initiation factor 4 gamma 1 (eIF4G1). As this N-terminal labelling mass spectrometry technique has been previously shown to aid in the identification of viral protease host substrates, it stands that it could provide a useful tool in the elucidation of the potential additional role of other viral proteases, such as the main protease (M^{pro}) and papain-like protease (PL^{pro}) of the recent coronavirus SARS-CoV-2.

Studies regarding PL^{pro} of the *Coronaviridae* family, including the severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1), have primarily focused on its role as a deubiquitinase (Liu et al., 2021; Shin et al., 2020; Sun et al., 2012), with some exceptions further discussed in **Chapter two** (Mohamud et al., 2021; Moustaqil et al., 2021; Reynolds et al., 2021). Ubiquitination refers to the conjugation of ubiquitin to proteins as a PTM (see ref (Komander & Rape, 2012) for a review). Ubiquitination plays an important role in the activation of cellular signaling related to the antiviral immune response, and its disruption aids in the evasion of this immune response (see ref (Chiang et al., 2021) for a review). This also applies to ISGylation with ubiquitin-like molecules, such as interferon-stimulated gene 15 (ISG15) (see ref (Perng & Lenschow, 2018) for a review), which structurally resembles di-ubiquitin (Narasimhan et al., 2005) and has 30% homology with ubiquitin (Dao & Zhang, 2005). However, it has been previously shown that PL^{pro} can cleave within proteins, not just to deubiquitinate or deISGylate proteins. Therefore, further investigation is required to elucidate the potential role PL^{pro} may serve not only as a deubiquitinase but also in the cleavage of host proteins.

1.1.3 Additional Post-translational modifications during viral infection

In addition to proteolysis, other PTMs have been shown to occur to host proteins by the host as a consequence of viral infection. For example, the infection of HeLa cells with influenza A, adenovirus, and vaccinia virus disrupts the proper function of the aminoacyl-tRNA responsible for loading methionine onto its corresponding tRNA, known as the methionyl-tRNA synthetase (MARS) (Netzer et al., 2009). More specifically, MARS can misload non-cognate tRNAs with methionine to an increased extent and this phenomenon can result in the incorrect incorporation of methionine during protein translation. This mistranslation is a consequence of the dual phosphorylation of MARS at two serine residues, which can occur as a result of viral infection and oxidative stress (J. Y. Lee et al., 2014). Therefore, viral infection and the consequential immune

and stress responses lead to PTMs that disrupt the proper function of MARS and result in mistranslation during protein synthesis.

1.1.4 The long-term effects of viral infection

Viral proteases have been shown in several cases to cleavage host proteins during infection to evade the host immune response and additional PTMs can result from viral infection. There are also more long-term consequences of viral infection. Particularly, the development and later onset of diseases. A common example is the link between human papillomavirus (HPV) infection and cervical cancer (see ref (Bosch et al., 2002) for a review). But there are also neurodegenerative diseases that have been suspected to develop as a result of earlier viral infections.

Multiple sclerosis is a neurodegenerative disease characterized by demyelination of neurons and inflammatory cell death, called pyroptosis (see refs (Karussis, 2014; McKenzie, Dixit, et al., 2020; Steinman, 2001) for reviews). The mechanism of pyroptosis involves the cleavage of the protein gasdermin-D by the inflammatory caspase proteases (He et al., 2015; Kayagaki et al., 2015; Shi et al., 2015). The N-terminal gasdermin-D product oligomerizes and forms pores at the plasma membrane allowing for inflammation and ultimately cellular rupture. While the exact causes of multiple sclerosis are unknown, it has been proposed that previous infections with the Epstein-Barr virus (EBV) can increase the risk of developing multiple sclerosis (Bjornevik et al., 2022). EBV viral infection is associated with infectious mononucleosis (Thorley-Lawson, 2001), which is known to increase the risks of developing multiple sclerosis (Thacker et al., 2006). Additionally, evidence of the virus has been identified in the lesions of human multiple sclerosis patients during some (Hassani et al., 2018; Moreno et al., 2018; Serafini et al., 2007), but not all (Sargsyan et al., 2010; Willis et al., 2009) pathological studies. EBV appears to be one part of a larger viral signature identified in multiple sclerosis based on investigations of antiviral antibodies from multiple sclerosis patients (Enose-Akahata et al., 2023). As such, infection with EBV may be a risk factor for the development of multiple sclerosis along with genetics and environmental factors, such as low levels of vitamin D and smoking (see ref (Olsson et al., 2017) for a review).

Overall, EBV is one of many neurotropic viruses capable of infecting the central nervous system (CNS), whose list includes but is not limited to Zika virus (Alimonti et al., 2018), Influenza A (Hara et al., 2021), SARS-CoV-2 (see ref (Tavčar et al., 2021) for a review), and Monkeypox virus (MPXV) (Chailangkarn et al., 2022).

1.1.5 Neurotropic viral infections

While many neurotropic viruses have been identified and studied throughout the 20th and 21st centuries, MPXV is a recent addition to this list (Chailangkarn et al., 2022). Several severe symptoms of MPXV infection, including seizure and encephalitis, have been reported and illustrate its potential neuroinvasive properties (see ref (Sepehrinezhad et al., 2023) for a review). MPXV was also identified to infect astrocytes but not neurons derived from human induced pluripotent stem cells (iPSC) (Chailangkarn et al., 2022). Neurons are the cells that transmit chemical and electrical signals within the CNS, while astrocytes are glial cells that primarily maintain homeostasis within the CNS. Understanding the biology of MPXV infection in the CNS lies in understanding why astrocytes, as opposed to neurons, are infected with the virus.

In comparison to MPXV, other neurotropic viral infections have been identified to infect neurons as well as astrocytes, such as the Rabies virus (RABV) (Feige et al., 2022), West Nile virus (WNV) (Quick et al., 2014; Shrestha et al., 2003), and even EBV (Jha et al., 2015; Menet et al., 1999), while others, such as SARS-CoV-2, preferentially infect astrocytes (Andrews et al., 2022; Tavčar et al., 2021). The targeting of specific cells for infection may explain the severity of symptoms observed ranging from dizziness and headaches to encephalitis, meningitis, seizures, paralysis, or even death (see refs (Fooks et al., 2017; Petersen et al., 2013; N. Zhang et al., 2022; Z. Zhou et al., 2020) for reviews).

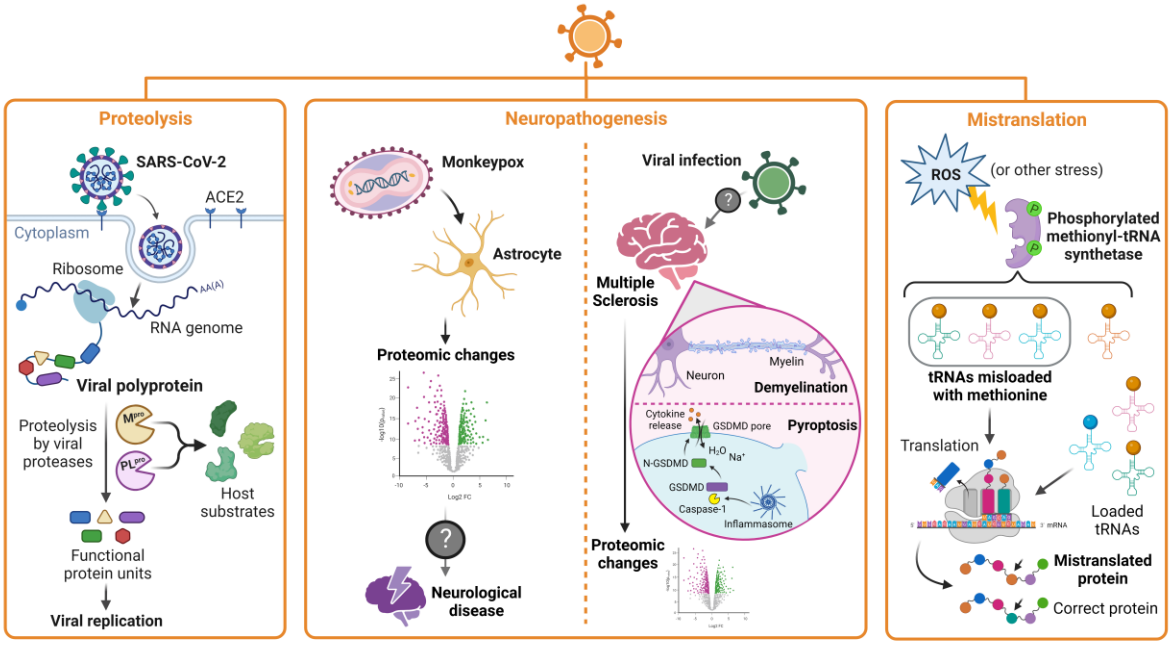


Figure 1.1. Consequences of viral infection on the human host proteome. The extent to which viral infection can change the host proteome can be studied in many contexts. During viral infection, following viral entry by SARS-CoV-2, the viral proteases are responsible for the proteolysis of viral polyproteins in function proteins which is necessary for viral replication and spread. Additionally, these viral proteases have been implicated in the cleavage of host protein substrates as a mechanism to evade the host immune response. Additionally, viral infection has implications for neuropathogenesis. Viruses such as MPXV have been previously identified to be capable of infecting cells within the CNS. As astrocytes play a key role in the maintenance of homeostasis within the CNS, proteomic changes to these cells following viral infection may provide key insights into the biology and why this virus may target these cells over other cells such as neurons. Previous viral infection has also been hypothesized to result in greater susceptibility to the development of neurodegenerative diseases such as multiple sclerosis. Multiple sclerosis is characterized by the demyelination of neurons within the brain and inflammatory cell death, called pyroptosis, whereby pores in the plasma membrane are formed by caspase-cleaved N-terminal gasdermin-D (N-GSDMD). The proteomic changes associated with disease and gasdermin-D (GSDMD) expression may highlight its role in the pathogenesis of multiple sclerosis. Lastly, viral infection and reactive oxidative stress have been implicated in the dual phosphorylation MARS. This PTM disrupts the normal function of MARS and allows it to misload methionine onto non-cognate tRNAs, leading to incorrect protein translation and increased incorporation of methionine.

1.2 Mass spectrometry is a powerful tool to study proteomic changes in disease

The success of proteomic studies can be dependent on the techniques and methodologies applied. One commonly utilized technology that continues to advance is the use of liquid chromatography tandem mass spectrometry (LC-MS/MS). LC-MS/MS can be utilized for the detection of peptides based on their mass-to-charge ratio (m/z). It can also be used to sequence the peptides identified based on fragment ions that include fragmented portions of these peptides. While there are several different techniques for LC-MS/MS, and optimal methods will depend on the instrument and detector itself, they share a common basic workflow (see ref (Glish & Vachet, 2003) for a review).

1.2.1 The basic workflow of liquid chromatography tandem mass spectrometry

The first step of LC-MS/MS involves the separation of a complex peptide mixture by liquid chromatography (LC) to allow for the analysis of a less complex mixture of peptides over time. This provides the instrument time to analyze the peptides with limited obscuring of lower abundance peptides by the higher abundance peptides. Reverse phase liquid chromatography (RFLC) on a C18 column of non-polar hydrocarbon can be used in combination with a polar reagent, typically acetonitrile, to elute the peptides at increasing concentration over a chosen time (Peng & Gygi, 2001).

The mixture of peptides is then injected into an instrument in an ionization step. One kind of ionization is called electrospray ionization (ESI) (Fenn et al., 1989; Yamashita & Fenn, 1984). The charged peptides are present in liquid droplets as they leave an ionizer. As they approach the instrument's heated capillary tube, the liquid evaporates such that only the charged peptides are injected. Under vacuum, the charged peptides can then be sorted in a mass analyzer by their m/z and subsequently detected in an MS^1 scan. Further fragmentation of these peptides into fragment ions can be conducted by collision induced dissociation (CID), whereby the collision of ions with neutral species generates the energy to fragment the charged peptides (McLuckey, 1992; Mitchell Wells & McLuckey, 2005). Subsequent detection of these fragment ions is referred to as tandem mass spectrometry (MS/MS) or the MS^2 scan. The results can be presented as both a chromatogram of retention time vs intensity or relative abundance as well as MS^1 and MS^2 spectra of m/z vs intensity. To subsequently analyze the data, previously built databases of protein sequences and their associated peptides can be compared to the experimentally obtained results. This is typically done by specialized software applications as significant computational analysis and comparisons are required.

1.2.2 Sample preparation and data acquisition

Despite the similar basic workflow, differences in LC-MS/MS methods can arise from the utilization of different sample preparation methods such as in-gel digestion (Shevchenko et al., 2006), in-solution digestion, and various forms of on-column digestion (Crowell et al., 2015; Ludwig et al., 2018) (see **section 3.2.3** for optimization of sample preparation). Different methods are typically optimized for different proteome coverage. With in-gel digestion and on-column digestion, harsher detergents can be used to lyse cells, and thus insoluble proteins such as membrane proteins are more likely to be detected by these methods than with gentler in-solution digestion. In addition to sample preparation, different MS/MS detection methodologies can be used.

Two methods of MS/MS allow for the detection of complex protein mixtures or shot-gun proteomics, data-dependent acquisition (DDA) and data-independent acquisition (DIA) (see ref (Hu et al., 2016) for a review). The primary difference is that DIA allows for the MS^2 fragmentation detection of all peptides identified at the MS^1 level whereas DDA only fragments some chosen peptides from the MS^1 scan based on the initial data acquired (see **section 3.2.4** for an explanation and analysis of DIA vs DDA). In addition to shotgun proteomic methods, utilizing DDA workflows allows for targeted mass spectrometry methods to be explored. Targeted mass spectrometry refers to the continuous monitoring of specific chosen peptides or fragment ions for their quantification. This group encompasses parallel reaction monitoring (PRM) and selective reaction monitoring or multiple reaction monitoring (SRM/MRM) (Peterson et al., 2012). Initial DDA is used as a discovery to identify the peptides present in a sample. Based on these results, select chosen peptides and their corresponding m/z, charge, and retention time are used to monitor these peptides by targeted methods. In both PRM and SRM/MRM, only specific precursors (charged peptides) that are selected are monitored in the MS^1 scan and fragmented. With PRM, all the fragment ions are monitored by MS^2 scan, while with SRM/MRM specific fragment ions are chosen to be monitored by MS^2. Ultimately, the technique chosen will depend on the purpose of the experiment.

Traditionally, DDA has not only been simpler to set up and analyze but also allows for the later use of targeted analysis, such as PRM and SRM/MRM, that offers more sensitive quantification albeit for a limited subset of chosen precursors. Additionally, labeling and internal control approaches have been used with DDA, including stable isotope labeling by amino acids in cell culture (SILAC) (Ong & Mann, 2006), isobaric tag for relative and absolute quantification (iTRAQ)

(Wiese et al., 2007), N-terminal acetylation (McDonald et al., 2005), and subtiligase-based N-terminomics (Weeks & Wells, 2018), to name a few. Some disadvantages of DDA include the level of bias introduced by the instrument's choice of which top precursors to fragment for MS/MS. This can often mean that lower abundance peptides become masked by higher abundance peptides. However, in terms of library and database generation, DDA is extremely useful as the sequential MS/MS peptide acquisition allows for the generation of assigned MS^2 spectra for each precursor. DIA can overcome part of the disadvantages of DDA as all the precursors are fragmented and analyzed by MS/MS. This reduces bias but also means that the MS spectra acquired are highly complex and require greater computational capabilities to deconvolute, analyze, and assign to each precursor. Recent software applications specialized for DIA, such as Spectronaut® by Biognosys, have incorporated complex algorithms and artificial intelligence (AI) to allow for better identification of precursors based on sample cross-comparative analysis. However, these software applications remain not only expensive but computationally demanding. Overall, both DIA and DDA have their advantages and disadvantages and thus the technique chosen for any experiment needs to be complementary to the goals of the study and the type of biological samples studied.

Chapter Two: Proteolysis during SARS-CoV-2 Viral Infection

2.1 Introduction to SARS-CoV-2

2.1.1 Coronaviruses: past, present, and future

Coronaviruses are a large family of RNA viruses named for their characteristic spherical shape with protruding surface proteins that resemble the corona of the sun. These viruses can infect both humans and animals with one of the earliest reports of coronavirus infections in the early 1930s identified in chickens with infectious bronchitis virus (IBV) (Beach & Schalm, 1936; Bushnell & Brandly, 1933; Schalk & Hawn, 1931). However, it wasn't until 1965 that the first human coronavirus was identified to cause the common cold.

In 1946, an effort known as the Common Cold Unit was founded to conduct research into the cause of the common cold. Directed by Dr. David Tyrrell until its closure in 1990, the research initiative studied nearly 9000 volunteers, published 1006 papers from 1948 to 1989 (Oransky, 2005), and involved efforts based in both Britain and the United States (D. Tyrrell, 1990). During the 1960s, the Common Cold Unit identified a virus from a nasal swab labeled B814 (D. A. J. Tyrrell & Bynoe, 1965) as one of the causes of the common cold. Around the same time, at the University of Chicago, Dr. Dorothy Hamre studied samples obtained from students with cold symptoms and identified another virus that was named 229E (Hamre & Beem, 1972; Hamre & Procknow, 1966). Similarly, Dr. McIntosh and his team at the National Institutes of Health discovered similar organ culture or "OC" viruses including OC43 and OC38 in their effort to identify the cause(s) of the common cold (McIntosh, Becker, et al., 1967; McIntosh, Dees, et al., 1967).

In collaboration with Dr. June Almeida, a virologist of international acclaim known for pioneering the technique of immune electron microscopy, the Common Cold Unit successfully imaged the first identified coronaviruses linked to the common cold ("Virology," 1968). In 1968, they compared the structures of IBV from chickens to that of B814, 229E, and other strains isolated from humans. Based on the structural similarity to the "solar corona" they suggested that this family of viruses be named coronaviruses. Dr. David Tyrrell, Dr. Dorothy Hamre, Dr. McIntosh and Dr. June Almeida are four of the eight virologists involved in the initial submission to *Nature* and coining the term "coronavirus", establishing the existence of a previously unknown viral family.

The time of the most recent common ancestor of coronaviruses has been suggested to be about 10,000 years ago based on molecular clock dating (Woo et al., 2012). Evolutionary models suggest that this value is an underestimation and that coronaviruses may have existed tens of millions of years ago if not longer in co-evolution with their bat and avian hosts (Wertheim et al., 2013). But despite the origins and history of these viruses, the modern coronaviruses responsible

for epidemics and pandemics alike, such as MERS, SARS-CoV-1, and now SARS-CoV-2, result in infections with symptoms exceeding the severity of the common cold.

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of COVID-19 and the resulting pandemic, totaling over 760 million confirmed cases and 6.8 million deaths globally as of March 16th, 2023 (*WHO Coronavirus (COVID-19) Dashboard*, 2023) - just over three years since cases began in December 2019. Similar to the Common Cold Unit, massive international efforts to study the causative virus amassed approximately 2 million papers mentioning “SARS-CoV-2” according to Google Scholar over three years. While the exact origin of the virus and its potential transmission from bats to humans is still a topic of study and discussion, the prophylactic and therapeutics efforts have made large steps forward, even paving the way for mRNA-based vaccines. Understanding the biology of the virus is key to the development of treatments not only for the initial strain originating in Wuhan, China but for the variants of concern that have since spread globally and will continue to arise in the future (CDC, 2020; Siddiqui et al., 2022).

2.1.2 SARS-CoV-2 and its viral proteases

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 (Lu et al., 2020; F. Wu et al., 2020; P. Zhou et al., 2020; Zhu et al., 2020). Two of the viral proteins, 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 that are important for viral replication, including helicase and RNA-dependent RNA polymerase (see ref (Ullrich & Nitsche, 2020) for a review). Due to the critical roles 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 (Vuong et al., 2020). Currently, Paxlovid (oral antiviral drug nirmatrelvir/ritonavir, Pfizer) is the only approved COVID-19 treatment targeting the SARS-CoV-2 viral protease (Abdelnabi et al., 2022; Boras et al., 2021; de Vries et al., 2021; Owen et al., 2021).

The two SARS-CoV-2 proteases are named according to their catalytic and structural similarities to other known enzymes. $NSP3^{pro}$ is also known as papain-like protease (PL^{pro}) and cleaves at three sites in the polyproteins pp1a and pp1ab. $NSP5^{pro}$ or picornaviral 3C-like protease ($3CL^{pro}$),

cleaves at eleven sites and is thus also referred to as the main protease (M^{pro}). Both SARS-CoV-2 proteases are cysteine proteases. The active site of SARS-CoV-2 M^{pro} contains a Cys145-His41 catalytic dyad. Based on its native cleavage consensus sequence in the polyproteins and its crystal structure (L. Zhang et al., 2020), M^{pro} preferentially cleaves after glutamine (P1=Gln, Schechter and Berger nomenclature) (Schechter & Berger, 1967), which allows stabilization in its S1 pocket by three hydrogen bonds (Rut et al., 2021). Studies on SARS-CoV-1 M^{pro} show that cleavage can also occur after histidine but with lower frequency (Chuck et al., 2010). PL^{pro} has a canonical cysteine protease catalytic triad Cys111-His272-Asp286, and is a multifunctional protein with both proteolytic and mainly deubiquitinating activities (Freitas et al., 2020; Gao et al., 2021). It cleaves almost exclusively after residues GlyGly at P1 and P2 positions, with a high preference for hydrophobic residues in P4 (Leu in particular) and broader specificity in P3 (Rut et al., 2020).

In addition to the proteolytic processing of viral polyproteins, viral proteases can cleave host substrates to modulate immune evasion and host gene expression shutoff (Hill et al., 2018; Jagdeo et al., 2018). Although the interactomes of SARS-CoV-2 viral proteins have been well studied (Gordon et al., 2020; May et al., 2022; Stukalov et al., 2021), 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 their subsequent degradation. 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 might contribute to common substrates and reported that ubiquitin-like interferon-stimulated gene 15 protein (ISG15) is cleaved by PL^{pro} (Shin et al., 2020). Systematic screening of interferon stimulatory genes and human innate immune pathway proteins showed that SARS-CoV-2 M^{pro} cleaves the E3 ligase BRE1A (*RNF20* gene) (S. Zhang et al., 2021), NLRP12, TAB1 (Moustaqil et al., 2021), and CARD8 (Tsu et al., 2022), and PL^{pro} also cleaves IRF3 to dysregulate the host innate immune response (Moustaqil et al., 2021). Other researchers examined disrupted cellular apoptosis and autophagy pathways. Wenzel et al. found that M^{pro} cleaves NEMO, an essential modulator of NF-kappa-B signaling in brain endothelial cells (Wenzel et al., 2021), while Mohamud et al. reported that PL^{pro} cleaves the protein kinase ULK1 (Mohamud et al., 2021). 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 PL^{pro} cleaves cardiac myosin proteins (MYH7 and MYH6), FOXP3, HER4, and PROS1 *in vitro* (Reynolds et al., 2021), and Miczi et al. showed that M^{pro} cleaves C-terminal-binding protein 1 (Miczi et al., 2020).

N-terminomics profiling of M^{pro} and PL^{pro} can facilitate the identification of human proteins potentially cleaved during SARS-CoV-2 infection. 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) (Meyer et al., 2021). Refining the cleavage sites to match viral protease specificities, they identified 14 putative M^{pro} and PL^{pro} substrates. Further biochemical analysis confirmed M^{pro} cleavage of pinin, phosphoribosylaminoimidazole carboxylase (*PAICS* gene) and golgin A3 (*GOLGA3* gene) showed that they are cleaved by M^{pro} whereas PL^{pro} 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) (Kleifeld et al., 2010), can be used to identify protease cleavage sites. Using this method, Koudelka et al. identified 318 unique protein substrates of SARS-CoV-2 M^{pro} in lung epithelial carcinoma cells and pulmonary microvascular endothelial cells, but did not validate these substrates in infected cells (Koudelka et al., 2021). Also using TAILS, Pablos et al. profiled 101 M^{pro} substrates in human embryonic kidney cells and lung epithelial cells treated with antiviral type I interferons (Pablos et al., 2021). They further characterized and performed functional studies on several of these M^{pro} substrates such as PTBP1 and the RNA polymerase, RPAP1, which are proteins involved in host transcription and translation. They confirmed the M^{pro} 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 M^{pro} and PL^{pro} in the human proteome (S. Y. Luo et al., 2023). Using subtiligase-mediated N-terminomics in cell lysates, we identified 191 and 16 putative substrates of SARS-CoV-2 M^{pro} and PL^{pro}, 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 in the gaps of uncharacterized M^{pro} and PL^{pro} interactomes that conventional interactomics approaches could not portray. 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. This work may also be more broadly applied to improve our understanding of other viruses.

2.2 Purification of a catalytically active SARS-CoV-2 PL^{pro} and M^{pro}

2.2.1 Purification of PL^{pro} and M^{pro}

PL^{pro} was purified from a GST-tagged PL^{pro} construct (Rut et al., 2020), a gift from Dr. Shaun Olsen. Following protein expression in *E. coli*, GST-PL^{pro} was isolated by affinity chromatography on a glutathione sepharose column (**Fig. 2.1a**). Subsequently, the GST tag was removed utilizing a cleavage site unique to the PreScission protease between the GST tag and PL^{pro}. PL^{pro} was then purified from the unbound fraction following affinity chromatography. Pure fractions of PL^{pro} were successfully collected (**Fig. 2.1b**). Purification and subsequent PL^{pro} experiments discussed were conducted by Eman Moussa.

Similarly, M^{pro} was purified from a recombinant His_6-GST-dual-tagged M^{pro} construct (L. Zhang et al., 2020), a gift from Dr. Rolf Hilgenfeld. The protein was expressed in *E. coli* and subsequently isolated by affinity chromatography on a nickel sepharose column first based on the His tag affinity binding (**Fig. 2.2a**). Again, utilizing the cleavage site unique to the PreScission protease, the His-GST tag was removed. M^{pro} was then purified from the unbound fraction following affinity chromatography on a tandem of glutathione sepharose and cobalt sepharose columns. Pure fractions of M^{pro} were successfully collected (**Fig. 2.2b**). Purification and subsequent M^{pro} experiments discussed were conducted by Shu Luo.

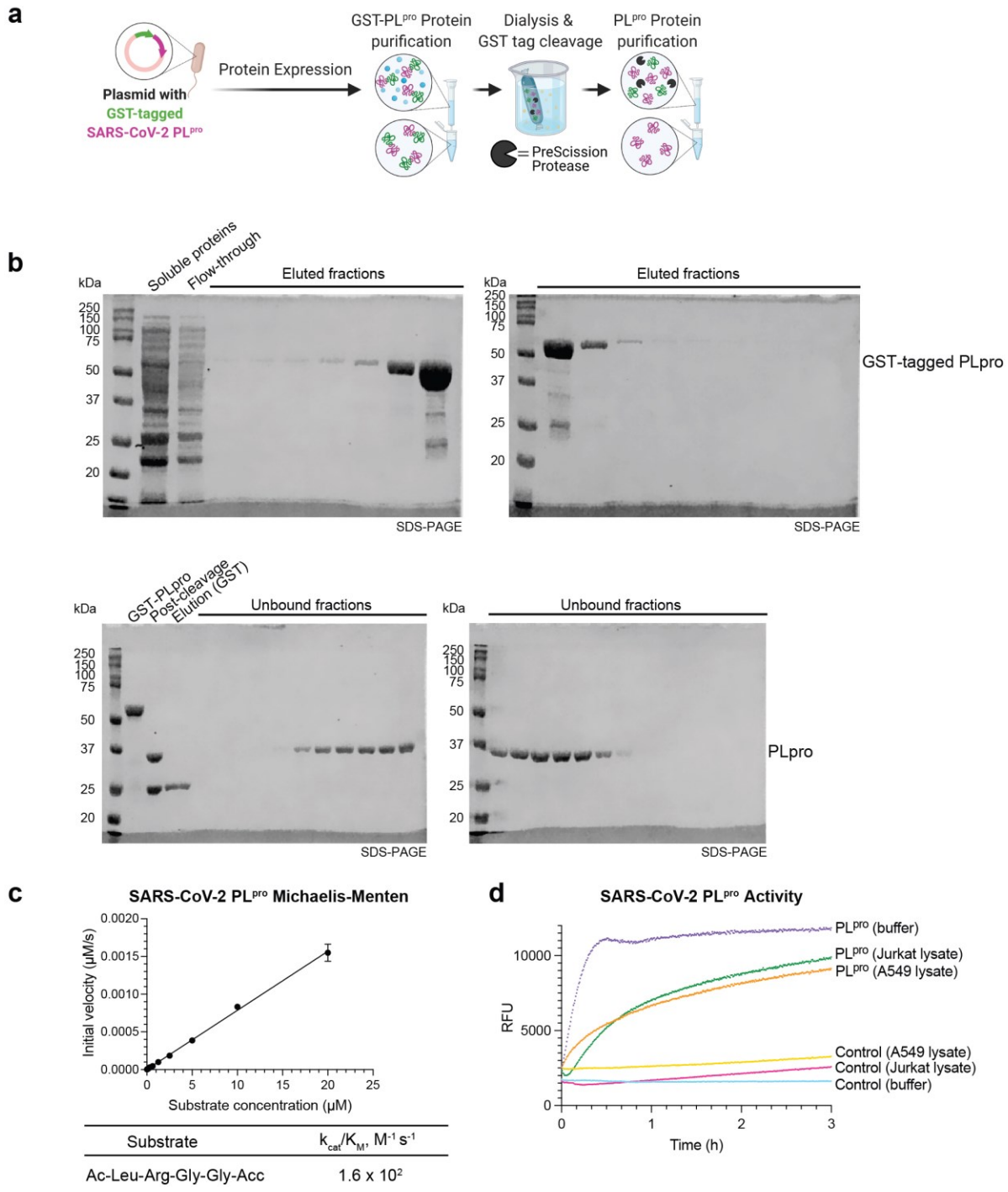


Figure 2.1. Protein expression, purification and activity of PL^{pro}. a) PL^{pro} expression and purification was conducted using a plasmid encoding for the protease domain of Nsp3 with an N-terminal GST tag. GST-PL^{pro} is purified from E. coli lysates using a glutathione sepharose column. The GST tag is removed in an overnight dialysis using a PreScission protease. **b)** Prior to removal of the GST tag, GST-PL^{pro} 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 PL^{pro} 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 PL^{pro}, and 0.78 to 20 µM coumarin substrate in 100 µL total assay volume. The k_{cat}/K_M was calculated using the linear region of the Michaelis-Menten curve k_{cat}/K_M = slope / [E] (Rut et al., 2020). **d)** The activity of 5 µM PL^{pro} was measured using 10 µM of the coumarin substrate in buffer, A549 and Jurkat cell lysates.

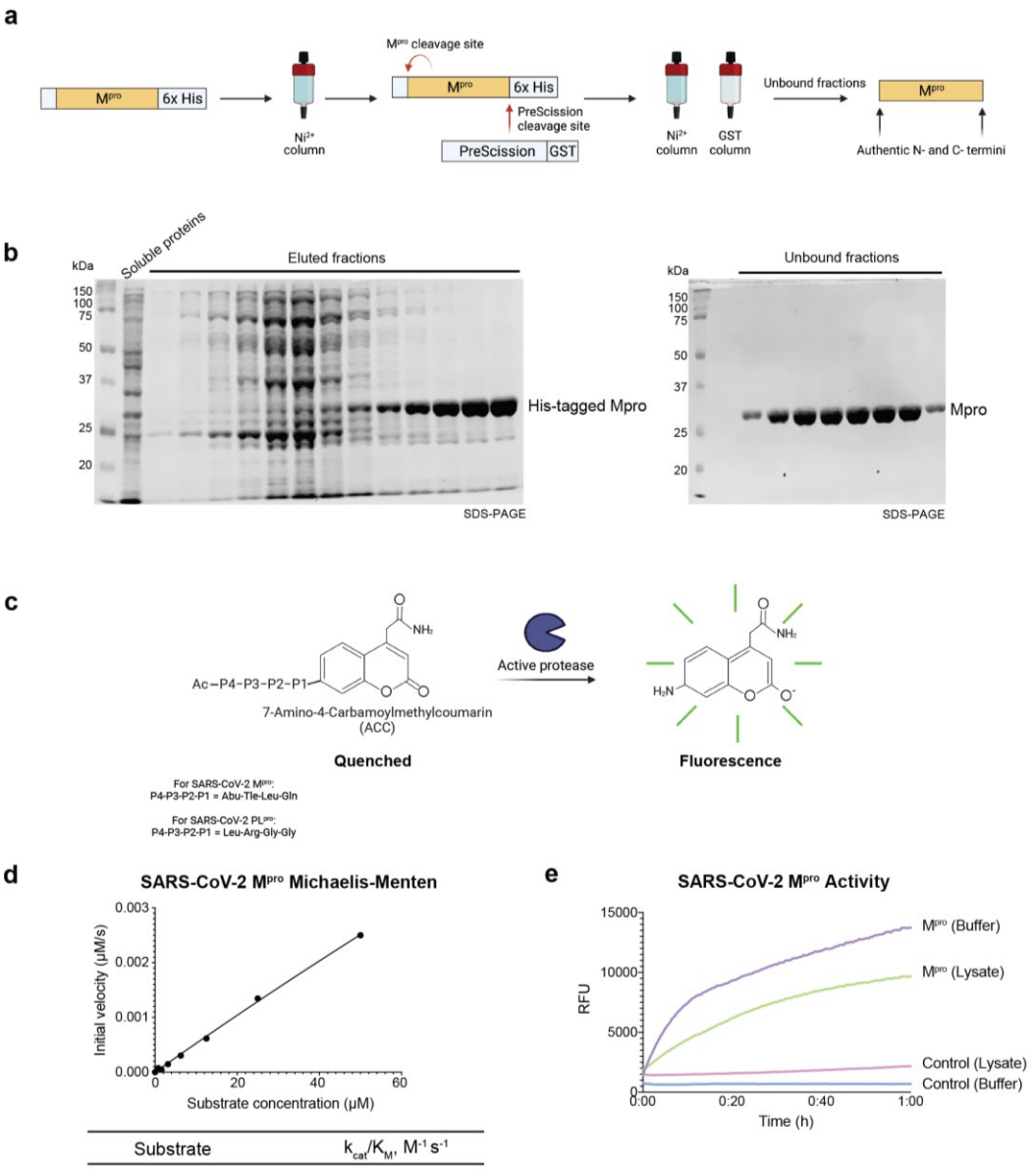


Figure 2.2. Plasmid construct, protein expression and purification of SARS-CoV-2 M^{pro}. a,b) SARS-CoV-2 M^{pro} was expressed and purified by affinity purification with authentic N- and C-termini. **c)** The fluorescence activity assay was carried out using the optimal coumarin substrate Ac-Abu-Tle-Leu-Gln-ACC. **d)** The enzyme kinetics assay was performed using 0.09 μM M^{pro}, and 0.78 to 50 μM coumarin substrate in 100 μL total assay volume. The k_{cat}/K_M was calculated using the linear region of the Michaelis-Menten curve k_{cat}/K_M = slope / [E] (Rut et al., 2020), and is consistent with previously reported value (Rut et al., 2021). **e)** The protease activity was monitored in parallel with the reverse N-terminomics, in cell-free conditions and cell lysates with 0.5 μM M^{pro} and 2 μM coumarin substrate, showing that it was proteolytically active.

2.2.2 The viral proteases are catalytically active

To measure the activity of the purified PL^{pro}, a fluorescence-based probe was used. The fluorescent probe includes the coumarin derivative 7-amino-4-carbamoylmethylcoumarin (ACC), in the P1′ position, quenched with a tetrapeptide of the preferred PL^{pro} cleavage sequence with N-terminal acetylation (Ac-Leu-Arg-Gly-Gly-ACC) (Rut et al., 2020). Upon cleavage between the P1 Gly and P1′ coumarin, the coumarin fluoresces and the fluorescence can be used to monitor the activity of PL^{pro}. The same method is utilized to measure the activity of the purified M^{pro} using the probe with sequence Ac-Abu-Tle-Leu-Gln-ACC (Rut et al., 2021) (**Fig. 2.2c**).

The Michaelis-Menten kinetics of each viral protease was measured in a cell-free environment (buffer) over a range of concentrations of the respective fluorescent probe. The k_{cat}/K_M of PL^{pro} measured was 1.6 $x10^2$ $M^{-1}s^{-1}$ (**Fig. 2.1c**) and approximately 5-fold less than the previously reported value (Rut et al., 2020). Additionally, the k_{cat}/K_M of M^{pro} measured was 5.9 $x10^2$ $M^{-1}s^{-1}$ (**Fig. 2.2d**) and is consistent with the previously reported value (Rut et al., 2021).

The assays were conducted in both a cell-free environment (buffer) and a complex protein-containing environment (cell lysate) for two cell lines, human lung carcinoma cells (A549) and human T lymphocyte cells (Jurkat). Both proteases were able to cleave the fluorescent probe and thus act with comparable activity in both the cell-free environment and complex protein-containing environment with higher activity observed in the former (**Fig. 2.1d, 2.2e**).

2.3 Subtiligase-based N-terminomics identifies novel human substrates of PL^{pro} and M^{pro}

To identify human host substrates of SARS-CoV-2 M^{pro} and PL^{pro}, as well as their corresponding cleavage site(s), we utilized subtiligase-based N-terminomics (**Fig. 2.3**). Cells were lysed and the resulting proteins were treated with the purified active protease (PL^{pro} or M^{pro}). Following cleavage of the host substrates, the newly formed N-termini were labeled with a biotin ester probe using the enzyme subtiligase. This probe features three main components: biotin, the cleavage site sequence recognized by the tobacco etch viral (TEV) protease, and the non-natural amino acid aminobutyric acid (Abu). Once labeled, the N-termini were enriched for using capture on streptavidin beads via the biotin tag. Subsequently, the cleaved protein segments were digested into peptides with trypsin and then released from the beads by TEV protease cleavage, thus removing the biotin tag while retaining the Abu label. The peptide mixture was analyzed by LC-MS/MS. To identify substrates and cleave sites corresponding to PL^{pro} and M^{pro}, the N-terminally

Abu labeled peptides were identified and focus was placed on those with the preferred cleavage site specificity for PL^{pro} and M^{pro}, P2-P1 of GG and P1 of Q/H respectively (Rut et al., 2020, 2021; L. Zhang et al., 2020). A549 and Jurkat lysates were investigated in duplicate.

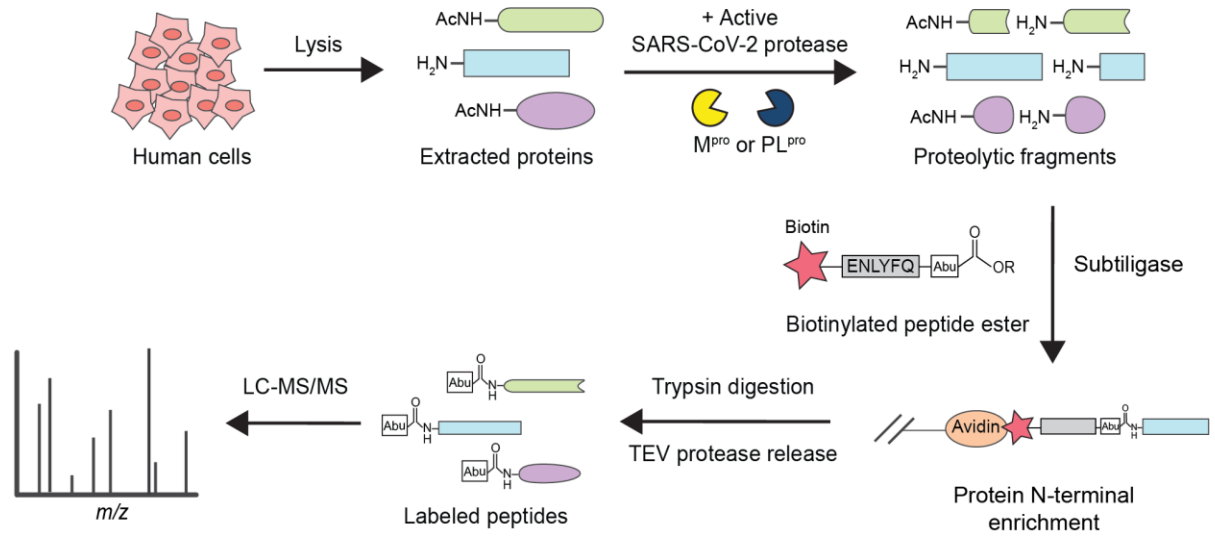


Figure 2.3. Identification of SARS-CoV-2 M^{pro} and PL^{pro} host substrates in *in vitro* subtiligase N-terminomics. Active recombinant M^{pro} or PL^{pro} was added to human cell lysate, generating protein cleavages that were labeled with a designed biotinylated peptide ester by subtiligase. After 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.

Experiments conducted for PL^{pro} in A549 cell lysate identified 3298 cleavage sites corresponding to 1723 proteins (**Fig. 2.4a**). Furthermore, 288 proteins with 380 cleavages corresponded to G↓ and 10 proteins with 11 cleavages corresponded to GG↓. This reflects an 11.5% and 0.33% enrichment of G↓ and GG↓, respectively. Additionally, in Jurkat cell lysate 1105 cleavage sites corresponding to 722 proteins were identified. Of these identifications, 101 proteins with 111 cleavages corresponded to G↓ and 15 proteins with 16 cleavages corresponded to GG↓. This reflects a 10.0% and 1.4% enrichment of G↓ and GG↓, respectively. In comparison, the Jurkat proteome has been shown to contain 4.06% and 0.23% of G↓ and GG↓ sites, respectively,(Crawford et al., 2013) thus reflecting an enrichment for these sequences at the cleavage site by greater than two-fold. Overall, in both A549 and Jurkat cell lysates, 3884 cleavage sites in 2000 proteins were identified and of these cleavage sites 438 cleavage sites and 330 proteins corresponded to G↓ and 22 cleavage sites and 20 proteins corresponded to GG↓ (**Fig. 2.5a**). Therefore, the overall enrichment was 11.3% and 0.57% of G↓ and GG↓, respectively. Of the 22 putative substrates, 5 were identified in both Jurkat and A549 (**Fig. 2.4c**).

M^{pro} was similarly investigated. In A549 cell lysate, a total of 1198 proteins with 2283 cleavage sites were identified, where 139 proteins or 152 cleavage sites corresponded to Q↓ and 57 proteins or 58 cleavage sites to H↓. In Jurkat cell lysate, a total of 600 proteins with 746 cleavage sites were identified, where 131 proteins or 139 cleavage sites corresponded to Q↓ and 15 proteins as well as cleavages sites to H↓. Overall, a total of 1554 proteins and 2870 cleavage sites were identified, with 241 proteins or 266 cleavage sites corresponded to Q↓ and 67 proteins or 68 cleavages sites to H↓ (**Fig. 2.5a**). The overall enrichment was 12% for Q/H↓.

Additionally, as a control, we compared the M^{pro} substrates identified to that of PL^{pro} to identify substrates unique to PL^{pro} or M^{pro}. Utilizing this method, we were able to filter out substrates that have evidence suggesting they could be targeted by multiple proteases, not just exclusively PL^{pro} or M^{pro}. We also hypothesized that while the viral proteases could be targeting previously identified sites for cleavage, a newly emerged virus such as SARS-CoV-2 may have evolved to allow its viral proteases to target and cleave never before observed cleavage sites and substrates. To this end, we utilized a database called the DegraBase (Crawford et al., 2013) containing previously observed cleavage sites in healthy and apoptotic cells from subtiligase-based N-terminomics experiments to isolate newly observed cleavage sites. This method also allows for the removal of potential background proteolysis due to incomplete protease inhibition in the lysate. Subsequently, of the 1238 total unique substrates for PL^{pro}, 200 corresponded to G↓ and 16 to

GG↓ (**Fig. 2.5a, Table. 2.1)**. For M^{pro}, of the total 1007 unique substrates, 148 corresponded to Q↓ and 46 to H↓.

IceLogos (Colaert et al., 2009) of the substrate cleavage site sequences for PL^{pro} demonstrated a preference for Leu in the P4 position when Gly was present in the P1 and P2 positions (**Fig. 2.4b, 2.5b**). Similarly, for M^{pro} a preference for Leu in the P2 position was observed when Gln or His were present in the P1 position. These observations reflect the preferred cleavage site sequence for PL^{pro} and M^{pro}, respectively (Rut et al., 2020, 2021; L. Zhang et al., 2020). Additionally, of the 16 putative cleavage sites for PL^{pro} 11 featured the LXGG preferred sequence, while in the M^{pro} dataset only one LXGG cleavage site was observed.

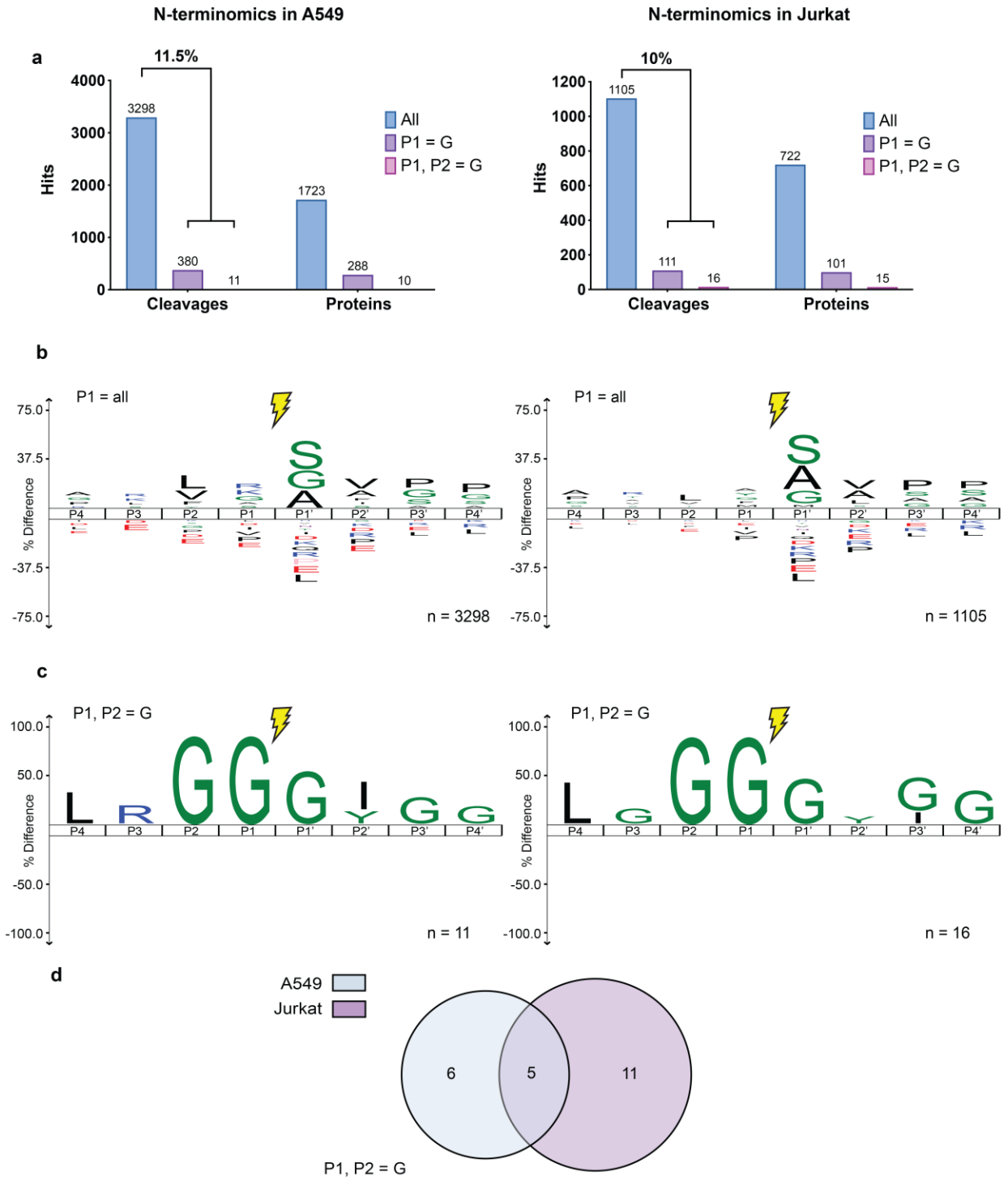


Figure 2.4. Identification of SARS-CoV-2 PLpro substrates. a) N-terminomics statistics of two A549 (left) and two Jurkat (right) replicates for PL^{pro}. In A549, 3298 unique labeled cleavages

were identified with 380 sites at P1=G and 11 sites at P1,P2=G in 288 and 10 proteins, respectively, showing an enhanced enrichment for P1=G at 11.5%. In Jurkat, 1105 unique labeled cleavages were identified with 111 sites at P1=G and 16 sites at P1,P2=G in 101 and 15 proteins, respectively, showing an enhanced enrichment for P1=G at 10%. **b)** IceLogo showing P4-P4′ residue enrichment in all labeled cleavage sites in A549 (left) and Jurkat (right), and **c)** in sites where P1,P2=G only. **d)** Venn diagram depicting the overlap in cleavage sites identified in A549 and Jurkat with P1,P2=G.

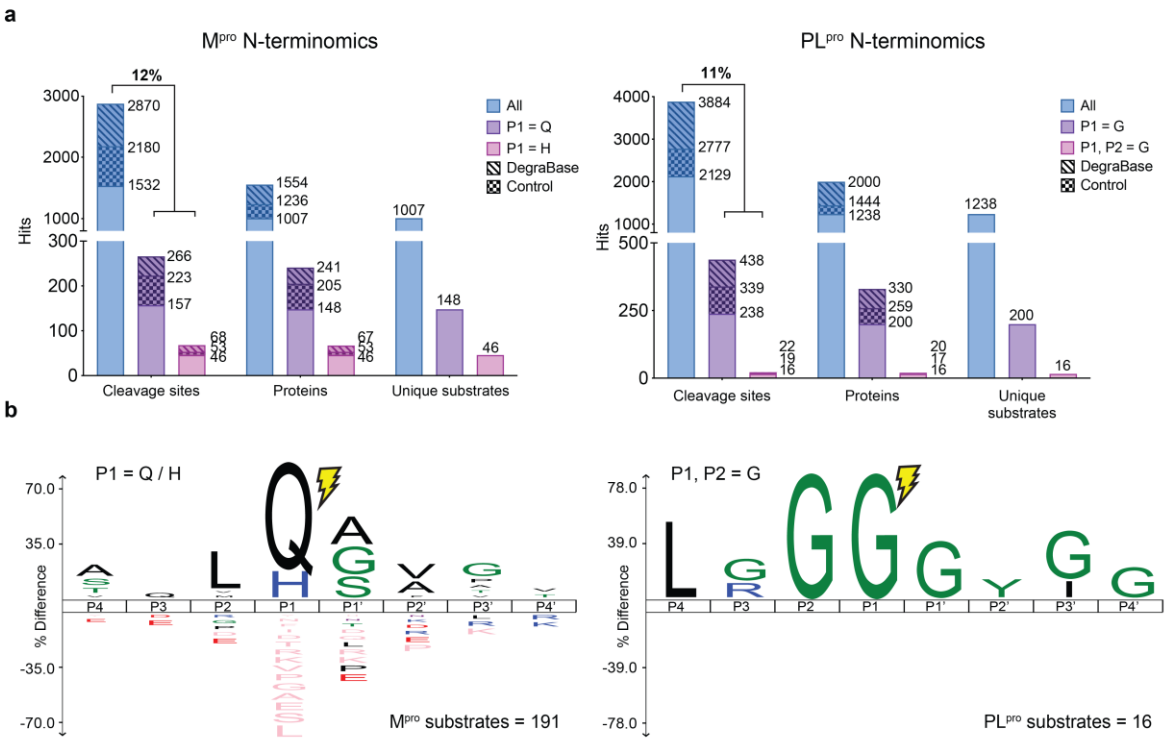


Figure 2.5. Identification of SARS-CoV-2 M^{pro} and PL^{pro} substrates using subtiligase N-terminal labeling. a) N-terminomics statistics of two Jurkat and two A549 replicates for each of M^{pro} (left) and PL^{pro} (right). For M^{pro}, 2870 unique labeled cleavages were identified with 334 sites at P1=Q/H in 308 proteins, showing an enhanced enrichment at 12%. For PL^{pro}, 3884 unique labeled cleavages were identified with 436 sites at P1=G and 22 sites at P1,P2=G in 330 and 20 proteins, respectively, show enrichment for P1=G at 11%. For each viral protease, the number of M^{pro} and PL^{pro} substrates unique to each protease is also reported. **b)** IceLogo showing P4-P4′ residue enrichment for M^{pro} with P1=Q/H (left) and for PL^{pro} with P1,P2=G (right).

Table 2.1. Selected putative substrates of SARS-CoV-2 PL^{pro}.

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
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

2.4 Further Investigation of PL^{pro} and M^{pro} host substrates

2.4.1 The splicing factor proline- and glutamine-rich is cleaved during infection

Of the 16 putative PL^{pro} substrates, we further investigated the splicing factor proline- and glutamine-rich (SFPQ) which was identified in both A549 and Jurkat cell lysate N-terminomics with the common cleavage site $LGGG_{637}\downarrow GGIG$ (**Fig. 2.6a**). SFPQ is a DNA- and RNA-binding protein. Previous investigation of SFPQ in influenza A infection identified that downregulation of SFPQ results in reduced viral replication, indicating a vital role for SFPQ in influenza A viral transcription (Landeras-Bueno et al., 2011). Additionally, SFPQ has also been shown to undergo proteolysis in human rhinovirus infection as part of a mechanism to promote viral replication (Flather et al., 2018), and is exploited by both the encephalomyocarditis virus (B. Zhou et al., 2019) and hepatitis delta virus (Greco-Stewart et al., 2006) during infection. Overall, this indicates a trend of SFPQ exploitation by several RNA viruses during infection.

A549 cells expressing the viral entry receptor (A549-ACE2) were infected with SARS-CoV-2 (**Fig. 2.6b**). At 24 and 48 hours post-infection, a band corresponding to the expected 69 kDa SFPQ cleavage product was observed only in the infected lysates in addition to the 76 kDa full-length SFPQ band observed in both infection and mock (control). Subsequently, we incubated Jurkat and A549-ACE2 lysates with PL^{pro} *in vitro* and observed full-length SFPQ but no cleavage product at the expected 69kDa molecular weight (**Fig. 2.6c, d)**. Similarly, no cleavage product was observed in the case of HEK293T-ACE2 cells with overexpression FLAG-SFPQ (**Fig. 2.6e**) nor was there any endogenous SFPQ cleavage with the expression of full-length NSP3 24 hours post-transfection (**Fig. 2.6g)**. However, following immunoprecipitation of FLAG-SFPQ, a cleavage product was observed with PL^{pro} incubation, although corresponding to a higher molecular weight than previously expected (**Fig. 2.6f**). In summary, while the infection study was consistent with our *in vitro* N-terminomics results, the subsequent PL^{pro} *in vitro* studies were not as consistent, indicating that while SFPQ is cleaved during SARS-CoV-2 infection it may undergo proteolysis by PL^{pro} and with additional processing by other proteases.

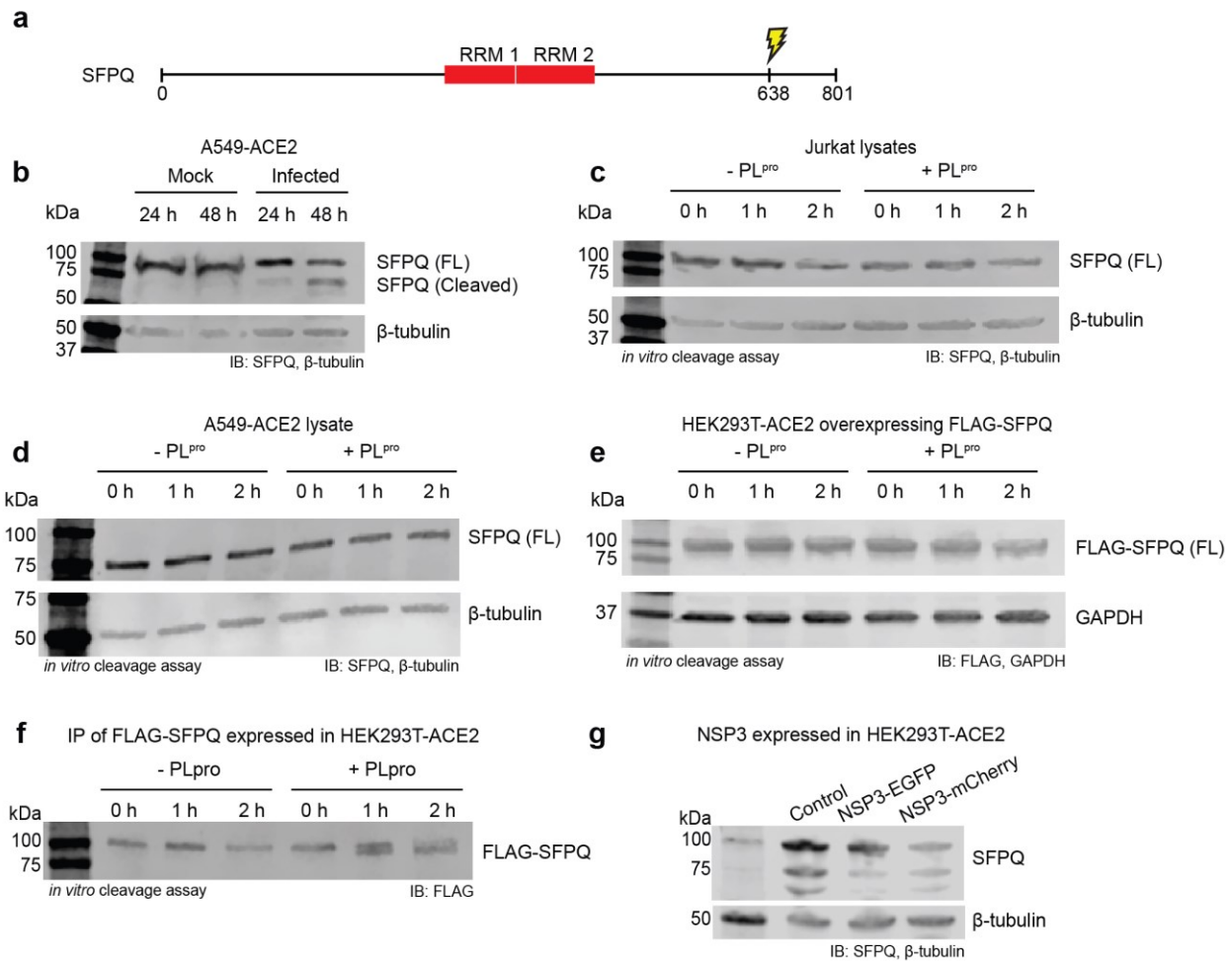


Figure 2.6. Proteolysis of SFPQ in SARS-CoV-2 infected cells. a) PL^{pro} cleavage of SFPQ at G637, C-terminal to the RNA recognition motifs. **b)** SFPQ was cleaved in A549-ACE2 cells infected with SARS-CoV-2. **c)** SFPQ cleavage by PL^{pro} could not be detected using immunoblotting in Jurkat cell lysates. **d)** Uninfected A549-ACE2 cell lysates were incubated with PL^{pro} and SFPQ cleavage by PL^{pro} could not be detected using immunoblotting. **e)** Cleavage of overexpressed FLAG-tagged SFPQ in HEK293T-ACE2 cells by PL^{pro} was also not detected on immunoblot. **f)** A potential cleavage product was observed when incubating immunoprecipitated FLAG-SFPQ with SARS-CoV-2 PL^{pro}. **g)** Expression of full-length NSP3 in HEK293T-ACE2 did not show distinct cleavage of SFPQ compared to the control.

2.4.2 Accounting for the role of PL^{pro} as a deubiquitinase

As PL^{pro} is a deubiquitinase, we hypothesized that inhibition of protein ubiquitination may allow for increased identification of PL^{pro} substrates. To pursue this avenue of investigation, we utilized an inhibitor of the ubiquitin-activating enzyme (UAE), called TAK-243 (Hyer et al., 2018). This inhibitor binds to ubiquitin and forms a TAK-243-ubiquitin adduct that inhibits the conjugation of ubiquitin to proteins.

Jurkat cells were treated with 1 μM of TAK-243 for 4 hours and imaged to monitor for cell death in media and in trypan blue to visualize dead cells (**Fig. 2.7a**). Overall, cell death with TAK-243 treatment was limited and comparable to the controls. Subsequently, it was shown by immunoblot that TAK-243 treatment successfully inhibited ubiquitination (**Fig. 2.7b**). In this instance, an anti-ubiquitin antibody that binds only conjugated ubiquitin was utilized to observe ubiquitin conjugated to proteins.

Subtiligase-based N-terminomics was conducted with the TAK-243-treated Jurkat cells. Jurkat cells not treated with PL^{pro}, both DMSO- and TAK-243-treated, were utilized as controls. Overall, 285 cleavage sites corresponding to 241 proteins were identified (**Fig. 2.8a**). Of these 241 proteins, 29 proteins/cleavages corresponded to G↓ and 3 proteins/cleavages corresponded to GG↓. This reflects a 10.2% and 1.1% enrichment of G↓ and GG↓, respectively. Additionally, cleavage sites previously identified in the Degrabase and M^{pro} datasets were utilized as additional controls as previously described. Therefore, the final statistics reflect 115 cleavage sites corresponding to 105 proteins identified with 14 proteins/cleavages corresponding to G↓, an enrichment in P1=G of 12.2%. None of the 3 GG↓ cleavage sites were identified in the Degrabase or M^{pro} datasets. The same methodology was applied for both the DMSO- and TAK-243-treated groups. Of the 3 cleavage sites identified, both protein CASC3 and SFPQ were previously identified in both the A549 and Jurkat PL^{pro} datasets while Rap guanine nucleotide exchange factor 1 (RAPGEF1) was not previously identified in any datasets including the M^{pro} dataset. Additionally, SFPQ was identified in both the DMSO- and TAK-243-treated groups, but not in the control. The consistent identification of protein CASC3 and SFPQ only in the PL^{pro}-treated cell lysates is a strong indication that these 2 putative substrates are more likely to be real substrates of PL^{pro}. Corresponding IceLogos, as well as additional comparisons of PL^{pro}-treated samples and controls, were conducted (**Fig. 2.8b-e**). Again, we observed the prevalence of Leu in the P4 position when Gly was present in the P1 or P1 and P2 positions (**Fig. 2.8d, e**). Comparison of the putative substrates identified in Jurkat, A549, and TAK-243-treated Jurkat cells reveals an overlap

of two putative substrates (SFPQ and protein CASC3) amongst all groups, as well as several putative substrates unique to each group (**Fig. 2.8f**). Cell lysates of TAK-243-treated Jurkat cells were incubated with PL^{pro}, but once again no SFPQ cleavage product was observed by immunoblot (**Fig. 2.8g**).

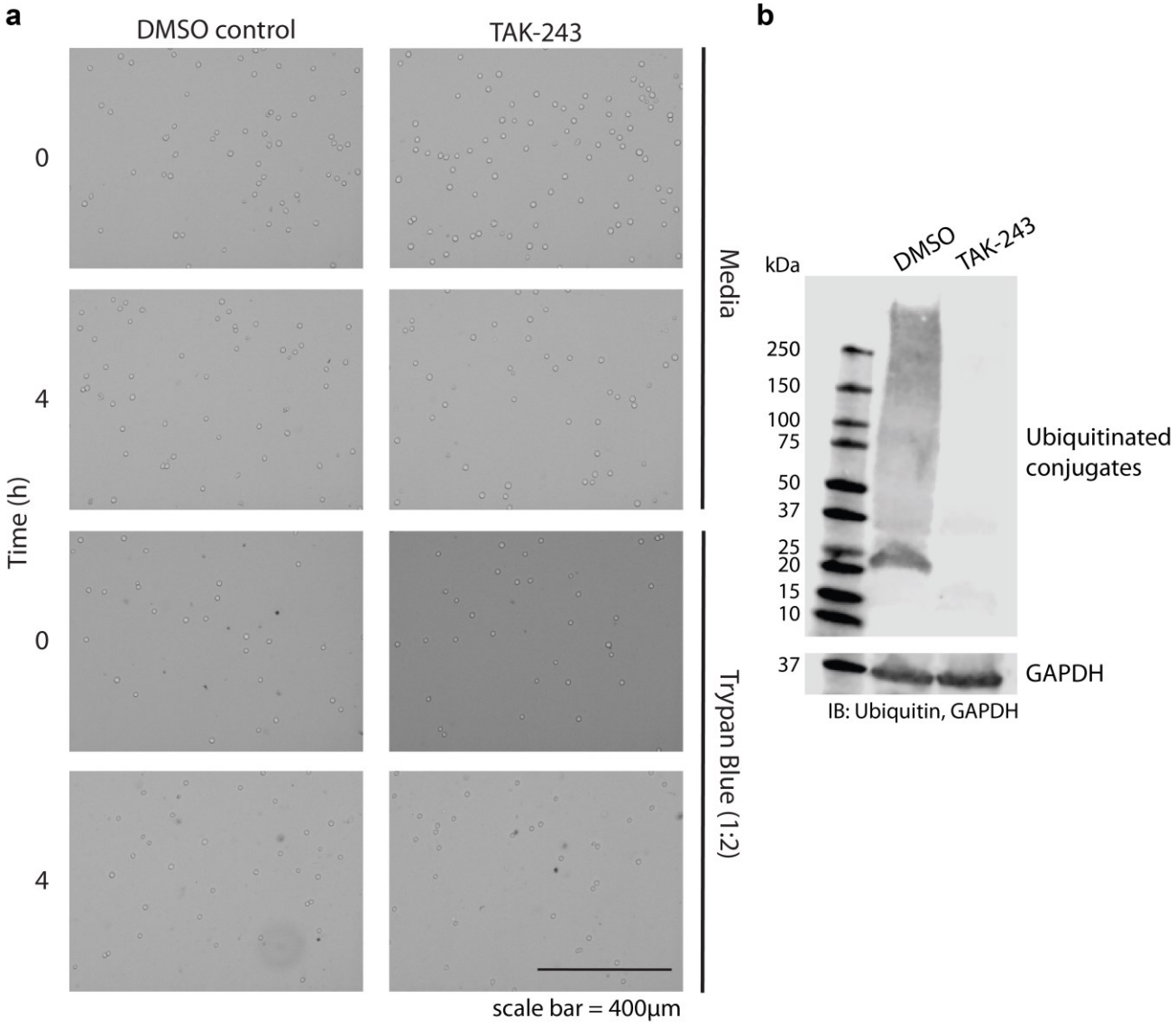


Figure 2.7. Inhibition of protein ubiquitination for further investigation of PL^{pro} substrates. Jurkat cells were treated with 1 µM TAK-243, an inhibitor of UAE, or DMSO as a control for 4 hours. **a)** Live cells were monitored at 0 and 4 hours in cell media as well as in trypan blue (1:2 dilution) to monitor for cell death. **b)** Jurkat cell lysate was monitored by immunoblot for protein ubiquitination utilizing an antibody that detects mono- and poly-ubiquitinated conjugates. TAK-243 treatment showed inhibition of protein ubiquitination compared to the DMSO control.

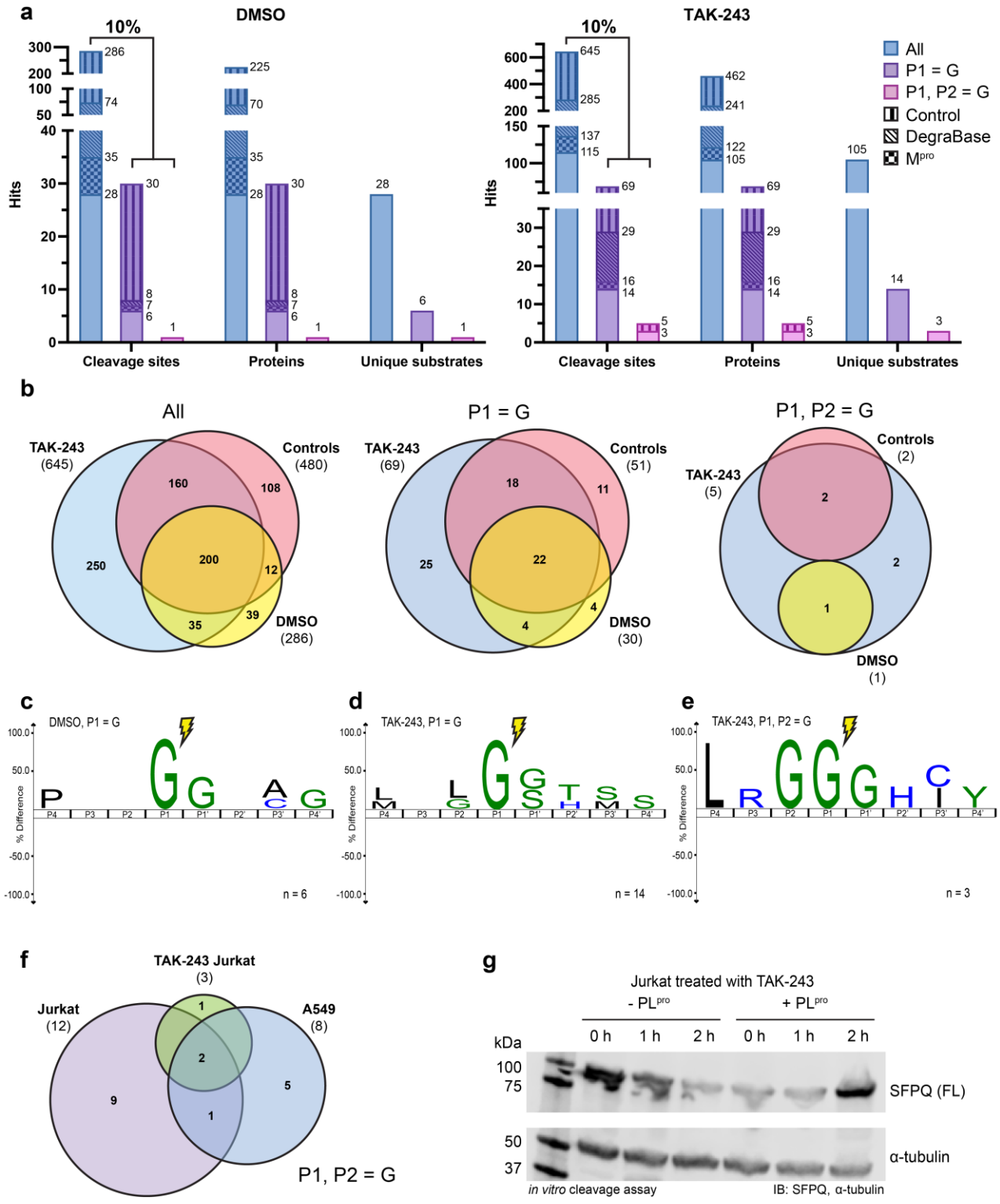

a
DMSO
TAK-243
10%
Hits
All
P1 = G
P1, P2 = G
Control
DegraBase
Mpro
Cleavage sites
Proteins
Unique substrates
b
All
P1 = G
P1, P2 = G
TAK-243
(645)
Controls
(480)
DMSO
(286)
TAK-243
(69)
Controls
(51)
DMSO
(30)
TAK-243
(5)
Controls
(2)
DMSO
(1)
c
DMSO, P1 = G
n = 6
d
TAK-243, P1 = G
n = 14
e
TAK-243, P1, P2 = G
n = 3
% Difference
f
TAK-243 Jurkat
(3)
Jurkat
(12)
A549
(8)
P1, P2 = G
g
Jurkat treated with TAK-243
- PLpro
+ PLpro
0 h
1 h
2 h
kDa
SFPQ (FL)
α-tubulin
in vitro cleavage assay
IB: SFPQ, α-tubulin

Figure 2.8. Identification of SARS-CoV-2 PL^{pro} substrates in the absence of ubiquitination using subtiligase-based N-terminal labeling. a) N-terminomics statistics of two DMSO (left) and two TAK-243 treated (right) Jurkat replicates for PL^{pro}. For DMSO-treated, 286 unique labeled cleavages were identified with 30 sites at P1=G and 1 site at P1,P2=G in 30 and 1 proteins, respectively. This shows enrichment for P1=G at 10.4%. For TAK-243 treated, 645 unique labeled cleavages were identified with 69 sites at P1=G and 5 sites at P1,P2=G in 69 and 5 proteins, respectively. This shows enrichment for P1=G at 10.6%. The number of unique PL^{pro} substrates is also reported. **b)** Venn diagrams displaying the overlap in cleavage sites identified between PL^{pro}-treated DMSO and TAK-243 groups and untreated groups. **C)** IceLogo showing P4-P4′ residue enrichment for unique PL^{pro} substrates in DMSO treated P1=G, as well as **d)** TAK-243 treated P1=G and **e)** P1,P2=G. **f)** The common and unique PL^{pro} P1,P2=G cleavage sites identified in Jurkat, A549, and TAK-243 treated Jurkat groups are shown with a Venn diagram. **g)** No cleavage product for SFPQ was identified by immunoblot of PL^{pro} incubated with lysate from Jurkat cells treated with TAK-243.

2.4.3 SARS-CoV-2 M^{pro} cleaves BRD2

Of the 194 putative M^{pro} substrates, we further investigated the bromodomain-containing protein 2 (BRD2) which was identified in our N-terminomics experiments in Jurkat lysate with the cleavage site $AALQ_{206}\downarrow GSVT$. BRD2 has been shown to play both a pro-viral and anti-viral role in SARS-CoV-2 infection. BRD2 is essential for the transcription of the angiotensin-converting enzyme-2 (ACE2) receptor in human lung epithelial cells (Samelson et al., 2022) which binds the SARS-CoV-2 spike protein for viral entry. BRD2 itself also binds the envelope E protein of the virus itself (Gordon et al., 2020). However, in an anti-viral capacity, the transcription of genes involved in type I interferon response is also promoted by BRD2 (Samelson et al., 2022). The loss of BRD2, BRD3, and BRD4 in A549-ACE2 was also identified to enhance SARS-CoV-2 viral replication and infection (I. P. Chen et al., 2022).

Similarly to SFPQ, we investigated BRD2 cleavage after SARS-CoV-2 viral infection in A549-ACE2 (**Fig. 2.9d**) and HEK293T-ACE2 cells (**Fig. 2.9e**). In both instances, at 24 hours and more so at 48 hours post-infection, full-length BRD2 depletion was observed. In addition to the depletion of full-length BRD2, we also observed the expected 23 kDa N-terminal cleavage product of BRD2 in Jurkat lysate incubated with recombinant M^{pro} (**Fig. 2.9b**). As an additional validation, we mutated the Gln at the P1 position of the cleavage site of BRD2 to produce the Q206A mutant GFP-tagged BRD2. Comparison of HEK293T lysates overexpressing wildtype GFP-BRD2 and mutant GFP-BRD2 Q206A showed the expected 50 kDa cleavage product only in the case of wildtype GFP-BRD2 with M^{pro} (**Fig. 2.9c**). Therefore, Q206 is necessary for the cleavage of BRD2 by M^{pro}, validating the cleavage site ($AALQ_{206}\downarrow GSVT$) identified in our N-terminomics experiments. Overall, BRD2 is cleaved by M^{pro} *in vitro* and during SARS-CoV-2 infection.

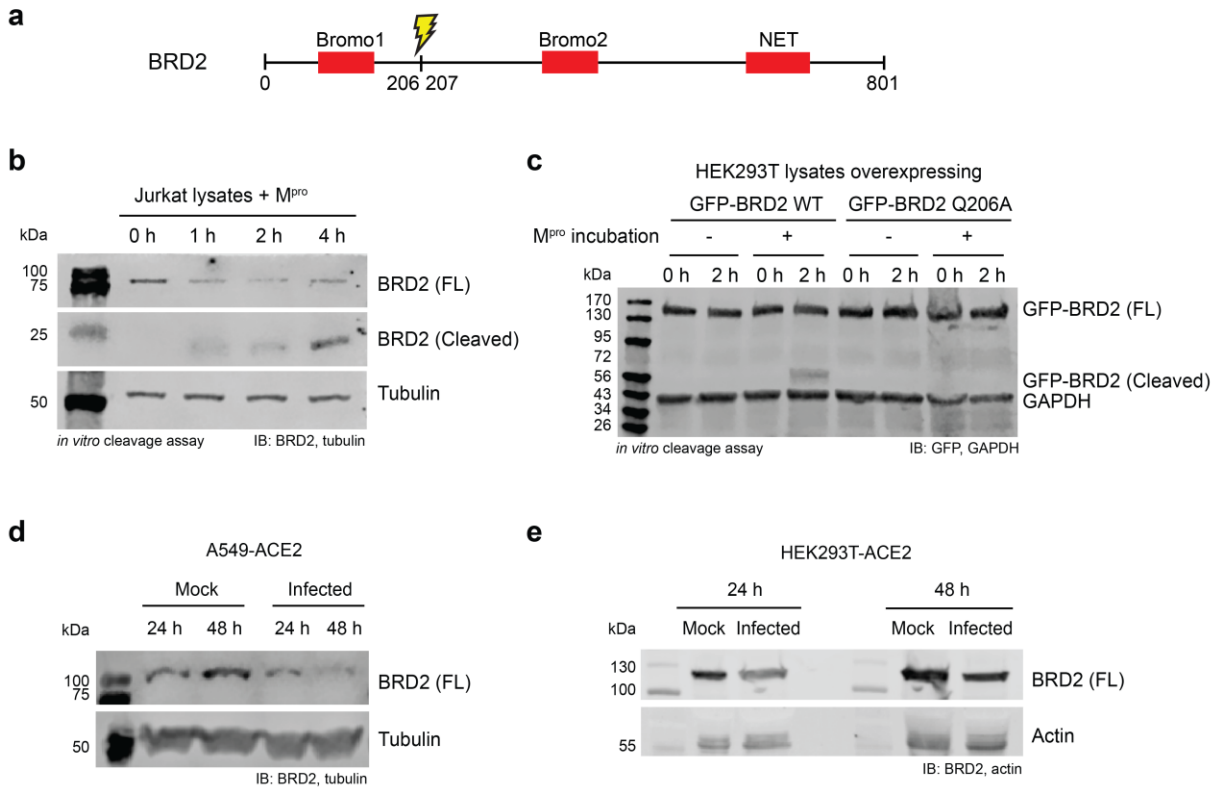


Figure 2.9. Proteolysis of BRD2 by M^{pro} *in vitro* and in SARS-CoV-2 infected cells. a) BRD2 and BRD4 share similar features of Bromo 1, Bromo 2, and N-terminal extra terminal (NET) domains. Cleavage events were observed at Q206 in BRD2, cleaving off Bromo 1, but at Q1077 in BRD4 releasing the C-terminal (CTD) region. **b)** 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 appeared with incubation time as the full-length BRD2 level decreased. **c)** 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 **d)** A549-ACE2 and **e)** HEK293T-ACE2 at 24 and 48 h.p.i.

2.5 The SARS-CoV-2 main protease cleaves galectin-8

Utilization of subtiligase-based N-terminomics and LC-MS/MS allowed for the identification of novel substrates of the SARS-CoV-2 viral proteases M^{pro} and PL^{pro}. In addition to this line of inquiry, we also investigated the previously reported M^{pro} substrate, galectin-8 (Gal8).

Pablos et al. identified the cleavage of galectin-8 (Gal8) by SARS-CoV-2 M^{pro} at LQ_{158} and proposed its role as a mechanism of viral immune response evasion (Pablos et al., 2021). Upon viral entry, Gal8 binds the spike protein as a means to activate antiviral-xenophagy. Cleavage of Gal8 by M^{pro} was demonstrated to disrupt binding to the spike protein and subsequent activation of antiviral-xenophagy, indicating a viral mechanism for evasion of the host immune response.

In addition to the cleavage event previously identified at LQ_{158}, Gal8 may be cleaved by M^{pro} at multiple sites to enhance the evasion of the host immune response. To investigate, we monitored the cleavage of Gal8 with SARS-CoV-2 M^{pro} over time by sodium dodecyl-sulfate polyacrylamide (SDS-PAGE) gel electrophoresis and LC-MS/MS.

Gal8 was incubated with SARS-CoV-2 M^{pro} for 2 hours. The assay was run on SDS-PAGE, where five cleavage products were observed (**Fig. 2.10a**) and prepared by in-gel trypsin digestion for subsequent LC-MS/MS. Three cleavage sites were identified for Gal8 at LQ_{9}, LQ_{158}, and LQ_{246} with corresponding MS/MS spectra (**Fig. 2.10b**).

2.5.1 Further investigation of galectin-8 cleavage by targeted mass spectrometry

Following the initial investigation of Gal8 cleavage, we further investigated the cleavage events associated with individual cleavage products using a targeted mass spectrometry technique called parallel reaction monitoring (PRM).

PRM is a method of targeted mass spectrometry that relies on an initial discovery, conducted with data-dependent acquisition (DDA) otherwise referred to as shotgun proteomics, where data is acquired for the precursors identified in a peptide mixture. Based on the initial discovery by DDA, the m/z, charge, and retention time identified can be used to select several precursors of interest from the larger initial list and to monitor the intensity of these precursors and their respective fragment ions. The discovery is also compared to the PRM chromatogram to ensure that the chromatogram observed at the expected retention time correctly corresponds to the chosen precursor. The parameters used to compare the discovery and PRM typically include the intensity of the fragment ions with respect to each other called the ion rank, the retention time, and the

distribution of the fragment ions shown as a dot product (dotp) from 0 to 1 where 1 is the most similar to the discovery.

While the initial discovery by DDA detects the precursor and its fragment ions at an initial time point, the targeted PRM allows for continuous monitoring of the selected precursor over a chosen retention time window. Continuous monitoring is necessary to obtain a whole chromatogram profile for each precursor which can subsequently be used for better quantification by chromatogram peak area.

Each Gal8 cleavage product band was analyzed individually and labeled 1 to 5 from highest to lowest molecular weight. The cleavage at LQ_{158} and LQ_{246} was successfully detected and monitored by quantification of the peptide C-terminal to the cleavage site (**Fig. 2.11a**). Quantification shows that bands 3, 4, and 5, corresponding to molecular weights of approximately 20, 18, and 17 kDa respectively, had the highest quantification. Bands 4 and 5 correspond more closely to the expected molecular weights of the N and C-terminal cleavage products resulting from cleavage at LQ_{158}, at 17.9 and 18 kDa respectively. Additionally, quantification of cleavage at LQ_{246} showed only bands 1 (approximately 34 kDa) and 4 (approximately 18 kDa) displayed similarity to the discovery with reasonable quantification. The cleavage product bands 4 and 5, at approximately 18 and 17 kDa respectively, were further investigated following 0.5, 1, 2, and 24-hour incubation of Gal8 with M^{pro} (**Fig. 2.11b**). Again, LQ_{158} quantification was consistent with the discovery, based on the high dotp, and showed an increase in cleavage product over the time course with a final quantification of just under 10^{10}. In contrast, LQ_{246} quantification showed a poorer comparison to the discovery with dotp as low as 0.55. At 24 hours, the final quantification was on the order of 10^6, four orders of magnitude less than that of LQ_{158}. Additionally, chromatogram profiles for each of LQ_{158} and LQ_{246} were compared to the initial discovery chromatograms based on retention time consistency and ion rank (**Fig. 2.11c-d**). While LQ_{158} is comparable to the discovery at all four of the time points, LQ_{246} is comparable at the 1, 2, and 24-hour time points with four to five of the top six fragment ions identified.

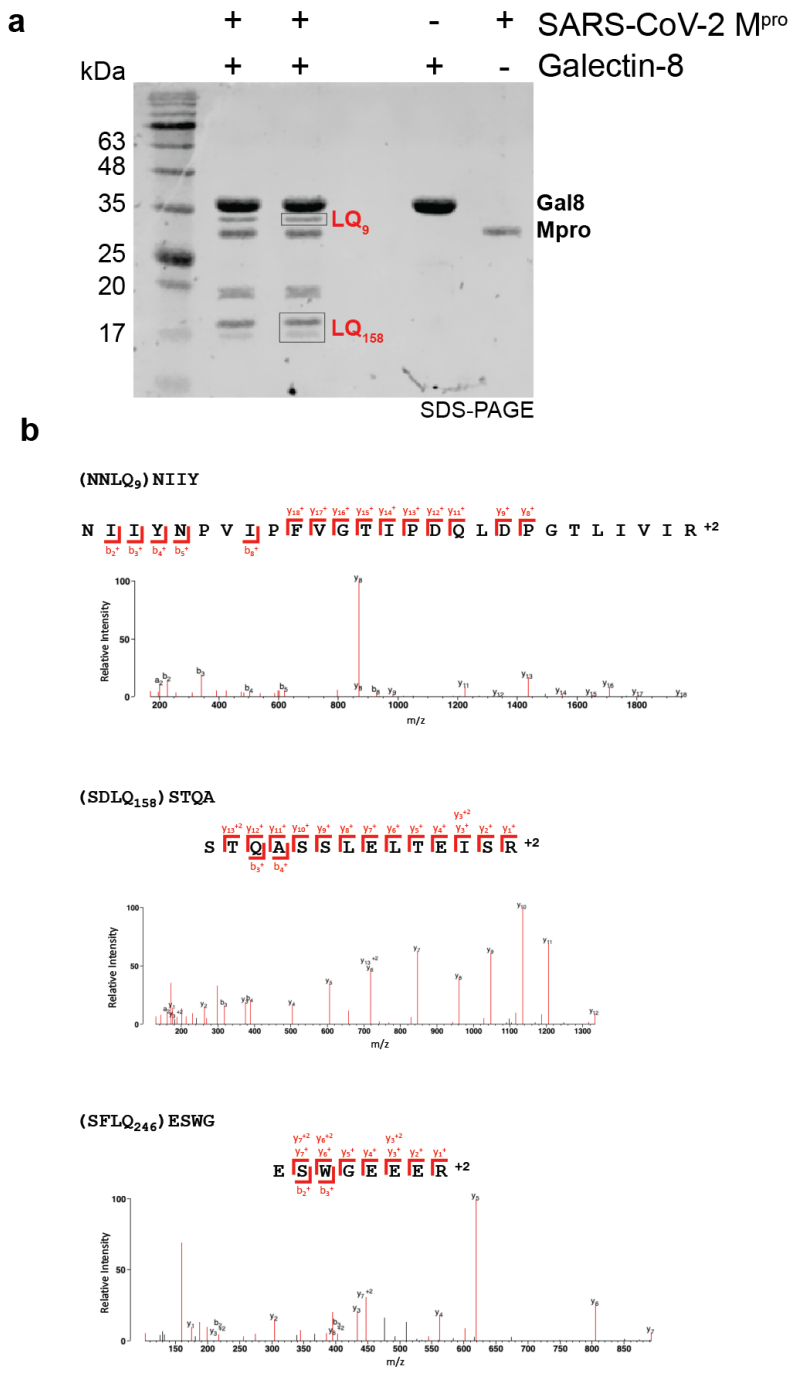


Figure 2.10. SARS-CoV-2 M^{pro} cleaves Gal8 primarily at LQ$_{158}$ as observed by mass spectrometry. a) Incubation of Gal8 with SARS-CoV-2 M^{pro} for 2 hours resulted in five identified Gal8 cleavage protein fragments on SDS-PAGE. Each band was analyzed individually using LC-MS/MS. **b)** In total, we identified three M^{pro} cleavage sites present: LQ$_9$, LQ$_{158}$, and LQ$_{246}$. The representative corresponding MS/MS spectra are shown for the identified peptides.

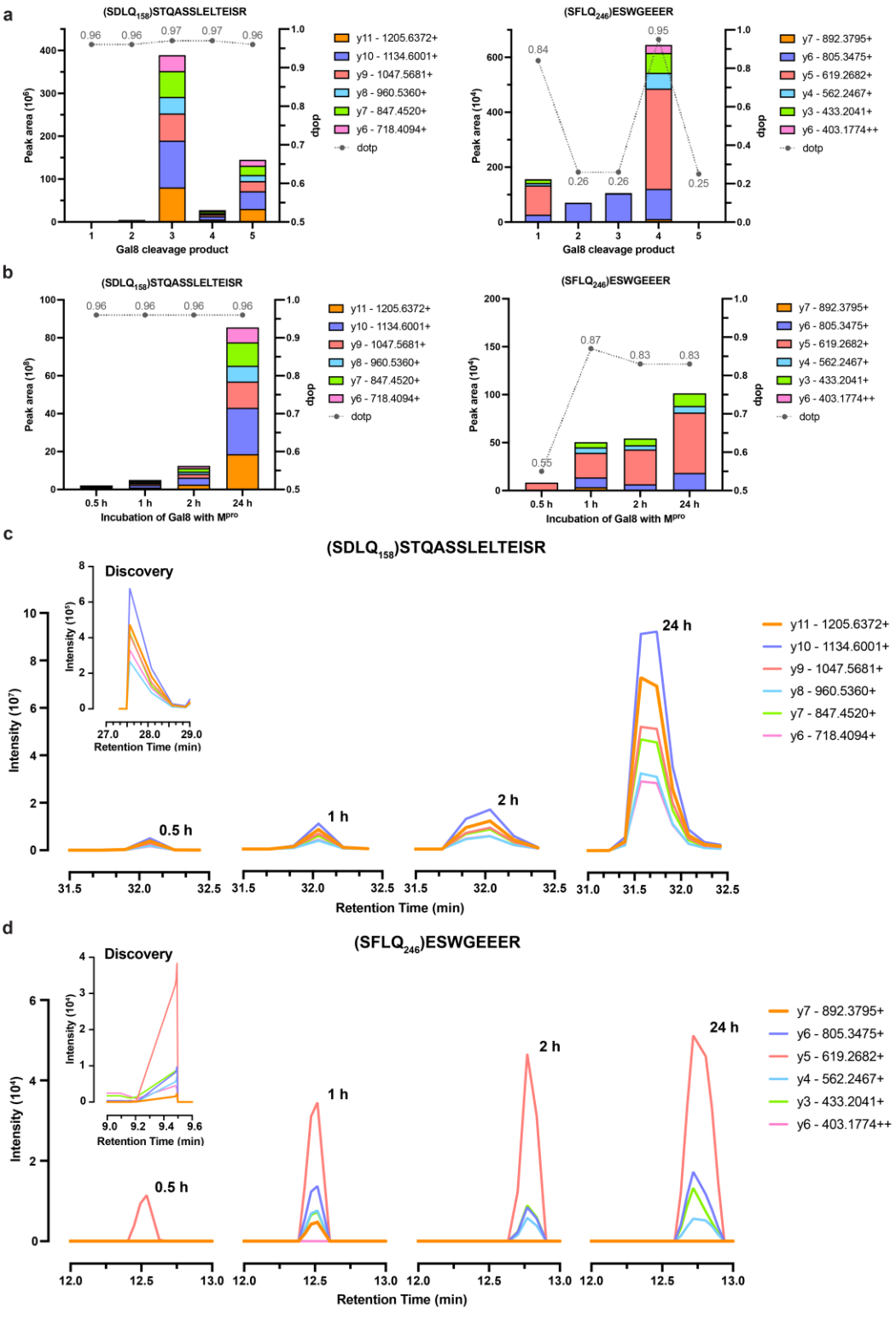


Figure 2.11. Quantification of Gal8 cleavage by targeted mass spectrometry. a) PRM was utilized to quantify the cleavage of Gal8 at LQ_{158} and LQ_{246}, where each of the five cleavage

product bands (see **Fig. 2.10**) were analyzed independently. The cleavage product bands were labeled 1 to 5 from highest to lowest molecular weight. The corresponding quantification by peak area for each cleavage product is shown. **b)** Subsequently, the two cleavage products of the lowest molecular weight (bands 4, and 5 from panel a) were quantified for cleavage of Gal8 at LQ_{158} and LQ_{246} with 0.5- to 24-hour incubations with M^{pro}. This was done by monitoring the resultant peptide C-terminal to the cleavage site. Quantification of the chromatogram peak area and the dotp that demonstrate the similarity in fragment ion distribution compared to the reference discovery are shown. Corresponding chromatograms for **c)** LQ_{158} and **d)** LQ_{246} are shown with the initial discovery.

2.6 Conclusions and Discussion

2.6.1 The potential role of SFPQ cleavage in viral infection.

The observation of SFPQ cleavage during SARS-CoV-2 viral infection (**Fig. 2.6b**) provides evidence that SFPQ may be cleaved during viral infection by PL^{pro}. Additionally, the consistent identification of SFPQ cleavage at $LGGG_{637}\downarrow GGIG$ by N-terminomics and mass spectrometry in only the Jurkat, A549, and TAK-243 treated Jurkat cell lysates treated with PL^{pro}, but not controls, suggests that SFPQ is cleaved by PL^{pro}. It remains difficult to detect SFPQ cleavage in the presence of PL^{pro} by immunoblot. Only immunoprecipitated FLAG-SFPQ cleavage was observed (**Fig. 2.6f**) with a different molecular weight than that observed with viral infection (**Fig. 2.6b**).

The difficulty in utilizing immunoblotting to validate mass spectrometry results is not a new challenge nor is the demand for such validation (Mehta et al., 2022). Immunoblotting is subject to several limitations, the first being the amount of starting material required on the order of micrograms. In contrast, mass spectrometry methods can detect peptides from nanograms worth of total peptide mixtures containing 4000 to 6000 peptides, if not more, with individual peptides then on the order of picograms and femtograms. The sensitivity of immunoblotting also relies on the quality of the antibody chosen, as not all antibodies are made equal. Monoclonal antibodies tend to yield more specific and reproducible results than polyclonal antibodies but are expensive and not necessarily available for all proteins of interest. Additionally, antibodies may not be tolerant of modifications made to the proteins that destroy or modify the epitopes they bind to. Overall, mass spectrometry is not only more sensitive and reproducible but all the peptides are also detected under the same conditions.

Overall, the combination of the results from LC-MS/MS and immunoblot analysis suggests that SFPQ is cleaved by PL^{pro} during SARS-CoV-2 infection and perhaps cleaved by additional proteases as well in cells. The quantity of cleaved SFPQ may be less abundant than that of full-length SFPQ. As an RNA-binding protein, SFPQ is primarily localized to the nucleus in the nucleoplasm, however, Lukong et al. demonstrated that the tyrosine phosphorylation of SFPQ in its C-terminal region alters its ability to bind RNA (Lukong et al., 2009). This leads to its subcellular localization to the cytoplasm and cell cycle arrest. SFPQ may be localized to the cytoplasm as a consequence of viral infection and subsequently cleaved by PL^{pro}. This mechanism may be an attempt by the cells to induce cell cycle arrest to prevent further viral replication and spread. If this is the case, then PL^{pro} cleaves SFPQ to promote viral replication. Alternatively, SFPQ has also been shown to play a role in the DNA-virus innate immune response (Morchikh et al., 2017).

As a paraspeckle protein, SFPQ binds within a complex that interacts with the cGAS-STING-IRF3 pathway called the HEXIM1-DNA-PK-paraspeckle components-ribonucleoprotein complex (HDP-RNP). Upon detection of DNA, HDP-RNP is remodeled such that SFPQ is released, interferon regulatory transcription factor 3 (IRF3) is activated, and interferon production is stimulated. The cGAS-STING-IRF3 pathway has also been implicated to play a role in the RNA-virus immune response, although the mechanism remains unclear (Ma et al., 2021). Additionally, SARS-CoV-1 PL^{pro} has been shown to target STING to prevent IRF3 activation (X. Chen et al., 2014). So perhaps SARS-CoV-2 PL^{pro} targets SFPQ as a mechanism to evade the immune response via the STING pathway.

Alternatively, SFPQ may act as a pro-viral factor as previously observed in other cases of RNA virus infection. These include the promotion of influenza A viral transcription (Landeras-Bueno et al., 2011), the promotion of human rhinovirus replication (Flather et al., 2018), and exploitation by both the encephalomyocarditis virus (B. Zhou et al., 2019) and hepatitis delta virus (Greco-Stewart et al., 2006) during infection. Regardless of the mechanism of action, based on LC-MS/MS and immunoblotting results, PL^{pro} appears to cleave SFPQ in viral infection but may also be cleaved by other proteases as well.

2.6.2 The possible roles of BRD2 cleavage in SARS-CoV-2.

Bromodomain and extra-terminal domain (BET) proteins interact with many viral proteins and modulate viral infections. In particular, BRD4 forms a complex with E2 for transcriptional silencing in human papillomaviruses (S. Y. Wu et al., 2006), and BRD2 interacts with latency-associated nuclear antigen 1 (LANA1) in Kaposi's sarcoma-associated herpesvirus (Viejo-Borbolla et al., 2005). BRD2 binds the SARS-CoV-2 envelope E protein (Gordon et al., 2020) 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 (Samelson et al., 2022). In another recent study, BRD2, 3, and 4 inactivation was shown to aggravate viral infection in cells and mice overexpressing ACE2 (I. P. Chen et al., 2022). Our N-terminomics and immunoblot studies showed BRD2 cleavage by M^{pro} after Q206 (AALQ↓GSVT). This viral protease cleavage removes the bromodomain I (BDI) (**Fig. 3a**), potentially disrupting BRD2 binding to the acetylated histones and thereby affecting host gene transcription (Filippakopoulos et al., 2012). 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 (I. P. Chen et al., 2022; Gordon et al., 2020). 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 (Bisgrove et al., 2007). 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.

2.6.3 Potential activation of other cellular proteases in cell lysates.

The putative SARS-CoV-2 M^{pro} and PL^{pro} cellular targets were subjected to pathway analyses using Metascape (Y. Zhou et al., 2019) to reveal how viral proteases potentially disrupt cellular processes during infection. M^{pro} cleavage of host substrates is predicted to affect the cell cycle and cellular gene expression (**Fig. A.1a**), and the enriched processes of PL^{pro} substrates highlight the metabolism of RNA (**Fig. A.1b**). 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 M^{pro} and PL^{pro}. 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 M^{pro} and PL^{pro} specificities, we have minimized the identification of non-related protease substrates. In addition, we can exclude substrates found in the PL^{pro} dataset matching M^{pro} 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 M^{pro} and PL^{pro} 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 resulting 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 (Fortelny et al., 2015), to investigate which endogenous proteases account for those cleavages (**Fig. A.2**). A majority of the cleavage sites correspond to granzyme M specificity (P1=L/M) (Mahrus et al., 2004). 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.

2.6.4 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 (Bojkova et al., 2020; Stukalov et al., 2021) or CRISPR screens (Mac Kain et al., 2022; Rebendenne et al., 2022), 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 (Flather et al., 2018).

2.6.5 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. 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 M^{pro} and PL^{pro} 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 overexpressed ACE2 might affect the degradation of these substrates *in vivo*, such as TRIM28 (Wang et al., 2021) and BRD2 (Samelson et al., 2022). 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 (Araya et al., 2021). 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.

2.6.6 SARS-CoV-2 M^{pro} cleaves Gal8 primarily at LQ_{158}

Based on molecular weight, peptide quantification and time course experiments, the main cleavage event of Gal8 clearly occurs at LQ_{158}, while LQ_9 and LQ_{246} are secondary events. The results identified the N- and C-terminal fragments from cleavage at LQ_{158}, observed at 18 and 17 kDa by SDS-PAGE. Based on molecular weight the cleavage product at approximately 35 kDa is likely the C-terminal fragment that results from cleavage at LQ_9. As the quantification of LQ_{246} by mass spectrometry was not as well-defined, it remains difficult to make any conclusions on which, if any of the cleavage products observed is due to this specific site. The remaining double band around 20 kDa could not be confidently assigned by mass spectrometry.

Additional investigation by Chen et al. identified a trend in cleavage site targeting utilizing SDS-PAGE and densitometry analysis (S. A. Chen et al., 2023). The results indicated that Gal8 is first cleaved at LQ_9 and LQ_{158} which produces two cleavage products, Gal-$8^{10–317}$ (34.8 kDa) and Gal-$8^{159–317}$ (17.9 kDa). Subsequent cleavage of Gal-$8^{10–317}$ (34.8 kDa) at LQ_{158} results in Gal-$8^{10\text{-}158}$ (16.8 kDa). Therefore, the main two cleavage products are Gal-$8^{10\text{-}158}$ (16.8 kDa) and Gal-$8^{159–317}$ (17.9 kDa) which arise primarily due to cleavage at LQ_{158} as supported by mass spectrometry analysis. These results suggest that cleavage of Gal8 at LQ_{158} serves as the primary mechanism for evasion of the host immune response by disruption of Gal8 binding to the SARS-CoV-2 spike protein and subsequent antiviral-xenophagy.

2.7 Materials and methods

2.7.1 Identification of SARS-CoV-2 M^{pro} and PL^{pro} host substrates by N-terminal labeling

SARS-CoV-2 M^{pro} expression and purification. The recombinant His_6-GST-dual-tagged SARS-CoV-2 M^{pro} expression plasmid in the pGEX-6P-1 vector was cloned and kindly gifted by Dr. Rolf Hilgenfeld's lab (L. Zhang et al., 2020). The plasmid was transformed into *E. 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 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.

SARS-CoV-2 PL^{pro} expression and purification. The GST-tagged SARS-CoV-2 PL^{pro} expression plasmid in the pGEX-6P-1 vector was cloned and graciously gifted by Dr. Shaun K. Olsen's lab (Rut et al., 2020). 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 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 7000 rpm 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.

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 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 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 AAPPTec recommendations. For Reagent A,

16.5 mg of 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 (Poreba et al., 2017). 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 M^{pro}, Ac-Abu-Tle-Leu-Gln-ACC (Rut et al., 2021). 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/H_2O/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 8000x*g* for 20 min. The pellet was resuspended in ACN/H_2O (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 PL^{pro} (Rut et al., 2020), and stored at -80°C in DMSO at a final concentration of 1 mM.

SARS-CoV-2 M^{pro} 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 M^{pro} 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 $\lambda_{excitation}$ of 355 nm and $\lambda_{emission}$ of 460 nm.

SARS-CoV-2 PL^{pro} 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 4mM 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 PL^{pro} 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 $\lambda_{excitation}$ of 355 nm and $\lambda_{emission}$ of 460 nm.

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.

Inhibition of protein ubiquitination. Jurkat cells were cultured to 75% confluency and treated with an inhibitor of the ubiquitin-activating enzyme (UAE), 1 μM TAK-243 (Selleck Chemicals #S8341) or 0.01% DMSO for 4 h at 37°C with 5% CO_2 (Hyer et al., 2018). Cellular images were taken with an EVOS FL fluorescence microscope at 10X magnification prior to and following treatment in medium alone and in a 1:2 dilution with 0.4% Trypan blue stain (Gibco #15250-061). For the purposes of N-terminal labeling and enrichment, replicates of 4 x 10^9 Jurkat cells were utilized.

N-Terminal labeling and enrichment. The expression constructs for subtiligase expression (WT and M222A mutants) were a gift from Jim Wells and Amy Weeks (Weeks & Wells, 2018). 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 M^{pro} 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, 4mM IAM subsequently quenched with 20mM DTT prior to addition of PL^{pro}] for PL^{pro}). In Jurkat cell lysates, we also added 20 μM z-VAD-fmk to irreversibly inhibit endogenous caspases prior to adding SARS-CoV-2 M^{pro}. Cell lysates were clarified by centrifugation. For PL^{pro}, 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 M^{pro} or 5 μM of purified SARS-CoV-2 PL^{pro} was added to the soluble cell lysates for 2 h incubation, with aliquots taken out to monitor protease activity in activity assay as a function of time. N-terminal labeling was then performed with 1 μM stabiligase WT, 1 μM subtiligase M222A, and 1 mM TEVest6 (Weeks & Wells, 2018) 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).

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 PL^{pro} Jurkat dataset) from 3.85% to 36.8% acetonitrile in 0.1% formic acid. The PL^{pro} Jurkat treated with TAK-243 samples were separated on an analytical column (Aurora Ultimate nanoflow UHPLC column 25 cm x 75 μm ID, 1.7 μm C18, 120 Å; IonOpticks inc.) over the course of 90 min from 0% to 36.8% acetonitrile in 0.1% formic acid. 2 replicates of M^{pro} 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 (M^{pro}), and MSV000090124, MSV000090125, and MSV000091994 (PL^{pro}).

***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, and harvested using 0.5 mM EDTA. Jurkat and A549 cells were cultured, harvested using 0.5 mM EDTA, 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 M^{pro} and PL^{pro}, 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.

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 (Kumar et al., 2021). 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.

Immunoblot. SARS-CoV-2 infected cell lysates and cell lysates incubated with SARS-CoV-2 M^{pro} and PL^{pro} were loaded on 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 A.1**.

2.7.2 Mass spectrometry investigation of galectin-8 cleavage by SARS-CoV-2 Mpro

Plasmid information and protein purification (S. A. Chen et al., 2023)**.** M^{pro} SARS-CoV-2 gene with N-terminal His-SUMO tag in pET SUMO expression vector (Invitrogen) was used for protein production. Purification of SARS-CoV-2 M^{pro} was performed as described earlier (Arutyunova et al., 2021). Gal-8-6X His gene in pET vector was transformed into Rosetta (DE3) *Escherichia coli* cells and the protein was expressed for 18 hours at 18 ℃. The harvested cells were resuspended in 50 mM HEPES, pH 7.0, 300 mM NaCl, 5 mM Imidazole buffer, lysed by sonication and the protein was purified using Ni-NTA affinity chromatography.

In-gel Trypsin Digestion. SARS-CoV-2 M^{pro} cleavage assays with Galectin-8 (1.84 µg Galectin-8 and 0.34 µg M^{pro}/lane, incubated for 0.5, 1, 2, or 24 h) were separated by 14% sodium dodecyl-sulfate polyacrylamide gel electrophoresis. The gels were fixed for 20 minutes (50% ethanol, 2% phosphoric acid), washed twice for 20 minutes each (ddH2O) and stained overnight with blue-sliver coomassie stain (20% ethanol, 10% phosphoric acid, 750 mM ammonium sulphate, 0.12% Coomassie Blue G-250) and washed twice for 10 minutes each (ddH2O). Each lane was separated into 2 fractions, cut into 1 mm pieces for analysis of the whole assay or individual bands were cut out into 1 mm pieces for analysis of the protein fragment bands. For specific protein fragment band analysis, the desired protein fragment bands were similarly cut out and processed. The gel bands were transferred to a round bottom 96-well plate and to each well 200 µL of destaining solution (50mM ammonium bicarbonate, 50% acetonitrile) was added. The plate was incubated at 37°C for 10 minutes. The solution was removed from the wells and the detaining was repeated 3 times. The solution was removed and replaced with acetonitrile. The plate was

incubated again at 37°C for 10 minutes. The dehydration was repeated until the gel bands became white (2 times). The solution was removed and the sample were dried at 37°C for 10 minutes. The gel bands were rehydrated with 200 µL of reducing solution (100mM ammonium bicarbonate, 11.4 mM 2-mercaptoethanol) and incubated at 37°C for 30 minutes. The reducing solution was removed, 200 µL of alkylating solution (100mM ammonium bicarbonate, 10 mg/mL iodoacetamide) was added, and the gel bands were incubated at 37°C for 30 minutes. The gel bands were then washed with 200 µL 100mM ammonium bicarbonate at 37°C for 10 minutes (repeated twice). The gel bands were dehydrated in 200 µL acetonitrile at 37°C for 10 minutes (repeated twice), the solution was removed, and the bands were dried at 37°C for 15 minutes. The gel bands were trypsinized (100 µL of 100 mM ammonium bicarbonate and 6 µg/mL Sequence Grade Modified Trypsin, Promega Inc.) overnight. The solutions containing tryptic peptides were transferred to a round bottom 96-well plate. Tryptic peptides were further extracted from the gel bands with extraction solution (2% acetonitrile, 1% formic acid) and incubated at 37°C for 1 hour. A final extraction was conducted using 50% acetonitrile and 0.5% formic acid and incubated at 37°C for 1 hour. All solutions containing tryptic peptides were transferred to a round bottom 96-well plate and lyophilized overnight. The samples were resuspended (1 fraction per lane) in 0.1% formic acid before analysis by liquid chromatography tandem mass spectrometry (LC-MS/MS).

Mass spectrometry analysis. 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). The peptide mixture underwent reverse phase separation by an analytical column (2 µm, 100 Å, 50 µm × 15 cm, PepMap RSLC C18; Thermo Fisher Scientific). Peptides were eluted over a 45 minute (whole assay) or a 30 minute (individual protein fragment bands) linear gradient from 0 to 36.8% acetonitrile in 0.1% formic acid. Data analysis was conducted using ProteinProspector (v5.22.1) against a concatenated database of the *Homo sapien* proteome (SwissProt.2017.11.01.random.concat). The search parameter included digestion by TrypsinPro with non-specificity at the N-termini, a maximum of 3 missed trypsin cleavages, a precursor change range of +2 +3 +4, a precursor mass tolerance of 15 ppm, a fragment mass tolerance of 0.8 Da, carbamidomethylation of Cys (constant modification), and oxidation of Met and deamidation of Asn and Gln (variable modifications). A decoy database search was conducted to evaluate the false-positive rates. The target false discovery rate was set at 2% at the protein level and 1% at the peptide level. MS data is available through MassIVE (MSV000091989).

Initial discovery experiments were conducted with data-dependent acquisition as described above, over a range of 375 to 1700 m/z, with a MS^1 resolution of 120, 000 m/z. Parallel reaction monitoring (PRM) was subsequently conducted following the same methodology of the discovery (whole assay) at a resolution of 15,000 m/z at the MS^2 level with a range based on a starting mass-to-charge ratio of 140 m/z. Isolation lists (mass-to-charge ratio, charge, and 10-minute retention time window) were based on initial discovery experiments and were selected specifically for peptides C-terminal to the LQ_9, LQ_{158}, and LQ_{246} cleavage sites. Data analysis was conducted using Skyline (v21.2) against a library based on the discovery results with the parameters listed previously and an ion match tolerance of 0.5 m/z. MS data is available through MassIVE (MSV000091990).

Chapter Three: Neurovirology: Neurodegenerative disease and the consequences of viral infection

3.1 Mistranslation mediated by the methionyl-tRNA synthetase

3.1.1 Viral infection and oxidative stress have been previously shown to lead to mistranslation

Aminoacyl-tRNA synthetases are the proteins responsible for loading the 20 amino acids onto their appropriate transfer RNAs (tRNAs) (Mohler & Ibba, 2017). Their proper function is essential for mRNA translation as they possess proofreading mechanisms for the correctly aminoacylated-tRNAs and no further verification is conducted at the ribosome during translation. Such an example is the spontaneous mutation of the alanyl-tRNA synthetase in mice which results in neurodegeneration observed as a loss of Purkinje cells in the cerebellum (J. W. Lee et al., 2006).

The methionyl-tRNA synthetase (MARS) is the aminoacyl-tRNA synthetase responsible for loading the amino acid methionine onto its tRNA in preparation for mRNA translation. Viral infection, activation of toll-like receptors (TLRs), and oxidative stress have been shown to result in the misacylation of methionine onto non-cognate tRNAs and their subsequent use during translation (Netzer et al., 2009). Infection of HeLa cells with influenza A, adenovirus, and vaccinia virus resulted in increased Met-misacylation of non-cognate cytosolic tRNAs. Additionally, Met-misacylation was still detected when cells were treated with UV-inactivated virus. Activation of TLR3, TLR4, and TLR9 by exposure to respective ligands also increased Met-misacylation. Furthermore, induced oxidative stress by treatment with arsenite, telluride, or hydrogen peroxide also resulted in Met-misacylation of approximately 10%, a substantial increase compared to the baseline of 1%. Further investigation showed that oxidative stress in cells results in the dual phosphorylation of MARS at two serine residues, 209 and 825, by extracellular signal-related kinase (ERK 1/2) (J. Y. Lee et al., 2014). This post-translational modification disrupts the proper function of this enzyme and allows for increased Met-misacylation of tRNAs responsible for amino acids such as lysine, valine, glycine, aspartate, glutamate, and leucine (J. Y. Lee et al., 2014; Netzer et al., 2009). Cumulatively, these results suggested that viral infection and the consequential immune and stress responses lead to disruption of the proper function of MARS by its dual phosphorylation and therefore Met-misacylation resulting in Met-misincorporation during protein synthesis.

Met-misacylation has been proposed as a protective mechanism in which increased incorporation of solvent-exposed methionine in proteins may react with reactive oxygen species in an endogenous antioxidant mechanism to restore redox homeostasis (Levine et al., 1996). Methionine is oxidized by the reactive oxygen species to methionine sulfoxide, which is

subsequently reduced (Levine et al., 1996; S. Luo & Levine, 2009; Oien & Moskovitz, 2007). The significance of methionine during oxidative stress has been investigated in *E. coli* where decreased methionine results in reduced survival under oxidative stress (S. Luo & Levine, 2009). However, the importance of such an antioxidant salvaging mechanism does not negate the potentially hazardous implication of Met-misacylation on protein translational fidelity and the downstream consequences of mistranslated proteins.

Oxidative stress has also been implicated during the progression of neurodegenerative diseases including Alzheimer's disease (Hensley et al., 1995) (see refs (Butterfield & Halliwell, 2019; Tönnies & Trushina, 2017) for reviews), which may present a link between viral infection and the onset of neurodegenerative disease. We hypothesize that in human neurons, the presence of proteins produced by mistranslation events eventually leads to protein aggregation and cell toxicity.

3.1.2 A mimetic system for the investigation of Met-mistranslation

To investigate methionine mistranslation, a cumate inducible vector of MARS was cloned as well as its phosphomimetic (MARS-SD) and non-phosphomimetic (MARS-SA) forms containing dual mutation of S209 and S825 to aspartate and alanine, respectively. The MARS-SD vector was transfected into a HEK293 cell line and MARS-SD was induced with cumate. The level of MARS expression was observed by immunoblot to be increased by approximately two-fold (**Fig. 3.1**).

Utilizing the cumate inducible MARS-SD system, HEK293 cells were analyzed by LC-MS/MS utilizing in-gel digestion and label-free quantification. The protein abundance was compared with and without MARS-SD expression. Gene ontology analysis of the increased and decreased proteins was also conducted using Metascape (Y. Zhou et al., 2019). The increased pathways identified included cellular amino acid metabolic process, cytosolic tRNA aminoacylation, amino acid metabolism, response to ER stress, and ribosomal large subunit biogenesis while the decreased pathways included mitochondrial translation initiation and ribosome assembly (**Fig. 3.2**). The proteomic changes with MARS-SD expression suggested a good system for further investigation of MARS-mediated mistranslation.

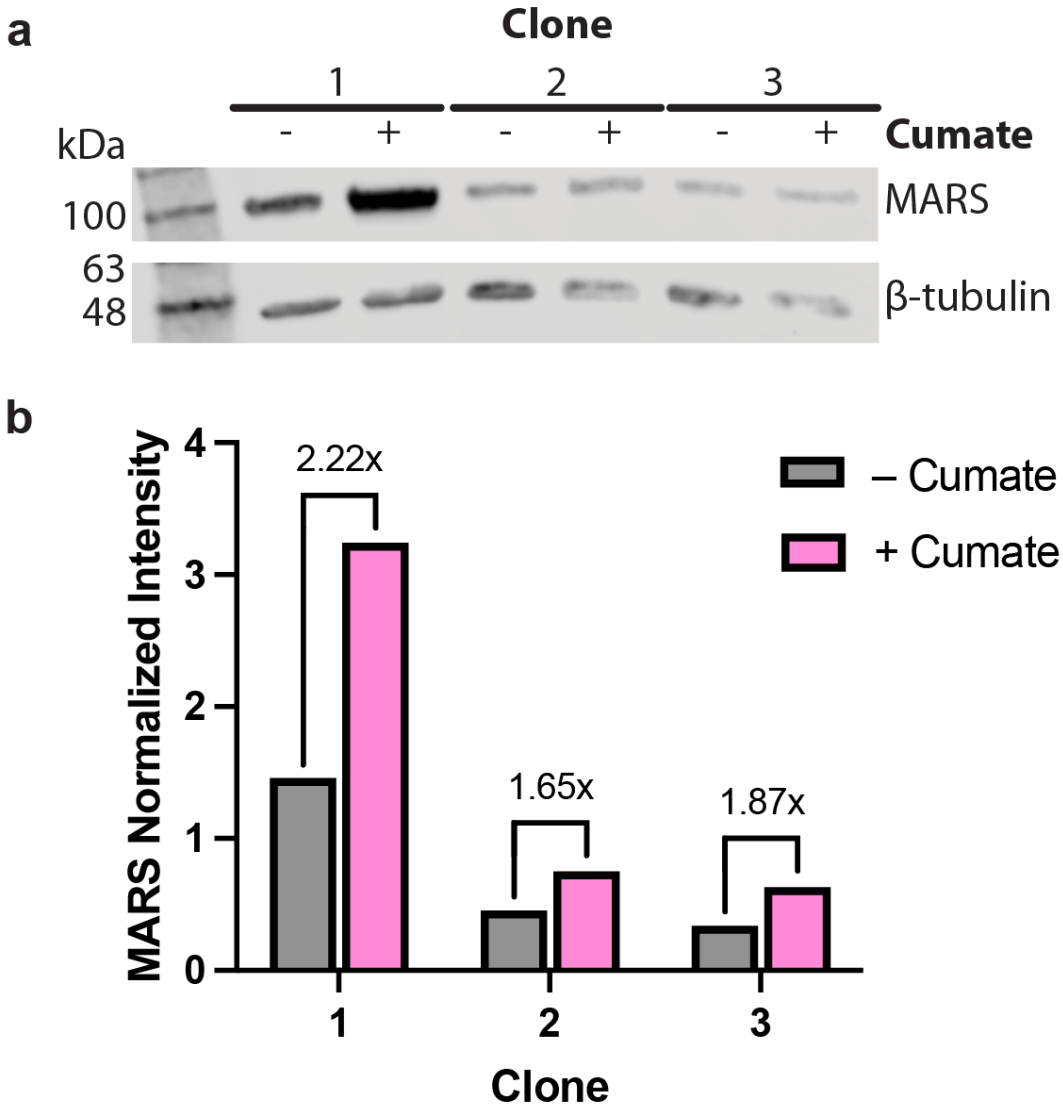


Figure 3.1. MARS overexpression was observed upon induction of MARS-SD expression in a single clone. a) Expression of MARS-SD in HEK293 P301L cells was induced with 300 µg/mL of cumate and monitored by immunoblot. **b)** The MARS band intensity was normalized on β-tubulin and clone 1 showed the greatest MARS expression with a greater than 2-fold increase in expression following MARS-SD expression with cumate induction.

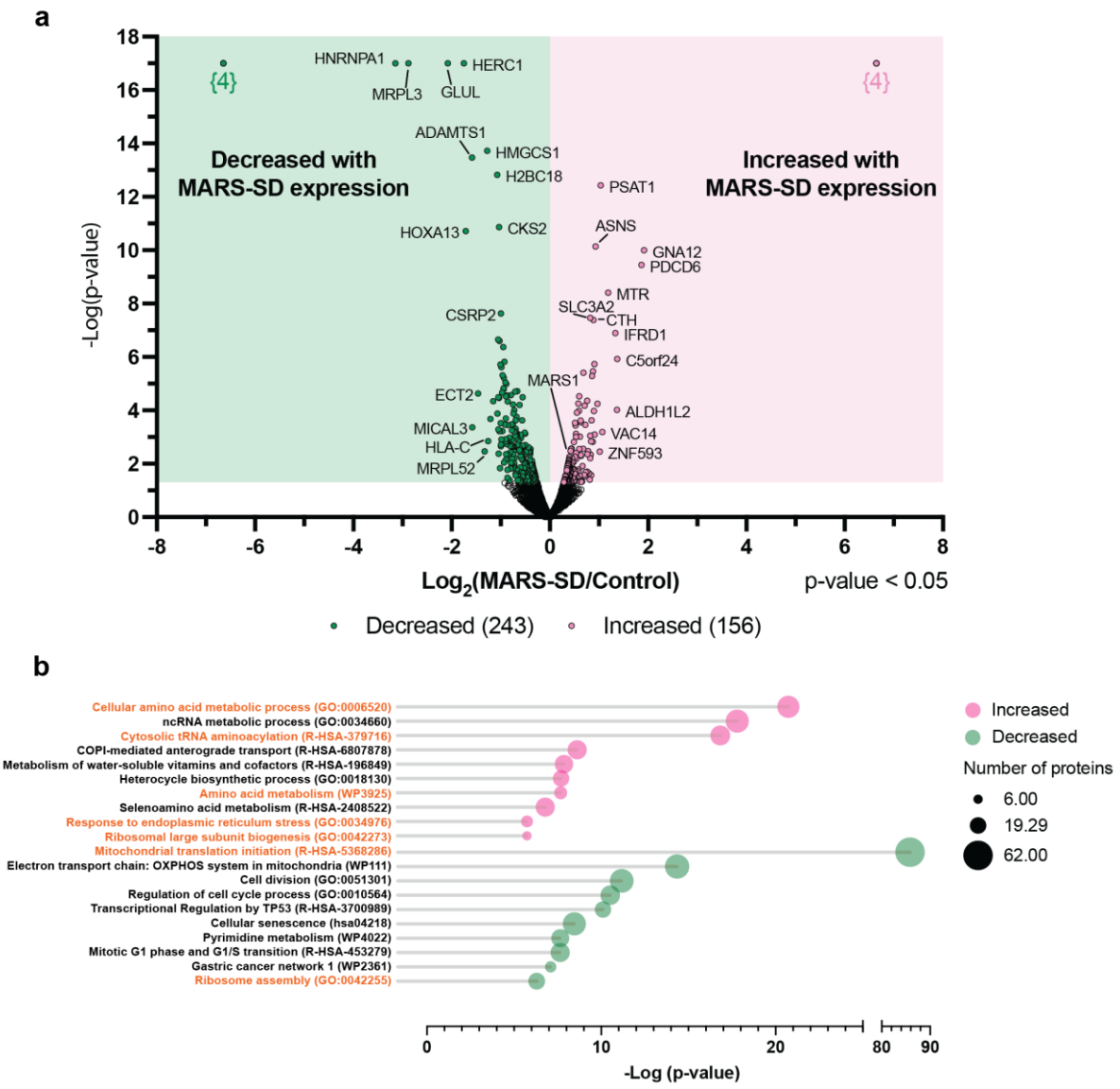


Figure 3.2. Expression of the MARS phospho-mimetic results in proteomic changes associated with translation and amino acid metabolism as observed by mass spectrometry. a) Volcano plot comparing the ratio of protein abundance to the p-value on a log scale with and without MARS phospho-mimetic expression (technical n=3). Proteins identified as decreased (243) appear on the left and increased (156) appear on the right with MARS phospho-mimetic expression. **b)** Gene ontology analysis of proteins identified as decreased (243) and increased (156) with MARS phospho-mimetic expression.

3.1.3 Data analytic techniques for the investigation of Met-mistranslation

Utilizing the previously acquired data from label-free quantification, met-mistranslation was investigated by looking for key mass shifts attributed to Met-misincorporation. We observed residues previously shown to exhibit this method of met-mistranslation including valine, glycine, lysine, aspartate, and glutamate (J. Y. Lee et al., 2014; Netzer et al., 2009), as well as other residues including proline, histidine, and alanine. We looked for mass shifts associated with a change from the initial residue to methionine. These shifts are present in the N-terminal and C-terminal fragment ions, b- and y-ions respectively, containing the residue of interest (**Fig. 3.3a**). The number of peptides that showed misincorporation of methionine with MARS-SD expression were compared to those without expression (control) and quantified as % mistranslated peptides (mistranslated/total peptides detected). Only valine displayed a greater average misincorporation of methionine with MARS-SD compared to the control, based on technical triplicates (**Fig. 3.3b**). Unfortunately, this method did not appear to be an appropriate method to analyze mass spectrometry data for met-mistranslation by the current workflow.

Mass spectrometry analysis relies on a database of proteins and their known sequences (**Fig. 3.4a**). This information can be analyzed by a software application based on a chosen digestion method to produce the associated peptides and their theoretical MS^1 and MS^2 spectra. These theoretical spectra are compared to the spectra acquired experimentally to identify proteins present in each sample dataset. As an alternative to investigating mistranslation by amino acid substitution mass shifts, we focused on modification of the proteome database to include met-mistranslation and thus a set of met-mistranslated proteins' theoretical spectra. Utilizing Python 3, a database of human protein sequences with met-mistranslation at a single amino acid (all possibilities or specific amino acids) can be constructed (**Fig. 3.4b, see Appendix B for code**). This database can be added to the existing human proteome database and utilized in future experiments for comparison of spectra from mass spectrometry data to theoretical spectra for met-mistranslated peptides.

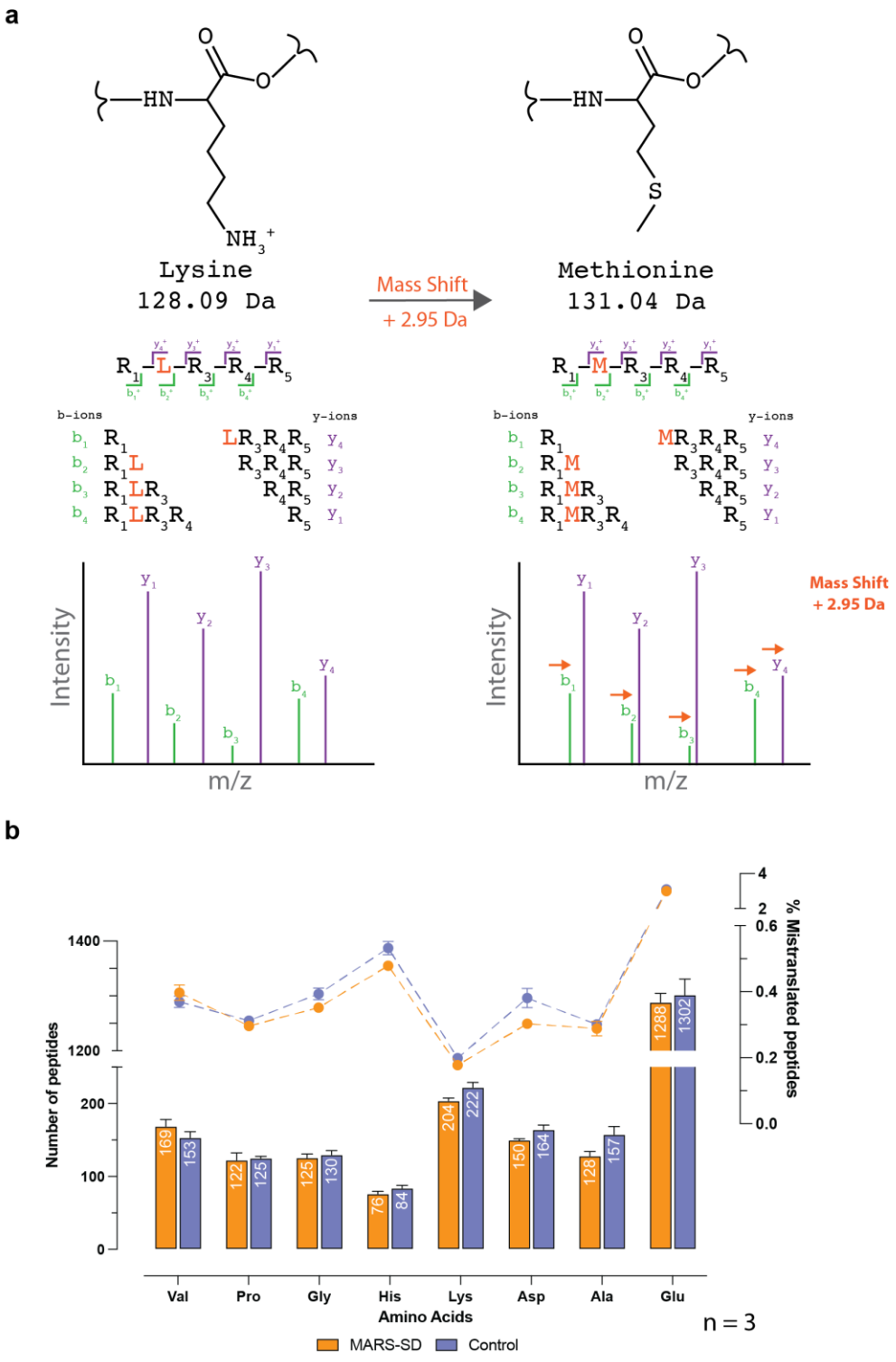


Figure 3.3. Methionine misincorporation was not identified by investigation of resultant mass shifts. a) Methionine was previously identified to be misloaded onto the lysine tRNA by a

MARS phospho-mimetic (J. Y. Lee et al., 2014). Lysine exhibits a molecular weight 2.95 Da less than methionine. Applying a variable mass shift of +2.95 Da allows for the detection of a residue change from lysine to methionine by the unique mass shift. The corresponding m/z shift can be calculated for the peptide fragment b and y ions, the N-terminal and C-terminal fragments respectively. **b)** MARS-SD expression in HEK293 was compared to a control. Mass shift analysis was used to investigate mistranslation of the residues valine, proline, glycine, histidine, lysine, aspartate, alanine, or glutamate to methionine (technical n=3). Only in the case of valine was met-mistranslation observed to be greater with MARS-SD expression compared to the control.

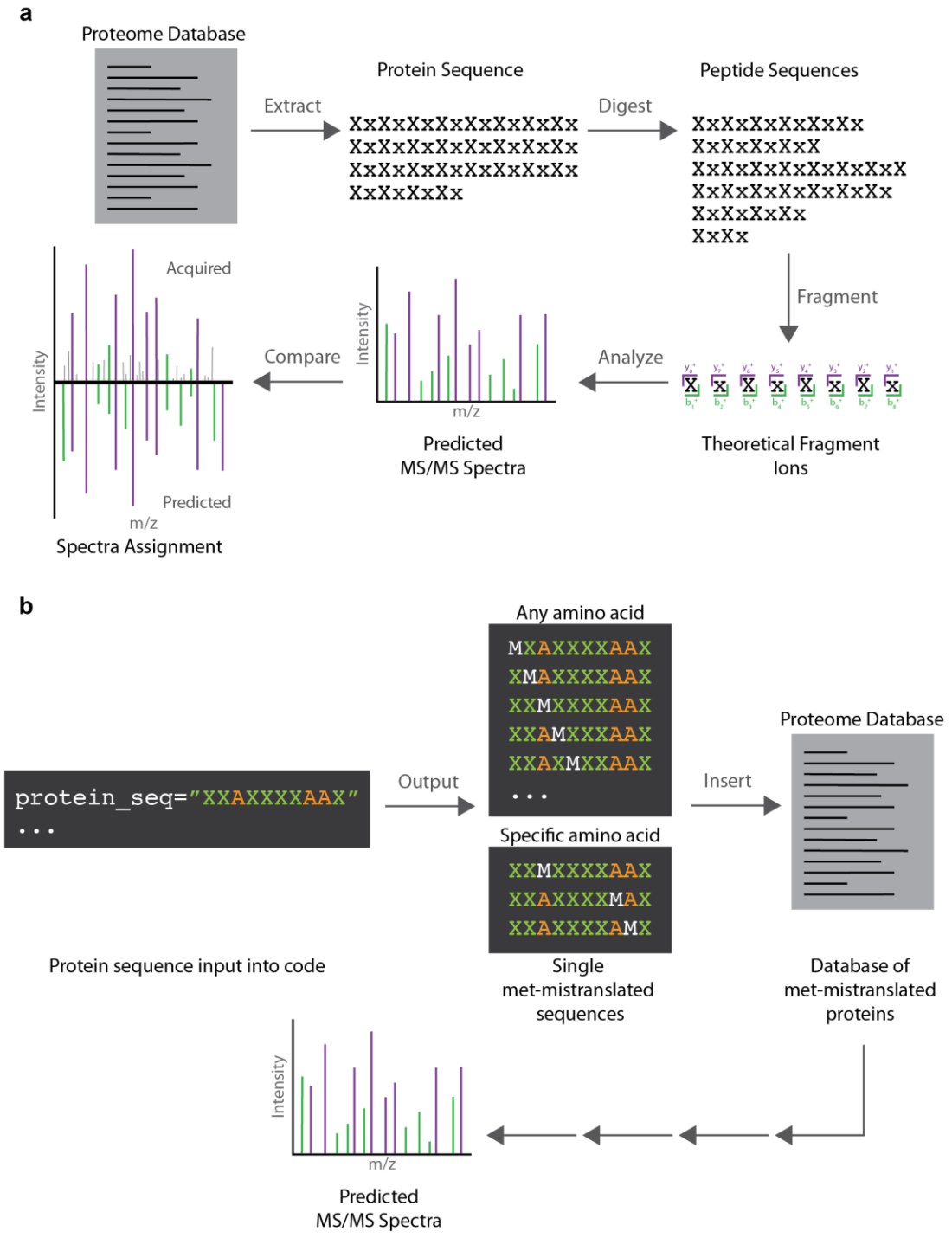


Figure 3.4. Design of a theoretical database containing mistranslated protein sequences for mass spectrometry analysis. a) Mass spectrometry analytical programs rely on a proteome database with individual protein sequences that can be digested into peptides by the chosen

enzymatic or chemical method. Peptide fragmentation then allows for the construction of a database of theoretical MS/MS spectra. The theoretical spectra are compared to acquired experimental mass spectra during data analysis. **b)** Utilizing a coding script a list of mistranslated protein sequences can be generated for single met-mistranslation at all possible residues and/or for specific amino acids. The mistranslated sequences can be incorporated into a database of met-mistranslated proteins used to predict MS/MS spectra as previously illustrated and compared to acquired experimental MS/MS spectra for the identification of met-mistranslation.

3.1.4 Investigation of mistranslation by an alternative protein digestion method

An advantage of LC-MS/MS at the MS^2 level is it can provide information on protein and thus peptide sequence when compared to a known database. This method is ideal when a database of protein sequences with met-mistranslation is available for reference, as previously discussed. However, due to the low occurrence of mistranslational events, these lower abundance mistranslated peptides may be difficult to detect and differentiate from the higher abundance peptides that lack met-mistranslation. In particular, this method thus relies heavily on the ability to detect peptide fragment ions containing the mass shift due to met-mistranslation. As a solution, we considered instead relying on m/z information acquired on the MS^1 level for peptides. However, to do so each peptide would then need to include methionine or rely on the presence of methionine in the sequence for protein digestion. Traditionally, we utilize the protease trypsin to produce tryptic peptides, but these may not all contain methionine. We aimed to use alternative protein digestion methods with mass spectrometry to act as an initial indicator of the increased presence of methionine due to met-mistranslation. To this end, we chose to utilize a digestion method that would cleave at methionine, and thus rely on increased incorporation of methionine in a protein sequence.

Cyanogen bromide (CNBr) can be used to cleave proteins C-terminal to methionine residues. CNBr is a chemical method of cleavage rather than an enzymatic method. With increased incorporation of methionine, as expected with met-mistranslation, an increased number of peptides per protein would result from digestion (**Fig. 3.5a**). Additionally, not all peptides are easily detected by mass spectrometry. Typically, the ideal peptide size for detection falls between 6 to 25 amino acids. In addition to increasing the number of peptides produced, met-mistranslation would in theory also increase the number of peptides of ideal length which can potentially be detected by mass spectrometry. Therefore, observation of the MS^1 spectra of a sample with met-mistranslation compared to a control without should show an increase in the number of peptides detected, indicating an increase in the number of methionine residues incorporated in the protein sequence.

We conducted preliminary analysis of our cyanogen bromide cleavage method with a purified isoform of Tau, 0N4R, with both wildtype and P301L mutant constructs. Initial results observed peptides in the corresponding MS^1 spectra (**Fig. 3.5b**), suggesting that the cleavage method was successful. In the future, this method will be tested to determine if it can be used as an indicator of increased methionine incorporation due to met-mistranslation.

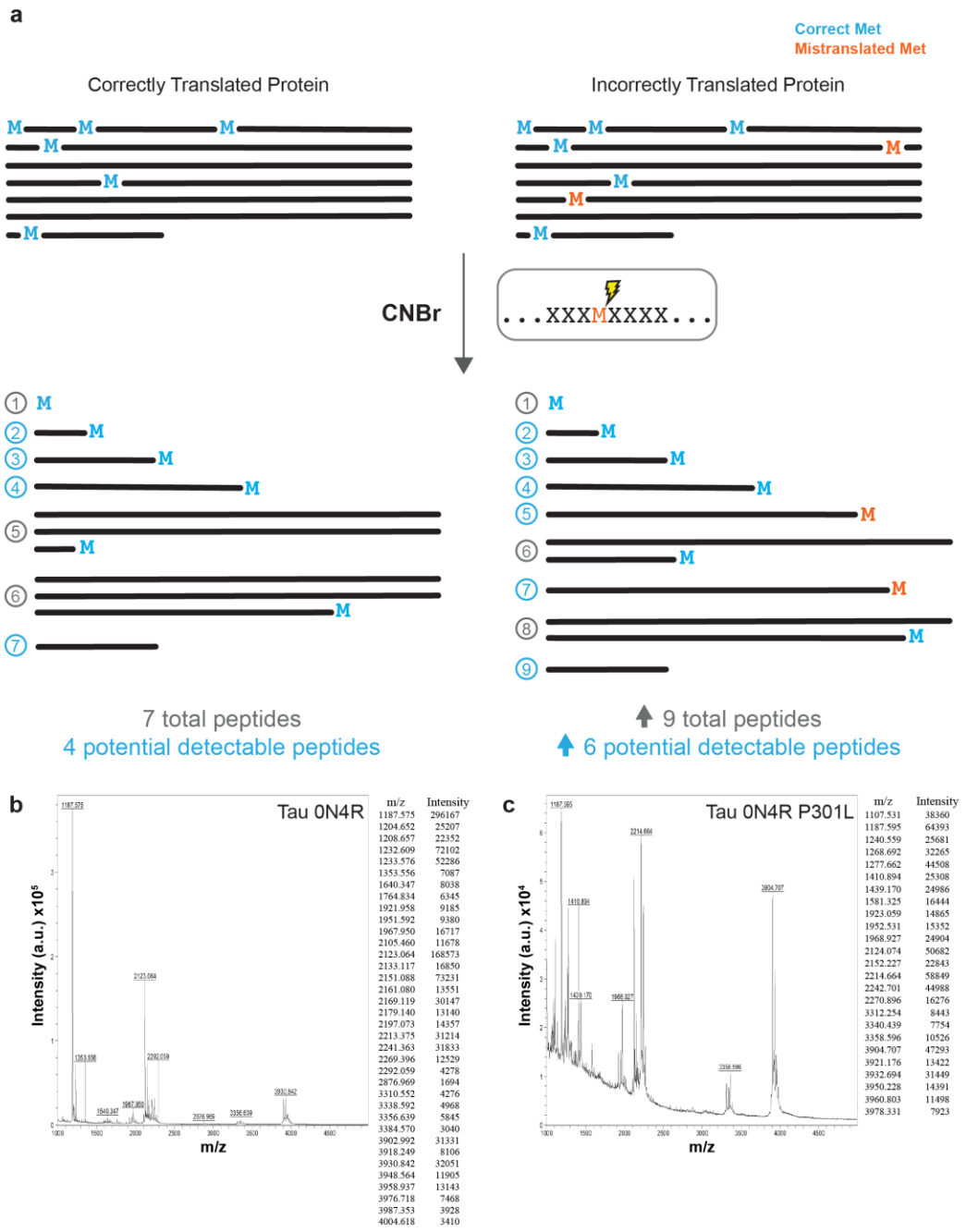


Figure 3.5. Cleavage of purified tau protein with CNBr shows detectable peptides by mass spectrometry. a) CNBr facilitates a chemical reaction that cleaves proteins C-terminal to methionine. Correctly translated proteins have few methionines and may produce some quantifiable peptides (ideally 6-25 amino acids). Incorrect translation of proteins via methionine mistranslation would result in the incorporation of additional methionines. This could allow for not only increased peptide count with CNBr cleavage but also an increase in the number of potentially

detectable peptides that fall in the ideal size range. Preliminary experiments were conducted utilizing in-gel CNBr cleavage on the purified proteins **b)** tau 0N4R wildtype and **c)** tau 0N4R P301L mutant. The MS spectra shown depict the identification of several peptides (m/z with associated intensity) for each protein.

3.1.5 Conclusions and future directions

Mistranslation due to the incorporation of misacylated tRNAs during protein synthesis occurs in mammalian cells due to oxidative stress and viral infection. However, its detection by mass spectrometry remains elusive. Utilizing a phosphomimetic of MARS, MARS-SD, to induce met-mistranslation in HEK293 cells, we were able to detect increased MARS expression (**Fig. 3.1, Fig. 3.2**) and proteomic changes that point to MARS-SD expression (**Fig. 3.2**). Despite this, we were unable to conclusively identify met-mistranslation by mass spectrometry (**Fig. 3.3**). In particular, the MARS-SD system utilized to induce mistranslation may not be the best model of use. Simple induction of oxidative stress via chemical agents, like arsenite or H_2O_2, may provide better results but will require future investigation. Additionally, the utilization of databases that include met-mistranslated proteins in the mass spectrometry analysis may aid in the identification of met-mistranslation (**Fig. 3.4, Appendix B**). We may not detect met-mistranslation due to its presence in the MS spectra that could not be assigned by the computational analysis. Either optimized databases or unbiased and less stringent search requirements can be applied in future studies to try to illuminate this data. Lastly, utilizing alternative sample preparation methods such as CNBr digestion, rather than trypsin, which cleaves peptides C-terminal to methionine residues has shown promising preliminary results (**Fig. 3.5**). This method may provide an alternative to better focus our analysis on the presence, or absence, of increased met-mistranslation.

3.2 Multiple sclerosis

3.2.1 Introduction to the central nervous system and multiple sclerosis

The central nervous system (CNS) is comprised of the brain and the spinal cord. Within the CNS, there are neurons and glial cells (see ref (Bigbee, 2023) for a review). Neurons are cells that are specialized to transmit chemical and electrical signals. Glial cells include the different supporting cells complementary to the neurons that aid in information transmission. The glial cells in the CNS include microglia, astrocytes, and oligodendrocytes. Astrocytes have many functions that maintain homeostasis within the CNS. They are connected to both the neurons and capillaries thus allowing them to supply the appropriate nutrients for neurons and neurotransmitter functions while maintaining the blood pressure and ion concentrations. Microglia function as the immune cells within the CNS, thus targeting foreign bodies and microorganisms as well as dead cells. Oligodendrocytes are responsible for the production of myelin in the CNS, which is a lipid-rich protein-containing material that surrounds the axons of neurons in sections called a myelin sheath. The myelin sheath acts to increase the rate at which nerve signals are transported. The

normal function of these cells in the CNS is essential, but several neuropathological diseases disrupt these normal functions. This includes neurodegenerative diseases, during which the loss of neuronal cell function and eventual cell death leads to neurological disability.

Multiple sclerosis is a neurodegenerative disease that affects over 2.8 million individuals globally (MS Society of Canada, 2020). The initial causes of the disease are unknown, yet it has been proposed that the risk factors for developing multiple sclerosis include viral infection, genetics, and environmental factors, such as low levels of vitamin D and smoking (see ref (Olsson et al., 2017) for a review). The top suspect for multiple sclerosis risk-associated viral infections is the Epstein-Barr virus (EBV) (Bjornevik et al., 2022). Following infection, the virus itself remains in a latent form in B lymphocytes and is a cause of infectious mononucleosis, commonly known as Mono (Thorley-Lawson, 2001). Infectious mononucleosis is known to increase the risk of multiple sclerosis development (Thacker et al., 2006). Additionally, EBV has been identified in some (Hassani et al., 2018; Moreno et al., 2018; Serafini et al., 2007), but not all multiple sclerosis lesions (Sargsyan et al., 2010; Willis et al., 2009). EBV appears to be one of many viruses that make up a larger viral signature identified in multiple sclerosis based on antiviral antibodies from multiple sclerosis patients (Enose-Akahata et al., 2023). Therefore, while the cause-and-effect relationship has yet to be determined, infection with EBV may be a risk factor for the development of multiple sclerosis.

While the cause is unknown, the biology of the disease has been studied significantly. It is characterized by inflammation in the CNS and the loss of the myelin sheath, a phenomenon known as demyelination, which results in the damage of neuronal axons (see refs (Karussis, 2014; Steinman, 2001) for reviews). As neurons are responsible for relaying electrical stimuli between neuronal bodies across the axons, myelin acts as an insulator to increase the rate at which these signals are transported. Demyelination therefore impacts the function of neurons in relaying signals and additionally, the resulting loss of axons and their cell bodies primarily determines the level of neurological disability observed in multiple sclerosis patients (Tallantyre et al., 2010).

Pyroptosis is a form of proinflammatory cell death that is lytic and involves the formation of pores in the plasma membrane that lead to cellular swelling and rupture (see ref (McKenzie, Dixit, et al., 2020) for a review). While regulatory cell death is important for brain development and the maintenance of homeostasis, pyroptosis and inflammation have been implicated in multiple sclerosis as well as other neurodegenerative diseases such as Alzheimer’s disease. Caspases are a family of proteases that are responsible primarily for the mechanism of cell death involved

in inflammation and apoptosis (Shalini et al., 2015). The family of inflammatory caspases includes caspases 1, 4, and 5 in humans (or caspases 1, and 11 in mice).

These inflammatory caspases are responsible for the cleavage of gasdermin-D (GSDMD), a known executioner of pyroptosis (Agard et al., 2010; He et al., 2015; Kayagaki et al., 2015; Shi et al., 2015). Cleavage of GSDMD is carried out by activated caspase-1 during the canonical inflammasome pathway and activated caspase-4/5 (or -11 in mice) in the non-canonical inflammasome pathway. Following the cleavage of GSDMD at Asp275 in humans (or Asp276 in mice), the 31 kDa N-terminal fragment is relocated to the plasma membrane where it oligomerizes and forms pores in the plasma membrane. These pores allow for the release of cytokines and the uptake of sodium ions and water. This disrupts osmosis leading to cellular inflammation and ultimately lysis. Caspase-1 cleavage and activation of the inflammatory cytokines IL-1β and IL-18 then allow for their release through these pores for the promotion of inflammation and the immune response. GSDMD is therefore a target of interest for the potential treatment of multiple sclerosis and its role in the disease requires further investigation.

3.2.2 Multiple CNS cell lines

As a preliminary investigation, we used LC-MS/MS to investigate the proteomic profiles of the specific cell types of neurons, astrocytes, microglia, and oligodendrocytes as well as whole human brain homogenate (**Fig. 3.6**). We identified a total of 4739 proteins in neurons, 4616 proteins in astrocytes, 3551 proteins in microglia, 5298 proteins in oligodendrocytes, 3605 proteins in brain homogenate A, and 3022 proteins in brain homogenate B.

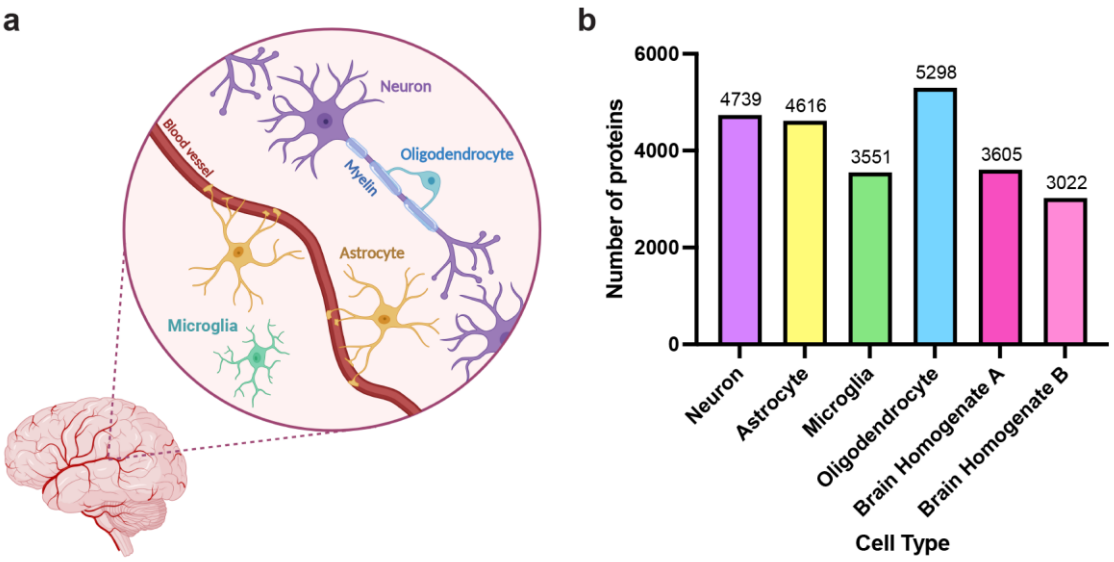


Figure 3.6. Identification of proteins in human cell types of the CNS. a) Neuronal and glial cells including astrocytes, microglia, and oligodendrocytes each play a unique role in the CNS. As depicted in the brain schematic, neurons are responsible for signal transmission. Astrocytes maintain homeostasis within the CNS facilitated by their connection to both the neurons and blood vessels. Microglia function as immune cells and oligodendrocytes myelinate the axons of neurons. **b)** Following LC-MS/MS, we identified a total of 4739 proteins in neurons, 4616 proteins in astrocytes, 3551 proteins in microglia, 5298 proteins in oligodendrocytes, 3605 proteins in brain homogenate A, and 3022 proteins in brain homogenate B.

3.2.3 Method optimization for LC-MS/MS sample preparation

Shot-gun proteomics or label-free quantification involves the lysis of cell or tissue homogenization to obtain a lysate containing a complex mixture of proteins, the proteome, of the respective cell type. While gentler lysis conditions, such as urea with tris to maintain the pH of 8 to 8.5, can be utilized these methods typically obtain soluble proteins and limit the extraction of proteins such as membrane-bound proteins. For some analysis, these conditions may suffice, typically when the targets of interest are known. However, when analysis of the whole proteome is required then harsher lysis conditions are needed. These typically include the use of Sodium dodecyl-sulfate (SDS) or SDS-containing buffers such as Radioimmunoprecipitation assay (RIPA) buffer. However, the presence of SDS in lysate samples presents a problem because such detergents can't be injected into the mass spectrometer. This is because they tend to disrupt the binding and elution of peptides during the LC step, as well as their ionization.

Traditionally, to remove the SDS, these protein lysate samples are separated by molecular weight on SDS-PAGE then digested within the gel and extracted as peptides that can be suspended in the appropriate buffer for mass spectrometer analysis (**Fig. 3.7a**). This method is called in-gel digestion. While this technique is well documented and utilized in the proteomics field, it can be time-consuming and costly especially when working with a greater number of samples. Therefore, we aimed to optimize an alternative technique for the removal of SDS and subsequent protein digestion.

The ProTrap XG is a column-based SDS-removal and protein digestion method that was previously developed and optimized for proteomics at Dalhousie University (Crowell et al., 2015). The basic principle of the column involves the removal of SDS by acetone-precipitation of proteins (**Fig. 3.7b**). The proteins can then be resuspended and digested into peptides with a chosen protease, typically trypsin. Additionally, the workflow includes a solid phase extraction (SPE) column for peptide de-salting before resuspension in the appropriate buffer and analysis by LC-MS/MS.

To optimize this method for our proteomics experiments, we first compared it to the traditional in-gel trypsin digestion in a side-by-side experiment. Lysates of MO3.13 cells, a human glial oligodendrocytic hybrid cell line, overexpressing gasdermin-B were prepared with either 20 or 40 µg of protein starting material. Typically, 20-30 µg is required as sufficient starting material for in-gel digestion. Using LC-MS/MS and starting with 20 µg of lysate, we identified 3243 proteins and 3262 proteins by in-gel digestion of samples 1 and 2, respectively, and 4641 proteins using the

ProTrap XG column workflow for both samples (**Fig. 3.7c**). These results indicated an increase in protein identification of approximately 42% with the use of the ProTrap XG columns. Additionally, when starting with 40 µg of lysate, no increase in protein identification was observed with the ProTrap XG, therefore 20 µg of protein lysate was determined as the appropriate starting material. We utilized SDS clean-up by ProTrap XG column for our subsequent experiments.

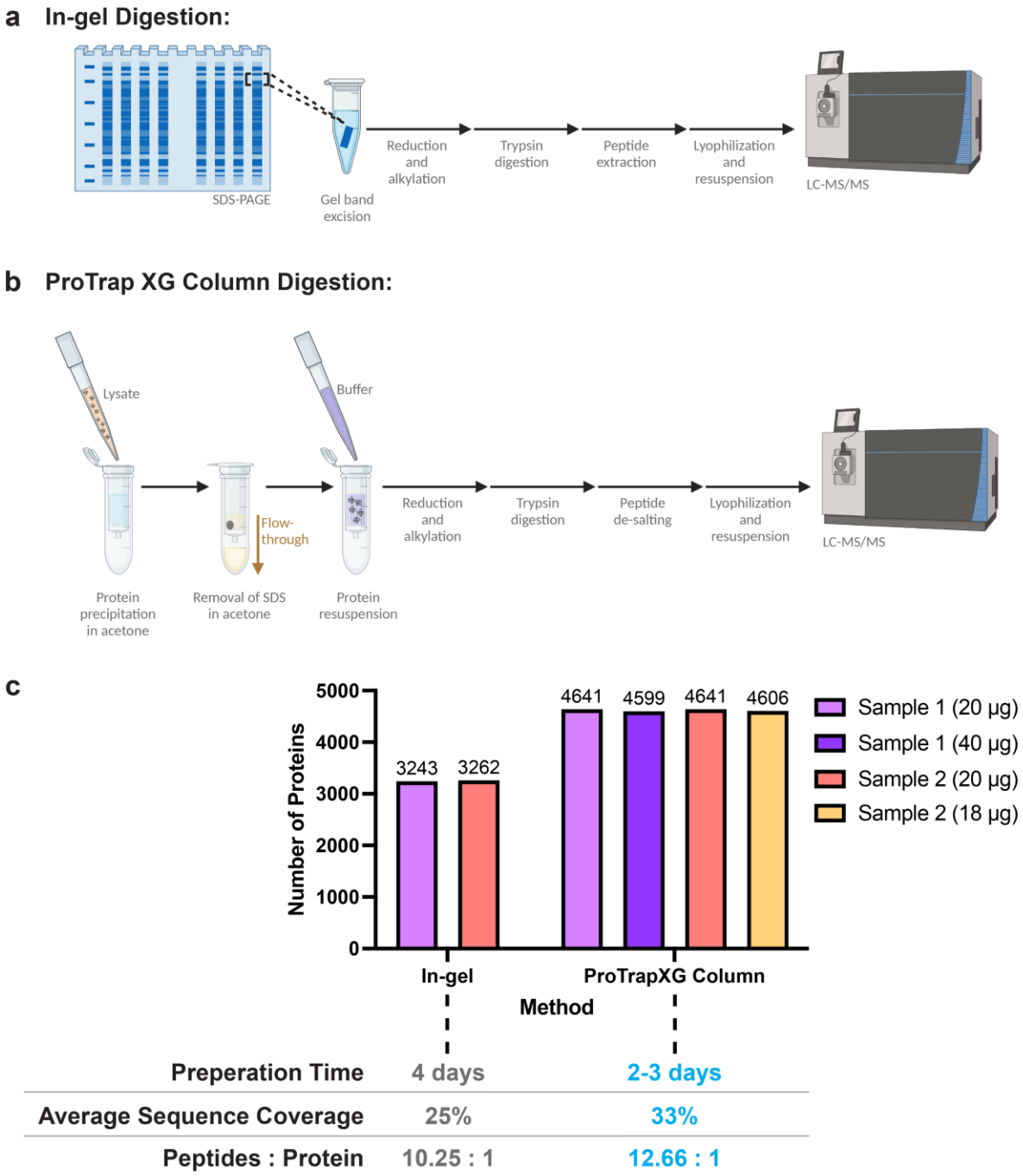


Figure 3.7. Method optimization shows that on-column preparation of mass spectrometry samples increased the number of proteins identified. a) The workflow for in-gel digestion involves using SDS-PAGE gel electrophoresis to fractionate proteins by molecular weight and extraction of peptides without detergent contamination. **b)** Alternatively, the workflow for ProTrap XG column digestion utilizes protein precipitation in acetone to remove detergent contaminates

followed by subsequent digestion to peptides. **c)** Following LC-MS/MS analysis, two samples prepared using either ProTrap XG column or in-gel digestion methods showed increased protein identifications, average sequence coverage, and peptides per protein quantified with ProTrap XG column digestion compared to in-gel digestion. Additionally, the on-column method requires less preparation time, 2 to 3 days instead of 4 days.

3.2.4 Data-dependent and data-independent acquisition in LC-MS/MS

In addition to optimization of our sample preparation workflow, we also investigated an alternative mass spectrometry acquisition method for better protein quantification (see ref (Hu et al., 2016) for a review). Traditionally and previously, we have utilized data-dependent acquisition (DDA) which relies on the precursor data collected at the MS^1 level to determine which precursors will be further fragmented for the MS^2 scan and data acquisition. In comparison, direct data-independent acquisition (DIA) operates independently of bias from the MS^1 scan. In this instance, we refer to the technique of direct DIA simply as DIA. In both DDA and DIA, a peptide mixture is injected into the instrument, and an MS^1 spectrum, detailing the m/z and relative intensity of the peptide is acquired (**Fig. 3.8a**). This is because the instrument scans in the MS^1 at each retention time over a range of m/z in both DDA and DIA (**Fig. 3.8b**). Of these precursors identified with DDA, some are selected for fragmentation and MS^2 scan to detect their fragment ions. However, with DIA all the precursors are fragmented, and their fragment ions are monitored. This kind of continuous monitoring is possible with DIA because short windows of MS^2 scans are generated with overlapping edges. This allows the instrument to scan at each retention time over the whole range of m/z selected while maintaining within the restrictions of resolution and acquisition time. Therefore, with DIA the resultant MS^2 spectra include the fragment ions, typically b- and y-ions, of multiple peptides (**Fig. 3.8c**). These spectra require software applications with the capabilities to deconvolute and resolve which spectra belong to which precursors. In contrast, with DDA the multiple MS^2 spectra tend to be cleaner, and each shows the fragment ions of a single peptide. Overall, with DDA the fragment ions for the selected precursor are monitored at a single retention time point with corresponding m/z and intensity (**Fig. 3.8d**). With DIA, the fragment ions for all precursors are monitored continuously at multiple retention time points with corresponding m/z and intensity. The resultant chromatograms show a peak for DDA where the maximum peak intensity is typically used for quantification (**Fig. 3.8e**). For DIA, the chromatogram ideally appears as a Gaussian-like curve and the peak area is used for quantification. As DDA does not allow for continuous monitoring of the ion fragments, the method of peak area quantification can't be applied due to the uncertainty of the peak "boundaries". To obtain the ideal chromatogram peak for DIA we aim for six points across the peak, which refers to the amount of detection points used to construct the peak. We also need to keep in mind that the number of data points detected is limited by the target resolution of the instrument scan and the speed of the instrument.

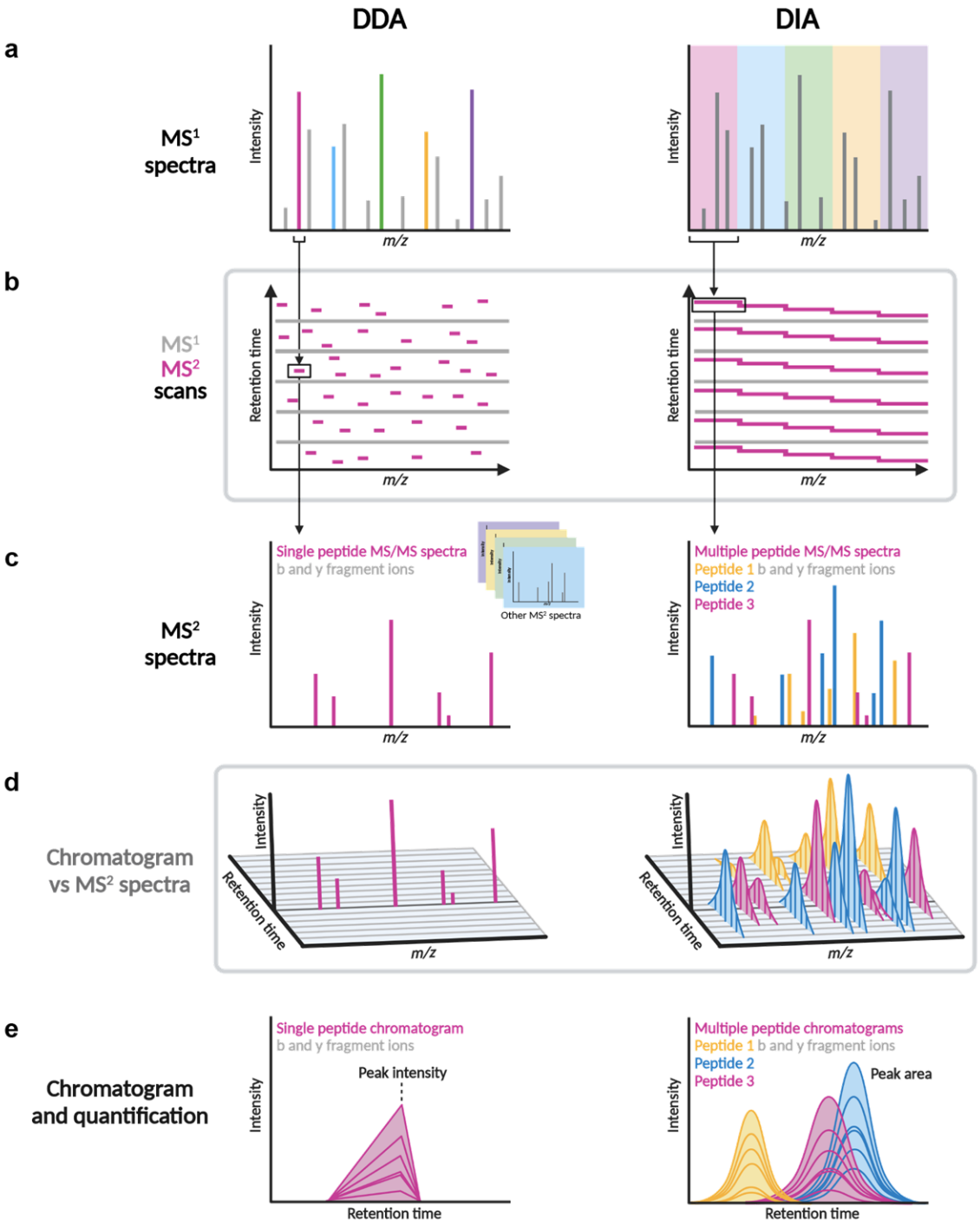


Figure 3.8. Comparison of DDA and DIA methods for mass spectrometry. a) Both DDA and DIA involve the initial acquisition of MS^1 spectra characterizing the m/z and intensity of the precursor peptides in a sample. **b)** These MS^1 scans span the total m/z range across the retention time. During DDA, precursors from the MS^1 spectra are selected for fragmentation and their MS^2

spectra are acquired. However, DIA selects all precursors from the MS^1 spectra for fragmentation and conducts MS^2 scans at each retention time in the form of m/z windows with overlapping overhangs between windows. These windows allow for the scanning of all precursors and their fragment ions with DIA instead of just select precursors as observed with DDA. **c)** The subsequent MS^2 (or MS/MS) spectra include the b and y fragment ions for either single or multiple precursors for DDA and DIA, respectively. **d)** This data can be observed considering the m/z, intensity, and retention time. With DDA, at a single retention time, the intensity of the fragment ions for a single peptide is acquired. In contracts, with DIA, the intensity of the fragment ions for multiple peptides is measured over time, with an ideal of six points across the peak. **e)** The result is a chromatogram peak with one detection time point for DDA that can be quantified based on the maximum peak intensity. In comparison, a chromatogram for DIA resembles a Gaussian curve, where the fragment ions differ in their maximal intensity but share a similar distribution. In the case of DIA, quantification is based on the peak area.

Previous studies have shown improved peptide identification and reproducibility with DIA compared to DDA (Bruderer et al., 2015; Kelstrup et al., 2018). Additionally, we have also observed the limitations of quantification by peak intensity as opposed to peak area in previous optimization of targeted mass spectrometry techniques (see **section 2.5, Appendix E**). Such observations have reinforced the need for continuous monitoring of precursor peptides and their fragment ions for sufficient quantification and observation of whole chromatogram peaks.

To compare DIA to DDA, we ran both methods on the same samples, further discussed in the following section on GSDMD. We compared the total number of protein groups, peptide precursors, increased or decreased protein groups (fold-change >2, p-value <0.05), and the number of proteins of interest identified (**Fig. 3.9a**). DIA identified a total of 6709 proteins and 140969 peptide precursors, with 536 increased and decreased proteins, and 29 proteins of interest. In comparison, DDA identified a total of 4704 proteins and 79801 peptide precursors, with 409 increased and decreased proteins, and 22 out of the 29 proteins of interest. Considering the overlap in proteins identified by the two methods, 4321 of the total proteins and 71 of the increased or decreased proteins were identified by both methods (**Fig. 3.9b-c**). Perhaps most significantly, this means that the DIA method identified 2388 proteins unaccounted for by DDA, a stark contrast to the 383 identified by DDA but not DIA. The same trend is observed by comparison of the common increased and decreased protein identified by each method. Additionally, by DIA we obtained a median of four points across the peak for all four conditions (**Fig. 3.9d**). As a result, the subsequent analysis was guided based on the proteomic data acquired by DIA.

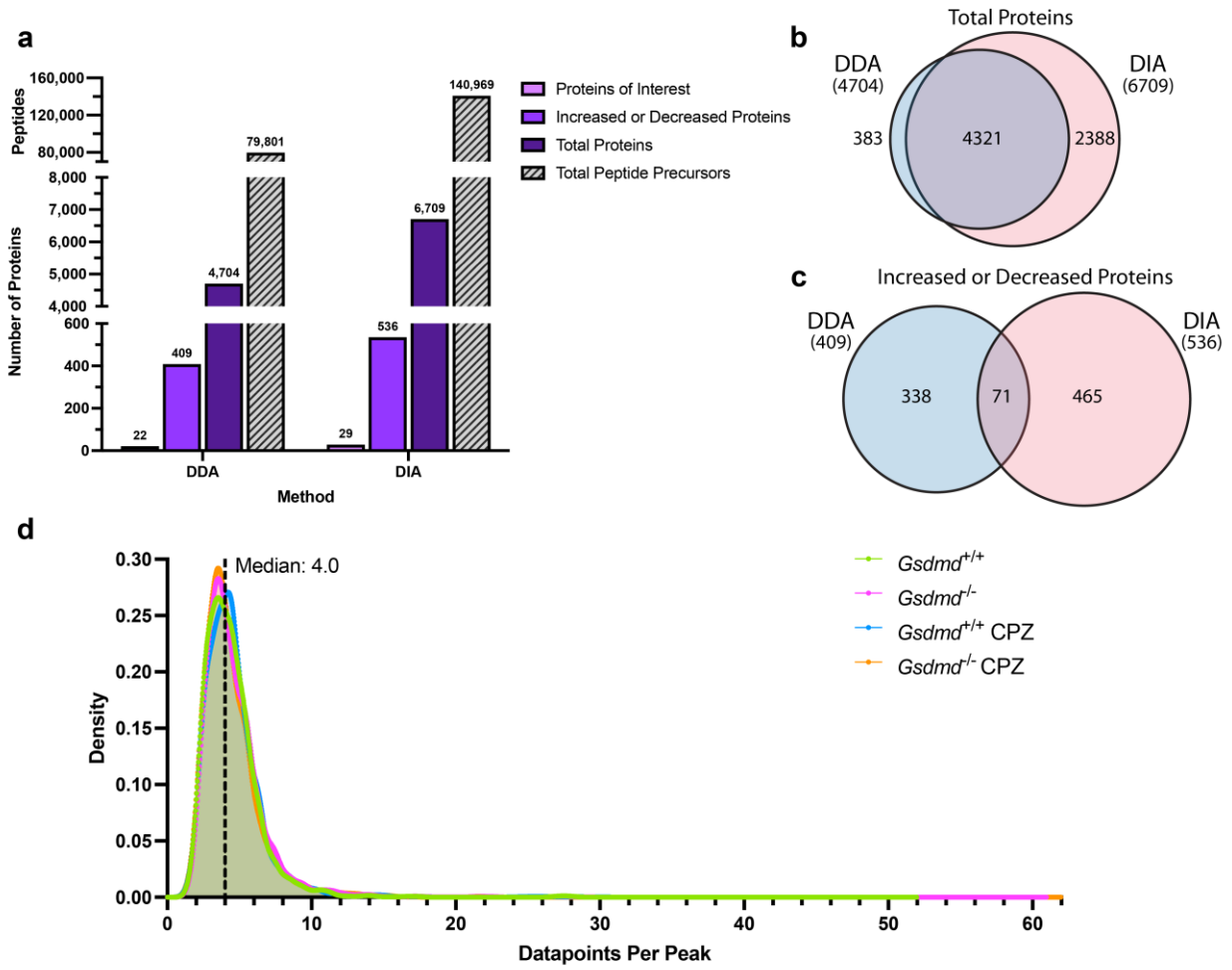


Figure 3.9. Comparison of proteomic data acquired by DDA and DIA mass spectrometry. **a)** Comparison of the number of proteins and peptide precursors identified by DDA and DIA shows a continuous trend of increased protein quantification by DIA compared to DDA. This includes proteins identified as increased and/or decreased (across multiple condition comparisons), and a list of 29 proteins of interest. Additionally, a comparison of the proteins quantified by DDA and DIA shows an overlap of **b)** 4321 total proteins and **c)** 71 increased or decreased proteins. **d)** Analysis of the DIA chromatogram data also reveals a median of 4.0 points across the peak.

3.2.5 Investigation of the role of gasdermin-D in demyelination

GSDMD has been identified to play a significant role in caspase-1,4,5 mediated pyroptosis. We aimed to investigate its role in the neuronal demyelination observed during multiple sclerosis disease pathology. To mimic multiple sclerosis disease pathology, mice were treated with a compound called cuprizone. Cuprizone is a copper chelator that has been shown to cause oligodendrocyte cell loss and subsequent demyelination in brain tissue when administered to rodents (see ref (Zirngibl et al., 2022) for a review). From its initial discovery as a demyelinating agent in the 1960s (Carlton, 1966), cuprizone now remains a commonly used multiple sclerosis model (see ref (Hooijmans et al., 2019) for a review and meta-analysis). Yet, its mechanism of action on oligodendrocytes and resultant demyelination remains largely unknown.

Recent results, from our collaborators in Dr. Christopher Power's group, suggest that the demyelination observed with cuprizone treatment in mice can be reduced with the knockout of GSDMD. Electron microscopy results show greater myelin thickness as well as the preservation of neuronal axons in the central corpus callosum of mice with GSDMD knockout compared to GSDMD wildtype when treated with cuprizone. To further investigate the role of GSDMD and the mechanism of action of cuprizone, we analyzed brain tissue samples taken from mice treated with or without cuprizone that additionally either expressed (*Gsdmd*$^{+/+}$) or had a knockout (*Gsdmd*$^{-/-}$) for GSDMD. These samples were prepared using the optimized SDS clean-up by ProTrap XG column and analyzed by DIA LC-MS/MS.

The proteomic quantification results were presented as the following ratios: *Gsdmd*$^{-/-}$/ *Gsdmd*$^{+/+}$, *Gsdmd*$^{-/-}$CPZ/ *Gsdmd*$^{+/+}$CPZ, *Gsdmd*$^{+/+}$CPZ/ *Gsdmd*$^{+/+}$, and *Gsdmd*$^{-/-}$CPZ/ *Gsdmd*$^{-/-}$. Corresponding volcano plots show the protein groups identified and the candidate increased and decreased proteins, with their corresponding fold change and p-values on a logarithmic scale (**Fig. 3.10a-d**). Comparison of GSDMD knockout to wildtype (*Gsdmd*$^{-/-}$/ *Gsdmd*$^{+/+}$) identified 27 decreased and 36 increased proteins while the comparison of GSDMD knockout to wildtype in the cuprizone multiple sclerosis model (*Gsdmd*$^{-/-}$CPZ/ *Gsdmd*$^{+/+}$CPZ) identified 123 decreased and 51 increased proteins. These values depict a greater increase in the number of decreased proteins with GSDMD knockout in the multiple sclerosis disease model compared to mice without the disease. Additionally, GSDMD knockout and wildtype conditions were compared to their cuprizone multiple sclerosis model counterparts. Respectively, 56 decreased and 71 increased proteins were identified in the wildtype (*Gsdmd*$^{+/+}$CPZ/ *Gsdmd*$^{+/+}$) and 156 decreased and 85 increased were identified in the knockout (*Gsdmd*$^{-/-}$CPZ/ *Gsdmd*$^{-/-}$). The GSDMD knockout ratio

in the case of disease reveals a greater number of decreased proteins in contrast to the GSDMD wildtype ratio.

The data were further analyzed for proteins typically associated with expression in neurons, microglia, astrocytes, and oligodendrocytes/myelin (**Fig. 3.10e**). The clearest trend was observed in proteins typically expressed by microglia, where a clear increase was observed in mice treated with cuprizone for demyelination in contrast to those without cuprizone. This list of 29 proteins identified in mice brains was also compared to those previously identified in multiple human CNS cell lines (**Table. C.1, C.2**).

Of the proteins of interest investigated, glial fibrillary acidic protein (GFAP) is a known biomarker for astrocyte activation (see (Petzold, 2015) for a review). The activation of astrocytes and macrophages in the CNS contributes to inflammation and injury in multiple sclerosis lesions (Absinta et al., 2021). A 2-fold increase in GFAP abundance was observed with cuprizone treatment compared to without cuprizone, with ratios of 2.38 and 2.20 for each of *Gsdmd*$^{+/+}$CPZ/ *Gsdmd*$^{+/+}$ and *Gsdmd*$^{-/-}$CPZ/ *Gsdmd*, respectively. In contrast, ratios of 1.20 and 1.11 were identified for *Gsdmd*$^{-/-}$/ *Gsdmd*$^{+/+}$ and *Gsdmd*$^{-/-}$CPZ/ *Gsdmd*$^{+/+}$CPZ, indicating a greater change in GFAP protein abundance associated with cuprizone treatment.

Additionally, complement C1q protein plays a role in the complement system key to the regulation of the immune response (see ref (Bohlson et al., 2014) for a review). C1q is particularly involved in the resolution of inflammation. A substantial increase in C1q subunits C1qa, C1qb, and C1qc is observed with cuprizone treatment based on the protein abundance ratios with *Gsdmd*$^{+/+}$CPZ/ *Gsdmd*$^{+/+}$ ratios of 4.20, 4.03, and 4.61 and *Gsdmd*$^{-/-}$CPZ/ *Gsdmd* ratios of 5.30, 5.69, and 7.42, respectively. A greater increase in abundance is also observed with GSDMD knockout in cuprizone-treated mice with *Gsdmd*$^{-/-}$CPZ/ *Gsdmd*$^{+/+}$CPZ ratios of 1.27, 1.46, and 1.54. This is in contrast to the lack of abundance change observed with only the knockout, seen in the *Gsdmd*$^{-/-}$/ *Gsdmd*$^{+/+}$ ratios of 1.01, 1.03, and 0.96, respectively.

Overall, these results are primarily indicative of the increased protein abundance observed for microglial proteins as well as astrocyte activation, suggesting the activation of such cells during the cuprizone multiple sclerosis disease model. Additionally, the proteomics data when combined with additional collaborator findings suggests that GSDMD plays a critical role in demyelination during multiple sclerosis (to be published).

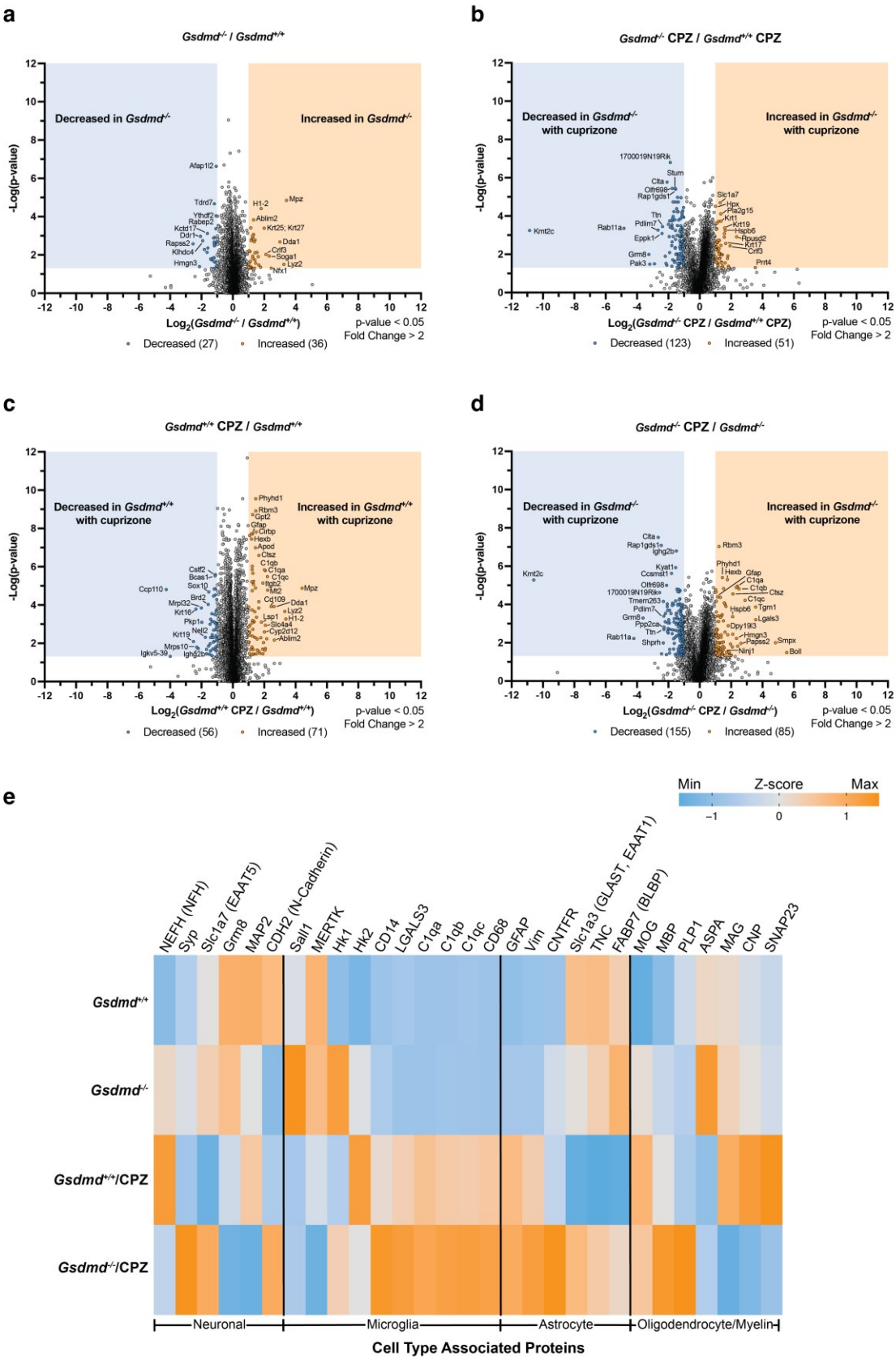

a
Gsdmd-/- / Gsdmd+/+
Decreased in Gsdmd-/-
Increased in Gsdmd-/-
-Log(p-value)
Log2(Gsdmd-/- / Gsdmd+/+)
p-value < 0.05
Fold Change > 2
Decreased (27)
Increased (36)
b
Gsdmd-/- CPZ / Gsdmd+/+ CPZ
Decreased in Gsdmd-/- with cuprizone
Increased in Gsdmd-/- with cuprizone
Log2(Gsdmd-/- CPZ / Gsdmd+/+ CPZ)
Decreased (123)
Increased (51)
c
Gsdmd+/+ CPZ / Gsdmd+/+
Decreased in Gsdmd+/+ with cuprizone
Increased in Gsdmd+/+ with cuprizone
Log2(Gsdmd+/+ CPZ / Gsdmd+/+)
Decreased (56)
Increased (71)
d
Gsdmd-/- CPZ / Gsdmd-/-
Decreased in Gsdmd-/- with cuprizone
Increased in Gsdmd-/- with cuprizone
Log2(Gsdmd-/- CPZ / Gsdmd-/-)
Decreased (155)
Increased (85)
e
Min
Z-score
Max
NEFH (NFH)
Syp
Slc1a7 (EAAT5)
Grm8
MAP2
CDH2 (N-Cadherin)
Sall1
MERTK
Hk1
Hk2
CD14
LGALS3
C1qa
C1qb
C1qc
CD68
GFAP
Vim
CNTFR
Slc1a3 (GLAST, EAAT1)
TNC
FABP7 (BLBP)
MOG
MBP
PLP1
ASPA
MAG
CNP
SNAP23
Gsdmd+/+
Gsdmd-/-
Gsdmd+/+/CPZ
Gsdmd-/-/CPZ
Neuronal
Microglia
Astrocyte
Oligodendrocyte/Myelin
Cell Type Associated Proteins

Figure 3.10. Proteomic profile of cuprizone-induced demyelination with and without GSDMD knockout. Volcano plots compare the fold-change ratio and p-value, on a logarithmic scale, for the protein groups quantified and reveal the increased and decreased protein groups associated with each comparison. These include **a)** GSDMD knockout and wildtype, **b)** GSDMD knockout and wildtype with cuprizone treatment, **c)** GSDMD wildtype with and without cuprizone treatment, and **d)** GSDMD knockout with and without cuprizone treatment. **e)** Of the protein groups identified, 29 proteins were categorized by cell type association, and their protein abundance was interpreted as a heatmap.

3.2.6 Conclusions and future directions

Optimization of our sample preparation and mass spectrometry acquisition workflows showed that a combination of on-column SDS clean-up and DIA produced increased protein identification and quantification by continuous monitoring of precursors. This methodology was applied to the investigation of multiple sclerosis via a cuprizone mouse model with a GSDMD knockout. The results identified key proteins with cellular-associated expressions such as GFAP and complement C1q proteins C1qa, C1qb, and C1qc. Overall, increased protein abundance was observed for microglial proteins as well as astrocyte activation in the disease model compared to controls. The combination of the proteomics data with collaborators' results highlights the role of GSDMD in demyelination during multiple sclerosis. Additionally, future optimization of the DIA method for specific sample types (mammalian cells, mouse samples, human samples, etc) may provide increased points across the peak closer or greater than the ideal six points.

3.3 Monkeypox infection in the cells of the central nervous system

3.3.1 Introduction to the monkeypox virus

The disease mpox emerged as a global outbreak that was declared a public health emergency of international concern by the World Health Organization in July of 2022 (*Mpox (Monkeypox)*, 2023). It is caused by the monkeypox virus (MPXV), an enveloped double-stranded DNA virus of the *Poxviridae* family, *Orthopoxvirus* genus (Lansiaux et al., 2022). The *Poxviridae* family of viruses also includes the variola virus, the causative agent of smallpox, and the vaccinia virus (VACV), used to vaccinate against smallpox due to its similarity to variola but less harmful infection (see ref (Jacobs et al., 2009) for a review).

As of May 2023, there have been 87, 377 confirmed cases of mpox and 140 deaths in 111 countries (*2022-23 Mpox (Monkeypox) Outbreak: Global Trends*, 2023). Typical symptoms of MPXV infection include an initial rash followed by any combination of common viral symptoms including fever, sore throat, headache, muscle aches, back pain, low energy, and swollen lymph nodes (*Mpox (Monkeypox)*, 2023). In addition to these symptoms, more severe cases of seizure, encephalitis, and additional symptoms indicative of the potential neuroinvasive properties of MPXV have been previously reported (see ref (Sepehrinezhad et al., 2023) for a review). In addition, Chailangkarn et al. observed that human induced pluripotent stem cell (iPSC) derived astrocytes were susceptible to MPXV infection but neurons were not susceptible (Chailangkarn et al., 2022). Therefore, to better understand the neuroinvasive properties of MPXV, more

specifically how infection modifies human astrocytes, we conducted proteomic investigation by mass spectrometry to characterize the proteomic changes observed with MPXV infection and similar virus, VACV.

3.3.2 Monkeypox virus infection modifies several mitochondrial pathways

The proteomic changes identified with MPXV and VACV infection of astrocytes by LC-MS/MS was presented as volcano plots (**Fig. 3.11a, b**). We highlighted proteins in these volcano plots with increased or decreased abundance. With VACV infection we identified 105 increased proteins and 170 decreased proteins, while with MPXV infection we identified 198 increased proteins and 289 decreased proteins. We further investigated the increased and decreased proteins by gene ontology pathway analysis with Metascape (Y. Zhou et al., 2019). Several mitochondrial pathways were found in the top 10 pathways identified for the proteins increased in abundance with MPXV infection in contrast to VACV infection (**Fig. 3.11c, d**). These include mitochondrial transmembrane transport, mitochondrial transport, and mitochondrial translation initiation. Additionally, the LC-MS/MS results were also investigated for the MPXV and VAVC viral proteins, with a total of 143 out of 174 MPXV proteins and 121 out of 199 VACV proteins identified in infected samples.

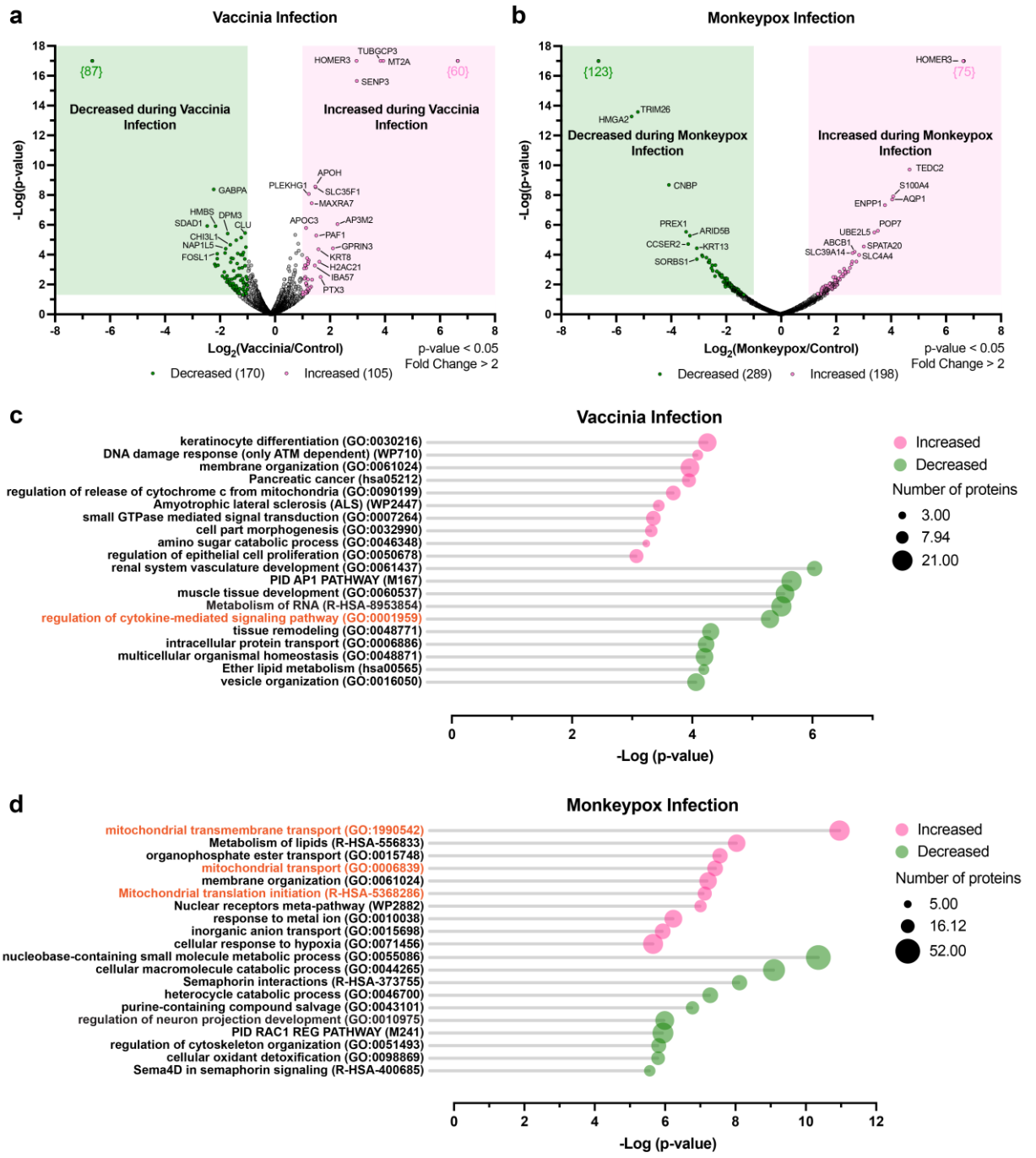


Figure 3.11. Proteomic profile of astrocytes infected with vaccinia or monkeypox virus. Volcano plots comparing the fold-change ratio and p-value, on a logarithmic scale, of the proteins quantified reveal the increased and decreased protein groups associated with **a)** vaccinia, or **b)** monkeypox viral infection of astrocytes. Gene ontology analysis of the increased and decreased proteins during **c)** vaccinia, or **d)** monkeypox viral infection shows the top 10 biological processes associated with these proteins.

3.3.3 Conclusions and future directions

A preliminary investigation of human astrocytes infected with MPXV and VACV by LC-MS/MS highlighted the impact of MPXV infection on mitochondrial protein functions, an effect not observed with VACV infection. Additional investigation of the proteins identified is required to understand the biology of MPXV infection and why astrocytes as opposed to neurons are susceptible to infection and potentially targeted by the virus.

3.4 Materials and methods

3.4.1 Methionine mistranslation

Plasmid information. The QM800A plasmid with cumate inducible expression contains MARS with mutations p.S209D (c.625 AGC>GAC) and p.S825D (c.2473 TCC>GAC), inserted at the 5' NheI and 3' SwaI restriction sites. Wildtype MARS was originally obtained as the plasmid 1533 pcDNA3 HA MARS gifted from William Sellers (Addgene #10716).

Transfection and stable cell line selection. HEK293 P301L cells were transfected with the QM800A MARS-SD plasmid by lentiviral transfection as previously described (McNamara, 2022). Antibody selection was conducted with puromycin for MARS-SD and doxycycline for P301L vectors.

Induced protein expression and cell harvesting. HEK293 P301L MAR-SD cells were cultured to 20-30% confluence in Dulbecco's Modified Eagle medium (DMEM, Gibco #11965-092) supplemented with 10% fetal bovine serum (FBS) and 1% Penicillin-Streptomycin (50 units/mL Penicillin, 50 µg/mL Streptomycin). MARS-SD expression was induced with 300 µg/mL cumate in fresh medium for 48 hours, with a replenishment of fresh 300 µg/mL cumate added at 24 hours post-induction. Medium alone was added for the control. Cells were pelleted 48 hours post-induction. The medium was aspirated from the flasks and the pellets were washed with 1X phosphate-buffered saline (PBS), then detached from the flask with 0.05% Trypsin-EDTA at 37°C for 3-5 minutes. The cells were diluted in fresh medium, centrifuged at 150*xg* for 15 minutes, resuspended in 1X PBS, counted, washed with 1X PBS, and centrifuged again. The 1X PBS was aspirated and the cell pellets were stored at -80°C.

Immunoblot. Harvested cell pellets of HEK293 P301L with MARS-SD expression were lysed with RIPA buffer and 15 µg of protein from cell lysate was loaded onto 4-20% SDS-PAGE gel (BioRad #4561094). Proteins were transferred to 0.45 µm nitrocellulose membranes (BioRad), blocked in 2.5% fish skin gelatin in TBS at RT for 1 h, washed with TBST for 5 min, then incubated with

primary antibody mouse anti-MetRS/MARS Mouse (Abcam ab50793) diluted to 1:100 at 4˚C overnight. The membrane was washed 3X with TBST for 5 min and incubated with 1:10000 anti-mouse IgG HRP-linked secondary antibody (Cell Signaling Technology #7076S) at RT for 1 h. The membrane was washed 3X with TBST which was replaced with a detection reagent (Pierce West Pico Plus) before viewing on the LI-COR Odyssey imaging system. The procedure was repeated for the loading control with primary antibody rabbit anti-Tubulin B-3 (BioLegend TUBB3) at 1:200 and secondary antibody 1:10000 anti-rabbit IgG HRP-linked secondary antibody (Cell Signaling Technology #7076S).

In-solution trypsin digestion. Cell pellets of HEK293 P301L with and without MARS-SD expression were lysed in 8 M urea, 100 mM Tris pH 8.5 with protease inhibitors (1 mM AEBSF, 5 mM EDTA, 1 mM PMSF, 4 mM IAM). The pellets were sonicated (Qsonica Q125) on ice at 20% amplitude, with 2 s on and 2 s off pulses, for 2 minutes, then centrifugated at 13 000 rpm for 10 minutes, and the supernatant was recovered. The protein concentration was measured by bicinchoninic acid (BCA) protein assay (Pierce Thermo Fisher Scientific). Aliquots of 100 µg (n=3 technical replicates) of protein were diluted 1:6 (final 1.3 M urea). The reduction was conducted with 5 mM DTT at 37C for 30 minutes and followed by alkylation with 10 mM IAM at R.T. for 30 minutes in the dark. The remaining IAM was quenched with 25 mM DTT and vortexed. The pH was adjusted to 8.5 with 1M bicine pH 9.2. The digestion was conducted with trypsin (Promega inc. #V5113) in a 1:100 (trypsin: protein) ratio overnight at R.T. Trypsin was inactivated and precipitated in 2.5% TFA at R.T. for 15 minutes. The peptide mixture was centrifugated at 13 000*xg* for 10 minutes and the supernatant was recovered. The peptide mixture was vacuum dried (Genevac EZ 2.3) and resuspended in 0.5% TFA the desalted on C18 zip tip (Millipore, cat #ZTC18S960). The peptide mixture was dried and resuspended in 0.1% formic acid for LC-MS/MS analysis.

Mass spectrometry analysis. Peptide separation was conducted using a nanoflow-HPLC (Thermo Scientific EASY-nLC 1200 System) coupled to Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific inc.). The peptide mixture underwent reverse phase separation on an analytical column (Aurora Ultimate nanoflow UHPLC column 25 cm x 75 µm ID, 1.7 µm C18, 120 Å; IonOpticks inc.). Peptides were eluted over a 120 min linear gradient from 0% to 36.8% acetonitrile in 0.1% formic acid. Data analysis was conducted using ProteomeDiscoverer (v2.4.1.15) software against a Homo sapien proteome reference (Proteome ID UP000005640, download date: 2020/02/18). The search parameter included a maximum of 3 missed trypsin cleavages, a precursor mass tolerance of 15 ppm, a fragment mass tolerance of

0.8 Da, carbamidomethylation of Cys (static modification), and oxidation of Met and deamidation of Asn and Gln (dynamic modifications). A decoy database search was conducted to evaluate the false-positive rates. The strict target false discovery rate was set at 0.01 and the relaxed false discovery rate was set at 0.05. Results reported include only the proteins identified at medium and high confidence, were identified by two or more peptides, were identified in two or more technical replicates, and were identified as master proteins (the top-ranking protein of a group). Proteins with a p-value less than 0.05 were identified as increased or decreased. MS data is available through MassIVE (MSV000092033).

Mistranslation database scripts. The library of mistranslated proteins was compiled utilizing two scripts (see **Appendix B**). All scripts were written in Python (v3.11) and tested in PyCharm community edition (v2022.3.2).

In-gel cyanogen bromide digestion. Purified tau 0N4R wildtype or P301L mutant protein (3 µg) was loaded onto SDS-PAGE gel and separation by electrophoresis was conducted only 2 cm into the resolving gel. The gel was stained in Coomassie blue for 10 min, destained twice (20 min, overnight), and rehydrated in ddH_2O. The gel was cut into 1 mm pieces, transferred to microfuge tubes, and detained again for 30 min with vortexing every 5 min. The gel pieces were washed with mass spectrometry grade water, dehydrated twice in acetonitrile, and dried. Gel pieces were incubated with cyanogen bromide (16 mg/mL in 70% formic acid) in the dark for 48 h. Peptides were extracted from the gel pieces with 25% acetonitrile/0.1% formic acid for 1 h and 50% acetonitrile/0.1% formic acid for another 1 h. The peptide mixture was vacuum dried (Genevac EZ 2.3) and resuspended in 10 µL acetonitrile for analysis by matrix-assisted laser desorption ionization (MALDI).

3.4.2 The CNS and the role of gasdermin-D in multiple sclerosis

Ethics statement. Use of autopsied brain tissue (Pro0002291) and human cell lines (Pro00027660) was approved by the University of Alberta Human Research Ethics Board (Biomedical). Written informed consent was received for all samples that included normal-appearing white matter and demyelinated lesions collected from matched multiple sclerosis and non-multiple sclerosis patients (McKenzie, Fernandes, et al., 2020; Roczkowsky et al., 2022). All animal experiments were performed according to the Canadian Council on Animal Care and approved by the University of Alberta Health Sciences Animal Care and Use Committee (AUP0002244).

Cell culture. MO3.13 cells were cultured for 5 days in high glucose DMEM with 1X penicillin-streptomycin and 100 nM phorbol-12-myristate-13-acetate (PMA) to differentiate cells to oligodendrocytes.

Human cell sample preparation and mass spectrometry analysis. Cell lysates (30µg protein/lane, n=1) were separated by SDS-PAGE gel electrophoresis in preparation for in-gel digestion. The gels were fixed for 20 minutes (50% ethanol, 2% phosphoric acid), washed twice for 20 minutes each (ddH2O), stained overnight with blue-sliver coomassie stain (20% ethanol, 10% phosphoric acid, 750 mM ammonium sulphate, 0.12% Coomassie Blue G-250) and washed twice for 10 minutes each (ddH2O). Each lane was separated into fractions, cut into 1 mm pieces. The gel bands were transferred to a round bottom 96-well plate and destained (50mM ammonium bicarbonate, 50% acetonitrile) at 37°C for 10 minutes. The solution was removed from the wells and the detaining was repeated 3 times. The gel bands were dehydrated in acetonitrile at 37°C for 10 minutes. The dehydration was repeated until the gel bands became white (2 times). The solution was removed, and the samples were dried at 37°C for 10 minutes. The proteins were reduced (100mM ammonium bicarbonate, 11.4 mM 2-mercaptoethanol) at 37°C for 30 minutes then alkylated (100mM ammonium bicarbonate, 10 mg/mL iodoacetamide) at 37°C for 30 minutes. The gel bands were then washed with 100mM ammonium bicarbonate at 37°C for 10 minutes (repeated twice), dehydrated in acetonitrile at 37°C for 10 minutes (repeated twice) and dried at 37°C for 15 minutes. The proteins were trypsinized (90 µL of 100 mM ammonium bicarbonate and 6 µg/mL Sequence Grade Modified Trypsin, Promega Inc.) overnight. The solutions containing tryptic peptides were transferred to a round bottom 96-well plate. Tryptic peptides were further extracted from the gel bands with extraction solution (2% acetonitrile, 1% formic acid) and incubated at 37°C for 1 hour. A final extraction was conducted using 50% acetonitrile and 0.5% formic acid and incubated at 37°C for 1 hour. All solutions containing tryptic peptides were transferred to a round bottom 96-well plate and lyophilized overnight. The samples were resuspended (3 fractions per lane) in 0.1% formic acid in preparation for analysis by LC-MS/MS. Mass spectrometry analysis was conducted with DDA as described in **section 3.4.3**, except a 90-minute gradient was used. MS data is available through MassIVE (MSV000092226).

Sample preparation and analysis for comparison of ProTrap XG and in-gel digestion. Two samples of MO3.13 cell lysates overexpressing gasdermin-B were prepared with either in-gel trypsin digestion as described for the human cell samples (above) or by ProTrap XG column as described for the GSDMD mice samples (below). In-gel trypsin digestion was conducted with 20 µg protein/lane (n=1) for each sample. ProTrap XG on-column preparation was conducted with

20 and 40 µg protein for sample 1 (n=1) and 18 and 20 µg protein for sample 2 (n=1). In-gel digestion and ProTrap XG samples were analyzed by LC-MS/MS in 3 fractions per sample and 1 fraction per sample, respectively. Mass spectrometry analysis was conducted with DDA as described in **section 3.4.3**, except 60-minute and 120-minute gradients were used for in-gel digestion and ProTrap XG samples, respectively. MS data is available through MassIVE (MSV000092227).

Mouse experiments and cuprizone treatment. 10-week-old female C57BL/6J mice (Jackson Laboratories #000664) or *Gsdmd*$^{-/-}$ mice (Jackson Laboratories #032663) (Roczkowsky et al., 2022; Saito et al., 2021) were fed Nutra-Gel diet (BioServ, 130 F4798-KIT) or Nutra-Gel diet containing 0.26% cuprizone (Sigma-Aldrich, cat: 14690) for 6 weeks. At the end of the experiment, mice were euthanized, cerebella were collected and frozen at −80˚C. Mouse cerebellum samples were homogenized in 10µl of cold RIPA buffer (Abcam, Cat: ab156034) with protease inhibitors (Millipore, Cat: 539134) per mg of tissue and were then centrifuged at 15,000*xg* for 15 min at 4°C. The protein concentration was measured by BCA protein assay (Pierce Thermo Fisher Scientific). All animal and human tissue samples were provided by Dr. Christopher Power.

GSDMD sample preparation for mass spectrometry. Samples we prepared using the ProTrapXG column (Proteoform Scientific inc.) for SDS clean-up and trypsin digestion. Mouse brain lysate (~20 ug) for each of *Gsdmd*$^{+/+}$, *Gsdmd*$^{-/-}$, *Gsdmd*$^{+/+}$/CPZ, and *Gsdmd*$^{-/-}$/CPZ (n=4) was loaded onto a filtration cartridge containing 400 µL of acetone, gently mixed and incubated for 30 minutes to allow the proteins to aggregate. The columns were centrifuged at 2500g for 2 minutes to pellet the protein aggregates and then centrifuged at 400g for 5 minutes to discard the supernatant. The pellets were washed with 400 µL of acetone and centrifuged again to discard the wash solvent. The proteins were then resolublized in 8 M urea, vortexed, sonicated, left at R.T. for 30 minutes, followed by addition of 100 mM Tris (pH 8). The reduction was carried out with 10 mM DTT at 37°C for 30 minutes followed by alkylation with 20 mM iodoacetamide for 1.5 hours in the dark. Trypsin (Promega inc. #V5113) was added in a 50:1 ratio (protein:trypsin) and incubated overnight with rotation. The peptides were recovered by centrifugation at 2500g for 5 minutes and the reaction was stopped by the addition of trifluoroacetic acid (TFA) to a final concertation of 1% and incubated for 10 minutes. The peptide mixture was centrifuged at 13000g for 10 minutes and the supernatant was collected. Peptide clean-up and desalting was conducted on the ProTrap XG SPE cartridge as per manufacturer recommendations. The peptide mixture was lyophilized overnight. The samples were resuspended in 0.1% formic acid for LC-MS/MS analysis.

Mass spectrometry analysis. Peptide separation was conducted using a nanoflow-HPLC (Thermo Scientific EASY-nLC 1200 System) coupled to Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific inc.). The peptide mixture underwent reverse phase separation on an analytical column (Aurora Ultimate nanoflow UHPLC column 25 cm x 75 µm ID, 1.7 µm C18, 120 Å; IonOpticks inc.). Peptides were eluted over a 120 min linear gradient from 0% to 36.8% acetonitrile in 0.1% formic acid.

DIA was conducted at the MS^1 level over a range of 350 to 2000 m/z at a resolution of 120,000 followed by acquisition of MS^2 over a range of 350 to 1400 m/z with isolation windows of 38.5 m/z including a 1 m/z overlap at a resolution of 30,000. The minimum desired points across the peak was set to 6. The data was analyzed on Spectronaut (v17.0.221202.55965) against a Uniprot database of the mouse proteome (Mus musculus Proteome ID UP000000589, download date: 06/21/2021). The BGS factory default settings were used except for the use of global normalization on the median. The search parameters included Trypsin/P cleavage with a maximum of 2 missed cleavages, fixed modification of carbamidomethylation (C), and variable modifications of deamidation (N/Q) and oxidation (M). The maximum number of variable modifications was set to 5. The peptide and protein group false discovery rate was set to 1% and precursor filtering was set to Q-value with a 1% cut-off. Quantification was conducted at the MS^2 level based on peak area. Differential abundance testing was conducted using an unpaired t-test with unequal variance using all MS level quantities. Proteins with a p-value less than 0.05 and a fold change threshold of 2.0 were identified as increased or decreased. MS data is available through MassIVE (MSV000091185).

DDA was conducted with a similar m/z range (375 to 1700 m/z) and resolution as described above. Data analysis was conducted using ProteomeDiscoverer (v2.4.1.15) software against a Homo sapien proteome reference (Proteome ID UP000005640, download date: 2020/02/18). The search parameter included a maximum of 3 missed trypsin cleavages, a precursor mass tolerance of 15 ppm, a fragment mass tolerance of 0.8 Da, carbamidomethylation of Cys (static modification), and oxidation of Met and deamidation of Asn and Gln (dynamic modifications). A decoy database search was conducted to evaluate the false-positive rates. The strict target false discovery rate was set at 0.01 and the relaxed false discovery rate was set at 0.05. Results reported were filtered to be comparable to DIA results and include only the proteins identified at high confidence, were identified by one or more unique peptides, were identified in two or more biological replicates, and were identified as master proteins (the top-ranking protein of a group).

Proteins with a p-value less than 0.05 and a fold change threshold of 2 were identified as increased or decreased. MS data is available through MassIVE (MSV000092228).

3.4.3 Viral infection of astrocytes

Cell culture and infection. Human fetal tissues were collected from 15 to 20-week-old electively terminated healthy pregnancies with written informed consent (Pro00027660), as approved by the University of Alberta Human Research Ethics Board (Biomedical). Human fetal astrocytes were prepared as previously reported (Na et al., 2011; Ramaswamy et al., 2013). The meninges were removed, and tissues were digested with 0.25% trypsin and 0.2 mg/mL DNase I (1 h, 37°C) then passes through a 70 µm cell strainer to prepare a single cell suspension. Cells were washed twice with fresh medium and plated in T-75 flasks at 6×10^7 to 8×10^7 cells/flask in Minimum Essential Medium (MEM) supplemented with 10% FBS, 2 mM L-glutamine, 1 mM sodium pyruvate, 1X MEM nonessential amino acids, 0.1% dextrose, 100 U/mL Penicillin, 0.5 µg/mL amphotericin B and 20 µg/mL gentamicin. The cells were passed at least once a week for 4-6 weeks to get pure mature astrocytes. The purity of these primary astrocyte cultures has been previously characterized (Na et al., 2011; Ramaswamy et al., 2013). For poxvirus infection experiments of primary human astrocytes, cells were seeded onto 6-well plates at 1×10^6 cells/well and exposed to MPXV (clade IIb lineage) or VACV (Acambis 2000) at MOI 10.0 for 6 h. For proteomic analyses, the cells were washed with PBS and collected with 100 µL RIPA buffer (Thermo Scientific, #89900) containing a protease inhibitor cocktail (Roche, #04693159001). Cells were homogenized and centrifuged (4°C, 15 min, 14,000*xg*), and the supernatants containing solubilized proteins were collected.

Sample preparation for mass spectrometry. Samples we prepared using the ProTrapXG column (Proteoform Scientific inc.) for SDS clean-up and trypsin digestion. Cell lysates (~20 ug) for each of MPXV, MPXV control, Vaccinia, and Vaccinia control (n=3) was loaded onto a filtration cartridge containing 400 µL of acetone, gently mixed and incubated for 30 minutes to allow the proteins to aggregate. The columns were centrifuged at 2500g for 2 minutes to pellet the protein aggregates and then centrifuged at 400g for 5 minutes to discard the supernatant. The pellets were washed with 400 µL of acetone and centrifuged again to discard the wash solvent. The proteins were then resolublized in 8 M urea, vortexed, sonicated, left at R.T. for 30 minutes, followed by addition of 100 mM Tris (pH=8). The reduction was carried out with 10 mM DTT at 37°C for 30 minutes followed by alkylation with 20 mM iodoacetamide for 1.5 hours in the dark. Trypsin (Promega inc. #V113) was added in a 50:1 ratio (protein:trypsin) and incubated overnight with rotation. The peptides were recovered by centrifugation at 2500g for 5 minutes and the

reaction was stopped by the addition of trifluoroacetic acid (TFA) to a final concertation of 1% and incubated for 10 minutes. The peptide mixture was centrifuged at 13000g for 10 minutes and the supernatant was collected. Peptide clean-up and desalting was conducted on the ProTrap XG SPE cartridge as per manufacturer recommendations. The peptide mixture was lyophilized overnight. The samples were resuspended in 0.1% formic acid for LC-MS/MS analysis.

Mass spectrometry analysis. Peptide separation was conducted using a nanoflow-HPLC (Thermo Scientific EASY-nLC 1200 System) coupled to Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific inc.). The peptide mixture underwent reverse phase separation on an analytical column (Aurora Ultimate nanoflow UHPLC column 25 cm x 75 µm ID, 1.7 µm C18, 120 Å; IonOpticks inc.). Peptides were eluted over a 120 min linear gradient from 0% to 36.8% acetonitrile in 0.1% formic acid. Data analysis was conducted using ProteomeDiscoverer (v2.4.1.15) software against a Homo sapien proteome reference (Proteome ID UP000005640, download date: 2020/02/18). The search parameter included a maximum of 3 missed trypsin cleavages, a precursor mass tolerance of 15 ppm, a fragment mass tolerance of 0.8 Da, carbamidomethylation of Cys (static modification), and oxidation of Met and deamidation of Asn and Gln (dynamic modifications). A decoy database search was conducted to evaluate the false-positive rates. The strict target false discovery rate was set at 0.01 and the relaxed false discovery rate was set at 0.05. Results reported include only the proteins identified at medium and high confidence, were identified by two or more peptides, were identified in two or more biological replicates, and were identified as master proteins (the top-ranking protein of a group). Proteins with a p-value less than 0.05 and a fold change threshold of 2 were identified as increased or decreased. Additionally, the data were searched against an MPXV (USA 2022 MA001) and VACV (ACAM2000) proteome of reference (Genbank, obtained 2023/04/04) as described previously. MS data is available through MassIVE (MSV000092229).

Chapter Four: Conclusions and future directions

4.1 Investigation of the SARS-CoV-2 viral proteases' human host substrates

We aimed to use mass spectrometry as a tool to identify the human host substrates of the SARS-CoV-2 viral proteases, M^{pro} and PL^{pro}. Utilizing subtiligase-based N-terminal labeling, we successfully identified a list of 16 potential substrates of PL^{pro} and 194 potential substrates of M^{pro} (S. Y. Luo et al., 2023). We determined, upon further investigation, that BRD2 is cleaved upon SARS-CoV-2 infection and during *in vitro* cleavage assays by M^{pro}. Interestingly, SFPQ was identified as a potential substrate of PL^{pro} in Jurkat, A549, and TAK-243-treated Jurkat cell lysates by mass spectrometry. Further investigation showed that SFPQ is cleaved upon SARS-CoV-2 infection, but cleavage was consistently difficult to observe following *in vitro* cleavage assays with PL^{pro}. Overall, our results suggest that BRD2 and SFPQ are targeted for cleavage during SARS-CoV-2 infection. BRD2 is identified as a human host substrate of M^{pro}, while SFPQ may be cleaved by PL^{pro} and other cellular proteases during infection.

As our study was conducted in Jurkat and A549 cell lysates, we were able to characterize the substrates of PL^{pro} and M^{pro} in the proteomes of these respective cell lines. In the future, we propose that additional cellular proteomes should be investigated. The suggestion that SARS-CoV-2 is a neurotropic virus (see ref (Tavčar et al., 2021) for a review), highlights the need for further investigation of the role that its viral proteases may play in the cells of the CNS, particularly the brain. Additionally, while we investigated some of the substrates identified by mass spectrometry, there remain many within the list of 16 and 194 potential substrates, for PL^{pro} and M^{pro} respectively, that require further investigation.

In addition to N-terminal labeling, we also demonstrated that DDA and PRM can be utilized to monitor the cleavage of a viral protease substrate. *In vitro* cleavage assays were combined with traditional DDA and PRM for targeted mass spectrometry to elucidate the cleavage events observed for Gal8 by SARS-CoV-2 M^{pro}. We identified three cleavage events of Gal8, at LQ_{9}, LQ_{158}, and LQ_{246} (S. A. Chen et al., 2023), that may disrupt the binding of Gal8 to the SARS-CoV-2 spike protein and prevent subsequent antiviral-xenophagy as previously suggested (Pablos et al., 2021). In future studies, PRM can be a tool to monitor the cleavage of viral protease substrates throughout infection. These studies may identify at what time during infection these substrates are targeted and at what cleavage sites.

4.2 Mistranslation mediated by the methionyl-tRNA synthetase

We aimed to detect the presence of mistranslated peptides due to the misincorporation of methionine residues in protein translation. We utilized a phosphor-mimetic of MARS, MARS-SD, with expression in mammalian cells to induce mistranslation. Mass spectrometry was used to detect protein mistranslational events. Unfortunately, despite various attempts to identify met-misincorporation in proteins, we were unable to detect mistranslation in our study. To aid in future investigations, we developed a script that can be used to encode for a database of mistranslated proteins, due to met-misincorporation.

In the future, a database of mistranslated proteins can be implemented in our mass spectrometry analysis to identify mistranslation. The use of unbiased and less stringent search requirements may also aid in our future analysis. Additionally, mistranslation can be induced by oxidative stress with chemical agents, such as arsenite or H_2O_2, and may provide an alternative model for this study. Lastly, utilizing alternative sample preparation methods such as CNBr digestion may allow for the focused analysis of the presence, or absence, of increased mistranslation.

4.3 Investigation of the role of gasdermin-D in demyelination

We used mass spectrometry as a tool to investigate the role of GSDMD in multiple sclerosis, with a cuprizone mouse model. We highlighted several proteins with cell-associated expression in neurons, astrocytes, microglia, and oligodendrocytes. Our results indicated, in the multiple sclerosis model compared to controls, an increase in protein abundance for several microglia-associated proteins as well as astrocyte activation. Additionally, our collaborators' results by electron microscopy suggest that GSDMD knockout alleviates the level of demyelination observed in multiple sclerosis. The combination of our proteomics data with collaborators' results highlights the role of GSDMD in demyelination during multiple sclerosis (to be published).

4.4 Monkeypox infection in the cells of the central nervous system

We investigated the infection of human astrocytes with MPXV and VACV utilizing mass spectrometry. In our preliminary analysis, we observed a distinct impact of MPXV infection on mitochondrial protein functions, which was not observed with VACV infection. Future investigation of the host proteins identified is required to better understand the biology of MPXV infection in astrocytes, in contrast to VACV. Additionally, further analysis of the viral proteins identified during

our infection study may provide insights into the different roles these viral proteins may play in the biological pathways observed.

4.5 Optimization of mass spectrometry techniques for proteomics

We optimized our mass spectrometry methodology for use in current and future studies. We found that on-column SDS clean-up using ProTrap XG, compared to in-gel digestion, increased protein, and peptide identification. It also reduced the time necessary for sample preparation. Combining our on-column preparation with LC-MS/MS, we found that DIA increased the number of proteins and peptides identified when compared to DDA. In the future, further optimization of our DIA method is necessary to increase our acquired points across the peak from four to the ideal 6 to more. Increasing our acquired points per peak can improve our quantification by DIA. Additionally, our optimization will need to account for different sample types, including but not limited to mammalian cells, mouse samples, and human samples.

Bibliography

2022-23 Mpox (Monkeypox) Outbreak: Global Trends. (2023, May 11). https://worldhealthorg.shinyapps.io/mpx_global/

Abdelnabi, R., Foo, C. S., Jochmans, D., Vangeel, L., De Jonghe, S., Augustijns, P., Mols, R., Weynand, B., Wattanakul, T., Hoglund, R. M., Tarning, J., Mowbray, C. E., Sjö, P., Escudié, F., Scandale, I., Chatelain, E., & Neyts, J. (2022). The oral protease inhibitor (PF-07321332) protects Syrian hamsters against infection with SARS-CoV-2 variants of concern. *Nat Commun*, *13*(1), 719. https://doi.org/10.1038/s41467-022-28354-0

Absinta, M., Maric, D., Gharagozloo, M., Garton, T., Smith, M. D., Jin, J., Fitzgerald, K. C., Song, A., Liu, P., Lin, J.-P., Wu, T., Johnson, K. R., McGavern, D. B., Schafer, D. P., Calabresi, P. A., & Reich, D. S. (2021). A lymphocyte–microglia–astrocyte axis in chronic active multiple sclerosis. *Nature*, *597*(7878), Article 7878. https://doi.org/10.1038/s41586-021-03892-7

Agard, N. J., Maltby, D., & Wells, J. A. (2010). Inflammatory Stimuli Regulate Caspase Substrate Profiles *. *Molecular & Cellular Proteomics*, *9*(5), 880–893. https://doi.org/10.1074/mcp.M900528-MCP200

Alimonti, J. B., Ribecco-Lutkiewicz, M., Sodja, C., Jezierski, A., Stanimirovic, D. B., Liu, Q., Haqqani, A. S., Conlan, W., & Bani-Yaghoub, M. (2018). Zika virus crosses an in vitro human blood brain barrier model. *Fluids and Barriers of the CNS*, *15*(1), 15. https://doi.org/10.1186/s12987-018-0100-y

Andrews, M. G., Mukhtar, T., Eze, U. C., Simoneau, C. R., Ross, J., Parikshak, N., Wang, S., Zhou, L., Koontz, M., Velmeshev, D., Siebert, C.-V., Gemenes, K. M., Tabata, T., Perez, Y., Wang, L., Mostajo-Radji, M. A., de Majo, M., Donohue, K. C., Shin, D., … Kriegstein, A. R. (2022). Tropism of SARS-CoV-2 for human cortical astrocytes. *Proceedings of the National Academy of Sciences*, *119*(30), e2122236119. https://doi.org/10.1073/pnas.2122236119

Araya, L. E., Soni, I. V., Hardy, J. A., & Julien, O. (2021). Deorphanizing Caspase-3 and Caspase-9 Substrates In and Out of Apoptosis with Deep Substrate Profiling. *ACS Chem Biol*, *16*(11), 2280–2296. https://doi.org/10.1021/acschembio.1c00456

Arutyunova, E., Khan, M. B., Fischer, C., Lu, J., Lamer, T., Vuong, W., van Belkum, M. J., McKay, R. T., Tyrrell, D. L., Vederas, J. C., Young, H. S., & Lemieux, M. J. (2021). N-Terminal Finger Stabilizes the S1 Pocket for the Reversible Feline Drug GC376 in the SARS-CoV-2 Mpro Dimer. *Journal of Molecular Biology*, *433*(13), 167003. https://doi.org/10.1016/j.jmb.2021.167003

Beach, J. R., & Schalm, O. W. (1936). A Filterable Virus, Distinct from that of Laryngotracheitis, the Cause of a Respiratory Disease of Chicks1. *Poultry Science*, *15*(3), 199–206. https://doi.org/10.3382/ps.0150199

Beauchamp, E., Yap, M. C., Iyer, A., Perinpanayagam, M. A., Gamma, J. M., Vincent, K. M., Lakshmanan, M., Raju, A., Tergaonkar, V., Tan, S. Y., Lim, S. T., Dong, W.-F., Postovit, L. M., Read, K. D., Gray, D. W., Wyatt, P. G., Mackey, J. R., & Berthiaume, L. G. (2020). Targeting N-myristoylation for therapy of B-cell lymphomas. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-18998-1

Bigbee, J. W. (2023). Cells of the Central Nervous System: An Overview of Their Structure and Function. In C.-L. Schengrund & R. K. Yu (Eds.), *Glycobiology of the Nervous System* (pp. 41–64). Springer International Publishing. https://doi.org/10.1007/978-3-031-12390-0_2

Bisgrove, D. A., Mahmoudi, T., Henklein, P., & Verdin, E. (2007). Conserved P-TEFb-interacting domain of BRD4 inhibits HIV transcription. *Proc Natl Acad Sci U S A*, *104*(34), 13690–13695. https://doi.org/10.1073/pnas.0705053104

Bjornevik, K., Cortese, M., Healy, B. C., Kuhle, J., Mina, M. J., Leng, Y., Elledge, S. J., Niebuhr, D. W., Scher, A. I., Munger, K. L., & Ascherio, A. (2022). Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science*, *375*(6578), 296–301. https://doi.org/10.1126/science.abj8222

Bohlson, S. S., O'Conner, S. D., Hulsebus, H. J., Ho, M.-M., & Fraser, D. A. (2014). Complement, C1q, and C1q-Related Molecules Regulate Macrophage Polarization. *Frontiers in Immunology*, *5*. https://www.frontiersin.org/articles/10.3389/fimmu.2014.00402

Bojkova, D., Klann, K., Koch, B., Widera, M., Krause, D., Ciesek, S., Cinatl, J., & Münch, C. (2020). Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature*, *583*(7816), Article 7816. https://doi.org/10.1038/s41586-020-2332-7

Boras, B., Jones, R. M., Anson, B. J., Arenson, D., Aschenbrenner, L., Bakowski, M. A., Beutler, N., Binder, J., Chen, E., Eng, H., Hammond, H., Hammond, J., Haupt, R. E., Hoffman, R., Kadar, E. P., Kania, R., Kimoto, E., Kirkpatrick, M. G., Lanyon, L., … Allerton, C. (2021). Preclinical characterization of an intravenous coronavirus 3CL protease inhibitor for the potential treatment of COVID19. *Nat Commun*, *12*(1), 6055. https://doi.org/10.1038/s41467-021-26239-2

Bosch, F. X., Lorincz, A., Muñoz, N., Meijer, C. J. L. M., & Shah, K. V. (2002). The causal relation between human papillomavirus and cervical cancer. *Journal of Clinical Pathology*, *55*(4), 244–265.

Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinović, S. M., Cheng, L.-Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., Vitek, O., Rinner, O., & Reiter, L. (2015). Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues *[S]. *Molecular & Cellular Proteomics*, *14*(5), 1400–1410. https://doi.org/10.1074/mcp.M114.044305

Bushnell, L. D., & Brandly, C. A. (1933). Laryngotracheitis in Chicks*. *Poultry Science*, *12*(1), 55–60. https://doi.org/10.3382/ps.0120055

Butterfield, D. A., & Halliwell, B. (2019). Oxidative stress, dysfunctional glucose metabolism and Alzheimer disease. *Nature Reviews Neuroscience*, *20*(3), Article 3. https://doi.org/10.1038/s41583-019-0132-6

CDC. (2020, March 28). *COVID Data Tracker*. Centers for Disease Control and Prevention. https://covid.cdc.gov/covid-data-tracker

Chailangkarn, T., Teeravechyan, S., Attasombat, K., Thaweerattanasinp, T., Sunchatawirul, K., Suwanwattana, P., Pongpirul, K., & Jongkaewwattana, A. (2022). Monkeypox virus productively infects human induced pluripotent stem cell-derived astrocytes and neural progenitor cells. *The Journal of Infection*, *85*(6), 702–769. https://doi.org/10.1016/j.jinf.2022.10.016

Chen, I. P., Longbotham, J. E., McMahon, S., Suryawanshi, R. K., Khalid, M. M., Taha, T. Y., Tabata, T., Hayashi, J. M., Soveg, F. W., Carlson-Stevermer, J., Gupta, M., Zhang, M. Y., Lam, V. L., Li, Y., Yu, Z., Titus, E. W., Diallo, A., Oki, J., Holden, K., … Ott, M.

(2022). Viral E protein neutralizes BET protein-mediated post-entry antagonism of SARS-CoV-2. *Cell Rep*, *40*(3), 111088. https://doi.org/10.1016/j.celrep.2022.111088

Chen, S. A., Arutyunova, E., Lu, J., Khan, M. B., Rut, W., Zmudzinski, M., Shahbaz, S., Iyyathurai, J., Moussa, E. W., Turner, Z., Bai, B., Lamer, T., Nieman, J. A., Vederas, J. C., Julien, O., Drag, M., Elahi, S., Young, H. S., & Lemieux, M. J. (2023). SARS-CoV-2 Mpro Protease Variants of Concern Display Altered Viral Substrate and Cell Host Target Galectin-8 Processing but Retain Sensitivity toward Antivirals. *ACS Central Science*, *9*(4), 696–708. https://doi.org/10.1021/acscentsci.3c00054

Chen, X., Yang, X., Zheng, Y., Yang, Y., Xing, Y., & Chen, Z. (2014). SARS coronavirus papain-like protease inhibits the type I interferon signaling pathway through interaction with the STING-TRAF3-TBK1 complex. *Protein & Cell*, *5*(5), 369–381. https://doi.org/10.1007/s13238-014-0026-3

Chiang, C., Liu, G., & Gack, M. U. (2021). Viral Evasion of RIG-I-Like Receptor-Mediated Immunity through Dysregulation of Ubiquitination and ISGylation. *Viruses*, *13*(2), Article 2. https://doi.org/10.3390/v13020182

Chuck, C.-P., Chong, L.-T., Chen, C., Chow, H.-F., Wan, D. C.-C., & Wong, K.-B. (2010). Profiling of Substrate Specificity of SARS-CoV 3CLpro. *PLOS ONE*, *5*(10), e13197. https://doi.org/10.1371/journal.pone.0013197

Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., & Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. *Nature Methods*, *6*(11), Article 11. https://doi.org/10.1038/nmeth1109-786

Crawford, E. D., Seaman, J. E., Agard, N., Hsu, G. W., Julien, O., Mahrus, S., Nguyen, H., Shimbo, K., Yoshihara, H. A., Zhuang, M., Chalkley, R. J., & Wells, J. A. (2013). The DegraBase: A database of proteolysis in healthy and apoptotic human cells. *Mol Cell Proteomics*, *12*(3), 813–824. https://doi.org/10.1074/mcp.O112.024372

Crowell, A. M. J., MacLellan, D. L., & Doucette, A. A. (2015). A two-stage spin cartridge for integrated protein precipitation, digestion and SDS removal in a comparative bottom-up proteomics workflow. *Journal of Proteomics*, *118*, 140–150. https://doi.org/10.1016/j.jprot.2014.09.030

Dao, C., & Zhang, D.-E. (2005). ISG15: A ubiquitin-like enigma. *Frontiers in Bioscience : A Journal and Virtual Library*, *10*, 2701–2722. https://doi.org/10.2741/1730

de Vries, M., Mohamed, A. S., Prescott, R. A., Valero-Jimenez, A. M., Desvignes, L., O'Connor, R., Steppan, C., Devlin, J. C., Ivanova, E., Herrera, A., Schinlever, A., Loose, P., Ruggles, K., Koralov, S. B., Anderson, A. S., Binder, J., & Dittmann, M. (2021). A comparative analysis of SARS-CoV-2 antivirals characterizes 3CL^{pro} inhibitor PF-00835231 as a potential new treatment for COVID-19. *J Virol*, JVI.01819-20. https://doi.org/10.1128/JVI.01819-20

Enose-Akahata, Y., Wang, L., Almsned, F., Johnson, K. R., Mina, Y., Ohayon, J., Wang, X. W., & Jacobson, S. (2023). The repertoire of CSF antiviral antibodies in patients with neuroinflammatory diseases. *Science Advances*, *9*(1), eabq6978. https://doi.org/10.1126/sciadv.abq6978

Feige, L., Kozaki, T., Dias de Melo, G., Guillemot, V., Larrous, F., Ginhoux, F., & Bourhy, H. (2022). Susceptibilities of CNS Cells towards Rabies Virus Infection Is Linked to Cellular Innate Immune Responses. *Viruses*, *15*(1), 88. https://doi.org/10.3390/v15010088

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., & Whitehouse, C. M. (1989). Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science*, *246*(4926), 64–71. https://doi.org/10.1126/science.2675315

Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J. P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Müller, S., Pawson, T., Gingras, A. C., Arrowsmith, C. H., & Knapp, S. (2012). Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell*, *149*(1), 214–231. https://doi.org/10.1016/j.cell.2012.02.013

Flather, D., Nguyen, J. H. C., Semler, B. L., & Gershon, P. D. (2018). Exploitation of nuclear functions by human rhinovirus, a cytoplasmic RNA virus. *PLoS Pathog*, *14*(8), e1007277. https://doi.org/10.1371/journal.ppat.1007277

Fooks, A. R., Cliquet, F., Finke, S., Freuling, C., Hemachudha, T., Mani, R. S., Müller, T., Nadin-Davis, S., Picard-Meyer, E., Wilde, H., & Banyard, A. C. (2017). Rabies. *Nature Reviews Disease Primers*, *3*(1), Article 1. https://doi.org/10.1038/nrdp.2017.91

Fortelny, N., Yang, S., Pavlidis, P., Lange, P. F., & Overall, C. M. (2015). 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*, *43*(Database issue), D290-7. https://doi.org/10.1093/nar/gku1012

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. (2020). Characterization and Noncovalent Inhibition of the Deubiquitinase and deISGylase Activity of SARS-CoV-2 Papain-Like Protease. *ACS Infect Dis*, *6*(8), 2099–2109. https://doi.org/10.1021/acsinfecdis.0c00168

Gao, X., Qin, B., Chen, P., Zhu, K., Hou, P., & Wojdyla…, J. A. (2021). Crystal structure of SARS-CoV-2 papain-like protease. *… Pharmaceutica Sinica B*.

Glish, G. L., & Vachet, R. W. (2003). The basics of mass spectrometry in the twenty-first century. *Nature Reviews Drug Discovery*, *2*(2), Article 2. https://doi.org/10.1038/nrd1011

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., O'Meara, M. J., Rezelj, V. V., Guo, J. Z., Swaney, D. L., Tummino, T. A., Hüttenhain, R., Kaake, R. M., Richards, A. L., Tutuncuoglu, B., Foussard, H., Batra, J., Haas, K., Modak, M., … Krogan, N. J. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature*, *583*(7816), Article 7816. https://doi.org/10.1038/s41586-020-2286-9

Greco-Stewart, V. S., Thibault, C. S.-L., & Pelchat, M. (2006). Binding of the polypyrimidine tract-binding protein-associated splicing factor (PSF) to the hepatitis delta virus RNA. *Virology*, *356*(1), 35–44. https://doi.org/10.1016/j.virol.2006.06.040

Hamre, D., & Beem, M. (1972). *VIROLOGIC STUDIES OF ACUTE RESPIRATORY DISEASE IN YOUNG ADULTS: V. CORONAVIRUS 229E INFECTIONS DURING SIX YEARS OF SURVEILLANCE*. *96*(2), 94–106.

Hamre, D., & Procknow, J. J. (1966). A New Virus Isolated from the Human Respiratory Tract. *Proceedings of the Society for Experimental Biology and Medicine*, *121*(1), 190–193. https://doi.org/10.3181/00379727-121-30734

Hara, H., Chida, J., Uchiyama, K., Pasiana, A. D., Takahashi, E., Kido, H., & Sakaguchi, S. (2021). Neurotropic influenza A virus infection causes prion protein misfolding into infectious prions in neuroblastoma cells. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-89586-6

Hassani, A., Corboy, J. R., Al-Salam, S., & Khan, G. (2018). Epstein-Barr virus is present in the brain of most cases of multiple sclerosis and may engage more than just B cells. *PLOS ONE*, *13*(2), e0192109. https://doi.org/10.1371/journal.pone.0192109

He, W., Wan, H., Hu, L., Chen, P., Wang, X., Huang, Z., Yang, Z.-H., Zhong, C.-Q., & Han, J. (2015). Gasdermin D is an executor of pyroptosis and required for interleukin-1β secretion. *Cell Research*, *25*(12), Article 12. https://doi.org/10.1038/cr.2015.139

Hensley, K., Hall, N., Subramaniam, R., Cole, P., Harris, M., Aksenov, M., Aksenova, M., Gabbita, S. P., Wu, J. F., Carney, J. M., Lovell, M., Markesbery, W. R., & Butterfield, D. A. (1995). Brain Regional Correspondence Between Alzheimer's Disease Histopathology and Biomarkers of Protein Oxidation. *Journal of Neurochemistry*, *65*(5), 2146–2156. https://doi.org/10.1046/j.1471-4159.1995.65052146.x

Hill, M. E., Kumar, A., Wells, J. A., Hobman, T. C., Julien, O., & Hardy, J. A. (2018). The Unique Cofactor Region of Zika Virus NS2B–NS3 Protease Facilitates Cleavage of Key Host Proteins. *ACS Chemical Biology*, *13*(9), 2398–2405. https://doi.org/10.1021/acschembio.8b00508

Hooijmans, C. R., Hlavica, M., Schuler, F. A. F., Good, N., Good, A., Baumgartner, L., Galeno, G., Schneider, M. P., Jung, T., de Vries, R., & Ineichen, B. V. (2019). Remyelination promoting therapies in multiple sclerosis animal models: A systematic review and meta-analysis. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-018-35734-4

Hu, A., Noble, W. S., & Wolf-Yadlin, A. (2016). Technical advances in proteomics: New developments in data-independent acquisition. *F1000Research*, *5*, F1000 Faculty Rev-419. https://doi.org/10.12688/f1000research.7042.1

Hyer, M. L., Milhollen, M. A., Ciavarri, J., Fleming, P., Traore, T., Sappal, D., Huck, J., Shi, J., Gavin, J., Brownell, J., Yang, Y., Stringer, B., Griffin, R., Bruzzese, F., Soucy, T., Duffy, J., Rabino, C., Riceberg, J., Hoar, K., … Bence, N. F. (2018). A small-molecule inhibitor of the ubiquitin activating enzyme for cancer treatment. *Nature Medicine*, *24*(2), Article 2. https://doi.org/10.1038/nm.4474

Jacobs, B. L., Langland, J. O., Kibler, K. V., Denzler, K. L., White, S. D., Holechek, S. A., Wong, S., Huynh, T., & Baskin, C. R. (2009). Vaccinia virus vaccines: Past, present and future. *Antiviral Research*, *84*(1), 1–13. https://doi.org/10.1016/j.antiviral.2009.06.006

Jagdeo, J. M., Dufour, A., Klein, T., Solis, N., Kleifeld, O., Kizhakkedathu, J., Luo, H., Overall, C. M., & Jan, E. (2018). N-Terminomics TAILS Identifies Host Cell Substrates of Poliovirus and Coxsackievirus B3 3C Proteinases That Modulate Virus Infection. *J Virol*, *92*(8), e02211-17. https://doi.org/10.1128/JVI.02211-17

Jha, H. C., Mehta, D., Lu, J., El-Naccache, D., Shukla, S. K., Kovacsics, C., Kolson, D., & Robertson, E. S. (2015). Gammaherpesvirus Infection of Human Neuronal Cells. *MBio*, *6*(6), e01844-15. https://doi.org/10.1128/mBio.01844-15

Karussis, D. (2014). The diagnosis of multiple sclerosis and the various related demyelinating syndromes: A critical review. *Journal of Autoimmunity*, *48–49*, 134–142. https://doi.org/10.1016/j.jaut.2014.01.022

Kayagaki, N., Stowe, I. B., Lee, B. L., O'Rourke, K., Anderson, K., Warming, S., Cuellar, T., Haley, B., Roose-Girma, M., Phung, Q. T., Liu, P. S., Lill, J. R., Li, H., Wu, J., Kummerfeld, S., Zhang, J., Lee, W. P., Snipas, S. J., Salvesen, G. S., … Dixit, V. M. (2015). Caspase-11 cleaves gasdermin D for non-canonical inflammasome signalling. *Nature*, *526*(7575), 666–671. https://doi.org/10.1038/nature15541

Kelstrup, C. D., Bekker-Jensen, D. B., Arrey, T. N., Hogrebe, A., Harder, A., & Olsen, J. V. (2018). Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics. *Journal of Proteome Research*, *17*(1), 727–738. https://doi.org/10.1021/acs.jproteome.7b00602

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. (2010). Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat Biotechnol*, *28*(3), 281–288. https://doi.org/10.1038/nbt.1611

Komander, D., & Rape, M. (2012). The Ubiquitin Code. *Annual Review of Biochemistry*, *81*(1), 203–229. https://doi.org/10.1146/annurev-biochem-060310-170328

Koudelka, T., Boger, J., Henkel, A., Schönherr, R., Krantz, S., Fuchs, S., Rodríguez, E., Redecke, L., & Tholey, A. (2021). N-Terminomics for the Identification of In Vitro Substrates and Cleavage Site Specificity of the SARS-CoV-2 Main Protease. *Proteomics*, *21*(2), e2000246. https://doi.org/10.1002/pmic.202000246

Kumar, A., Ishida, R., Strilets, T., Cole, J., Lopez-Orozco, J., Fayad, N., Felix-Lopez, A., Elaish, M., Evseev, D., Magor, K. E., Mahal, L. K., Nagata, L. P., Evans, D. H., & Hobman, T. C. (2021). SARS-CoV-2 Nonstructural Protein 1 Inhibits the Interferon Response by Causing Depletion of Key Host Signaling Factors. *J Virol*, *95*(13), e0026621. https://doi.org/10.1128/JVI.00266-21

Landeras-Bueno, S., Jorba, N., Pérez-Cidoncha, M., & Ortín, J. (2011). The Splicing Factor Proline-Glutamine Rich (SFPQ/PSF) Is Involved in Influenza Virus Transcription. *PLOS Pathogens*, *7*(11), e1002397. https://doi.org/10.1371/journal.ppat.1002397

Lansiaux, E., Jain, N., Laivacuma, S., & Reinis, A. (2022). The virology of human monkeypox virus (hMPXV): A brief overview. *Virus Research*, *322*, 198932. https://doi.org/10.1016/j.virusres.2022.198932

Lee, J. W., Beebe, K., Nangle, L. A., Jang, J., Longo-Guess, C. M., Cook, S. A., Davisson, M. T., Sundberg, J. P., Schimmel, P., & Ackerman, S. L. (2006). Editing-defective tRNA synthetase causes protein misfolding and neurodegeneration. *Nature*, *443*(7107), Article 7107. https://doi.org/10.1038/nature05096

Lee, J. Y., Kim, D. G., Kim, B.-G., Yang, W. S., Hong, J., Kang, T., Oh, Y. S., Kim, K. R., Han, B. W., Hwang, B. J., Kang, B. S., Kang, M.-S., Kim, M.-H., Kwon, N. H., & Kim, S. (2014). Promiscuous methionyl-tRNA synthetase mediates adaptive mistranslation to protect cells against oxidative stress. *Journal of Cell Science*, *127*(19), 4234–4245. https://doi.org/10.1242/jcs.152470

Levine, R. L., Mosoni, L., Berlett, B. S., & Stadtman, E. R. (1996). Methionine residues as endogenous antioxidants in proteins. *Proceedings of the National Academy of Sciences*, *93*(26), 15036–15040. https://doi.org/10.1073/pnas.93.26.15036

Liu, G., Lee, J.-H., Parker, Z. M., Acharya, D., Chiang, J. J., van Gent, M., Riedl, W., Davis-Gardner, M. E., Wies, E., Chiang, C., & Gack, M. U. (2021). ISG15-dependent activation of the sensor MDA5 is antagonized by the SARS-CoV-2 papain-like protease to evade host innate immunity. *Nature Microbiology*, *6*(4), Article 4. https://doi.org/10.1038/s41564-021-00884-1

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., … Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *The Lancet*, *395*(10224), 565–574. https://doi.org/10.1016/S0140-6736(20)30251-8

Ludwig, K. R., Schroll, M. M., & Hummon, A. B. (2018). Comparison of In-Solution, FASP, and S-Trap Based Digestion Methods for Bottom-Up Proteomic Studies. *Journal of Proteome Research*, *17*(7), 2480–2490. https://doi.org/10.1021/acs.jproteome.8b00235

Lukong, K. E., Huot, M.-É., & Richard, S. (2009). BRK phosphorylates PSF promoting its cytoplasmic localization and cell cycle arrest. *Cellular Signalling*, *21*(9), 1415–1422. https://doi.org/10.1016/j.cellsig.2009.04.008

Luo, S., & Levine, R. L. (2009). Methionine in proteins defends against oxidative stress. *The FASEB Journal*, *23*(2), 464–472. https://doi.org/10.1096/fj.08-118414

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. (2023). Identification of Human Host Substrates of the SARS-CoV-2 Mpro and PLpro Using Subtiligase N-Terminomics. *ACS Infectious Diseases*, *9*(4), 749–761. https://doi.org/10.1021/acsinfecdis.2c00458

Ma, Y., Wang, X., Luo, W., Xiao, J., Song, X., Wang, Y., Shuai, H., Ren, Z., & Wang, Y. (2021). Roles of Emerging RNA-Binding Activity of cGAS in Innate Antiviral Response. *Frontiers in Immunology*, *12*, 741599. https://doi.org/10.3389/fimmu.2021.741599

Mac Kain, A., Maarifi, G., Aicher, S. M., Arhel, N., Baidaliuk, A., Munier, S., Donati, F., Vallet, T., Tran, Q. D., Hardy, A., Chazal, M., Porrot, F., OhAinle, M., Carlson-Stevermer, J., Oki, J., Holden, K., Zimmer, G., Simon-Lorière, E., Bruel, T., … Roesch, F. (2022). Identification of DAXX as a restriction factor of SARS-CoV-2 through a CRISPR/Cas9 screen. *Nat Commun*, *13*(1), 2442. https://doi.org/10.1038/s41467-022-30134-9

Mahrus, S., Kisiel, W., & Craik, C. S. (2004). Granzyme M is a regulatory protease that inactivates proteinase inhibitor 9, an endogenous inhibitor of granzyme B. *J Biol Chem*, *279*(52), 54275–54282. https://doi.org/10.1074/jbc.M411482200

May, D. G., Martin-Sancho, L., Anschau, V., Liu, S., Chrisopulos, R. J., Scott, K. L., Halfmann, C. T., Díaz Peña, R., Pratt, D., Campos, A. R., & Roux, K. J. (2022). A BioID-Derived Proximity Interactome for SARS-CoV-2 Proteins. *Viruses*, *14*(3), Article 3. https://doi.org/10.3390/v14030611

McDonald, L., Robertson, D. H. L., Hurst, J. L., & Beynon, R. J. (2005). Positional proteomics: Selective recovery and analysis of N-terminal proteolytic peptides. *Nature Methods*, *2*(12), Article 12. https://doi.org/10.1038/nmeth811

McIntosh, K., Becker, W. B., & Chanock, R. M. (1967). Growth in suckling-mouse brain of “IBV-like” viruses from patients with upper respiratory tract disease. *Proceedings of the*

National Academy of Sciences, *58*(6), 2268–2273. https://doi.org/10.1073/pnas.58.6.2268

McIntosh, K., Dees, J. H., Becker, W. B., Kapikian, A. Z., & Chanock, R. M. (1967). Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proceedings of the National Academy of Sciences*, *57*(4), 933–940. https://doi.org/10.1073/pnas.57.4.933

McKenzie, B. A., Dixit, V. M., & Power, C. (2020). Fiery Cell Death: Pyroptosis in the Central Nervous System. *Trends in Neurosciences*, *43*(1), 55–73. https://doi.org/10.1016/j.tins.2019.11.005

McKenzie, B. A., Fernandes, J. P., Doan, M. A. L., Schmitt, L. M., Branton, W. G., & Power, C. (2020). Activation of the executioner caspases-3 and -7 promotes microglial pyroptosis in models of multiple sclerosis. *Journal of Neuroinflammation*, *17*(1), 253. https://doi.org/10.1186/s12974-020-01902-5

McLuckey, S. A. (1992). Principles of collisional activation in analytical mass spectrometry. *Journal of the American Society for Mass Spectrometry*, *3*(6), 599–614. https://doi.org/10.1016/1044-0305(92)85001-Z

McNamara, E. (2022). *Tau seeding in a HEK cell model is influenced by aggregation inducers and cellular chaperones*. https://doi.org/10.7939/r3-p15t-d551

Mehta, D., Ahkami, A. H., Walley, J., Xu, S.-L., & Uhrig, R. G. (2022). The incongruity of validating quantitative proteomics using western blots. *Nature Plants*, *8*(12), Article 12. https://doi.org/10.1038/s41477-022-01314-8

Menet, A., Speth, C., Larcher, C., Prodinger, W. M., Schwendinger, M. G., Chan, P., Jäger, M., Schwarzmann, F., Recheis, H., Fontaine, M., & Dierich, M. P. (1999). Epstein-Barr Virus Infection of Human Astrocyte Cell Lines. *Journal of Virology*, *73*(9), 7722–7733.

Meyer, B., Chiaravalli, J., Gellenoncourt, S., Brownridge, P., Bryne, D. P., Daly, L. A., Grauslys, A., Walter, M., Agou, F., Chakrabarti, L. A., Craik, C. S., Eyers, C. E., Eyers, P. A., Gambin, Y., Jones, A. R., Sierecki, E., Verdin, E., Vignuzzi, M., & Emmott, E. (2021). Characterising proteolysis during SARS-CoV-2 infection identifies viral cleavage sites and cellular targets with therapeutic potential. *Nat Commun*, *12*(1), 5553. https://doi.org/10.1038/s41467-021-25796-w

Miczi, M., Golda, M., Kunkli, B., Nagy, T., Tőzsér, J., & Mótyán, J. A. (2020). Identification of Host Cellular Protein Substrates of SARS-COV-2 Main Protease. *Int J Mol Sci*, *21*(24), E9523. https://doi.org/10.3390/ijms21249523

Mitchell Wells, J., & McLuckey, S. A. (2005). Collision-Induced Dissociation (CID) of Peptides and Proteins. In *Methods in Enzymology* (Vol. 402, pp. 148–185). Academic Press. https://doi.org/10.1016/S0076-6879(05)02005-7

Mohamud, Y., Xue, Y. C., Liu, H., Ng, C. S., Bahreyni, A., Jan, E., & Luo, H. (2021). The papain-like protease of coronaviruses cleaves ULK1 to disrupt host autophagy. *Biochem Biophys Res Commun*, *540*, 75–82. https://doi.org/10.1016/j.bbrc.2020.12.091

Mohler, K., & Ibba, M. (2017). Translational fidelity and mistranslation in the cellular response to stress. *Nature Microbiology*, *2*(9), Article 9. https://doi.org/10.1038/nmicrobiol.2017.117

Morchikh, M., Cribier, A., Raffel, R., Amraoui, S., Cau, J., Severac, D., Dubois, E., Schwartz, O., Bennasser, Y., & Benkirane, M. (2017). HEXIM1 and NEAT1 Long Non-coding RNA Form a Multi-subunit Complex that Regulates DNA-Mediated Innate Immune Response. *Molecular Cell*, *67*(3), 387-399.e5. https://doi.org/10.1016/j.molcel.2017.06.020

Moreno, M. A., Or-Geva, N., Aftab, B. T., Khanna, R., Croze, E., Steinman, L., & Han, M. H. (2018). Molecular signature of Epstein-Barr virus infection in MS brain lesions. *Neurology - Neuroimmunology Neuroinflammation*, *5*(4). https://doi.org/10.1212/NXI.0000000000000466

Moustaqil, M., Ollivier, E., Chiu, H. P., Van Tol, S., Rudolffi-Soto, P., Stevens, C., Bhumkar, A., Hunter, D. J. B., Freiberg, A. N., Jacques, D., Lee, B., Sierecki, E., & Gambin, Y. (2021). 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*, *10*(1), 178–195. https://doi.org/10.1080/22221751.2020.1870414

Mpox (monkeypox). (2023, April 18). https://www.who.int/news-room/fact-sheets/detail/monkeypox

MS Society of Canada. (2020). *Atlas of MS Report Shows 2.8 Million People Worldwide Live with Multiple Sclerosis—MS Society of Canada*. MS Society of Canada. https://mssociety.ca/resources/news/article/atlas-of-ms-report-shows-28-million-people-worldwide-live-with-multiple-sclerosis

Na, H., Acharjee, S., Jones, G., Vivithanaporn, P., Noorbakhsh, F., McFarlane, N., Maingat, F., Ballanyi, K., Pardo, C. A., Cohen, É. A., & Power, C. (2011). Interactions between human immunodeficiency virus (HIV)-1 Vpr expression and innate immunity influence neurovirulence. *Retrovirology*, *8*(1), 44. https://doi.org/10.1186/1742-4690-8-44

Narasimhan, J., Wang, M., Fu, Z., Klein, J. M., Haas, A. L., & Kim, J.-J. P. (2005). Crystal Structure of the Interferon-induced Ubiquitin-like Protein ISG15 *. *Journal of Biological Chemistry*, *280*(29), 27356–27365. https://doi.org/10.1074/jbc.M502814200

Netzer, N., Goodenbour, J. M., David, A., Dittmar, K. A., Jones, R. B., Schneider, J. R., Boone, D., Eves, E. M., Rosner, M. R., Gibbs, J. S., Embry, A., Dolan, B., Das, S., Hickman, H. D., Berglund, P., Bennink, J. R., Yewdell, J. W., & Pan, T. (2009). Innate immune and chemically triggered oxidative stress modifies translational fidelity. *Nature*, *462*(7272), Article 7272. https://doi.org/10.1038/nature08576

Oien, D. B., & Moskovitz, J. (2007). Substrates of the Methionine Sulfoxide Reductase System and Their Physiological Relevance. In *Current Topics in Developmental Biology* (Vol. 80, pp. 93–133). Academic Press. https://doi.org/10.1016/S0070-2153(07)80003-2

Olsson, T., Barcellos, L. F., & Alfredsson, L. (2017). Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nature Reviews Neurology*, *13*(1), Article 1. https://doi.org/10.1038/nrneurol.2016.187

Ong, S.-E., & Mann, M. (2006). A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature Protocols*, *1*(6), Article 6. https://doi.org/10.1038/nprot.2006.427

Oransky, I. (2005). David Tyrrell, Obituary. *The Lancet*, *365*(9477), 2084. https://doi.org/10.1016/S0140-6736(05)66722-0

Owen, D. R., Allerton, C. M. N., Anderson, A. S., Aschenbrenner, L., Avery, M., Berritt, S., Boras, B., Cardin, R. D., Carlo, A., Coffman, K. J., Dantonio, A., Di, L., Eng, H., Ferre, R., Gajiwala, K. S., Gibson, S. A., Greasley, S. E., Hurst, B. L., Kadar, E. P., … Zhu, Y. (2021). An oral SARS-CoV-2 M^{pro} inhibitor clinical candidate for the treatment of COVID-19. *Science*, *374*(6575), 1586–1593. https://doi.org/10.1126/science.abl4784

Pablos, I., Machado, Y., de Jesus, H. C. R., Mohamud, Y., Kappelhoff, R., Lindskog, C., Vlok, M., Bell, P. A., Butler, G. S., Grin, P. M., Cao, Q. T., Nguyen, J. P., Solis, N., Abbina, S., Rut, W., Vederas, J. C., Szekely, L., Szakos, A., Drag, M., … Overall, C. M. (2021).

Mechanistic insights into COVID-19 by global analysis of the SARS-CoV-2 3CLpro substrate degradome. *Cell Reports*, *37*(4), 109892. https://doi.org/10.1016/j.celrep.2021.109892

Pacylex Pharmaceuticals. (2022). *Phase I Trial of PCLX-001 in Relapsed/Refractory B-cell Non-Hodgkin Lymphoma and Advanced Solid Malignancies* (Clinical Trial Registration NCT04836195.). clinicaltrials.gov. https://clinicaltrials.gov/ct2/show/NCT04836195.

Peng, J., & Gygi, S. P. (2001). Proteomics: The move to mixtures. *Journal of Mass Spectrometry*, *36*(10), 1083–1091. https://doi.org/10.1002/jms.229

Perng, Y.-C., & Lenschow, D. J. (2018). ISG15 in antiviral immunity and beyond. *Nature Reviews Microbiology*, *16*(7), Article 7. https://doi.org/10.1038/s41579-018-0020-5

Petersen, L. R., Brault, A. C., & Nasci, R. S. (2013). West Nile Virus: Review of the Literature. *JAMA*, *310*(3), 308–315. https://doi.org/10.1001/jama.2013.8042

Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., & Coon, J. J. (2012). Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics *. *Molecular & Cellular Proteomics*, *11*(11), 1475–1488. https://doi.org/10.1074/mcp.O112.020131

Petzold, A. (2015). Glial fibrillary acidic protein is a body fluid biomarker for glial pathology in human disease. *Brain Research*, *1600*, 17–31. https://doi.org/10.1016/j.brainres.2014.12.027

Poreba, M., Salvesen, G. S., & Drag, M. (2017). Synthesis of a HyCoSuL peptide substrate library to dissect protease substrate specificity. *Nat Protoc*, *12*(10), 2189–2214. https://doi.org/10.1038/nprot.2017.091

Quick, E. D., Leser, J. S., Clarke, P., & Tyler, K. L. (2014). Activation of Intrinsic Immune Responses and Microglial Phagocytosis in an Ex Vivo Spinal Cord Slice Culture Model of West Nile Virus Infection. *Journal of Virology*, *88*(22), 13005–13014. https://doi.org/10.1128/JVI.01994-14

Ramaswamy, V., Walsh, J. G., Sinclair, D. B., Johnson, E., Tang-Wai, R., Wheatley, B. M., Branton, W., Maingat, F., Snyder, T., Gross, D. W., & Power, C. (2013). Inflammasome induction in Rasmussen's encephalitis: Cortical and associated white matter

pathogenesis. *Journal of Neuroinflammation*, *10*(1), 918. https://doi.org/10.1186/1742-2094-10-152

Rebendenne, A., Roy, P., Bonaventure, B., Chaves Valadão, A. L., Desmarets, L., Arnaud-Arnould, M., Rouillé, Y., Tauziet, M., Giovannini, D., Touhami, J., Lee, Y., DeWeirdt, P., Hegde, M., Urbach, S., Koulali, K. E., de Gracia, F. G., McKellar, J., Dubuisson, J., Wencker, M., … Goujon, C. (2022). Bidirectional genome-wide CRISPR screens reveal host factors regulating SARS-CoV-2, MERS-CoV and seasonal HCoVs. *Nat Genet*, *54*(8), 1090–1102. https://doi.org/10.1038/s41588-022-01110-2

Reynolds, N. D., Aceves, N. M., Liu, J. L., Compton, J. R., Leary, D. H., Freitas, B. T., Pegan, S. D., Doctor, K. Z., Wu, F. Y., Hu, X., & Legler, P. M. (2021). The SARS-CoV-2 SSHHPS Recognized by the Papain-like Protease. *ACS Infect Dis*, *7*(6), 1483–1502. https://doi.org/10.1021/acsinfecdis.0c00866

Roczkowsky, A., Doan, M. A. L., Hlavay, B. A., Mamik, M. K., Branton, W. G., McKenzie, B. A., Saito, L. B., Schmitt, L., Eitzen, G., Di Cara, F., Wuest, M., Wuest, F., Rachubinski, R., & Power, C. (2022). Peroxisome Injury in Multiple Sclerosis: Protective Effects of 4-Phenylbutyrate in CNS-Associated Macrophages. *The Journal of Neuroscience*, *42*(37), 7152–7165. https://doi.org/10.1523/JNEUROSCI.0312-22.2022

Rut, W., Groborz, K., Zhang, L., Sun, X., Zmudzinski, M., Pawlik, B., Wang, X., Jochmans, D., Neyts, J., Młynarski, W., Hilgenfeld, R., & Drag, M. (2021). SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging. *Nature Chemical Biology*, *17*(2), Article 2. https://doi.org/10.1038/s41589-020-00689-z

Rut, W., Lv, Z., Zmudzinski, M., Patchett, S., Nayak, D., Snipas, S. J., El Oualid, F., Huang, T. T., Bekes, M., Drag, M., & Olsen, S. K. (2020). Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti–COVID-19 drug design. *Science Advances*, *6*(42), eabd4596. https://doi.org/10.1126/sciadv.abd4596

Saito, L. B., Fernandes, J. P., Smith, M. J., Doan, M. A. L., Branton, W. G., Schmitt, L. M., Wuest, M., Monaco, M. C., Major, E. O., Wuest, F., & Power, C. (2021). Intranasal anti-caspase-1 therapy preserves myelin and glucose metabolism in a model of progressive multiple sclerosis. *Glia*, *69*(1), 216–229. https://doi.org/10.1002/glia.23896

Samelson, A. J., Tran, Q. D., Robinot, R., Carrau, L., Rezelj, V. V., Kain, A. M., Chen, M., Ramadoss, G. N., Guo, X., Lim, S. A., Lui, I., Nuñez, J. K., Rockwood, S. J., Wang, J.,

Liu, N., Carlson-Stevermer, J., Oki, J., Maures, T., Holden, K., … Kampmann, M. (2022). BRD2 inhibition blocks SARS-CoV-2 infection by reducing transcription of the host cell receptor ACE2. *Nat Cell Biol*, *24*(1), 24–34. https://doi.org/10.1038/s41556-021-00821-8

Sargsyan, S. A., Shearer, A. J., Ritchie, A. M., Burgoon, M. P., Anderson, S., Hemmer, B., Stadelmann, C., Gattenlöhner, S., Owens, G. P., Gilden, D., & Bennett, J. L. (2010). Absence of Epstein-Barr virus in the brain and CSF of patients with multiple sclerosis(e–Pub ahead of print). *Neurology*, *74*(14), 1127–1135. https://doi.org/10.1212/WNL.0b013e3181d865a1

Schalk, A. F., & Hawn, M. C. (1931). *An apparently new respiratory disease of baby chicks*. *78*, 413–422.

Schechter, I., & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications*, *27*(2), 157–162. https://doi.org/10.1016/S0006-291X(67)80055-X

Sepehrinezhad, A., Ashayeri Ahmadabad, R., & Sahab-Negah, S. (2023). Monkeypox virus from neurological complications to neuroinvasive properties: Current status and future perspectives. *Journal of Neurology*, *270*(1), 101–108. https://doi.org/10.1007/s00415-022-11339-w

Serafini, B., Rosicarelli, B., Franciotta, D., Magliozzi, R., Reynolds, R., Cinque, P., Andreoni, L., Trivedi, P., Salvetti, M., Faggioni, A., & Aloisi, F. (2007). Dysregulated Epstein-Barr virus infection in the multiple sclerosis brain. *Journal of Experimental Medicine*, *204*(12), 2899–2912. https://doi.org/10.1084/jem.20071030

Shalini, S., Dorstyn, L., Dawar, S., & Kumar, S. (2015). Old, new and emerging functions of caspases. *Cell Death & Differentiation*, *22*(4), Article 4. https://doi.org/10.1038/cdd.2014.216

Shevchenko, A., Tomas, H., Havli, J., Olsen, J. V., & Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols*, *1*(6), Article 6. https://doi.org/10.1038/nprot.2006.468

Shi, J., Zhao, Y., Wang, K., Shi, X., Wang, Y., Huang, H., Zhuang, Y., Cai, T., Wang, F., & Shao, F. (2015). Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. *Nature*, *526*(7575), Article 7575. https://doi.org/10.1038/nature15514

Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., Schulz, L., Widera, M., Mehdipour, A. R., Tascher, G., Geurink, P. P., Wilhelm, A., van der Heden van Noort, G. J., Ovaa, H., Müller, S., Knobeloch, K.-P., Rajalingam, K., Schulman, B. A., Cinatl, J., … Dikic, I. (2020). Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature*, *587*(7835), Article 7835. https://doi.org/10.1038/s41586-020-2601-5

Shrestha, B., Gottlieb, D., & Diamond, M. S. (2003). Infection and Injury of Neurons by West Nile Encephalitis Virus. *Journal of Virology*, *77*(24), 13203–13213. https://doi.org/10.1128/JVI.77.24.13203-13213.2003

Siddiqui, S., Alhamdi, H. W. S., & Alghamdi, H. A. (2022). Recent Chronology of COVID-19 Pandemic. *Frontiers in Public Health*, *10*. https://www.frontiersin.org/articles/10.3389/fpubh.2022.778037

Steinman, L. (2001). Multiple sclerosis: A two-stage disease. *Nature Immunology*, *2*(9), Article 9. https://doi.org/10.1038/ni0901-762

Stukalov, A., Girault, V., Grass, V., Karayel, O., Bergant, V., Urban, C., Haas, D. A., Huang, Y., Oubraham, L., Wang, A., Hamad, M. S., Piras, A., Hansen, F. M., Tanzer, M. C., Paron, I., Zinzula, L., Engleitner, T., Reinecke, M., Lavacca, T. M., … Pichlmair, A. (2021). Multilevel proteomics reveals host perturbations by SARS-CoV-2 and SARS-CoV. *Nature*, *594*(7862), Article 7862. https://doi.org/10.1038/s41586-021-03493-4

Sun, L., Xing, Y., Chen, X., Zheng, Y., Yang, Y., Nichols, D. B., Clementz, M. A., Banach, B. S., Li, K., Baker, S. C., & Chen, Z. (2012). Coronavirus Papain-like Proteases Negatively Regulate Antiviral Innate Immune Response through Disruption of STING-Mediated Signaling. *PLOS ONE*, *7*(2), e30802. https://doi.org/10.1371/journal.pone.0030802

Tallantyre, E. C., Bø, L., Al-Rawashdeh, O., Owens, T., Polman, C. H., Lowe, J. S., & Evangelou, N. (2010). Clinico-pathological evidence that axonal loss underlies disability in progressive multiple sclerosis. *Multiple Sclerosis Journal*, *16*(4), 406–411. https://doi.org/10.1177/1352458510364992

Tavčar, P., Potokar, M., Kolenc, M., Korva, M., Avšič-Županc, T., Zorec, R., & Jorgačevski, J. (2021). Neurotropic Viruses, Astrocytes, and COVID-19. *Frontiers in Cellular Neuroscience*, *15*. https://www.frontiersin.org/articles/10.3389/fncel.2021.662578

Thacker, E. L., Mirzaei, F., & Ascherio, A. (2006). Infectious mononucleosis and risk for multiple sclerosis: A meta-analysis. *Annals of Neurology*, *59*(3), 499–503. https://doi.org/10.1002/ana.20820

Thorley-Lawson, D. A. (2001). Epstein-Barr virus: Exploiting the immune system. *Nature Reviews Immunology*, *1*(1), Article 1. https://doi.org/10.1038/35095584

Tönnies, E., & Trushina, E. (2017). Oxidative Stress, Synaptic Dysfunction, and Alzheimer's Disease. *Journal of Alzheimer's Disease*, *57*(4), 1105–1121. https://doi.org/10.3233/JAD-161088

Tsu, B. V., Agarwal, R., Gokhale, N. S., Kulsuptrakul, J., Ryan, A. P., Castro, L. K., Beierschmitt, C. M., Turcotte, E. A., Fay, E. J., Vance, R. E., Hyde, J. L., Savan, R., Mitchell, P. S., & Daugherty, M. D. (2022). Host specific sensing of coronaviruses and picornaviruses by the CARD8 inflammasome. *BioRxiv*, 2022.09.21.508960. https://doi.org/10.1101/2022.09.21.508960

Tyrrell, D. (1990). The Origins of the Common Cold Unit. *Journal of the Royal College of Physicians of London*, *24*(2), 137–140.

Tyrrell, D. A. J., & Bynoe, M. L. (1965). Cultivation of a Novel Type of Common-cold Virus in Organ Cultures. *British Medical Journal*, *1*(5448), 1467–1470.

Ullrich, S., & Nitsche, C. (2020). The SARS-CoV-2 main protease as drug target. *Bioorganic & Medicinal Chemistry Letters*, *30*(17), 127377. https://doi.org/10.1016/j.bmcl.2020.127377

Viejo-Borbolla, A., Ottinger, M., Brüning, E., Bürger, A., König, R., Kati, E., Sheldon, J. A., & Schulz, T. F. (2005). 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*, *79*(21), 13618–13629. https://doi.org/10.1128/JVI.79.21.13618-13629.2005

Virology: Coronaviruses. (1968). *Nature*, *220*(5168), 650. https://doi.org/10.1038/220650b0

Vuong, W., Khan, M. B., Fischer, C., Arutyunova, E., Lamer, T., Shields, J., Saffran, H. A., McKay, R. T., van Belkum, M. J., Joyce, M. A., Young, H. S., Tyrrell, D. L., Vederas, J. C., & Lemieux, M. J. (2020). Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat Commun*, *11*(1), 4282. https://doi.org/10.1038/s41467-020-18096-2

Wang, L., Fan, X., Bonenfant, G., Cui, D., Hossain, J., Jiang, N., Larson, G., Currier, M., Liddell, J., Wilson, M., Tamin, A., Harcourt, J., Ciomperlik-Patton, J., Pang, H., Dybdahl-Sissoko, N., Campagnoli, R., Shi, P. Y., Barnes, J., Thornburg, N. J., … Zhou, B. (2021). Susceptibility to SARS-CoV-2 of Cell Lines and Substrates Commonly Used to Diagnose and Isolate Influenza and Other Viruses. *Emerg Infect Dis*, *27*(5), 1380–1392. https://doi.org/10.3201/eid2705.210023

Weeks, A. M., & Wells, J. A. (2018). Engineering peptide ligase specificity by proteomic identification of ligation sites. *Nature Chemical Biology*, *14*(1), Article 1. https://doi.org/10.1038/nchembio.2521

Wenzel, J., Lampe, J., Müller-Fielitz, H., Schuster, R., Zille, M., Müller, K., Krohn, M., Körbelin, J., Zhang, L., Özorhan, Ü., Neve, V., Wagner, J. U. G., Bojkova, D., Shumliakivska, M., Jiang, Y., Fähnrich, A., Ott, F., Sencio, V., Robil, C., … Schwaninger, M. (2021). The SARS-CoV-2 main protease M^{pro} causes microvascular brain pathology by cleaving NEMO in brain endothelial cells. *Nat Neurosci*, *24*(11), 1522–1533. https://doi.org/10.1038/s41593-021-00926-1

Wertheim, J. O., Chu, D. K. W., Peiris, J. S. M., Kosakovsky Pond, S. L., & Poon, L. L. M. (2013). A Case for the Ancient Origin of Coronaviruses. *Journal of Virology*, *87*(12), 7039–7045. https://doi.org/10.1128/JVI.03273-12

WHO Coronavirus (COVID-19) Dashboard. (2023, March 16). WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int

Wiese, S., Reidegeld, K. A., Meyer, H. E., & Warscheid, B. (2007). Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *PROTEOMICS*, *7*(3), 340–350. https://doi.org/10.1002/pmic.200600422

Willis, S. N., Stadelmann, C., Rodig, S. J., Caron, T., Gattenloehner, S., Mallozzi, S. S., Roughan, J. E., Almendinger, S. E., Blewett, M. M., Brück, W., Hafler, D. A., & O'Connor, K. C. (2009). Epstein–Barr virus infection is not a characteristic feature of multiple sclerosis brain. *Brain*, *132*(12), 3318–3328. https://doi.org/10.1093/brain/awp200

Woo, P. C. Y., Lau, S. K. P., Lam, C. S. F., Lau, C. C. Y., Tsang, A. K. L., Lau, J. H. N., Bai, R., Teng, J. L. L., Tsang, C. C. C., Wang, M., Zheng, B.-J., Chan, K.-H., & Yuen, K.-Y. (2012). Discovery of Seven Novel Mammalian and Avian Coronaviruses in the Genus

Deltacoronavirus Supports Bat Coronaviruses as the Gene Source of Alphacoronavirus and Betacoronavirus and Avian Coronaviruses as the Gene Source of Gammacoronavirus and Deltacoronavirus. *Journal of Virology*, *86*(7), 3995–4008. https://doi.org/10.1128/JVI.06540-11

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), Article 7798. https://doi.org/10.1038/s41586-020-2008-3

Wu, S. Y., Lee, A. Y., Hou, S. Y., Kemper, J. K., Erdjument-Bromage, H., Tempst, P., & Chiang, C. M. (2006). Brd4 links chromatin targeting to HPV transcriptional silencing. *Genes Dev*, *20*(17), 2383–2396. https://doi.org/10.1101/gad.1448206

Yamashita, M., & Fenn, J. B. (1984). Electrospray ion source. Another variation on the free-jet theme. *The Journal of Physical Chemistry*, *88*(20), 4451–4459. https://doi.org/10.1021/j150664a002

Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K., & Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science*, *368*(6489), 409–412. https://doi.org/10.1126/science.abb3405

Zhang, N., Zuo, Y., Jiang, L., Peng, Y., Huang, X., & Zuo, L. (2022). Epstein-Barr Virus and Neurological Diseases. *Frontiers in Molecular Biosciences*, *8*. https://www.frontiersin.org/articles/10.3389/fmolb.2021.816098

Zhang, S., Wang, J., & Cheng, G. (2021). Protease cleavage of RNF20 facilitates coronavirus replication via stabilization of SREBP1. *Proc Natl Acad Sci U S A*, *118*(37), e2107108118. https://doi.org/10.1073/pnas.2107108118

Zhou, B., Wu, F., Han, J., Qi, F., Ni, T., & Qian, F. (2019). Exploitation of nuclear protein SFPQ by the encephalomyocarditis virus to facilitate its replication. *Biochem Biophys Res Commun*, *510*(1), 65–71. https://doi.org/10.1016/j.bbrc.2019.01.032

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., … Shi, Z.-L. (2020). A pneumonia outbreak associated with a

new coronavirus of probable bat origin. *Nature*, *579*(7798), Article 7798. https://doi.org/10.1038/s41586-020-2012-7

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*, *10*(1), 1523. https://doi.org/10.1038/s41467-019-09234-6

Zhou, Z., Kang, H., Li, S., & Zhao, X. (2020). Understanding the neurotropic characteristics of SARS-CoV-2: From neurological manifestations of COVID-19 to potential neurotropic mechanisms. *Journal of Neurology*, *267*(8), 2179–2184. https://doi.org/10.1007/s00415-020-09929-7

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., & Tan, W. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. *The New England Journal of Medicine*, *382*(8), 727–733. https://doi.org/10.1056/NEJMoa2001017

Zirngibl, M., Assinck, P., Sizov, A., Caprariello, A. V., & Plemel, J. R. (2022). Oligodendrocyte death and myelin loss in the cuprizone model: An updated overview of the intrinsic and extrinsic causes of cuprizone demyelination. *Molecular Neurodegeneration*, *17*(1), 34. https://doi.org/10.1186/s13024-022-00538-8

Appendix A: Additional information from Chapter 2 on proteolysis during SARS-CoV-2 infection

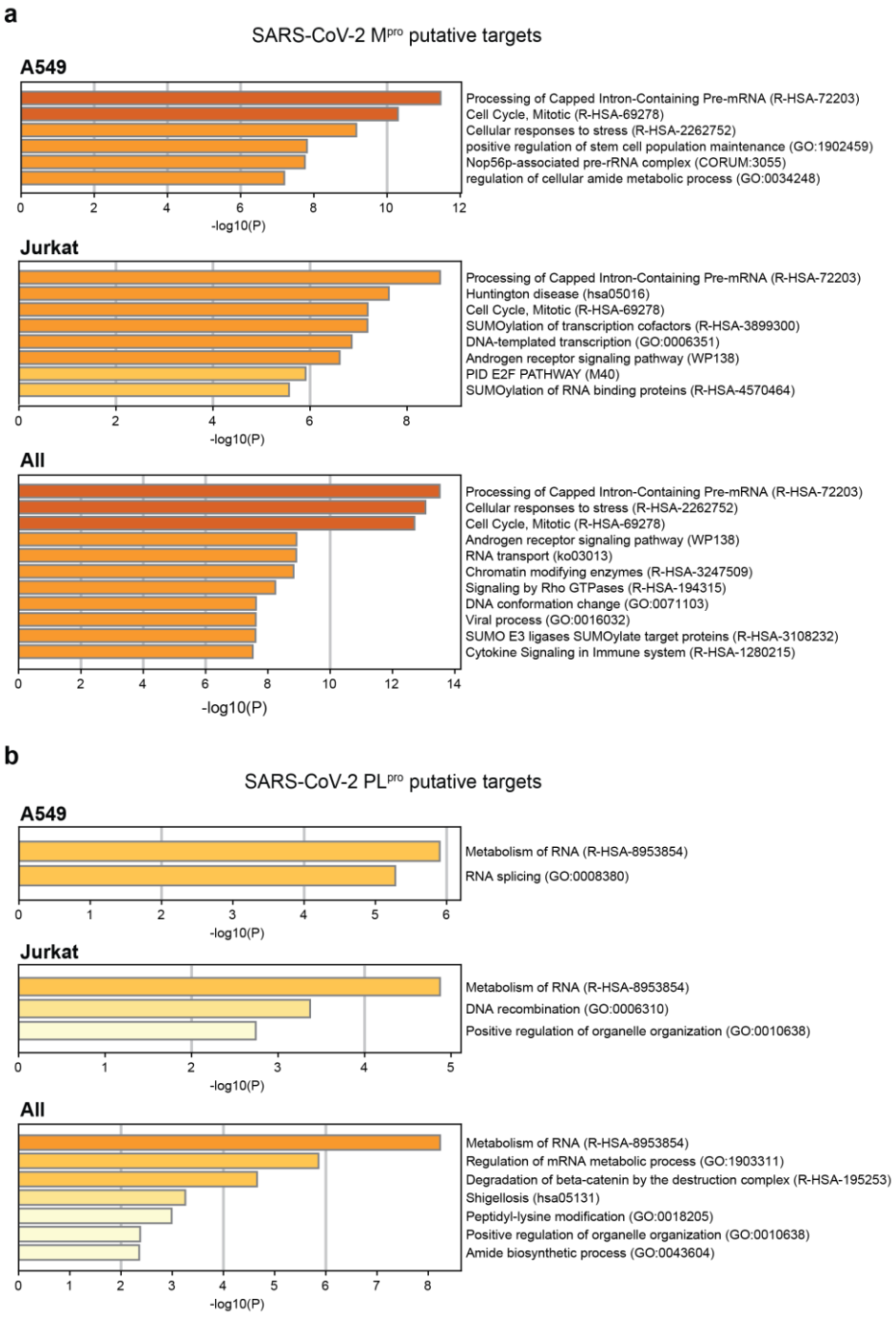


Figure A.1. Gene Ontology analysis of a) Mpro and b) PLpro putative substrates in A549, Jurkat, and both A549 and Jurkat cell lysates using Metascape (Y. Zhou et al., 2019).

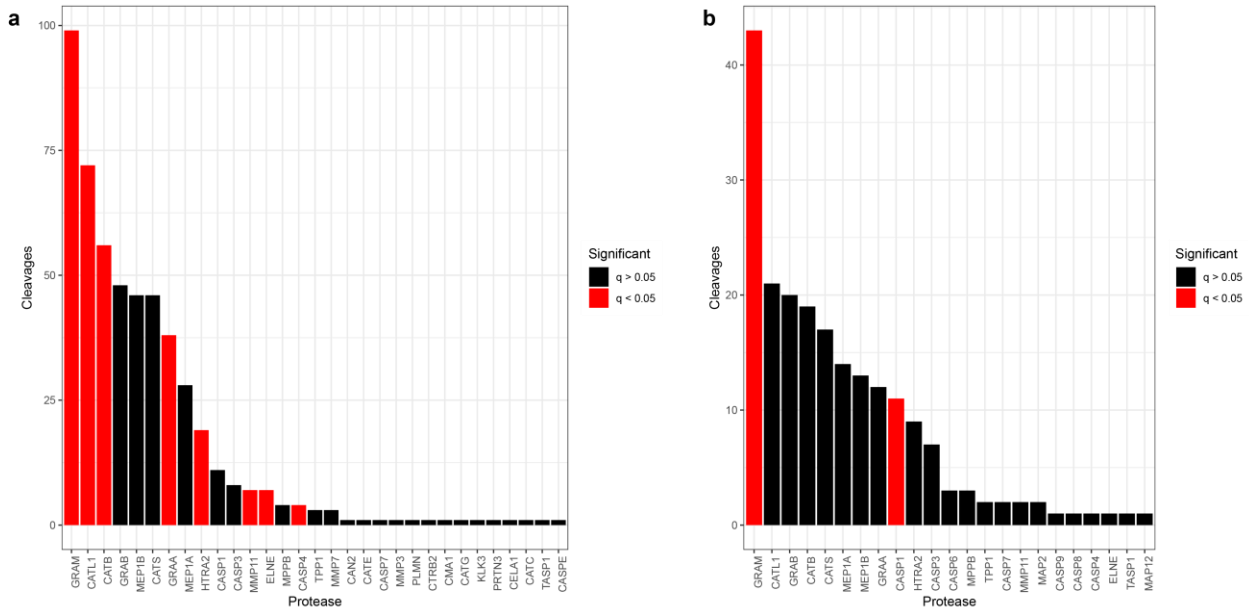


Figure A.2. TopFind analysis of all labeled cleavage sites in **a**) M^{pro} and **b**) PL^{pro} subtiligase N-terminomics experiments.

Table A.1 Antibodies and plasmids used in the study.

Acc #	Gene Name	Sources	Catalog #	Plasmids for overexpression studies
P25440	BRD2	Abcam	ab139690	https://www.addgene.org/65376/
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	-
P0CG48 P0CG47	Ubiquitin (FK2)	Enzo Life Sciences	BML-PW8810	-

Appendix B: Data analytic techniques for the investigation of Met-mistranslation

The following code was designed in Python 3 to create a database of met-mistranslated proteins. It is designed to substitute residues in a protein sequence with methionine. An example input with corresponding output is provided for demonstration. See section 3.1.3 for details.

```
""" Replacement of specific amino
acid:
case 1 replaces each amino acid
individually or
with the previous occurrences """

seqa = "abcabcabcabc"
seq = list(seqa)
i=seq.index("a")
print("Case 1 output:")
while i<= len(seq):
    seq[i]="M"
    print("".join(seq))
    seqb=list(seqa)
    if "a" in seq:
        i = seq.index("a")
        seqb[i] = "M"
        print("".join(seqb))
    else:
        break
```

```
Case 1 output:
Mbcabcabcabc
abcMbcabcabc
MbcMbcabcabc
abcabcMbcabc
MbcMbcMbcabc
abcabcabcMbc
MbcMbcMbcMbc
```

```
""" Replacement of specific amino
acid:
case 2 replaces only one residue """

seqa = "abcabcabcabcabc"
seq = list(seqa)
i=seq.index("a")
seq[i]="M"
print("\nCase 2 output:")
print("".join(seq))
while i<= len(seq):
     seq[i]="M"
     seqb=list(seqa)
     if "a" in seq:
        i = seq.index("a")
        seqb[i] = "M"
        print("".join(seqb))
     else:
         break
```

```
Case 2 output:
Mbcabcabcabcabc
abcMbcabcabcabc
abcabcMbcabcabc
abcabcabcMbcabc
abcabcabcabcMb
```

Appendix C: Cross-comparison of select proteins identified in mice brains and human cells of the CNS

Table C.1. Summary of 29 proteins of interest with cell-associated expression identified in mice brains with GSDMD wildtype (WT) or knockout (KO) treated with cuprizone (CPZ)

Mouse								
Cell Type Assoc.	**Acc #**	**Gene**	**Protein description**	**# Unique peptides**	**Z-score WT**	**Z-score KO**	**Z-score WT CPZ**	**Z-score KO CPZ**
Neuronal	P19246	Nefh	Neurofilament heavy polypeptide	81	-1.07	0.21	1.28	-0.42
Neuronal	Q62277	Syp	Synaptophysin	7	-0.57	-0.08	-0.79	1.43
Neuronal	Q8JZR4	Slc1a7	Excitatory amino acid transporter 5	1	0.02	0.34	-1.37	1.01
Neuronal	P47743;Q05BD6	Grm8	Metabotropic glutamate receptor 8;Glutamate receptor, metabotropic 8	3	0.91	0.64	-0.24	-1.31
Neuronal	A0A668KLC6	Map2	Microtubule-associated protein	154	0.92	-0.01	0.48	-1.39
Neuronal	D3YYT0;P15116	Cdh2	Cadherin-2	18	0.71	-1.06	-0.63	0.98
Microglia	Q6P5E3;Q9ER74	Sall1	Sal-like 1 (Drosophila);Sal-like protein 1	1	-0.07	1.44	-0.72	-0.64
Microglia	Q60805	Mertk	Tyrosine-protein kinase Mer	1	0.83	0.69	-0.17	-1.35

Microglia	G3UVV4;P17710	Hk1	Hexokinase;Hexokinase-1	80	-1.02	1.24	-0.56	0.34
Microglia	E9Q5B5;O08528	Hk2	Hexokinase;Hexokinase-2	4	-1.21	-0.04	1.24	0.01
Microglia	P10810	Cd14	Monocyte differentiation antigen CD14	2	-0.86	-0.63	0.12	1.36
Microglia	P16110;Q8C253	Lgals3	Galectin-3;Galectin	4	-0.74	-0.86	0.34	1.26
Microglia	P98086	C1qa	Complement C1q subcomponent subunit A	8	-0.85	-0.84	0.59	1.11
Microglia	P14106	C1qb	Complement C1q subcomponent subunit B	5	-0.83	-0.81	0.43	1.21
Microglia	Q02105	C1qc	Complement C1q subcomponent subunit C	6	-0.81	-0.82	0.40	1.23
Microglia	A0A0R4J1C8;P31996	Cd68	Macrosialin	1	-0.83	-0.83	0.49	1.17
Astrocyte	P03995	Gfap	Glial fibrillary acidic protein	59	-0.98	-0.73	0.69	1.02
Astrocyte	P20152	Vim	Vimentin	63	-0.88	-0.74	0.39	1.23
Astrocyte	O88507	Cntfr	Ciliary neurotrophic factor receptor subunit alpha	10	-0.84	-0.21	-0.40	1.45

Astrocyte	P56564	Slc1a3	Excitatory amino acid transporter 1	10	0.63	0.06	-1.43	0.74
Astrocyte	Q80YX1	Tnc	Tenascin	20	0.68	0.36	-1.49	0.44
Astrocyte	E9Q0H6;P51880	Fabp7	Fatty acid-binding protein, brain;Fatty acid-binding protein, brain	6	0.37	0.86	-1.44	0.21
Oligodendrocyte/ Myelin	Q3UY21;Q61885	Mog	Myelin-oligodendrocyte glycoprotein	17	-1.43	0.08	0.83	0.51
Oligodendrocyte/ Myelin	P04370	Mbp	Myelin basic protein	24	-1.04	-0.33	0.02	1.35
Oligodendrocyte/ Myelin	P60202	Plp1	Myelin proteolipid protein	13	-0.28	-0.55	-0.65	1.48
Oligodendrocyte/ Myelin	Q8R3P0	Aspa	Aspartoacylase	20	0.21	1.31	-0.96	-0.56
Oligodendrocyte/ Myelin	P20917	Mag	Myelin-associated glycoprotein	24	0.15	0.30	0.96	-1.40
Oligodendrocyte/ Myelin	P16330	Cnp	2',3'-cyclic-nucleotide 3'-phosphodiesterase	66	-0.15	0.04	1.27	-1.17
Oligodendrocyte/ Myelin	O09044;Q9D3L3	Snap23	Synaptosomal-associated protein 23;Synaptosomal-associated protein	7	-0.34	-0.22	1.44	-0.88

Table C.2. Comparison of 29 proteins identified in mice brains (Table C.1) with their corresponding identification in human CNS cell types

Human									
				Found in:					
Acc #	Gene	Seq. Coverage (%)	# Unique Peptides	Neuron	Microglia	Astrocyte	Oligodendrocyte	Brain Lysate A	Brain Lysate B
P12036	NEFH	21	18	Peak Found	Peak Found	Peak Found	Peak Found	High	High
P08247	SYP	38	8	High	Peak Found	Peak Found	Peak Found	High	High
	Slc1a7	Not Found							
	Grm8	Not Found							
P11137	MAP2	70	77	High	High	High	High	High	High
P19022	CDH2	26	14	High	High	High	High	High	High
Q9NSC2	SALL1	3	1	Not Found	Not Found	Not Found	Not Found	High	Not Found
	Mertk	Not Found							
P19367	HK1	52	50	High	High	High	High	High	High
P52789	HK2	28	17	Peak Found	High	High	High	High	Peak Found
P08571	CD14	45	12	Peak Found	High	Peak Found	Peak Found	High	High
P17931	LGALS3	34	8	High	High	High	Peak Found	High	High
	C1qa	Not Found							
P02746	C1QB	14	3	Not Found	High	Not Found	Peak Found	Peak Found	High
P02747	C1QC	21	4	Peak Found	High	Peak Found	Peak Found	Peak Found	High
	Cd68	Not Found							
P14136	GFAP	84	6	Peak Found	High	High	Peak Found	High	High
P08670	VIM	97	85	High	High	High	High	High	High
P26992	CNTFR	25	6	High	Peak Found	Peak Found	High	High	High

P43003	SLC1A3	23	9	High	High	High	Not Found	High	High
P24821	TNC	49	2	High	High	Peak Found	Not Found	High	High
O15540	FABP7	87	12	High	High	High	High	High	High
Q16653	MOG	43	17	High	Peak Found	Peak Found	High	High	High
P02686	MBP	52	1	Not Found	High	Peak Found	Not Found	Peak Found	Peak Found
P60201	PLP1	26	9	High	High	High	High	High	High
P45381	ASPA	55	14	Peak Found	Peak Found	Peak Found	Not Found	High	High
P20916	MAG	31	17	High	Peak Found	Peak Found	High	High	High
P09543	CNP	88	63	High	High	High	High	High	High
O00161	SNAP23	64	9	High	High	High	High	High	High

Appendix D: Substrates of the N-myristoyltransferases NMT1 and NMT2 may play a role in cancer

Introduction

N-myristoyltransferases attach myristate (the 14-carbon fatty acid) on N-terminal glycine to regulate proteins involved in signal transduction and protein-membrane binding interactions. The N-myristoyltransferases NMT1 and NMT2 perform this function in mammalian cells. The pan-NMT inhibitor PCLX-001 targets and kills hematological cancer cells *in vitro* and *in vivo* (Beauchamp et al., 2020) and entered phase 1 clinical trials for B-cell Non-Hodgkin Lymphoma and Advanced Solid Malignancies (Pacylex Pharmaceuticals, 2022). In collaboration with the Berthiaume lab, we aim to elucidate the mechanism of PCLX-001 action by characterization of the substrates of NMT1 and NMT2.

Results and Conclusions

Collaborators in the Berthiaume lab constructed knockout cell lines for NMT1, NMT2, and a double knockout. Additionally, HAP1 cells were treated with the pan-NMT inhibitor PCLX-001 at effective concentrations of EC50 or EC90. Utilizing in-gel trypsin digestion LC-MS/MS, samples were analyzed by label-free quantification. Several myristoylated proteins were identified with NMT knockout as well as inhibition assays and the results are summarized as part of Beauchamp *et al* (under review).

Methods

Sample preparation for mass spectrometry. Samples were provided by Erwan Beauchamp. Cell lysates (30µg protein/lane, n=3) were separated by SDS-PAGE gel electrophoresis in preparation for in-gel digestion. The gels were fixed for 20 minutes (50% ethanol, 2% phosphoric acid), washed twice for 20 minutes each (ddH_2O), stained overnight with blue-sliver coomassie stain (20% ethanol, 10% phosphoric acid, 750 mM ammonium sulphate, 0.12% Coomassie Blue G-250) and washed twice for 10 minutes each (ddH_2O). Each lane was separated into 11 fractions, cut into 1 mm pieces. The gel bands were transferred to a round bottom 96-well plate and destained (50mM ammonium bicarbonate, 50% acetonitrile) at 37°C for 10 minutes. The

solution was removed from the wells and the detaining was repeated 3 times. The gel bands were dehydrated in acetonitrile at 37°C for 10 minutes. The dehydration was repeated until the gel bands became white (2 times). The solution was removed, and the samples were dried at 37°C for 10 minutes. The proteins were reduced (100mM ammonium bicarbonate, 11.4 mM 2-mercaptoethanol) at 37°C for 30 minutes then alkylated (100mM ammonium bicarbonate, 10 mg/mL iodoacetamide) at 37°C for 30 minutes. The gel bands were then washed with 100mM ammonium bicarbonate at 37°C for 10 minutes (repeated twice), dehydrated in acetonitrile at 37°C for 10 minutes (repeated twice) and dried at 37°C for 15 minutes. The proteins were trypsinized (90 µL of 100 mM ammonium bicarbonate and 6 µg/mL Sequence Grade Modified Trypsin, Promega Inc.) overnight. The solutions containing tryptic peptides were transferred to a round bottom 96-well plate. Tryptic peptides were further extracted from the gel bands with extraction solution (2% acetonitrile, 1% formic acid) and incubated at 37°C for 1 hour. A final extraction was conducted using 50% acetonitrile and 0.5% formic acid and incubated at 37°C for 1 hour. All solutions containing tryptic peptides were transferred to a round bottom 96-well plate and lyophilized overnight. The samples were resuspended (5 fractions per lane) in 0.1% formic acid in preparation for analysis by LC-MS/MS.

Mass spectrometry analysis. 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). The peptide mixture underwent reverse phase separation by 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) for the NMT knockouts and controls, or an analytical column (Aurora Ultimate nanoflow UHPLC column 25 cm x 75 µm ID, 1.7 µm C18, 120 Å; IonOpticks inc.) for the PCLX-001 treated samples and controls. Peptides were eluted over a 60-minute linear gradient from 0 to 36.8% acetonitrile in 0.1% formic acid. Data analysis was conducted using ProteomeDiscoverer (v2.4.1.15) software with a Homo sapien proteome (Proteome ID UP000005640, download date: 2020/02/18). The search parameters included a maximum of 3 missed trypsin cleavages, a precursor mass tolerance of 15 ppm, a fragment mass tolerance of 0.8 Da, carbamidomethylation of Cys (static modification), and oxidation of Met and deamidation of Asn and Gln (dynamic modifications). A decoy database search was conducted to evaluate the false-positive rates. The strict target false discovery rate was set at 0.01 and the relaxed false discovery rate was set at 0.05. Additional filters included proteins identified at medium and high confidence, with two or more peptides, in two or more biological replicates, and were master

proteins (the top-ranking protein of a group). Proteins with a p-value less than 0.05 and a fold change threshold of 2 compared to the wildtype control were identified as increased or decreased. MS data is available through MassIVE (MSV000091913).

Figures

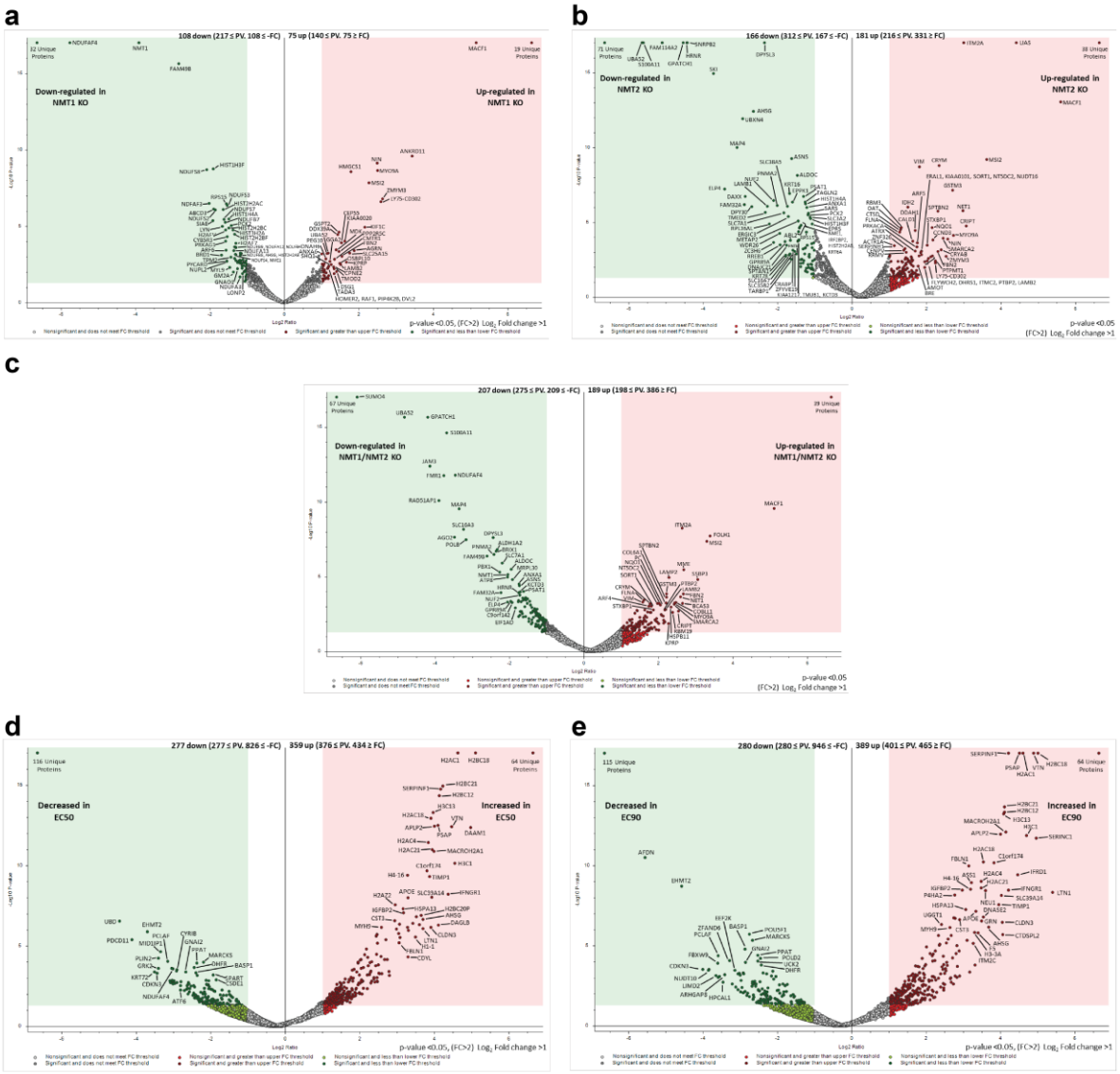


Figure D.1. Summary of proteomic data from NMT1, NMT2, or double knockout and PCLX-001 treatment. Volcano plots with the $\log_2$ of the fold change ratio and the $-\log_{10}$ of the p-value provide a summary of the proteins identified with those increased or decreased defined by a threshold of p-value < 0.05 and fold change > 2 or < 0.5. Fold change ratios are plotted as **a)** NMT1 knockout, **b)** NMT2 knockout, or **c)** double knockout over wildtype control. Additionally, **d)** PCLX-001 treatment at EC50, or **e)** EC90 over the respective untreated control.

Appendix E: Optimization of targeted mass spectrometry techniques for parallel reaction monitoring (PRM) of Tau

Introduction

To investigate met-mistranslation of the tau protein, we aimed to optimize a targeted mass spectrometry technique that could identify and differentiate between two forms of tau in a mixture (wildtype and the P301L mutant). We hypothesized that such a method could aid in the identification of a single residue modification within the protein, such as that observed with met-mistranslation.

Results and Conclusions

Tau samples contained a mixture of purified Tau 0N4R wildtype and P301L mutant protein at a known concentration. These samples were prepared by reduction and alkylation of cysteines, followed by trypsin addition to obtain peptides. The samples were analyzed by LC-MS/MS using DDA. The minimum threshold for identification of wildtype or P301L mutant tau was between 1 pmol and 0.5 pmol. To increase the sensitivity of detection, we developed a method to detect Tau using PRM.

For the targeted method, the samples were prepared by reduction and alkylation of cysteines followed by trypsin addition to obtain peptides. The addition of Pierce retention time calibration standard peptides was used to account for the shift in retention time between runs. The samples were analyzed by LC-MS/MS using DDA (discovery) and a list of tau 0N4R peptides was identified (m/z, charge, and 10-minute retention time window). This list was used to specifically monitor these tau peptides using PRM. The peptides were confirmed by comparing the ion rank and retention times of the PRM to the initial discovery. We were able to detect tau at 5 pmol and 0.5 pmol but were unable to identify the form (wildtype or mutant) at 0.5 pmol. Therefore, the PRM did not increase significantly the sensitivity of tau detection. Potential sample loss during sample preparation may account for the lower peptide yield and in the future, carrier proteins such as BSA may be explored as a method to limit sample loss during the preparation of low protein amounts.

Methods

Sample preparation for mass spectrometry. Samples of a tau mixture (wildtype 0N4R: P301L 0N4R) were prepared utilizing purified tau (provided by Angelle Britton, Sue-Ann Mok lab). Samples were reduced (5 mM TCEP, boiled at 95°C for 15 min) and alkylated (4mM IAM, 1 h in the dark, quenched with 10mM DTT). Samples were suspended in 50 mM Ammonium bicarbonate (pH 8), and trypsinized (Promega, #V5113) at a 1:100 ratio of trypsin: protein overnight. TFA was added to 2.5% and peptides were desalted using C18 Ziptips (Millipore). Following resuspension in 0.1% formic acid, samples were spiked with a retention time calibration peptide mixture (Pierce^{TM}, #88321) to a final concentration of 500 fmol per 5µL injection. A 5 pmol:5 pmol sample was used for initial discovery experiments. PRM was conducted for samples ranging from 1000 to 0.001 pmol of total tau (1:1).

Mass spectrometry analysis. 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 45 min from 3.85% to 36.8% acetonitrile in 0.1% formic acid. The data was analyzed using ProteinProspector (v5.22.1) against the concatenated database of the human proteome (SwissProt.2017.11.01.random.concat), with a maximum false discovery rate of 1% for peptides. The peptides were searched at a maximum of 3 missed trypsin cleavages with Trypsin. Search parameters included a precursor mass tolerance of 15 ppm, a fragment mass tolerance of 0.8 Da, a precursor charge range of 2-5, with the constant modification carbamidomethylation (C), and variable modifications of deamidated (N/Q), and oxidation (M). The maximum number of variable modifications was set to 2.

Initial discovery experiments were conducted with DDA as described above, over a range of 375 to 1700 m/z, with an MS^1 resolution of 120, 000 m/z. PRM was subsequently conducted following the same methodology of the discovery at a resolution of 15,000 m/z at the MS^2 level with a range based on a starting mass-to-charge ratio of 140 m/z. Isolation lists (mass-to-charge ratio, charge, and 10-minute retention time window) were based on initial discovery experiments. Data analysis was conducted using Skyline (v21.1) against a library based on the discovery results with the parameters listed previously and an ion match tolerance of 0.5 m/z. MS data is available through MassIVE (MSV000092231).

Figures

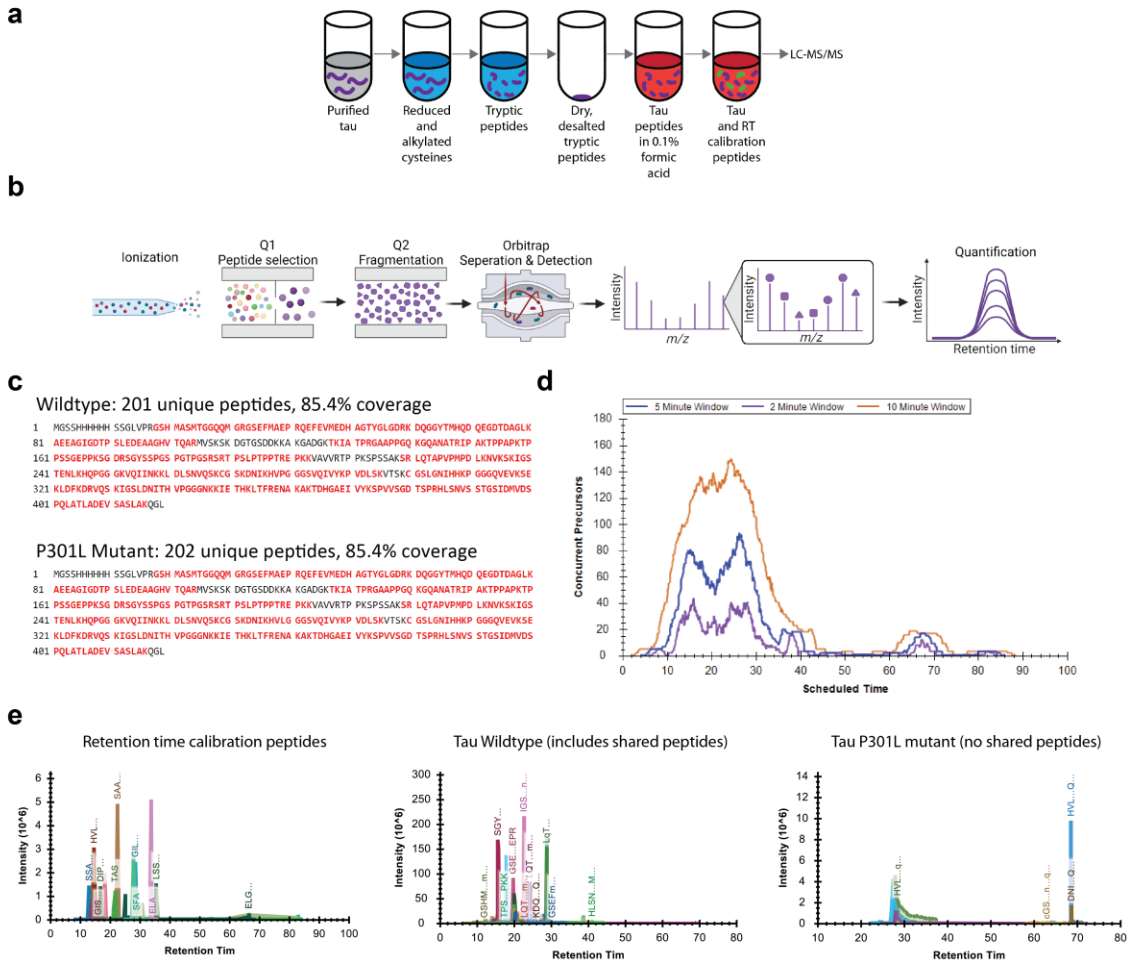


Figure E.1. Optimization of targeted mass spectrometry for the detection of tau. a) Samples of purified tau 0N4R underwent sample preparation by reduction and alkylation followed by trypsinization. The dried tryptic peptides were resuspended in a buffer compatible with mass spectrometry. A mixture of retention time calibration peptides was spiked into the sample before analysis by LC-MS/MS. **b)** The PRM workflow was differentiated from the discovery shot-gun analysis by the selection of specific ionized peptides for fragmentation, detection, and subsequent quantification by chromatogram peak area. **c)** The initial discovery identified approximately 200 tau peptides (wildtype or mutant) providing an 85.4% sequence coverage. **d)** Based on the discovery results, a 10-minute retention time window allows for the monitoring of a maximum of 150 precursors. **e)** Chromatograms show that subsequent PRM successfully monitored for a list of retention time calibration peptides, wildtype tau (including those shared with the mutant) peptides, and P301L mutant (unique) peptides.

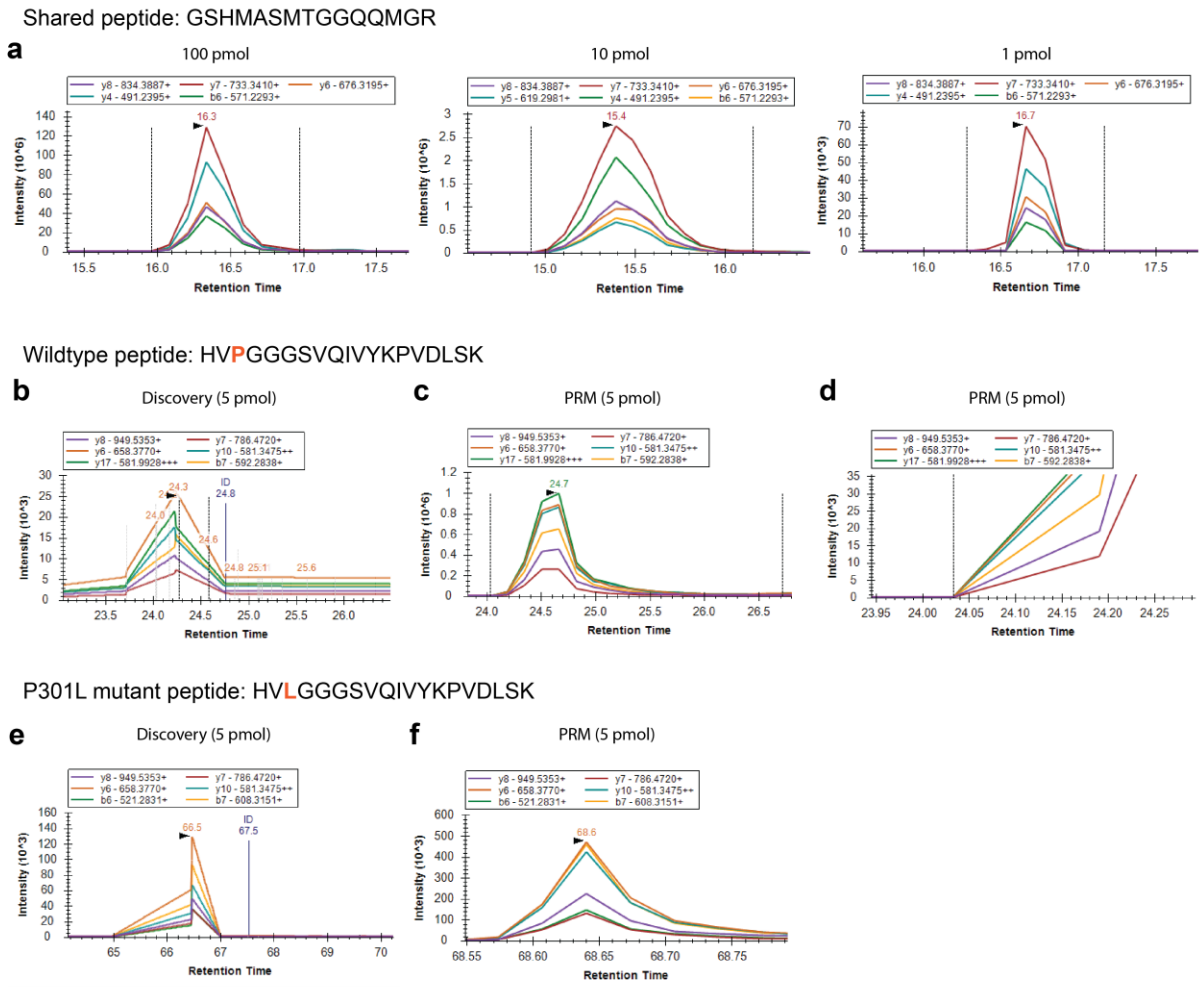


Figure E.2. PRM successfully identified wildtype and mutant tau as shown by chromatograms. a) Several peptides shared by wildtype and P301L tau were monitored at varying amounts of total tau. One such peptide is shown at 100 pmol, 10 pmol, and 1 pmol of total tau. We observed increasing maximal peak intensity (and thus area) with increasing amounts of tau. **b)** Wildtype tau was monitored for by the unique tryptic peptide containing P301. The initial discovery chromatogram was compared to the **c)** PRM-acquired chromatogram based on retention time and ion rank similarity. We confirmed successful peptide identification. **d)** The PRM chromatogram was subsequently observed on the same intensity scale as the discovery and revealed only the base of the peak, highlighting the necessity for continuous monitoring for quantification. **e)** Similarly, P301L mutant tau was monitored for by the unique tryptic peptide containing L301, and the discovery was compared to the **f)** PRM chromatogram.

Appendix F: Two HEK293 clones display differences in tau seeding and proteomic profiles

Introduction

Following transfection of HEK293-TauP301L cells with a vector expressing DNAJA2, of three clones, one clone showed inconsistent expression of DNAJA2 as well as differences in GFP-tau expression with tau seeding (McNamara, 2022). We aimed to use label-free quantification mass spectrometry to investigate the differences in protein expression.

Results and Conclusions

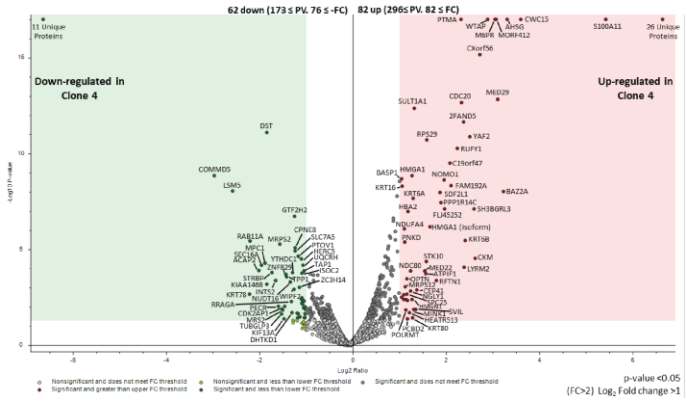
Clone 4 was identified to express inconsistent amounts of DNAJA2 and present differences following tau seeding compared to clones 6 and 7. Clones 4 and 7 were analyzed by in-gel trypsin digestion LC-MS/MS for label-free quantification. Subsequently, tau was identified to be expressed approximately 2-fold as much in clone 4 compared to clone 7.

Methods

Cell Lysis. Frozen HEK293T cells were provided by Emily McNamara and subsequently lysed (100 mM bicine pH 8.5, 5% sodium dodecyl-sulfate) in the presence of protease inhibitors (5 mM Ethylenediaminetetraacetic acid (EDTA), 1 mM 4-benzenesulfonyl fluoride hydrochloride (AEBSF), 1 mM phenylmethylsulfonyl fluoride (PMSF), 4 mM iodoacetamide (IAM)) and incubated at 4°C in the dark for 15 minutes. The lysates were sonicated for 2 minutes (20% amplitude, 2 seconds on, 5 seconds off) using a Qsonica sonicator (Mandel inc.), then incubated at 4°C in the dark for another 15 minutes. The total protein concentration was determined by Pierce™ BCA protein assay (Thermo Fisher Scientific). To quench the IAM, 20 mM dithiothreitol (DTT) was added. Loading buffer (containing sodium dodecyl-sulfate and 2-mercaptoethanol) was added to the cell lysates and they were boiled at 95°C for 5 minutes.

Sample preparation and mass spectrometry analysis. Cell lysates (24 µg protein/sample, n=4) were prepared and analyzed as previously described in **Appendix D**. MS data is available through MassIVE (MSV000092375).

Figures



Accession	Gene Symbol	Abundance Ratio: (Clone 4) / (Clone 7)	Abundance Ratio P-Value: (Clone 4) / (Clone 7)
P10636-6	MAPT	1.929	1.63846E-08

Figure F.1. Mass spectrometry reveals differences in tau protein abundance in two HEK293 cell clones. Based on protein quantification, tau protein abundance in clone 4 was approximately twice as much as that in clone 7. The proteomic data is summarized in the provided volcano plot, where increased or decreased proteins are defined by a threshold of p-value < 0.05 and fold change > 2 or < 0.5.

Appendix G: Proteomic changes with iPSCs differentiation into Neurons and cardiomyocytes

Results and Conclusions

To investigate the chaperone network in different cell types following differentiation, iPSCs were differentiated into neurons and cardiomyocytes. The cellular proteomes were compared using in-gel digestion LC-MS/MS for label-free quantification.

Methods

Sample preparation for mass spectrometry. Cell lysates of iPSCs, iPSC-derived neurons (9 weeks old), and iPSC-derived cardiomyocytes (26 days old) (20 µg lysate, n=4) were provided by Angelle Britton (Sue-Ann Mok lab) and prepared by ProTrap XG as previously described in **Chapter 3**.

Mass spectrometry analysis. Peptide separation was conducted using a nanoflow-HPLC (Thermo Scientific EASY-nLC 1200 System) coupled to Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific inc.). The peptide mixture underwent reverse phase separation on an analytical column (Aurora Ultimate nanoflow UHPLC column 25 cm x 75 µm ID, 1.7 µm C18, 120 Å; IonOpticks inc.). Peptides were eluted over a 120 min linear gradient from 0% to 36.8% acetonitrile in 0.1% formic acid. Data analysis was conducted using ProteomeDiscoverer (v2.4.1.15) software with a Homo sapien proteome (Proteome ID UP000005640, download date: 2020/02/18). The search parameters included a maximum of 3 missed trypsin cleavages, a precursor mass tolerance of 15 ppm, a fragment mass tolerance of 0.8 Da, carbamidomethylation of Cys (static modification), and oxidation of Met and deamidation of Asn and Gln (dynamic modifications). A decoy database search was conducted to evaluate the false-positive rates. The strict target false discovery rate was set at 0.01 and the relaxed false discovery rate was set at 0.05. Additional filters included proteins identified at medium and high confidence, with two or more peptides, in two or more biological replicates, and were master proteins (the top-ranking protein of a group). Proteins with a p-value less than 0.05 and a fold change threshold of 2 compared to iPSC samples were identified as increased or decreased. MS data is available through MassIVE (MSV000092230).

Figures

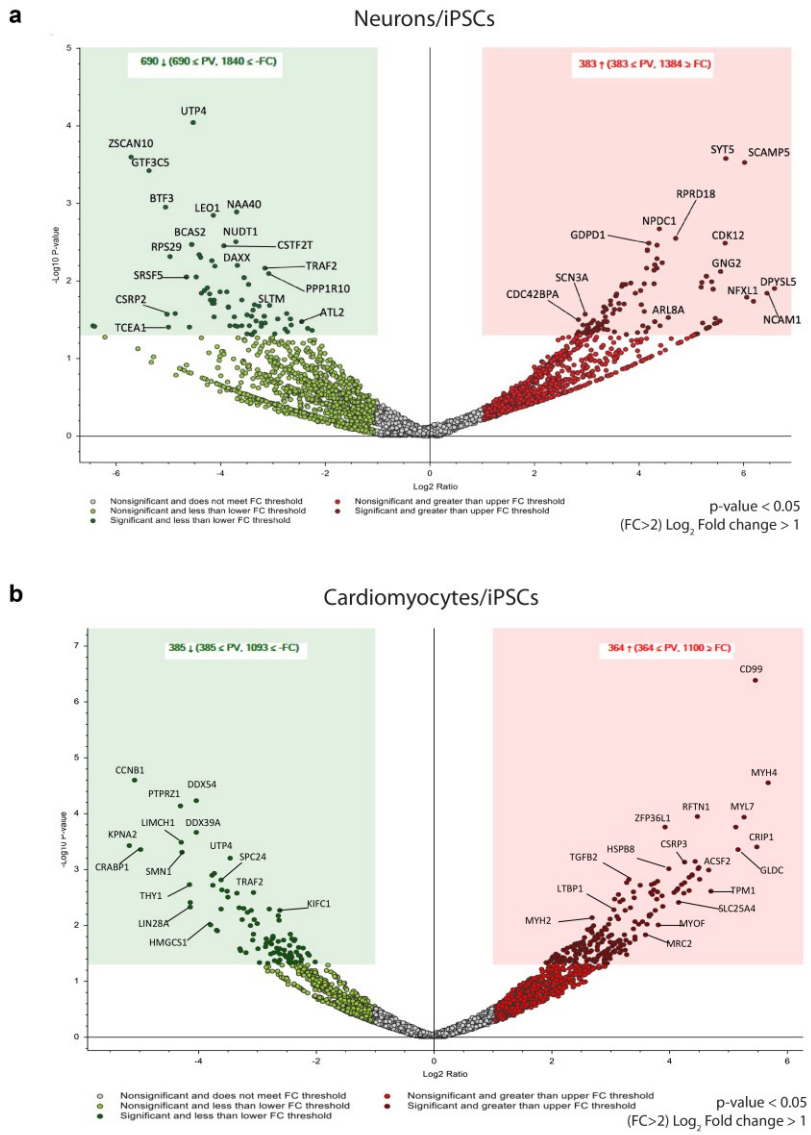


Figure G.1. Mass spectrometry reveals differences in protein abundance among neurons and cardiomyocytes derived from iPSCs. Volcano plots with the $\log_2$ of the fold change ratio and the $-\log_{10}$ of the p-value provide a summary of the proteins identified with those increased or decreased defined by a threshold of p-value < 0.05 and fold change > 2 or < 0.5. Fold change ratios are plotted as **a)** neurons/iPSCs and **b)** cardiomyocytes/iPSCs. Volcano plots do not include the unique proteins identified (data points at a fold change of 100 or 0.01 and p-value 10^{-17}. Proteomic analysis and visualization were conducted with Kolden Van Baar.