GALLARDO, IGNACIO A., M.S. Identification of SARS-CoV-2 Main Protease Cleavage Sites in Host Cellular Selenoproteins and Glutathione-Related Proteins (2022)
Directed by Dr. Ethan W. Taylor. 51 pp.

SARS-CoV-2 (severe acute respiratory syndrome-coronavirus-2) continues to be a global threat. Even as vaccines and therapeutics become increasingly available, COVID-19 (Coronavirus Disease-2019) infections continue to rise, leading to the proliferation of even more transmissible variants. Although the current Omicron variant appears to be milder in pathogenicity when compared to previous strains of the virus, it is said to be about three times more infectious than its predecessor, the Delta variant, thus greatly increasing the chances of more breakthrough infections.[1] Therefore, it has become increasingly important that we understand the mechanisms by which this virus operates in order to find ways to decrease the rates of infection/replication.

One viral mechanism of interest is that of the SARS-CoV-2 Main Protease (M^{pro}), also known as the 3CL Protease. This protease plays vital roles in viral gene expression and replication via the proteolytic processing of two overlapping polyproteins encoded by the virus' replicase gene.[2] However, recent studies, including the one from our group, have shown that M^{pro} may also be targeting certain host cellular proteins that possess short stretches of homologous host-pathogen protein sequences (SSHHPS) corresponding to the canonical M^{pro} cleavage sites.[3] Of particular interest are those with sequences resembling the P1, P2, P1', and P4 positions of SARS-CoV-2's non-structural proteins (NSPs), as those are the ones in direct contact with M^{pro}'s active site.

Relating to the established link between selenium and viral pathogenesis, the goal of this study was to assess whether SARS-CoV-2 M^{pro} is targeting essential host selenoproteins and other glutathione-related proteins. Various *in silico* methods, including NetCorona, PROSPER,

Procleave, and 3D modeling were utilized in the prediction of six potential M^{pro} cleavage sites. The validity of these sites was assessed by incubating recombinant SARS-CoV-2 M^{pro} with 10-12-mer peptides corresponding to the proposed cleavage sites in each selected protein, and analyzing their cleavage via mass-spectrometry-based methods. Definitive *in vitro* cleavage was observed in those within thioredoxin reductase-1 (TXNRD1), glutamate-cysteine ligase catalytic subunit (GCLC), Selenoprotein F (SelenoF), and Selenoprotein P (SelenoP). These results, not only support our hypothesis, but have broad implications for the ways we view prevention, treatment, and the evolution of SARS-CoV-2.

IDENTIFICATION OF SARS-COV-2 MAIN PROTEASE CLEAVAGE SITES IN HOST CELLULAR SELENOPROTEINS AND GLUTATHIONE-RELATED PROTEINS

by

Ignacio A. Gallardo

A Thesis

Submitted to

the Faculty of The Graduate School at

The University of North Carolina at Greensboro

in Partial Fulfillment

of the Requirements for the Degree

Master of Science

Greensboro

2022

Approved by

Dr. Ethan W. Taylor
Committee Chair

APPROVAL PAGE

This thesis written by Ignacio A. Gallardo has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair

Dr. Ethan W. Taylor

Committee Members

Dr. Norman Chiu

Dr. Jonathan Chekan

June 14, 2022

Date of Acceptance by Committee

June 14, 2022

Date of Final Oral Examination

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

I.A. BACKGROUND

I.A.1. Overview of COVID-19

The world has been in crisis since the beginning of 2020 over the ensuing Coronavirus Disease-2019 (COVID-19) pandemic brought on by the emergence of the Severe Acute Respiratory Syndrome Virus 2 (SARS-CoV-2). While initially discovered in Wuhan, China in December of 2019, the WHO did not officially deem it a pandemic until March, 11 of 2020 as cases began to spread globally at an alarming pace.[4] Since then, the world has faced massive health crises, indefinite lockdowns, numerous business closures, increasing social division, and crippling uncertainty.

As of June 5th, 2022, COVID-19 has infected over 531,640,861 individuals, 6,298,056 of whom succumbed to the disease, resulting in about a 1.2% mortality rate.[5] Those infected experience symptoms ranging from nothing for the asymptomatic cases, flu-like symptoms and agnosia/ageusia for the mild/moderate cases, and pneumonia/acute respiratory distress syndrome (ARDS) for the more severe cases. While most severe cases befall the elderly and those with various co-morbidities, such as high blood pressure and immunosuppression, many young and healthy individuals have succumbed to this disease, as well. In addition, the CDC has revealed that some people also experience post-COVID conditions, reporting various long-term respiratory, neurologic, and cardiovascular issues, even those who were originally asymptomatic.[6]

Fortunately, the United States has been able to fast-track the development of various COVID-19 vaccines, including the mRNA vaccines from Pfizer-BioNTech and Moderna, the

virus vector vaccines from Johnson & Johnson and Astrazeneca, and a protein subunit vaccine from Novavax.[7] Currently, about 11,660,691,546 vaccines/boosters have been administered world-wide, providing preventative measures for much of the global population.[5] In addition, various therapeutics have also been developed to treat the disease, such as the antiviral drug Remdesivir and several monoclonal antibodies. However, new, and more virulent strains of SARS-CoV-2 have also emerged that have evolved to be more transmissible and resistant.

Currently, the South African-borne Omicron variant, which was first discovered in November of 2021, is now the dominant variant, accounting for over 94% of new cases worldwide.[8] Therapeutics and vaccines that were once effective against the previous variants are now rendered less effective due to Omicron's reduced sensitivity to antibody neutralization.[9] In fact, a recent study has found that, due to Omicron's highly mutated spike protein, the new variant exhibits a 26-34 fold resistance to vaccination with both doses of either the Pfizer or the Moderna vaccines when compared to previous variants.[10] In addition, the study also showed that some therapeutic monoclonal antibodies, like Regeneron, have lost their ability to bind and thus neutralize the SARS-CoV-2 spike proteins of the Omicron variant.[10] Therefore, it has become evident that a more holistic approach is necessary to efficiently combat this coronavirus as it continues to evolve.

I.A.2. Overview of SARS-CoV-2

As a coronavirus, SARS-CoV-2 belongs to the Coronaviridae family, which is a group of enveloped, positive-sense, single-stranded RNA viruses that are broken down into four genera, consisting of α, β, γ, and δ.[11] Of interest are the β-coronaviruses, which include SARS-CoV-1, MERS-CoV, and SARS-CoV-2, all of which are associated with severe respiratory diseases in mammals, sharing about 80% and 53% overall nucleotide sequence identity, respectively.[12]

Furthermore, they generally share greater sequence identity for the major enzymes and structural proteins.[12] Thus, preceding data generated from studies of other β-coronaviruses provide valuable insights to our understanding of SARS-CoV-2.

Like most coronaviruses, SARS-CoV-2 is composed of four structural proteins, including spike proteins (S), envelope proteins (E), membrane proteins (M) and nucleocapsid proteins (N), along with a positive-sense, single-stranded RNA.[13] Each of these structures play vital roles in the viral life cycle of SARS-CoV-2. The notorious spike protein is critical for viral entry into the host cell via the Angiotensin-Converting Enzyme II (ACE2) receptor.[14] Additionally, as previously noted, the spike protein is the target for vaccine development and various therapeutics.[14] Envelope proteins, on the other hand, are integral transmembrane proteins that form ion channels essential for viral assembly and release from the host cell.[15] While membrane proteins are also integral transmembrane proteins, their role is in attenuating the host immune response via inhibition of various proteins, including Nuclear Factor Kappa B (NFκB), cyclooxygenase-2 (COX-2), and 3-phosphoinositide-dependent protein kinase 1 (PDK1).[15] Finally, the most abundant of all the structural proteins are the nucleocapsid proteins, which bind to the genomic RNA, forming a ribonucleoprotein. This helps the genomic RNA enter the host cell upon viral fusion and aids in its interaction with host cellular structures responsible for its translation.[15]

The coronavirus genome is known to be one of the largest of all RNA viral genomes, ranging between 26 and 32 kilobases (kb) in length.[13] Specifically, the SARS-CoV-2 genome runs about 30kb in length, containing about 14 open reading frames (ORFs), coding for 16 non-structural proteins, four structural proteins, and nine accessory proteins.[16] **Figure 1** shows that of the entire genomic RNA, only the replicase gene, composed of the overlapping open reading

frames, ORF1a and ORF1b, is directly translated upon entry into the host cell. The rest of the genes are transcribed discontinuously by a product of the replicase gene, called the replicase-transcriptase complex, synthesizing subgenomic mRNAs (sgmRNAs) that code for the structural and accessory proteins.[17] Therefore, proper translation and posttranslational processing of polyproteins 1a and 1ab are essential for viral proliferation.

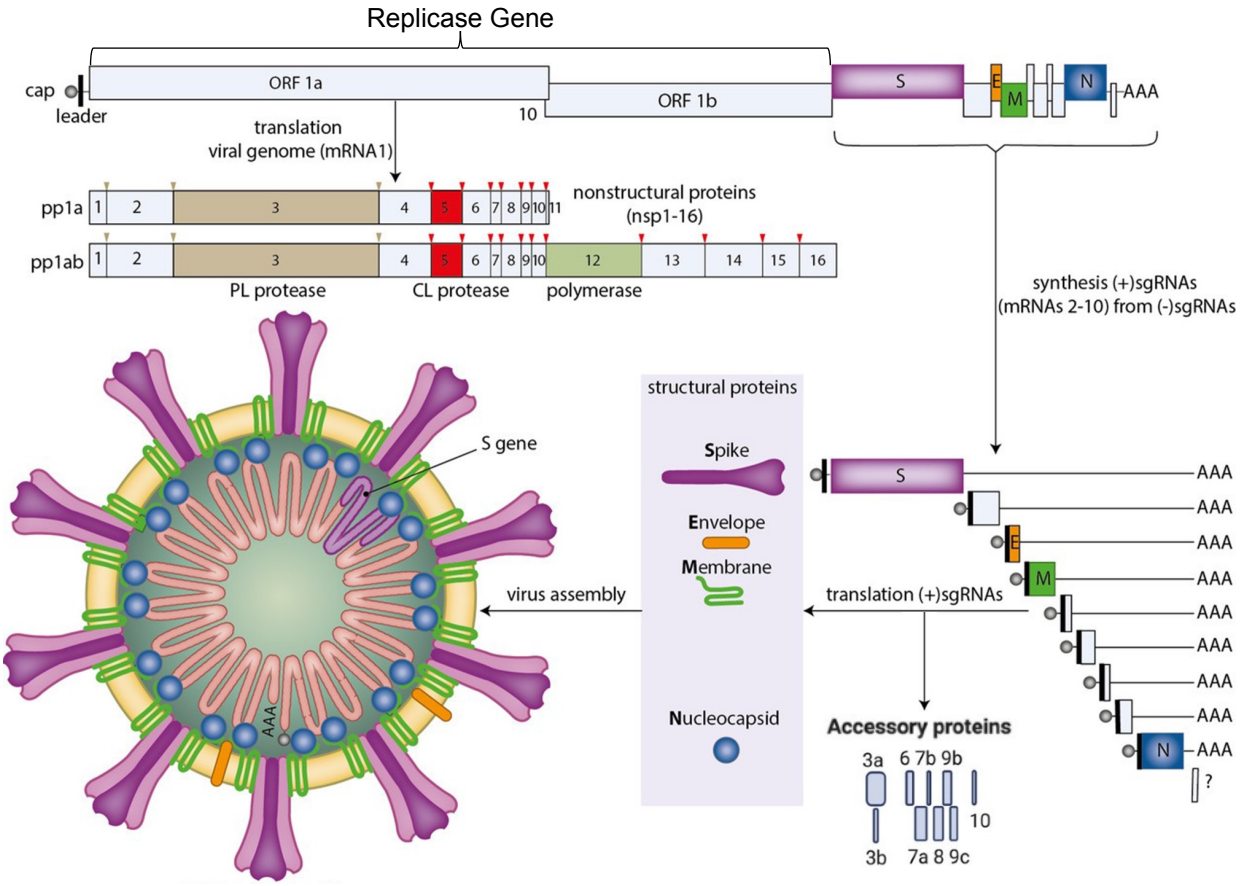

Figure 1. Overview of SARS-CoV-2 Gene Expression, Translation, and Overall Structure. Depicted at the top of this figure is SARS-CoV-2's entire genome, consisting of multiple open reading frames (ORFs). These ORFs code for various proteins, including the virus' non-structural (left), structural (right), and accessory (right) proteins. The bottom left shows how the structural proteins make up the virus' overall structure. Figure from de Vries et al.[18]

As shown in **Figure 1**, polyprotein 1a (pp1a) is encoded solely by ORF1a while polyprotein 1ab (pp1ab) is encoded by both ORFs 1a and 1b via a -1 ribosomal frameshift.[19] Each of these polyproteins are composed of various proteins, called nonstructural proteins

(NSP); pp1a containing NSPs 1-11 and pp1ab containing NSPs 1-10 and 12-16. Each of the 16 individual NSPs play key roles in the viability of the virus, ranging from pathogenicity to replication. However, in order for any of them to carry out these functions, they must all be cleaved from their respective polyprotein by two important viral proteases, the papain-like protease (PL^{pro}) and the 3C-Like protease ($3CL^{pro}$), also known as the Main Protease (M^{pro}).[2] Understandably, as non-structural proteins themselves (NSP3 and NSP5, respectively) PL^{pro} and M^{pro} must undergo autoproteolysis from their respective polyproteins before being able to hydrolyze any other proteins.[2]

Once proteolytically processed, the rest of the nonstructural proteins can carry out their vital functions. NSPs 1 and 2 are responsible for host modulation, suppressing the interferon pathway, resulting in the notorious cytokine storm.[20] NSPs 4 and 6 are involved in the double membrane vesicle (DMV) formation, which increases the efficiency of replication.[20] Lastly, NSPs 7-16 are involved in the replication of the genomic RNA and transcription of the sgmRNAs, ultimately resulting in the translation of the structural and accessory proteins.[20]

I.A.3. SARS-CoV-2 M^{pro} Structure and Function

SARS-CoV-2 M^{pro} is a homo-dimeric cysteine protease responsible for the cleavage of NSPs 5-16 and one end of NSP4 (PL^{pro} cleaves NSPs 1-3 and the other end of NSP4) in both pp1a and pp1ab.[2] As illustrated in **Figure 2a**, a single protomer of M^{pro} consists of three domains, each playing important roles in its proteolytic activity. The active site can be found in a cleft between domains I and II, harboring the catalytic dyad, consisting of histidine 41 and cysteine 145.[21] Additionally, the dimerization of two Main Protease protomers, which is essential for the existence of a functional M^{pro}, is known to occur between domains II and III.[22] Specifically, the N-Terminal finger (residues 1-7) of one protomer squeezes between domains II

and III of the other, forming the various intermolecular interactions necessary for dimerization.[22] These interactions also contribute to the formation of the S1 subsite of the adjacent protomer's active site.[23] However, it is important to note that while dimerization is required for optimal proteolytic activity, only one protomer is active at a time. The reason for this is that when the average volume of the active site increases in one protomer, it decreases in the other, thus hindering the substrate-binding capability of that protomer.[23]

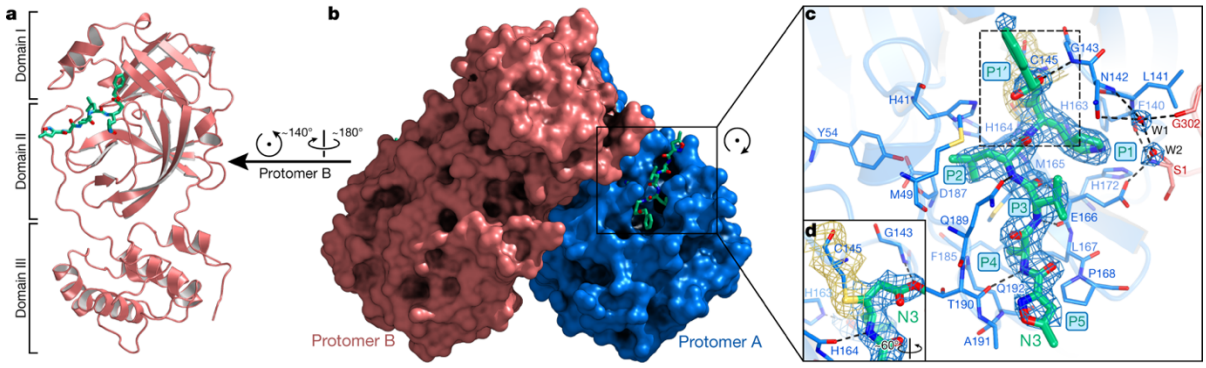

Figure 2. Overall Structure of SARS-CoV-2 M^{pro}. (a) Ribbon model of a single protomer of SARS-CoV-2 M^{pro}, designating the three domains and the location of the substrate-binding cleft. (b) Space model of a functional M^{pro} in its homo-dimeric form, designating the location of the active site. (c) A close-up view of the M^{pro} active site with a bound inhibitor. (d) A close-up view of the irreversible covalent bond formed between the inhibitor and the catalytic Cys145. Figure from Jin et al.[22]

Figures 2b and 2c portray how a substrate (in this case, an inhibitor) binds to the active site of a mature M^{pro}. Notably, the active site is composed of four subsites: S1, S2, S4, and S1', which ultimately dictate the protein sequences that "fit" into this pocket.[23] The S1 subsite consists of side chains from F140, N142, E166, H163 and H172 of the active protomer and those of the S1 residue of the inactive protomer via the aforementioned N-Terminal Finger.[22] Consequently, this subsite only allows for the presence of a glutamine at the P1 position of the substrate.[23] The S2 subsite consists of side chains from H41, M49, Y54, and M165 of the active protomer, along with the alkyl moiety located on D187 of the inactive protomer.[22] Overall, this results in the formation of a hydrophobic pocket, allowing for the binding of residues with non-

polar side chains at the P2 position, primarily Leucine[22]. The S4 subsite forms a smaller hydrophobic pocket, consisting of M165, L167, F185, and Q192[22]. Therefore, the residues at the P4 position will primarily be those that are smaller than those of the P2 position, such as valine, threonine, alanine, and proline. Finally, the S1' subsite contains mainly hydrophobic residues, including L27, V42, and C145[23]. This is quite a shallow subsite that primarily includes smaller amino acids, such as serine and alanine, and will sparingly include asparagine and glycine.[22] Overall, these subsites function to stabilize the substrate in the active site via various hydrogen bonds and hydrophobic interactions.

Predictably, the canonical, M^{pro} cleavage sites present on pp1a and pp1ab all fall in line with the amino acid residues prescribed for each of the four major subsites. However, they are not all cleaved with the same catalytic efficiency. **Table 1** shows that the cleavage sites at NSP4/5 and NSP5/6 have much greater catalytic efficiencies than the rest of the cleavage sites, which makes sense as these are the autocleavage sequences used to free NSP5 from the polypeptides. Furthermore, NSP4/5 has a 100% relative catalytic efficiency because the N-Terminal end of NSP5 is the first to be cleaved, stimulating the autolytic cleavage of the C-Terminal end, which has a 41% catalytic efficiency.[24] The rest of the NSPs seem to undergo hydrolysis at varying efficiencies, ranging from 0.2%, for the N-Terminal end of the RNA-dependent RNA Polymerase, (NSP12) to 28%, for the N-terminal end of the endoribonuclease, (NSP15).

Table 1. SARS-CoV-2 M^{pro} Cleavage Sites and Relative Catalytic Efficiencies.[3]

M^{pro} Cleavage Site	P5	P4	P3	P2	P1		P1'	P2'	P3'	P4'	P5'	Relative K_{cat}/K_m
Nsp4/5	S	A	V	L	Q	X	S	G	F	R	K	**100%**
Nsp5/6	G	V	T	F	Q	X	S	A	V	K	R	**41%**
Nsp6/7	V	A	T	V	Q	X	S	K	M	S	D	**3%**
Nsp7/8	R	A	T	L	Q	X	A	I	A	S	E	**5%**
Nsp8/9	A	V	K	L	Q	X	N	N	E	L	S	**2%**
Nsp9/10	T	V	R	L	Q	X	A	G	N	A	T	**22%**
Nsp10/12	E	P	L	M	Q	X	S	A	D	A	Q	**0.2%**
Nsp12/13	H	T	V	L	Q	X	A	V	G	A	C	**8%**
Nsp13/14	V	A	T	L	Q	X	A	E	N	V	T	**9%**
Nsp14/15	F	T	R	L	Q	X	S	L	E	N	V	**28%**
Nsp15/16	Y	P	K	L	Q	X	S	S	Q	A	W	**27%**

I.A.4. Proteolytic Mechanism of SARS-CoV-2 M^{pro}

As a cysteine protease, the hydrolytic mechanism of SARS-CoV-2 follows two basic steps: acylation, in which the enzyme-substrate complex is formed, and de-acylation, in which the products are released and the apoenzyme reforms. **Figure 3** demonstrates how the reaction is initiated by the catalytic dyad, consisting of H41 and Cys145. Initially, the presence of the substrate in the active site leads to the formation of the Michaelis-Complex, in which the substrate is stabilized in the substrate-binding pocket and the catalytic dyad is in its neutral form[25]. However, for this reaction to progress, the catalytic dyad must be in its ionized state, which is facilitated by the presence of a water molecule, termed H_2O_{cat}, D187, and H164[25]. H41 forms a hydrogen bond with H_2O_{cat}, which in turn is stabilized by the side chains of D187 and H164[25]. This results in an increase in the pKa of H41, allowing it to act as a base and subsequently abstract the proton from Cys145. The newly formed ion pair is stabilized by hydrogen bonds to water molecules and the side chain of P1'. Interestingly, experimental analysis of substrate specificity indicated that having either alanine or glycine in the P1' position,

instead of serine, resulted in a diminished rate constant by one order of magnitude, implying the importance of the hydroxyl group in the stabilization of the ion pair[25].

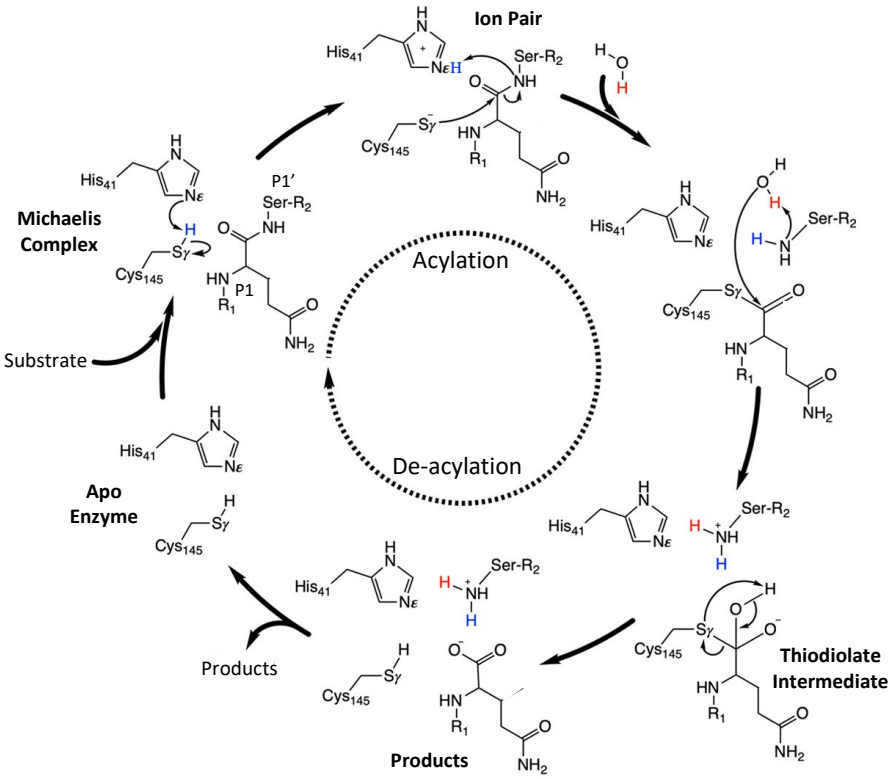

Figure 3. SARS-CoV-2 M^{pro} Mechanism. This mechanism shows the acylation and de-acylation steps necessary for cleavage to occur. The blue hydrogens originate from the catalytic Cys145 while the red hydrogens originate from the incoming water molecule. Figure adapted from Ramos-Guzmán et. al.[25]

The reaction now proceeds with the acylation step. Computational simulations and transition state studies suggest that this step begins with a proton transfer from H41 to the N-terminal nitrogen of the P1' residue as Cys145 gets closer to the substrate.[25] Subsequently, the negatively charged Sγ of Cys145 performs a nucleophilic attack on the carbonyl carbon of P1 (glutamine), causing the cleavage of the scissile peptide bond, and thus the release of P1' (serine).[25] Again, hydrogen bonding to the hydroxyl group of serine results in stabilization, but in this case of the acylation step's transition state, re-emphasizing the preference for serine at the P1' position.[25]

The de-acylation step begins with the introduction of a water molecule, whose proton is abstracted by the N-terminal nitrogen of the, now cleaved, P1' residue. This is immediately followed by a nucleophilic attack on the carbonyl carbon of P1 by the hydroxyl group, ultimately forming a thiodiolate intermediate.[25] Finally, the cleavage of the C(P1)-Sγ bond occurs when the Sγ atom abstracts a proton from the hydroxyl group of the C-terminal end of the P1 and allows for the carbonyl to reform, thus regenerating the enzyme back to its neutral form. It is important to note that the products formed are oppositely charged ions that could potentially form salt bridges. Therefore, it is suggested that water molecules are bound tightly between the products, allowing for their separate release from the active site, and the subsequent reformation of the apo-enzyme.[25]

I.A.5. Short Stretches of Homologous Host-Pathogen Protein Sequences (SSHHPS)

While the primary purpose of SARS-CoV-2 M^{pro} is to hydrolyze pp1a and pp1ab, recent studies have demonstrated the viral cleavage of several host proteins, as well. Furthermore, various *in silico* methods have confirmed the existence of various potential M^{pro} cleavage sites within even more host proteins. Therefore, it is possible that SARS-CoV-2 has the ability to recognize and cleave short pieces of host sequences, which are known as short stretches of homologous host-pathogen protein sequences (SSHHPS).[26] **Figure 4** illustrates this phenomenon by showing how the presence of SSHHPS in the polyproteins of SARS-CoV-2 and in the host protein, results in cleavage of the host protein. This occurrence is thought to be due to the fact that viruses co-evolve with their hosts. Thus, they may have adopted the appropriate sequences into their genomes to allow for the knockdown of proteins that hinder their proliferation.

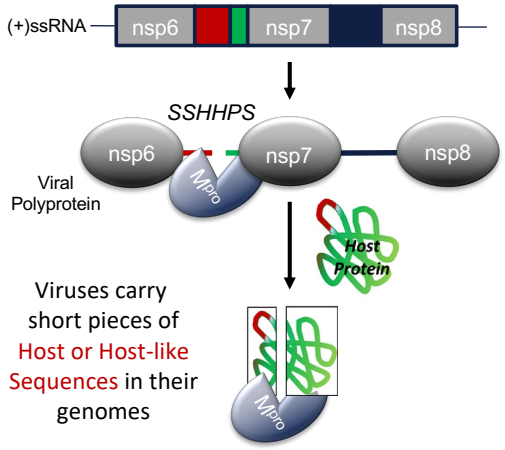

Figure 4. Representation of SSHHPS Involvement in Host Cellular Protein Cleavage by SARS-CoV-2 M^{pro}. The red and green blocks shown at the top represent the RNA sequences that code for the SSHHPS within the viral polyprotein. These SSHHPS make up the M^{pro} cleavage site nsp6/7. In this case, the host protein also contains these SSHHPS, as is portrayed by its red and green color, resulting in the cleavage of both the viral polyprotein and the host protein by SARS-CoV-2 M^{pro}. Figure adapted from Reynolds et al.[26]

As of now, only four host proteins have been confirmed as targets for SARS-CoV-2 M^{pro}. These include three proteins involved in host innate immune responses, namely, NLR Family Pyrin Domain Containing 12 (NLRP12), Interleukin-1 Receptor-Associated Kinase 1 (IRAK1), and TGF-Beta Activated Kinase 1 (TAB1), and one involved in controlling cell development, oncogenesis, and apoptosis, called C-Terminal-Binding Protein 1 (CTBP1).[27,28] Predictably, the proteolytic cleavage of any of these proteins can have disastrous effects on the host, with the most notable being the induction of the cytokine storm, a symptom unnervingly familiar to those who've had COVID-19. Therefore, a positive correlation exists to support the existence of SSHHPS in these cases, and their contribution to the cleavage of these proteins by SARS-CoV-2 M^{pro}.

As depicted in **Table 2**, each proven cleavage site either closely or loosely resembles a canonical cleavage site at the four key positions, P1, P2, P1', and P4. The cleavage sites within NLRP-b, TAB-1, and CTBP1 each contain identical residues to their respective homologs at these positions. Furthermore, the only mismatches in the cleavage sites within TAB1-b and

IRAK1 at these positions are actually homologous to their respective homologs due to the similarities in size and polarity of the residues at P4. Finally, the one with the most mismatches is that of NLRP-a, which has a mismatch with NSP5/6 at P1' and P4. However, since glycine is found in various SARS-CoV-1 cleavage sites at the P1' position and is very small in size, glycine is actually thought to be allowed in the P1' position. The only complete mismatch, therefore, is at the P4 position, but this is thought to be a less important position than the other three key positions.[3] Overall, the host protein cleavage site sequences are consistent with the most important viral cleavage site residues.

Table 2. Known Host Protein SSHHPS/NSPs

Viral Protease	Known Protein SSHHPS/NSP	Cleavage Sites
SARS-CoV-2 M^{pro}	NLRP-a NSP5/6	KLFQ/G VTFQ/S
	NLRP-b NSP9/10	VVLQ/A VRLQ/A
	TAB1-a NSP4/5	ASLQ/S AVLQ/S
	TAB1-b NSP4/5	LTLQ/S AVLQ/S
	CTBP1 NSP14/15	TRVQ/S TRLQ/S
	IRAK1 NSP12/13	STLQ/A TVLQ/A

Notes: The colored letters represent the P1, P2, P4, and P1' positions. The residues that are identical to specific canonical SARS-CoV-2 cleavage sites are in yellow, accepted residues, according to the consensus sequence, are in purple, homologous residues are in green, and unaccepted residues are not highlighted.

These cleavage sites were identified using *in silico* analyses, including NetCorona and the Basic Alignment Search Tool (BLAST). However, not all hits from these tools resulted in a confirmed cleavage, as depicted in Miczi et al.'s study[27]. They initially had four hits: PTK6, PLMN, IRAK1, and CTBP1, but as previously mentioned, were only able to confirm the

cleavages of IRAK1 and CTBP1.[27] Conversely, Moustaqil et al. were able to confirm the cleavages of both hits, NLRP12 and TAB1, from their database screening.[28] This suggests that merely the existence of SSHHPS may not be enough for protein cleavage by SARS-CoV-2 M^{pro}.

Notably, both studies approached incubation and cleavage analysis slightly differently. Upon expression and purification of their recombinant SARS-CoV-2 M^{pro}, Miczi et al. incubated their protease with their target proteins, a His_6-tagged, positive-control (based on the NSP4/5 cleavage site sequence, TSAVLQ/SGFRKM), and a negative control (Bovine Serum Albumin, BSA) for at least one hour at 37º C.[27] They then used sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) to confirm a successful cleavage of the positive control and the targeted proteins. Moreover, they used matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) to identify the actual cleavage positions by utilizing His_6-tagged 12-mer peptides corresponding to the proposed cleavage sites.[27] Alternatively, Moustaqil et al. incubated their reaction for 2.5 hours at 27ºC before using SDS-PAGE to detect a cleavage. They also infected ACE2 cells with SARS-CoV-2 and used western blot to visualize the drop in target protein levels, thus indirectly detecting protein cleavage by SARS-CoV-2 M^{pro}.[28]

I.B. BASIS FOR STUDY

I.B.1. SARS-CoV-2's Role in the Downregulation of Selenoproteins and Other Glutathione-Related Proteins

The Taylor group has long been interested in the link between selenium and viral pathogenesis. Various studies, primarily in China, have shown that people with lower selenium levels tend to exhibit more severe pathogenicity from various RNA viruses, including SARS CoV-2.[3] Biologically, selenium plays key roles in immune and thyroid function, DNA synthesis,

and detoxification. This is due to its incorporation in certain vital proteins, known as selenoproteins, which are heavily involved in essential redox reactions within the body. These special proteins contain at least one selenocysteine, an analogue of cysteine, in which selenium takes the place of sulfur, generating the 21st amino acid. Recent evidence has shown that SARS-CoV-2 causes a downregulation in host cellular selenoproteins, like selenoprotein P (SELENOP), but a tangible mechanism for such suppression by SARS-CoV-2 is only now emerging.[3]

Of interest is a 2020 study by Gordon et al. depicting high-confidence protein-protein interactions between SARS-CoV-2 proteins and human cellular proteins, particularly that of SARS-CoV-2 M^{pro} and GPX1.[3] The researchers found that GPX1 bound strongly to the inactive C145A mutant of SARS-CoV-2 M^{pro}, but showed no interaction with wild-type SARS-CoV-2 M^{pro}.[3] This suggests that GPX1 was cleaved by the wild-type protease because the cleaved protein fragments would no longer have sufficient affinity for SARS-CoV-2 M^{pro}'s active site.[3]Additionally, GPX1 contains a short stretch of amino acids that closely resembles that of the P4-P1' positions of SARS-CoV-2's NSP13/14 cleavage site.

Our hypothesis is that SARS-CoV-2 M^{pro} is targeting selenoproteins, like GPX1, and other glutathione and redox-related proteins, based on the homology between short stretches of their amino acid sequences and those of the canonical cleavage sites at the four key positions discussed previously (AKA their SSHHPS), as well as the accessibility of those residues to M^{pro}. Furthermore, we believe that the targeting of these proteins is due to a viral evolutionary adaptation, allowing for greater proliferation of SARS-CoV-2.

In one of Dr. Taylor's recent publications, he discussed the various *in silico* methods he and his collaborator utilized to further support the proposed cleavage of GPX1 and identify five

more potential host cellular protein targets of SARS-CoV-2, based on their structures. Four of these were selenoproteins, including thioredoxin reductase 1, glutaredoxin 1, selenoprotein F and selenoprotein P. The remaining protein was another glutathione-related protein, like GPX1, called glutamate-cysteine ligase catalytic subunit, the product of which (glutathione) plays a key role in the activation of ribonucleotide reductase, along with TXNRD1 and GLRX-1.[3] Therefore, the goal of this study is to show definitive *in vitro* cleavage of the six proposed host cellular proteins (including GPX1) via incubation with recombinant SARS-CoV-2 M^{pro}.

I.B.2. Structural Basis for Proposed GPX1 Cleavage Site

Taylor et al. used various *in silico* methods to identify at least one potential M^{pro} cleavage site within GPX1. As discussed in the Background, a valid approach was to identify short stretches of sequences resembling the canonical cleavage sites of SARS-CoV-2 M^{pro}. This involved the use of a combination of three protease cleavage prediction tools: NetCorona, Procleave, and PROSPER (the Protease Specificity Prediction Server). Of them, only NetCorona can specifically predict M^{pro} cleavage sites in coronaviruses. It provides a score ranging from 0 to 1 based on sequence homology and structural similarity to known coronaviruses, where anything below 0.5 is unsatisfactory. The other two are limited to predictions of generic cysteine proteases, such as those of the cathepsin family, as shown in **Figure 5**. Procleave compares all the possible sites within a protein to various cysteine protease cleavage sites and ranks them based on both sequence and 3D structural information. While PROSPER works on a scoring system, similar to NetCorona, where anything above 1.0 is said to be a satisfactory score.[29] That being said, any sequences predicted by the latter two prediction tools required manual screening against the consensus sequence, which included all the accepted amino acids for positions P5-P5', as shown in **Figure 6**.

	Site	P4-P1'	*Relative k_{cat}/k_m*	*Protease cleavage prediction methods used to identify and/or confirm site*
A	GCLC	AVLQ/G		**Procleave** #7/592 for cathepsin S
	NSP4/5	AVLQ/S	100%	
	NSP5/6	VTFQ/S	41%	
	NSP6/7	ATVQ/S	3%	
B	SELENOP	ALLQ/A		**Procleave** #3/374 for cathepsin L
	NSP7/8	ATLQ/A	5%	
C	GLRX-1	VSLQ/Q		**PROSPER** (1.30); **Procleave** #1 of 99 for cathepsin S
	NSP8/9	VKLQ/N	2%	
	NSP9/10	VRLQ/A	22%	
	NSP10/11	PMLQ/S	0.2%	
D	TXNRD1	SILQ/A		**PROSPER** (1.14); **NetCorona** (0.640, rank #1)
	NSP12/13	TVLQ/A	8%	
E	SELENOF	TVLQ/A		**PROSPER** (1.02); **NetCorona** (0.846, rank #1)
	NSP13/14	ATLQ/A	9%	
F	GPX1	ASLU/G		**Procleave** #2/196 for cathepsin S
	NSP14/15	TRLQ/S	28%	
	NSP15/16	PKLQ/S	27%	

Figure 5. Protease Cleavage Prediction Methods Used to Identify and/or Confirm Site. The proposed cleavage sites are compared to known cleavage sites within the viral NSPs. The residues that are identical to specific canonical SARS-CoV-2 cleavage sites are in yellow, accepted residues, according to the consensus sequence, are in purple, homologous residues are in green, and unaccepted residues are not highlighted. Figure adapted from Taylor et al.[3]

Unfortunately, since GPX1 has a selenocysteine (U) in the P1 position, instead of the putative glutamine residue, NetCorona failed to identify a site. However, Procleave was able to distinguish the GPX1 active site sequence, ASLU/GTTV, as a highly ranked cleavage site for a similar cysteine protease, called cathepsin S, offering some corroboration.[3] Further screening showed that the P6-P1' sequences of GPX1's active site (NVASLU/G) displayed significant similarity to that of the NSP13/14 cleavage site of SARS-CoV-2 (NVATLQ/A).[3] Although not a direct match, much of the mismatches are quite similar in size and polarity. For example, the residues in the P3 position (serine and threonine) and the P1' position (glycine and alanine) of each corresponding sequence differ by only a methyl group. Furthermore, as discussed previously, glycine is allowed in the P1' position, and thus still falls in line with the consensus

sequence. Conversely, the selenocysteine and the glutamine in the P1' positions are not quite as similar. However, they are both midrange in size and polar in nature. Moreover, several intermediates in the GPX1 mechanism have selenium attached to either an oxygen or a nitrogen.[3] When attached to a nitrogen, it is known as a selenenylamide, which is much closer in structure to glutamine than it is to selenocysteine, thus providing more confidence that this is a valid M^{pro} cleavage site.[3]

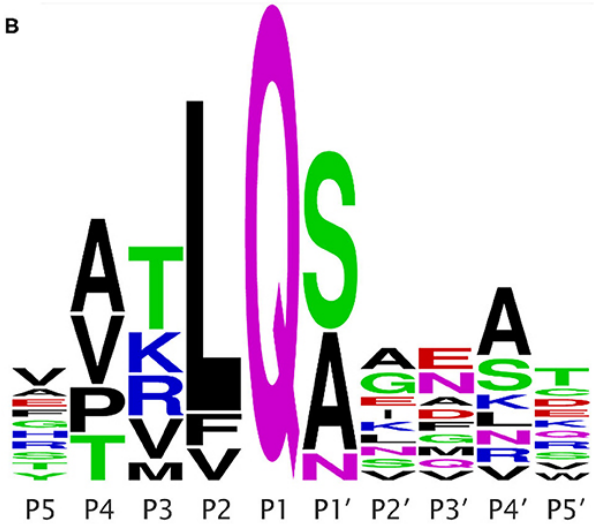

Figure 6. SARS-CoV-2 M^{pro} Consensus Sequence. Depicted are the amino acids that comprise the canonical SARS-CoV-2 M^{pro} cleavage sites. The size of each letter reflects the probability of the alignment. Figure from Taylor et al.[3]

To further increase the validity of NVASLU/G as an M^{pro} cleavage site, Taylor et al. also used 3D modeling to determine whether the proposed site would even be accessible to the protease for cleavage. Ideally, it would be located on the surface of the protein, where it would be easily accessible, rather than in the interior, where the likelihood of cleavage would diminish dramatically regardless of the SSHHPS. Fortunately, the proposed site was shown to be located on the surface of GPX1, as highlighted in **Figure 7**, thus further supporting NVASLU/G as a potential cleavage site and GPX1 as a target for SARS-CoV-2 M^{pro}.

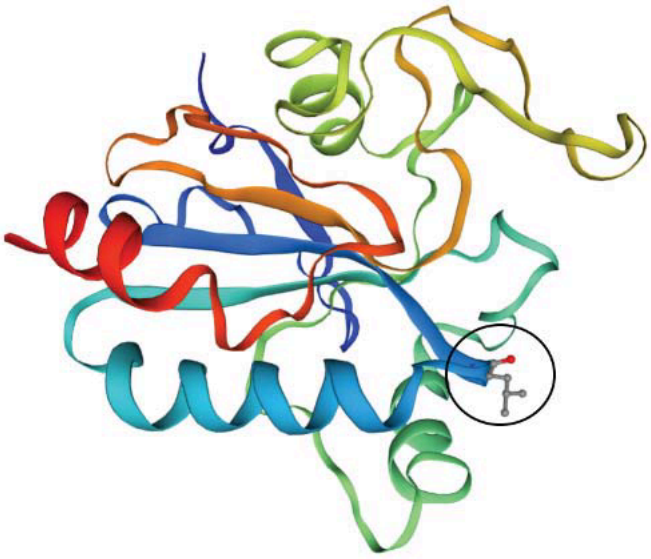

Figure 7. X-Ray Structure of GPX1. The proposed M^{pro} cleavage site (ASLU/G) is circled and the Leucine at P2 is displayed. Figure from Taylor et al.[3]

I.B.3. Selection and Identification of the Five Remaining Proteins and their Proposed Cleavage Sites

Apart from GCLC, the remaining proteins examined in this research were selected based on a combination of their selenoprotein character and their homology to the canonical SARS-CoV-2 M^{pro} cleavage sites. This approach was used to further support the idea that SARS-CoV-2 M^{pro} was targeting many selenoproteins. Upon realizing that some of the proposed proteins were involved in the activation of ribonucleotide reductase, various other proteins in that pathway were screened for potential cleavage sites, resulting in the proposal of GCLC as a potential target, as well.

The identification of potential M^{pro} cleavage sites within these proteins followed a similar approach as for that of GPX1. The sequences of those proteins were run through a combination of prediction methods, discussed above, in order to determine the best hits and the comparative viral NSP sequences. Notably, if NetCorona failed to determine a cleavage site, giving a score of less than 0.5, then Procleave and/or PROSPER were used. However, because they only predicted

cleavage sites in generic cysteine proteases, as shown in **Figure 5**, further screening against the consensus sequence was needed. If similar enough, as with the P3 and P1' sites of GPX1 discussed previously, that sequence was determined as a potential cleavage site. Finally, before being ready for *in vitro* assessment, 3D models of the proteins were obtained/constructed to ensure that the proposed cleavage site was even accessible to M^{pro} for cleavage.

I.B.4. Structural Basis for Proposed TXNRD1 Cleavage Site

When screening for potential cleavage sites in thioredoxin reductase 1, PROSPER and NetCorona both yielded high scores for the sequence, SILQ/A. NetCorona yielded a score of 0.640 and PROSPER a score of 1.14. Upon examination of this site, one could see that it most closely resembled the NSP12/13 cleavage site (TVLQ/A) of SARS-CoV-2 M^{pro}, with the only mismatches being at the P4 and P3 positions. However, looking closely at these mismatches, it was evident that the SI sequence in TXNRD1 is isosteric to the TV sequence in NSP12/13.[3] This is due to the fact that serine has one less methyl group than threonine and isoleucine has one more methyl group than valine. Therefore, they would occupy the same volume in the active site. Additionally, the proven host M^{pro} target, IRAK1, also has a Ser in the P4 position; in fact, the IRAK1 and predicted TXNRD1 sequences only differ in the P3 position (STLQ/A vs SILQ/A) Furthermore, the X-ray structure of TXNRD1 showed that SILQA was located at the surface of the protein, as shown in **Figure 8**, making it an excellent target.[3]

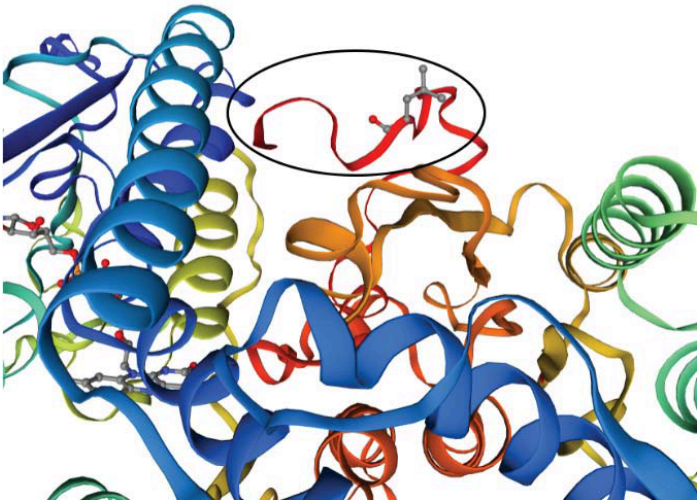

Figure 8. X-Ray Structure of TXNRD1. The proposed M^{pro} cleavage site (SILQ/A) is circled and the Leucine at P2 is displayed. Figure from Taylor et al.[3]

I.B.5. Structural Basis for Proposed GLRX-1 Cleavage Site

When screening for potential cleavage sites in glutaredoxin-1, PROSPER and Procleave each generated high rankings for the sequence VSLQ/Q. PROSPER gave it a score of 1.30 and Procleave ranked it #1 of the 99 potential sites within GLRX-1 when comparing it to the cysteine protease, cathepsin S, as depicted in **Figure 5**.[3] Due to the lack of NetCorona information, this site had to be further screened, and as depicted in Figure 7, it most closely resembled NSP8/9 (VKLQ/N), where the only mismatches were at the P1' position, and the less important P3 position. At the P1' position GLRX-1 has a glutamine, while NSP8/9 has an asparagine, which are both homologous, differing by one methyl group, thus making this a likely cleavage site. Moreover, as depicted in the NMR structure of **Figure 9**, VSLQQ was also shown to be located at GLRX-1's surface.

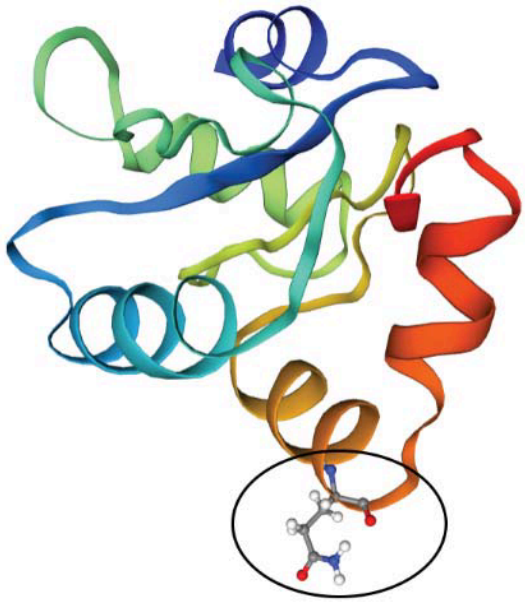

Figure 9. NMR Structure of GLRX-1. The proposed M^{pro} cleavage site (VSLQ/Q) is circled and the Glutamine at P1 is displayed. Figure from Taylor et al.[3]

I.B.6. Structural Basis for Proposed GCLC Cleavage Site

When screening for potential cleavage sites in GCLC, Procleave ranked the sequence AVLQ/G as #7/592 for cathepsin S.[3] While it was not the top ranked sequence when compared to cathepsin S, it was nearly identical to the NSP4/5 cleavage site of SARS-CoV-2 M^{pro} (AVLQ/S), with the only mismatch being in the P1' position. Although serine and glycine are not homologous, and glycine is technically not included in the consensus sequence, there are a few arguments for why this is still an acceptable residue at this position. Firstly, glycine is homologous to alanine, differing by only one methyl group, and alanine is included in the consensus sequence, as shown in **Figure 6**. Secondly, the S1' subunit is a very shallow pocket, as was highlighted when discussing the M^{pro} structure. Therefore, having a small residue, like glycine in this position would be advantageous. Finally, glycine is actually included in the consensus sequence for SARS-CoV-1 M^{pro}, which is highly similar to that of SARS-CoV-2.

This, thus increases the possibility that SARS-CoV-2 M^{pro} would allow glycine at this position. Furthermore, AVLQG is located at the surface of GCLC, as shown in **Figure 10**.[3]

Figure 10. Homology Model of GCLC. The proposed M^{pro} cleavage site (AVLQ/G) is circled and the Glutamine at P1 is displayed. Figure from Taylor et al.[3]

I.B.7. Structural Basis for Proposed SELENOP Cleavage Site

When screening for potential cleavages sites in SELENOP, the sequence ALLQ/A ranked #3/374 for cathepsin L via Procleave. More importantly, this sequence was nearly identical to that of NSP7/8, with the only mismatch being at the least important, P3 position. Therefore, this sequence is highly likely to be a SARS-CoV-2 M^{pro} cleavage site. Although the 3D model for this particular site was not available, it is thought to be located right before a redox center active site, thus suggesting a high probability of surface accessibility.[3]

I.B.8. Structural Basis for Proposed SELENOF Cleavage Site

When screening for potential cleavages sites in SELENOF, the sequence TVLQ/A scored high for both Prosper (1.02) and NetCorona (0.846).[3] In fact, it ranked #1 for NetCorona and the reasoning is evident. This sequence is 100% identical to that of NSP12/13 (TVLQ/A). Therefore, of all the potential cleavage sites stated, this is by far the most likely to be cleaved by SARS-CoV-2 M^{pro}. Unfortunately, along with SELENOP, the 3D model for this site was not available, but this site is near the N-terminal ER signal peptide cleavage site of SELENOF, thus also suggesting a high probability of surface accessibility.[3]

I.C. SIGNIFICANCE

I.C.1. Public Health

The emergence of new SARS-CoV-2 variants has made it evident that a more holistic approach is necessary to efficiently combat this virus as it continues to evolve. Instead of focusing on one mutable characteristic of SARS-CoV-2, such as the spike protein, which has undergone many mutations, we must also uncover and subsequently disrupt the various other mechanisms and machinery that drive its proliferation and pathogenicity.

Our study seeks to uncover unknown viral mechanisms driven by SARS-CoV-2 M^{pro}. While primarily used in the hydrolysis of two overlapping polyproteins (pp1a and pp1ab) encoded by the virus' replicase gene, our group is investigating its possible proclivity for a selected group of host proteins based on their sequence similarity to viral M^{pro} cleavage sites. Each of the selected host selenoproteins/glutathione-related proteins play vital roles in essential host processes, some of which impact fundamental viral processes. Therefore, positive results for

any of the chosen host proteins could have broad implications for the ways we view prevention, treatment, and the evolution of SARS-CoV-2.

I.C.2. Disruption of DNA Synthesis and Repair

Three of the proposed host proteins play vital roles in the redox regulation of ribonucleotide reductase via the thioredoxin and glutaredoxin systems. These include the selenoproteins, glutaredoxin-1 and thioredoxin reductase 1, and the enzyme responsible for glutathione synthesis, GCLC. Glutaredoxin-1 and GCLC are part of the glutaredoxin system while thioredoxin reductase 1 is a part of the thioredoxin system, each of which are essential for the function of ribonucleotide reductase (RNR). RNR is the rate-limiting enzyme in DNA synthesis responsible for catalyzing the reduction of ribonucleotides to deoxyribonucleotides.[30] However, in order for this to occur, the enzyme must be provided with reducing equivalents via either the thioredoxin or glutaredoxin systems.

The reduction of ribonucleotides by RNR involving the thioredoxin pathway is a short process, utilizing three key molecules: NADPH, TXNRD1, and thioredoxin (TRX). It begins with the reduction of thioredoxin by NADPH, catalyzed by thioredoxin reductase 1. The newly reduced thioredoxin is what then reduces RNR, thus activating it for the reduction of ribonucleotides. The oxidized thioredoxin is then recycled, allowing for the cycle to begin again. RNR reduction via the glutaredoxin system is more complicated, utilizing nine key molecules: glutamate (Glu), cysteine (Cys), glycine (Gly), GCLC, glutathione synthetase (GSS), glutathione (GSH), GLRX-1, glutathione reductase (GR), and NADPH. This process begins with the biosynthesis of two glutathione molecules generated via the ligation of glutamate and cysteine, forming γ-glutamylcysteine (γ-GC), followed by the coupling of γ-GC with glycine.[31] GCLC catalyzes the Glu-Cys ligation while GSS catalyzes the γ-GC-Glycine coupling, but GCLC is the

rate-limiting enzyme for this process.[31] The resulting glutathione molecules are then used to reduce the oxidized form of GLRX-1, which subsequently reduces RNR, and like thioredoxin is recycled in the system. As depicted in **Figure 11**, the reduced form of glutathione (GSH) is necessary for this cycle to continue. Therefore, due to GSH's utilization as a cofactor in many biological processes, this system has two ways of replenishing it in the case that it is allocated to another biochemical system. First, is the aforementioned biosynthesis of GSH catalyzed by GCLC and the second is the reduction of the newly oxidized glutathione molecules (GSSG) by NADPH, which is catalyzed by GR.

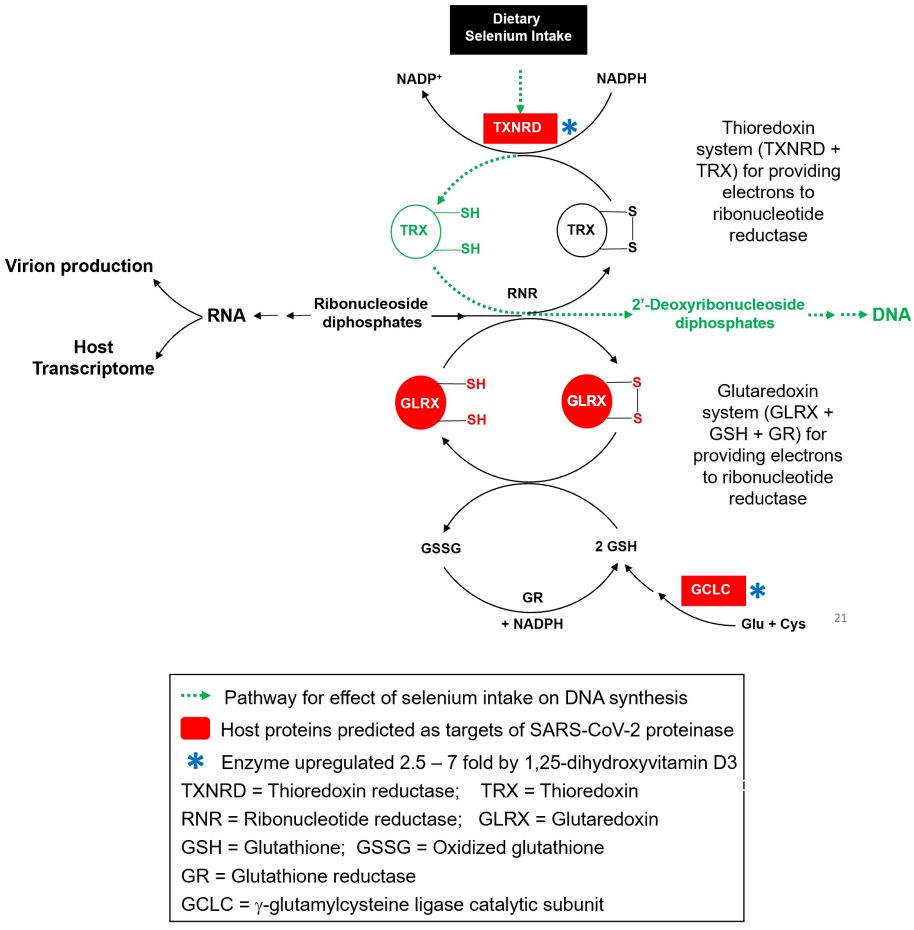

Figure 11. Role of Proposed Proteins in RNR Activation. Two systems are responsible for the reduction of RNR, including the thioredoxin system (above) and the glutaredoxin system (below). TXNRD1 is essential to the thioredoxin system while GLRX and GCLC are essential to the glutaredoxin system. Also shown are the roles that dietary selenium and vitamin D supplementation play in the upregulation of selenoproteins and GCLC. Figure adapted from Drs. Taylor and Chiu.

As portrayed in their respective mechanisms, the thioredoxin and glutaredoxin systems function independently of each other. This provides a fail-safe for the reduction of RNR in the case that one of them is inhibited. Yet, if both systems are inhibited, as is proposed by the proteolytic knockdown of TXNRD1, GLRX-1, and GCLC by SARS-CoV-2 M^{pro}, RNR will not have the reducing power necessary to convert ribonucleotides to deoxyribonucleotides, resulting in impaired DNA synthesis/repair. In fact, the proposed cleavage site of TXNRD1 is located at its C-terminal redox center, thus unquestionably inhibiting its ability to activate RNR.[3] This would lead to increased levels of oxidative stress within the cell, eventually inducing apoptosis (programmed cell death)[30]. Although apoptosis is a natural process that occurs regularly in healthy individuals, too much can lead to tissue death, exposing individuals to a wide range of health issues.

One can also deduce that SARS-CoV-2 confers a significant advantage when RNR is inhibited due to its reliance on host ribonucleotides for RNA replication. By diminishing the allocation of ribonucleotides for DNA synthesis/repair, SARS-CoV-2 can utilize the greater supply of ribonucleotides to further enhance its proliferation within the cell, as depicted in **Figure 11.**[3] Therefore, the proteolytic knockdown proposed would not only lead to greater pathogenicity but also to enhanced viral proliferation.

I.C.3. Disruption of Glycoprotein Folding

SelenoF plays a considerable role in the regulation of glycoprotein folding in the endoplasmic reticulum (ER) by enhancing the enzymatic activity of UDP-glucose:glycoprotein glucosyltransferase (UGGT).[32] Consequently, its knockdown would result in significant ER stress, inducing the cytokine storm associated with COVID-19 infection. Remarkably, Wang et al. demonstrated that cells infected with SARS-CoV-2 resulted in a downregulation of SelenoF

mRNA, by 75.9%.[33] Therefore, knockdown at the gene and potentially the protein level, suggests that SARS-CoV-2 confers some kind of advantage from its downregulation. Such an advantage could stem from the fact that nearly all of the key molecules involved in the innate and adaptive immune response depend on glycoproteins for proper functioning.[34] For example, certain glycoforms are necessary for the assembly of peptide-loaded major histocompatibility complex (MHC) antigens and the T-Cell receptor complex.[34] As a result, inhibition of SelenoF could potentially provide SARS-CoV-2 with uninterrupted access to host cells.

I.C.4. Disruption of Hydrogen Peroxide Detoxification

GPX1, along with many other selenoproteins, plays a vital role in redox regulation within the cell.[35] Specifically, it is known as one of the most important antioxidant enzymes in the human body due to its role in the detoxification of hydrogen peroxide.[35] The proteolytic knockdown of GPX1 would therefore leave the human cell susceptible to reactive oxygen species (ROS), thus increasing the levels of oxidative stress, eventually resulting apoptosis.

I.C.5. Disruption of Selenium Transport and Reduction of Phospholipid Hydroperoxides

The final host protein of interest is SelenoP, whose knockdown could theoretically inhibit the activity of all the aforementioned selenoproteins (TXNRD1, GLRX-1, SelenoF, and GPX1). This is due to its key role in selenium transport within the body. SelenoP functions as a selenium carrier in human plasma, thus supplying the selenium necessary to maintain adequate levels of various selenoenzymes found in human tissues, including those described above.[36] Additionally, SELENOP is also essential for the reduction of phospholipid hydroperoxides.[36] Therefore, if proven as a SARS-CoV-2 M^{pro} target, COVID-19 infection could potentially inhibit DNA synthesis/repair, regulation of glycoprotein folding, and protection from reactive oxygen species,

conferring multiple advantages for SARS-CoV-2 and several disadvantages for infected individuals.

I.C.6. Potential Measures Taken to Counteract Proposed Effects of SARS-CoV-2 M^{pro}

The extent of the pathogenicity and viral proliferation generated via cleavage of the six proposed proteins may be mediated by micronutrient supplementation. Studies have shown that vitamin D acts as a "potent activator" of TXNRD1 and GCLC, thereby upregulating their production in the cell, as shown in **Figure 11**.[3] Additionally, since TXNRD1, GLRX1, GPX1, SelenoF, and SelenoP are selenoproteins, dietary intake of selenium may also result in their upregulation. Thus, supplementation with dietary Selenium and vitamin D may be able to partially offset the proteolytic knockdown of these proteins, allowing for the neutralization/diminution of viral proliferation and pathogenicity.

CHAPTER II: EXPERIMENTAL

II.A. APPROACH

II.A.1. Utilization of Peptides for Preliminary Assessment of Proposed Cleavage Sites

10-12-mer peptide sequences incorporating the proposed cleavage sites of each chosen protein were used initially to assess the validity of the cleavage site itself. This acted as a "proof of concept," allowing for another researcher in the Taylor group, David Stubbs, to analyze the cleavage of the full proteins, (apart from TXNRD1), via western blot. One of the reasons for employing this approach instead of directly analyzing all the proposed proteins was due to its cost-effectiveness. Buying all six proteins would have been exceptionally expensive while providing me with a minute quantity of substrate to work with. Conversely, the utilization of 10-mer/12-mer peptides afforded me with the ability to optimize my protocol through various trials, using different concentrations of each substrate. This also allowed me to account for human error and drops in protease activity in order to definitively rule out a proposed cleavage site.

II.A.2. Replacement of Selenocysteine in TXNRD1 for Serine

Although the WT TXNRD1 protein has a selenocysteine at the P4' position of its proposed cleavage site (ASILQ/AGCUG), it had to be mutated for this research in order for TXNRD1 to be expressed in bacteria. This is due to the complicated eukaryotic co-translational mechanism necessary for the incorporation of selenocysteines at UGA codons that only a few labs have the capability to replicate. Therefore, the U was substituted out for an S in both the TXNRD1 peptide and protein.

The reasons for choosing serine over another amino acid, sharing greater similarities to Sec are two-fold. First, it is known that serine is sometimes inserted at UGA codons under low Se conditions, thus being the most natural choice. Secondly, is the presence of the geminal cysteine residue that would have led to the possibility of both disulfide and dithiol forms being present, thus confounding the mass spectrometry results.

II.A.3. Utilization of UPLC-MS for Cleavage Analysis

While western blot is effective at detecting a cleavage, it is not as effective at definitively identifying a cleavage site. UPLC-MS, on the other hand, is remarkably precise at identifying specific fragments based on their monoisotopic masses and charge states due to its remarkable sensitivity and selectivity. For example, when analyzing the cleavage of the GCLC, RDAVLQ / GMFYFR, the fragments within the sample solution injected would separate based on their affinity to the column and should therefore each have their own retention times (RT). This provides enhanced selectivity, allowing for one to distinguish between sample components. Therefore, when searching for the mass-to-charge ratio (m/z) of either fragment, one can distinguish said fragment from another molecule in the sample with the same m/z value based on differing retention times. Additionally, one can use that fragment's characteristic retention time to compare the chromatograms of the control sample (peptide sample not incubated with M^{pro}) against that of the test sample (peptide sample incubated with M^{pro}). If the base peak for a particular m/z value was observed at the same retention time, then the molecule observed most likely did not correspond to cleavage by M^{pro}. Furthermore, the high pressure (about 15,000 psi) with which the UPLC operates allowed for increased sensitivity, thus allowing it to detect fragments at miniscule concentrations.

II.A.4. Sole Analysis of the TXNRD1 Protein

Of the six proposed proteins, only TXNRD1 is fully analyzed. The reason for this is due to the location of the proposed cleavage site of TXNRD1 compared to those of the remaining five proteins. As depicted in **Figure 12**, the proposed cleavage site of TXNRD1 is located five residues away from its C-Terminus while the rest have cleavage sites near the middle of their sequences. A more central cleavage site poses a problem it results in two large fragments that will ultimately be broken down into various fragments when analyzed via UPLC-MS. due to the Mass Spec process. Therefore, although not impossible, it will be quite difficult to distinguish between the fragments cleaved by M^{pro} and those fragmented by the mass spectrometer. Conversely, the cleavage of TXNRD1 will result in a large P fragment and small P' fragment. In fact, the P' fragment will be the same as that which was analyzed in the preliminary assessment (AGCSG). Thus, the cleavage can be easily confirmed via detection of the same P' fragment as that of the TXNRD1 peptide.

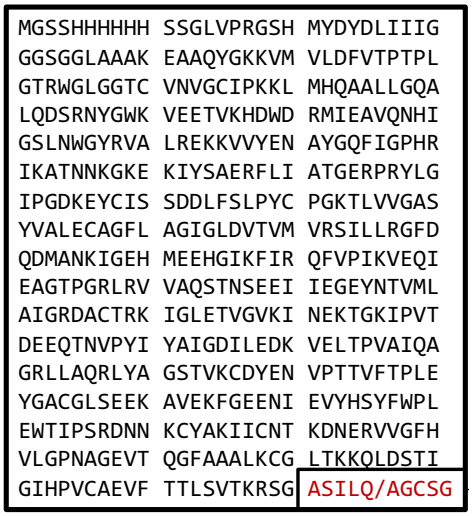

Figure 12. Proposed Cleavage Site of TXNRD1. The blocked-in region highlighted in red represents the P5-P5' positions of the proposed M^{pro} cleavage site.

The issue involving can be easily resolved via utilization of western blot, which will result in either the presence of one band, suggesting no cleavage occurred, or two bands, suggesting a cleavage did occur. Therefore, the implementation of both analytical methods is imperative, UPLC-MS will confirm the proposed cleavage sites in the chosen peptides, and western blot will confirm the cleavage of the corresponding proteins. Together, these two pieces of information will lend profound credibility to our research.

II.B. METHODOLOGY

II.B.1. Assessment of Proposed Cleavage Sites

The assessment of the chosen peptides as targets for SARS-CoV-2 M^{pro} took place in six steps: buffer preparation, M^{pro} preparation, substrate (peptide) preparation, sample preparation, and analysis via UPLC-MS. For simplicity, this procedure is described assuming only one proposed peptide is being assessed at a time.

To begin, 100 mL of buffer was prepared using 20 mM of Tris-HCl at a pH of 7.5, 110 mM NaCl, 1mM DTT, and 1mM EDTA, all dissolved in about 100 mL of UltraPure water. Next, a µL solution of M^{pro} solution was prepared. 2 µL of SARS-CoV-2 M^{pro}, reconstituted at a concentration of 1 mg/mL, was serially diluted in order to achieve an overall concentration of 5 ng/µL. The serial dilution was done in a 1:10, 2:20, 3:30, 16:80, and 60:180 stepwise fashion, resulting in a dilution factor of 15,000. Subsequently, a 200 µM solution of each substrate was prepared, including the positive control and the proposed peptide. Each peptide, including the positive control, was synthesized by GenScript. Since each peptide has a different molecular weight, the 200 µM solutions are prepared using slightly different masses of peptide and volume of UltraPure water, but on average are 1 mg and 3 mL, respectively.

Seven, 50 µL samples were then prepared in 0.2 mL microcentrifuge tubes, incorporating the four solutions described above. These include two blanks, one SARS-CoV-2 M^{pro} sample, one positive control sample with M^{pro}, one positive control sample without M^{pro}, one proposed peptide sample with M^{pro}, and one proposed peptide sample without M^{pro}. The reason for this was to assure that any background noise detected in the UPLC-MS analysis was not mistaken for cleaved fragments of the assessed peptides. The compositions of the samples are detailed in **Table 3**. Upon completion of the sample prep, they were each incubated overnight in a PCR Thermal Cycler at 37º to mimic physiological conditions.

Table 3. Sample Compositions for Peptide Cleavage Assessment

Sample	Composition
Blank 1	50 µL of buffer solution
SARS-CoV-2 M^{pro} Sample	30 µL of SARS-CoV-2 M^{pro} solution and 20 µL of buffer solution
Positive Control Sample w/o M^{pro}	10 µL of the positive control solution and 40 µL of the buffer solution
Positive Control Sample w/ M^{pro}	10 µL of the positive control solution, 30 µL of the SARS-CoV-2 M^{pro} solution, and 10 µL of buffer
Proposed Peptide w/o M^{pro}	10 µL of the proposed peptide solution and 40 µL of the buffer solution
Proposed Peptide w/ M^{pro}	10 µL of the proposed peptide solution, 30 µL of the SARS-CoV-2 M^{pro} solution, and 10 µL of buffer
Blank 2	50 µL of buffer solution

Finally, the samples were analyzed via UPLC-MS. In this step, all samples were transferred to autosampler vials, which were then placed in a solvent tray before ultimately being inserted into the Q-Exactive Orbitrap Mass Spectrometer for analysis. The Liquid Chromatography set-up was a BEH- C18 column (50 mm x 2.1 mm) with 0.1% formic acid in water (A) and acetonitrile (B) for the mobile phase. Ultimately, the Mass Spectrometry software, XCalibur, was used to analyze the results.

II.B.2. Assessment of TXNRD1 Protein Cleavage

The procedure for the detection of TXNRD1's cleavage was nearly identical to that of the proposed peptides. The main differences were in the synthesis of the substrate, concentration of the substrate, and the identification of the fragment via UPLC-MS analysis. Initially, the TXNRD1 protein was synthesized by SeLENOZYME because they were one of the only companies that could synthesize TXNRD1 with the characteristic selenocysteine at an affordable price. However, due to the miniscule quantity provided, only one trial was possible. Therefore, for subsequent assays recombinant TXNRD1 was heterologously expressed in-house, but with the selenocysteine substituted out for a serine for reasons previously discussed.

This was a multi-step process, beginning with designing the primers to be synthesized by Twist Bioscience. LB Agar plates were then streaked with strands of DNA and incubated overnight. The next step was to grow 5 mL of starter culture and incubate overnight. The Promega Wizard Genomic DNA Purification Kit was then used to obtain the genomic DNA, which was then run in the PCR, along with the primers. Gel electrophoresis was then utilized on the PCR products, followed by PCR purification, using the AIQuick PCR Purification Kit. Gel electrophoresis was then run again, and the intensities were calculated. The next step was to clone the gene into pET-28a plasmids using TOP10 then transforming them into competent *E. coli* cells and incubating them overnight. Subsequently 5 mL starter cultures were grown and incubated overnight, followed by plasmid purification, using the QIAprep Spin Mini Prep Kit. Then PCR was performed on the purified plasmid, followed by gel electrophoresis on the PCR products. The next step was to clone this DNA into BL21* and incubate overnight. After that, 1 L of culture was induced and spun down at 6500 rpm, after which the pellets were stored in the -80. The pellets were then suspended in 10mL of 1x binding buffer, and the cells were lysed via

sonication on ice. The products were then spun down at 16000g and the supernatant was filtered with syringe filters. Finally, the resulting sample was columned to obtain the protein.

The concentration of TXNRD1 used for the peptide sample in the assay was much lower than the 50 μM used prior because the aforementioned process only yielded about 1-2 mg/mL. In attempts to remedy this, no M^{pro} buffer was used, and instead 10 more μL of the protein sample was used in its place. Therefore, the sample contained 30 μL of the M^{pro} solution and 20 μL of the TXNRD1 protein. As discussed in the Approach, the cleavage of TXNRD1 was analyzed via UPLC-MS, but instead of searching for two fragments in the chromatogram, the cleavage was based on only the P' fragment (AGCSG). Otherwise, the protocol for the analysis of TXNRD1 was the same as for that of the peptides.

CHAPTER III: RESULTS AND DISCUSSION

III.A. Confirmed *In Vitro* Cleavage

III.A.1. Positive Control

A 12-mer positive control for this assay was synthesized by GenScript to emulate the NSP4/NSP5 cleavage site, TSAVLQ/SGFRKM. This was the ideal positive control for all analyses in this research due to the 100% relative catalytic efficiency with which SARS-CoV-2 M^{pro} cleaves this site, as discussed previously.

In order to accurately detect the cleavage for this sample, the ion chromatogram illustrated in **Figure 13** focused on the base peaks of the in-tact peptide (M_{mi} = 1324.71) and the two cleaved fragments, TSAVLQ (M_{mi} = 618.35) and SGFRKM (M_{mi} = 725.38). This was done by inputting specific m/z ranges into XCalibur that corresponded to the highest relative abundance for each monoisotopic mass, depicted above. The mass-to-charge ratio that corresponded to the highest relative abundance for the in-tact Positive Control was 662.86, meaning it was double-charged. Therefore, the m/z ranges used to illustrate the drop in the in-tact peptide abundance upon incubation with recombinant SARS-CoV-2 M^{pro} were 662.85-662.87.

The ion chromatograms in **Figure 13A** illustrate a vast difference in the heights and areas of the base peaks when analyzing the In-Tact Peptide sample without protease versus that with protease at a normalization level of 1.04E8. A comparison of the two areas shows that the peptide sample containing the protease exhibited about a 91.2% drop in abundance, making this an ideal positive control for this study. Furthermore, the mass spectrum, depicted at the bottom of **Figure 15A**, shows the peaks observed at the t_R range of 3.51-3.67 minutes, corresponding to

the base peaks in the ion chromatogram. Each peak observed is either the single, double, or triple-charged states of the In-Tact Peptide, thus confirming the identity of each base peak.

Figures 15B and C show that the P (m/z = 618.35) and the P' (m/z = 363.19) fragments, were observed in the + Protease sample, but not in the – Protease sample. This information allowed us to definitively confirm that not only did a cleavage occur, but it occurred at the correct position. The peaks in the mass spectra below their respective chromatograms further confirm the identities of these fragments.

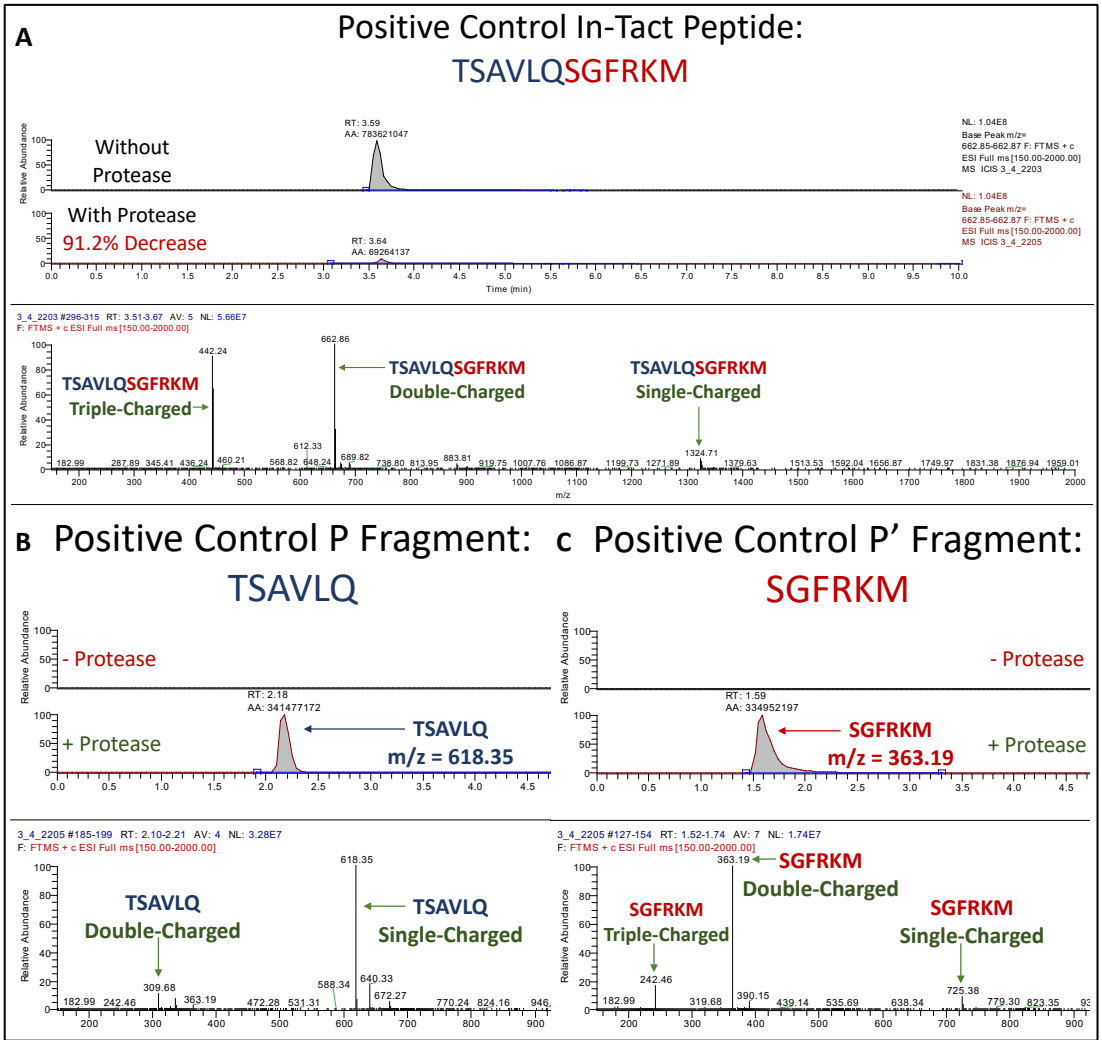

Figure 13. LC-MS Data for Positive Control. Depicted are the ion chromatograms for the samples with and without protease (top) and the mass spectra (bottom) for the (A) In-Tact Peptide, (B) the P Fragment, and (C) the P' Fragment. The mass spectra were added to corroborate the identities of each base peak in the corresponding Ion-Chromatogram.

III.A.2. TXNRD1 Peptide

The results generated from the analysis of the TXNRD1 peptide were promising. Although, as shown in **Figure 14A**, incubation with the protease resulted in only a fairly small decrease (21.4%) in the amount of In-Tact Peptide in the test sample (with protease) when compared to that of the control sample (without protease). However, this small decrease can be explained in three ways. Firstly, as stated previously, this sequence (SILQA) had mismatches when compared to that of NSP12/13 (TVLQ/A) at the P4 and P3 positions. Although the residues at these sites are both homologous and isosteric, as previously discussed, serine is not included in the consensus sequence at P4, which could potentially interfere with the cleavage. Secondly, NSP12/13 is known to be cleaved by SARS-CoV-2 M^{pro} with an extremely low catalytic efficiency (8%). Therefore, it is possible that the small drop could be due to M^{pro}'s low efficiency at cleaving NSP12/13. Lastly, it is possible that the substitution of the selenocysteine to serine may have also caused M^{pro} to cleave the peptide inefficiently. Or it could also have been a combination of each of these reasons.

More importantly, **Figures 14B and C** show that the P and P' fragments (m/z= 531.31 and 394.14, respectively) for this peptide were observed at high intensities in the test sample and not at all in the control. Additionally, when looking at the mass spectrum of **Figure 14C**, it is evident that the peak corresponding to AGCSG is much lower than the base peak. This is due to the number of other compounds that were eluting off of the column at the same retention time, as shown in **Figure 14D**. Therefore, the miniscule relative abundance of the P' fragment eluting off at this RT when compared to those compounds makes it appear that it is barely present in the sample, when in-reality it is present at similar quantities as that of the P fragment.

Overall, this information provided confirmation of the TXNRD1 peptide's cleavage by the protease at the proposed cleavage site, thus providing support for our selenoprotein hypothesis and that of RNR inhibition. Proteolytic cleavage at this position could potentially inhibit the thioredoxin system, leaving the bulk of RNR activation to the glutaredoxin system.

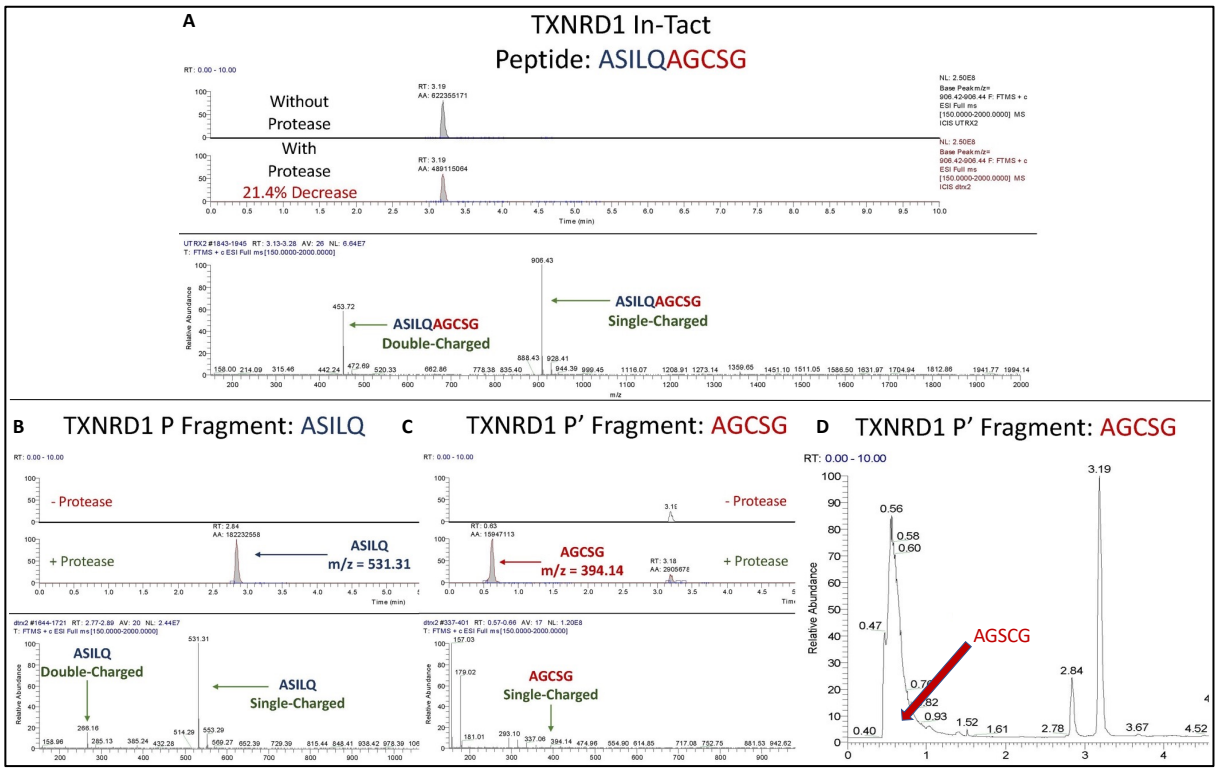

Figure 14. LC-MS Data for TXNRD1 Peptide. Depicted are the ion chromatograms for the samples without protease and with protease (top) and the mass spectra (bottom) for the (A) In-Tact Peptide, (B) the P Fragment, and (C) the P' Fragment. The base peaks for the P and P' Fragments had normalization levels of 2.5E8 and 6.15E7, respectively. The low AGCSG peak in the mass spectrum is explained by (D) the large number of compounds eluting off at the same retention time. The mass spectra were added to corroborate the identities of each base peak in the corresponding Ion-Chromatogram.

III.A.3. GCLC Peptide

The results for the GCLC peptide turned out to be even more promising that those of TXNRD1. **Figure 15A** depicts a 40.8% decrease in the amount of In-Tact peptide when compared to the control sample, suggesting that M^{pro} cleaved this peptide with great efficiency.

This makes sense since the GCLC peptide sequence (AVLQ/G), like the positive control, most closely resembles that of NSP4/5 (AVLQ/S), which has a 100% catalytic efficiency. Of course, when compared to the results of the positive control, the drop was lower due to the mismatch at the P1' position. As discussed in the proteolytic mechanism section, having a serine in the P1' position leads to the most efficient cleavage due to the stabilization offered via hydrogen bonding. Furthermore, glycine is not present in the consensus sequence for SARS-CoV-2 M^{pro}.

Figures 15B and C provide further confirmation that a cleavage took place at the proposed site due to the intense presence of the anticipated P and P' fragments, whose m/z values are 351.20 and 410.69, respectively, and the complete absence of these fragments in the control samples. This information provides further proof that SARS-CoV-2 is inhibiting RNR via both the thioredoxin and glutaredoxin systems.

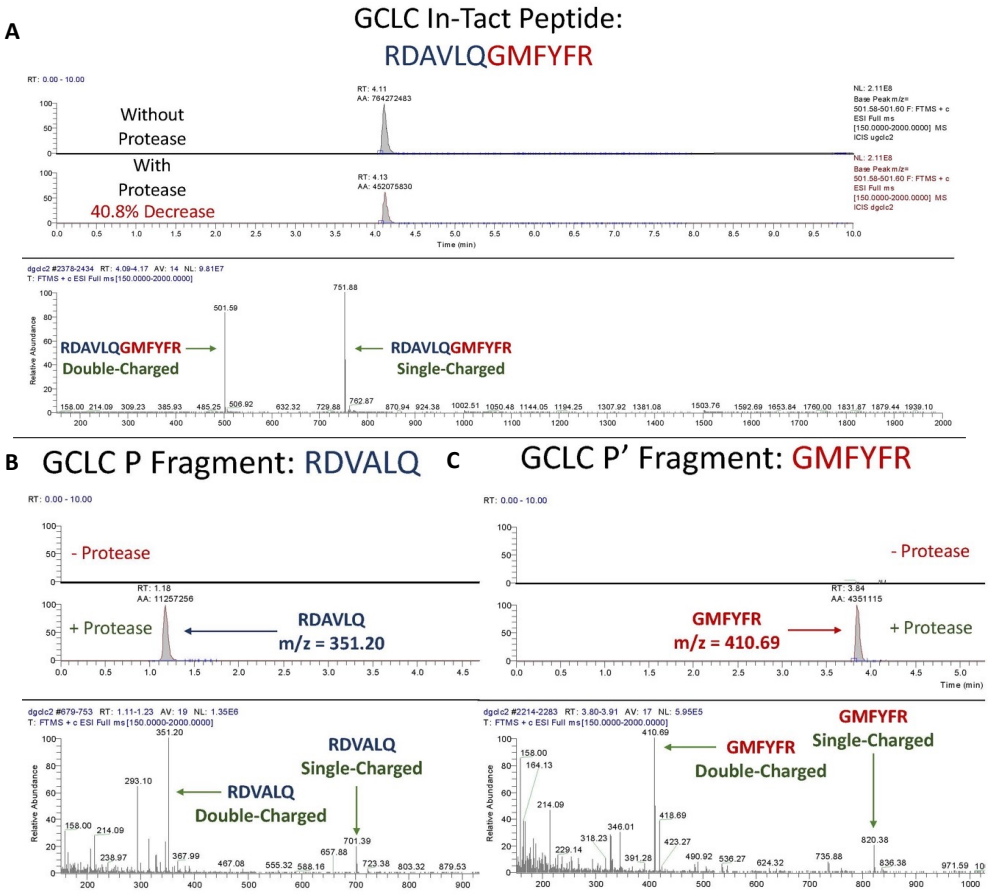

Figure 15. LC-MS Data for GCLC Peptide. Depicted are the ion chromatograms for the samples without protease and with protease (top) and the mass spectra (bottom) for the (A) In-Tact Peptide, (B) the P Fragment, and (C) the P' Fragment. The base peaks for the P and P' Fragments had normalization levels of 3.20E6 and 1.40E6, respectively. The mass spectra were added to corroborate the identities of each base peak in the corresponding Ion-Chromatogram.

III.A.4. SelenoP Peptide

The results for the SelenoP Peptide were similar to those of TXNRD1. **Figure 16A** shows a 21.4% decrease in the amount of In-Tact Peptide in the test sample when compared to the control. This again reflects a rather small knockdown of the peptide but can also be explained due to the various efficiencies by which M^{pro} cleaves its canonical cleavage sites. The SelenoP Peptide sequence (ALLQA) most closely resembles that of NSP7/8 (ATLQ/A), which is cleaved by M^{pro} with a relative catalytic efficiency of 5%. Therefore, even though the only mismatch

between the two sequences is in the least important P3 position, the efficiency by which it is cut is so low that it accounts for the small drop in In-Tact SelenoP. Nonetheless, the mismatch at the P3 position could have further decreased the efficiency by which it was cut, as well.

Furthermore, **Figures 16B and C** both show base peaks at similar intensities corresponding to the P and P' fragments of SelenoP (m/z values = 643.28 and 642.42, respectively). Once again, a cleavage site within another selenoprotein is confirmed. In this case, the cleavage has broad ramifications for the existence of all other selenoproteins in the human body.

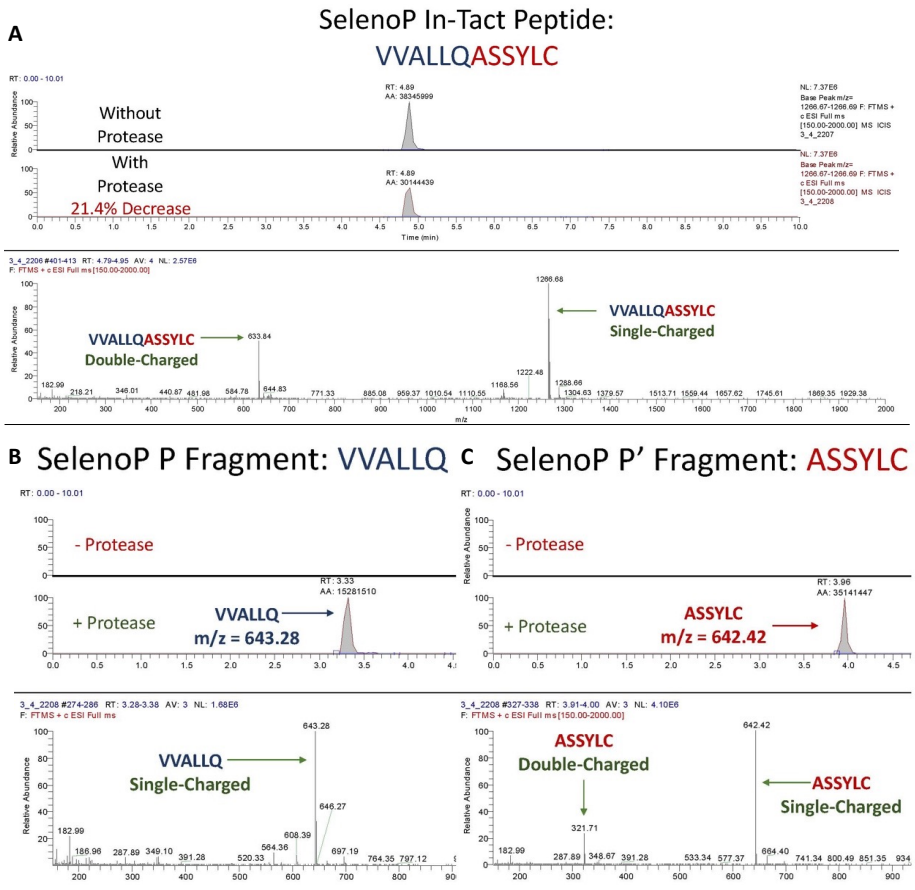

Figure 16. LC-MS Data for SelenoP Peptide. Depicted are the ion chromatograms for the samples without protease and with protease (top) and the mass spectra (bottom) for the (A) In-Tact Peptide, (B) the P Fragment, and (C) the P' Fragment. The base peaks for the P and P' Fragments had normalization levels of 2.81E6 and 8.56E6, respectively. The mass spectra were added to corroborate the identities of each base peak in the corresponding Ion-Chromatogram.

III.A.5. SelenoF Peptide

The results for SelenoF trump all of those discussed so far. **Figure 17A** shows that incubation with M^{pro} caused a 43.5% decrease in SelenoP levels when compared to the control sample. The reason for this is evident. The proposed cleavage site for SelenoF is 100% identical to that of NSP12/13. Although NSP12/13 is cleaved by M^{pro} with an 8% catalytic efficiency, as was the case for TXNRD1, the fact that there were no mismatches allowed for it to be cleaved at a much higher efficiency than TXNRD1 was. This information accounts for the drops in proteolytic cleavage of all the aforementioned peptides.

As with the other confirmed cleavage sites, base peaks for the proposed P and P' fragments of SelenoP (m/z values = 644.40 and 551.20, respectively) were observed at high intensities, and were absent in the control samples. Therefore, *in vitro* cleavage for the proposed SelenoF cleavage site was confirmed, as well, suggesting that SARS-CoV-2 is targeting SelenoF, a protein that is potentially heavily involved in the function of our immune system.

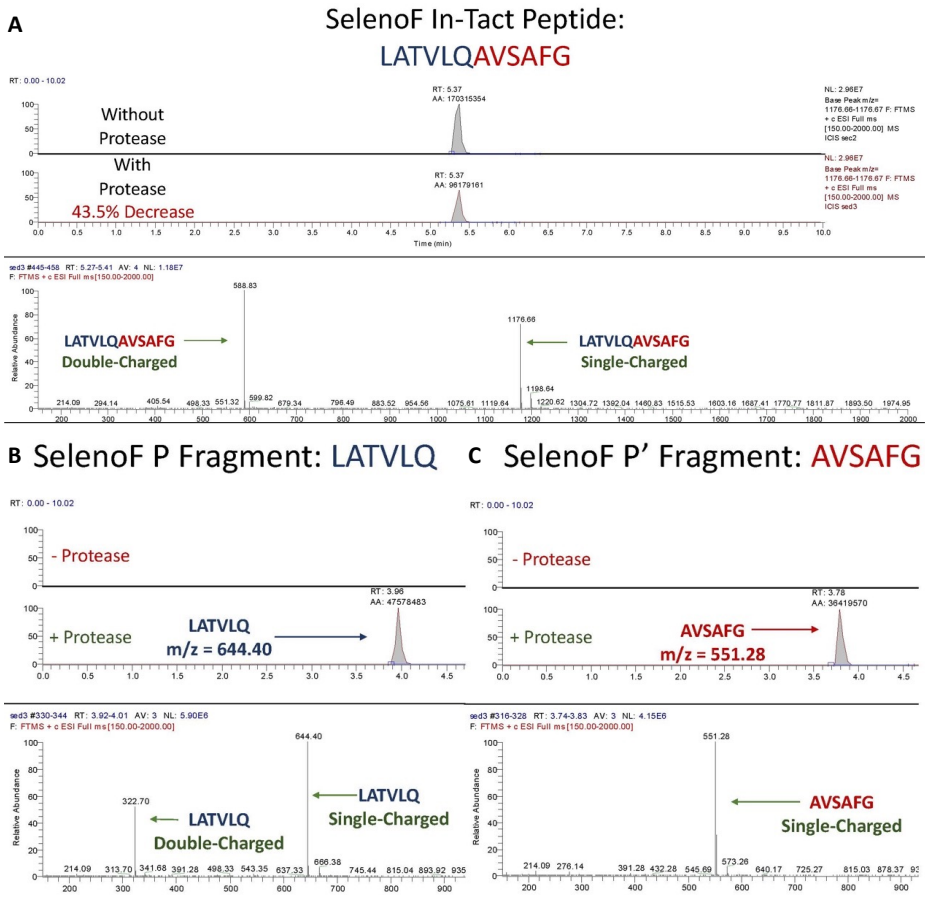

Figure 17. LC-MS Data for SelenoF Peptide. Depicted are the ion chromatograms for the samples without protease and with protease (top) and the mass spectra (bottom) for the (A) In-Tact Peptide, (B) the P Fragment, and (C) the P' Fragment. The base peaks for the P and P' Fragments had normalization levels of 2.96E7 and 1.12E7, respectively. The mass spectra were added to corroborate the identities of each base peak in the corresponding Ion-Chromatogram.

III.B. Unsuccessful/Inconclusive Results

III.B.1. Inconclusive Cleavage of GPX1 Peptide

Although GPX1 was the protein used as the basis for this study, no clear cleavage was observed. However, it is regarded as inconclusive cleavage for multiple reasons. Firstly, due to the fact that this was the first protein I had worked with, I had run out of it before I could optimize my protocol. Additionally, since the selenocysteine was in the key P1 position, mutating it to a serine was not an option, for it would have still provided inconclusive evidence,

even if it had cleaved. Therefore, the synthesis of this peptide was the most expensive and took the most time to obtain. We did not try purchasing it again because we realized that the only chance for it to be cleaved by M^{pro} would be via the utilization of cell cultures. This is because M^{pro} needs glutamine at the P1 position in order for it to cleave, and the only way for selenocysteine to even resemble glutamine is for it to go through a series of reactions to form a selenenylamide. Unfortunately, we did not have the capabilities to perform such reactions, but it provided us with ideas for future studies involving cell-based cleavage of GPX1.

III.B.2. Inconclusive Cleavage of TXNRD1 Protein

Although we did not observe any cleavage of the TXNRD1 protein when performing the assays, I still regard it as inconclusive because of the miniscule TXNRD1 concentrations available for each sample, and the drop in the activity of our protease in the latest experiment. As previously described in the methodology section, the peptide concentration utilized in each sample was about 50 μM. However, due to the low yield of our recombinant TXNRD1, the concentrations in each TXNRD1 sample were less than 1 μM. Therefore, the protocol was not nearly optimized for this assay. Furthermore, the latest experiment showed that our M^{pro} had lost nearly all of its activity, barely even cleaving the positive control. Therefore, this assay will need to be performed again using much higher TXNRD1 concentrations and M^{pro} with optimal activity.

III.B.3. Unsuccessful Cleavage of GLRX-1 Peptide

Unlike the results for the GPX1 Peptide and the TXNRD1 Protein, the results for GLRX-1 are conclusive. This is due to the fact that it was assessed multiple times in many different ways, and in batches of other peptides that did successfully cleave. The reason for the

unsuccessful cleavage is most likely due to the size of the glutamine in the P1' position. Although it was homologous to the asparagine in the P1' position of NSP8/9, it most likely did not fit into the active site of SARS-CoV-2 M^{pro}. As discussed earlier, the S1' subunit of SARS-CoV-2 M^{pro} is quite shallow, meaning that typically only smaller residues will fit into the pocket. This is the reason that asparagine, which is the largest residue in the consensus sequence for this position, was only found in the P1' position for one canonical cleavage site. Therefore, the fact that glutamine is an even bigger residue than asparagine, helps to explain why glutaredoxin-1 did not cleave.

CHAPTER IV: CONCLUSIONS

IV.A. Impact of Cleavage Site Confirmation

Four of the six proposed cleavage sites were definitively confirmed in this study, three of which were selenoproteins and the other, another essential, glutathione-related protein. These results, not only support our hypothesis, but have broad implications for the ways we view prevention, treatment, and the evolution of SARS-CoV-2.

Firstly, confirmation of the cleavage sites within TXNRD1 and GCLC supports our claim that SARS-CoV-2 is inhibiting RNR activation, thus inhibiting DNA synthesis and repair. This has serious ramifications on the health of our cells due to the increased induction of apoptosis, and therefore, the overall function of our organs. Furthermore, the inhibition of RNR leads to and increased allocation of ribonucleotides for virion production, thus conferring an evolutionary advantage for SARS-CoV-2. Additionally, the facts that SARS-CoV-2 M^{pro} cleaved SelenoF at such high levels and SelenoF mRNA has been proven to be downregulated by SARS-CoV-2, suggest that SARS-CoV-2 gains a vital advantage from the knockdown of SelenoF. This is most likely in the form of immune function inhibition, making it more difficult for individuals to combat the virus once infected. Finally, the confirmation of a cleavage site within SelenoP poses a threat to the levels of all selenoproteins in the body due to its critical role in selenium transport. Therefore, a large-scale knockdown of this protein could potentially inhibit DNA synthesis/repair, regulation of glycoprotein folding, and protection from reactive oxygen species, conferring multiple advantages for SARS-CoV-2 and several disadvantages for infected individuals.

Fortunately, selenoproteins levels have been shown to be upregulated when host selenium levels increase. Furthermore, TXNRD1 and GCLC levels have also been proven to be upregulated when vitamin D levels are high. Therefore, through supplementation with selenium and vitamin D, the effects of the proteolytic knockdown could potentially be counteracted, thus providing cost-effective therapeutics to be taken in conjunction with medically validated therapies/vaccinations.

IV.B. Inconclusive Results and Future Works

Although the GPX1 Peptide and the TXNRD1 Protein did not cleave in this study, that does not mean that they are not SARS-CoV-2 M^{pro} targets. Both compounds had reasonable explanations for not cleaving other than not having valid cleavage sites. In fact, as demonstrated in this research, TXNRD1 now has a confirmed cleavage site, and GPX1 has much data to support its cleavage, ranging from its interaction with Mutant SARS-CoV-2 M^{pro} and its massive downregulation upon infection with SARS-CoV-2. They just were not assessed in the manner that was necessary to induce a cleavage. However, future studies in the Taylor lab will aim to prove these cleavages as soon as Summer of 2022. Thanks to Stella de Lima Camargo of the Chekan Lab, I now have a high enough concentration of TXNRD1 protein to perform the assay properly and re-analyze via UPLC-Mass Spectrometry. Furthermore, Moukbel Nasr of the Chiu lab has agreed to use his expertise in dealing with cell cultures to assess the cleavage of GPX1 in cells. Therefore, the revelations from this study are far from over.

REFERENCES

(1) Variants of the Virus That Causes COVID-19. *Coronavirus*.

(2) *Coronavirus Main Proteinase (3CLpro) Structure: Basis for Design of Anti-SARS Drugs*. https://www.science.org/doi/abs/10.1126/science.1085658 (accessed 2021-09-02).

(3) Taylor, E. W.; Radding, W. Understanding Selenium and Glutathione as Antiviral Factors in COVID-19: Does the Viral Mpro Protease Target Host Selenoproteins and Glutathione Synthesis? *Front. Nutr.* **2020**, *7*, 143. https://doi.org/10.3389/fnut.2020.00143.

(4) *WHO Coronavirus (COVID-19) Dashboard*. https://covid19.who.int (accessed 2021-05-03).

(5) *COVID-19 Map*. Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/map.html (accessed 2021-10-09).

(6) CDC. *COVID-19 and Your Health*. Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html (accessed 2021-10-11).

(7) *Different types of COVID-19 vaccines: How they work - Mayo Clinic*. https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/different-types-of-covid-19-vaccines/art-20506465 (accessed 2021-10-11).

(8) Linnane, C. *BA.2 subvariant of omicron accounted for almost 94% of global cases in latest week, the latest sign of how dominant it has become*. MarketWatch. https://www.marketwatch.com/story/ba-2-subvariant-of-omicron-accounted-for-almost-94-of-global-cases-in-latest-week-in-latest-sign-of-how-dominant-it-has-become-11649254817 (accessed 2022-06-05).

(9) Planas, D.; Veyer, D.; Baidaliuk, A.; Staropoli, I.; Guivel-Benhassine, F.; Rajah, M. M.; Planchais, C.; Porrot, F.; Robillard, N.; Puech, J.; Prot, M.; Gallais, F.; Gantner, P.; Velay, A.; Le Guen, J.; Kassis-Chikhani, N.; Edriss, D.; Belec, L.; Seve, A.; Courtellemont, L.; Péré, H.; Hocqueloux, L.; Fafi-Kremer, S.; Prazuck, T.; Mouquet, H.; Bruel, T.; Simon-Lorière, E.; Rey, F. A.; Schwartz, O. Reduced Sensitivity of SARS-CoV-2 Variant Delta to Antibody Neutralization. *Nature* **2021**, *596* (7871), 276–280. https://doi.org/10.1038/s41586-021-03777-9.

(10) Tada, T.; Zhou, H.; Dcosta, B. M.; Samanovic, M. I.; Chivukula, V.; Herati, R. S.; Hubbard, S. R.; Mulligan, M. J.; Landau, N. R. Increased Resistance of SARS-CoV-2 Omicron Variant to Neutralization by Vaccine-Elicited and Therapeutic Antibodies. *eBioMedicine* **2022**, *78*. https://doi.org/10.1016/j.ebiom.2022.103944.

(11) Naqvi, A. A. T.; Fatima, K.; Mohammad, T.; Fatima, U.; Singh, I. K.; Singh, A.; Atif, S. M.; Hariprasad, G.; Hasan, G. M.; Hassan, Md. I. Insights into SARS-CoV-2 Genome, Structure, Evolution, Pathogenesis and Therapies: Structural Genomics Approach. *Biochim. Biophys. Acta Mol. Basis Dis.* **2020**, *1866* (10), 165878. https://doi.org/10.1016/j.bbadis.2020.165878.

(12) Padhan, K.; Parvez, M. K.; Al-Dosari, M. S. Comparative Sequence Analysis of SARS-CoV-2 Suggests Its High Transmissibility and Pathogenicity. *Future Virol.* **2021**, *16* (3), 245–254. https://doi.org/10.2217/fvl-2020-0204.

(13) Romano, M.; Ruggiero, A.; Squeglia, F.; Maga, G.; Berisio, R. A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping. *Cells* **2020**, *9* (5), 1267. https://doi.org/10.3390/cells9051267.

(14) Wu, J.; Yuan, X.; Wang, B.; Gu, R.; Li, W.; Xiang, X.; Tang, L.; Sun, H. Severe Acute Respiratory Syndrome Coronavirus 2: From Gene Structure to Pathogenic Mechanisms and Potential Therapy. *Front. Microbiol.* **2020**, *11*, 1576. https://doi.org/10.3389/fmicb.2020.01576.
(15) Satarker, S.; Nampoothiri, M. Structural Proteins in Severe Acute Respiratory Syndrome Coronavirus-2. *Arch. Med. Res.* **2020**, *51* (6), 482–491. https://doi.org/10.1016/j.arcmed.2020.05.012.
(16) Michel, C. J.; Mayer, C.; Poch, O.; Thompson, J. D. Characterization of Accessory Genes in Coronavirus Genomes. *Virol. J.* **2020**, *17* (1), 131. https://doi.org/10.1186/s12985-020-01402-1.
(17) *Continuous and Discontinuous RNA Synthesis in Coronaviruses | Annual Review of Virology*. https://www.annualreviews.org/doi/10.1146/annurev-virology-100114-055218 (accessed 2021-10-12).
(18) de Vries, A. A. F. SARS-CoV-2/COVID-19: A Primer for Cardiologists. *Neth. Heart J.* **2020**, *28* (7), 366–383. https://doi.org/10.1007/s12471-020-01475-1.
(19) Khailany, R. A.; Safdar, M.; Ozaslan, M. Genomic Characterization of a Novel SARS-CoV-2. *Gene Rep.* **2020**, *19*, 100682. https://doi.org/10.1016/j.genrep.2020.100682.
(20) GmbH, antibodies-online. *SARS-CoV-2 / COVID-19 Antibodies, Proteins, Kits | www.antibodies-online.com*. https://www.antibodies-online.com/areas/infectious-disease/covid-19/ (accessed 2021-10-13).
(21) Goyal, B.; Goyal, D. Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy. *ACS Comb. Sci.* **2020**, *22* (6), 297–305. https://doi.org/10.1021/acscombsci.0c00058.
(22) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors. *Nature* **2020**, *582* (7811), 289–293. https://doi.org/10.1038/s41586-020-2223-y.
(23) Mengist, H. M.; Dilnessa, T.; Jin, T. Structural Basis of Potential Inhibitors Targeting SARS-CoV-2 Main Protease. *Front. Chem.* **2021**, *9*, 7. https://doi.org/10.3389/fchem.2021.622898.
(24) Muramatsu, T.; Kim, Y.; Nishii, W.; Terada, T.; Shirouzu, M.; Yokoyama, S. Autoprocessing Mechanism of Severe Acute Respiratory Syndrome Coronavirus 3C-like Protease (SARS-CoV 3CL pro) from Its Polyproteins. *Febs J.* **2013**, *280* (9), 2002–2013. https://doi.org/10.1111/febs.12222.
(25) Ramos-Guzmán, C. A.; Ruiz-Pernía, J. J.; Tuñón, I. Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods. *ACS Catal.* **2020**, *10* (21), 12544–12554. https://doi.org/10.1021/acscatal.0c03420.
(26) 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. The SARS-CoV-2 SSHHPS Recognized by the Papain-like Protease. *ACS Infect. Dis.* **2021**, *7* (6), 1483–1502. https://doi.org/10.1021/acsinfecdis.0c00866.
(27) Miczi, M.; Golda, M.; Kunkli, B.; Nagy, T.; Tőzsér, J.; Mótyán, J. A. Identification of Host Cellular Protein Substrates of SARS-COV-2 Main Protease. *Int. J. Mol. Sci.* **2020**, *21* (24), 9523. https://doi.org/10.3390/ijms21249523.

(28) 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. 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.

(29) Song, J.; Tan, H.; Perry, A. J.; Akutsu, T.; Webb, G. I.; Whisstock, J. C.; Pike, R. N. PROSPER: An Integrated Feature-Based Tool for Predicting Protease Substrate Cleavage Sites. *PLoS ONE* **2012**, *7* (11), e50300. https://doi.org/10.1371/journal.pone.0050300.

(30) Sengupta, R.; Holmgren, A. Thioredoxin and Glutaredoxin-Mediated Redox Regulation of Ribonucleotide Reductase. *World J. Biol. Chem.* **2014**, *5* (1), 68–74. https://doi.org/10.4331/wjbc.v5.i1.68.

(31) Chen, Y.; Shertzer, H. G.; Schneider, S. N.; Nebert, D. W.; Dalton, T. P. Glutamate Cysteine Ligase Catalysis: DEPENDENCE ON ATP AND MODIFIER SUBUNIT FOR REGULATION OF TISSUE GLUTATHIONE LEVELS*. *J. Biol. Chem.* **2005**, *280* (40), 33766–33774. https://doi.org/10.1074/jbc.M504604200.

(32) Ren, B.; Liu, M.; Ni, J.; Tian, J. Role of Selenoprotein F in Protein Folding and Secretion: Potential Involvement in Human Disease. *Nutrients* **2018**, *10* (11), 1619. https://doi.org/10.3390/nu10111619.

(33) Wang, Y.; Huang, J.; Sun, Y.; Stubbs, D.; He, J.; Li, W.; Wang, F.; Liu, Z.; Ruzicka, J. A.; Taylor, E. W.; Rayman, M. P.; Wan, X.; Zhang, J. SARS-CoV-2 Suppresses MRNA Expression of Selenoproteins Associated with Ferroptosis, Endoplasmic Reticulum Stress and DNA Synthesis. *Food Chem. Toxicol.* **2021**, *153*, 112286. https://doi.org/10.1016/j.fct.2021.112286.

(34) Rudd, P. M.; Elliott, T.; Cresswell, P.; Wilson, I. A.; Dwek, R. A. Glycosylation and the Immune System. *Science* **2001**, *291* (5512), 2370–2376. https://doi.org/10.1126/science.291.5512.2370.

(35) Chen, S.; Su, X.; Mi, H.; Dai, X.; Li, S.; Chen, S.; Zhang, S. Comprehensive Analysis of Glutathione Peroxidase-1 (GPX1) Expression and Prognostic Value in Three Different Types of Renal Cell Carcinoma. *Transl. Androl. Urol.* **2020**, *9* (6), 2737–2750. https://doi.org/10.21037/tau-20-1398.

(36) Saito, Y. Selenium Transport Mechanism via Selenoprotein P—Its Physiological Role and Related Diseases. *Front. Nutr.* **2021**, *8*.