UNIVERSITY OF CALIFORNIA SAN DIEGO

# Evolution-Guided Discovery of Cleavage of Human Proteins by Viral Proteases

A Thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Biology

by

Ian Richard Robbins

Committee in charge:

Professor Matthew Daugherty, Chair
Professor Justin Meyer
Professor Elina Zuniga

2020

The Thesis of Ian Richard Robbins is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

---

---

---

Chair

University of California San Diego

2020

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Matthew Daugherty for his patience and guidance throughout my academic career. I would also like to acknowledge the work of Brian Tsu as crucial to the success of my project, and the advice and support he gave to help me persevere.

ABSTRACT OF THE THESIS

Evolution-Guided Discovery of Cleavage of Human Proteins by Viral Proteases

by

Ian Richard Robbins

Master of Science in Biology

University of California San Diego, 2020

Professor Matthew Daugherty, Chair

Coxsackie virus B3 and encephalomyocarditis virus, which are relevant to human health research, are members of a group called positive-sense single-stranded RNA (+ssRNA) viruses, which is known to translate their RNA genomes as monocistronic polyproteins. These polyproteins contain proteases that are known to fulfill two functions: Cleaving the viral polyprotein into functional protein subunits, and cleaving host proteins. By examining the cleavage sites found within viral polyproteins, homologous sites can be found within host proteins as well. Since infection by viruses can be treated as selection pressure on host immunity proteins, the host proteome can be bioinformatically filtered based on sites of rapid evolution across primate

genomes. The predictive power of this pipeline was tested by inserting potential cleavage sites, eight amino acids in size, into a novel reporter system. This evolution-guided model proved sufficient to confer significant vulnerability to cleavage to the reporter construct to CVB3 3C protease, but not to EMCV 3C protease, validating known antagonization targets of CVB3 and discovering new ones as well. Additionally, these two viral proteases exhibited a range of cleavage efficiencies on human proteins and on their own polyproteins, but largely did not cleave each other's polyproteins. Further investigation and refinement of this model would strengthen its predictive power and allow better insight into the evolution and mechanisms of these disease-causing agents.

# INTRODUCTION:

## The Host-Pathogen Arms Race:

There is a constant battle between pathogens and the organisms that they infect. Parasitic pathogens, like viruses, are known to employ a vast array of effectors that allow for entering into a host cell, commandeering of the host machinery, and replicating itself until the host is destroyed. For their part, hosts possess an arsenal of innate immunity proteins that recognize non-self macromolecules, eliminate them, and alert surrounding cells of infection (Chaplin, 2010). It is now well-known that, in response, viruses overcome these obstacles to their life cycles by including antagonizing agents within their own arsenals of effectors, allowing them to evade host defenses (Kikkert, 2019; Rustagi & Gale, 2014). These two players on the theater represent selection pressures for each other, driving an ongoing evolutionary "arms race" in which pathogenic effectors and host immunities leapfrog each other from species to species, one amino acid at a time (Rouzine & Rozhnova, 2018; Voskarides et al., 2018). By examining the diversity of pathogenic effectors and innate immunities across species, the ebb and flow of this arms race can come to be better understood, revealing the intricacies of virus evolution and replication.

In the specific case of viruses, one particular group that is of interest to human health research is that of positive-sense single-stranded RNA viruses. This group contains the families *Flaviviridae* and *Picornaviridae*, which are responsible for a wide range of human diseases. These viruses are encoded by small, polyadenylated, positive-sense RNA genomes, facilitating a cytoplasmic stage in their replication cycles in which the entirety of the (+)ssRNA genome mimics host mRNA (Gelderblom, 1996). However, as their name implies, these (+)ssRNA genomes are composed of a single continuous strand, and eukaryotic hosts like humans are not adapted for processing polycistronic RNA. The (+)ssRNA viruses that infect eukaryotic hosts have adapted to

this limitation by enabling their genomes to undergo monocistronic translation, similarly to the host's mRNA, but this process produces a large polyprotein composed of multiple nonfunctional protein subunits. These viruses have further adapted to this life cycle by including a functional protease within their polyprotein, but these proteases have become prominent in the arms race as more than just tools for the processing of the polyprotein.

It is now well-understood that the proteases encoded by (+)ssRNA genomes can serve dual functions: Processing the polyproteins into functional protein subunits (Narayanan et al., 2002), and post-processing antagonization of host proteins to improve viral replication efficacy. It has been shown that Coxsackie virus B3 (CVB3), a member of the family *Picornaviridae*, antagonizes mitochondrial antiviral signaling protein (MAVS) and Toll/IL-1 receptor domain-containing adaptor inducing interferon-beta (TRIF), which are both innate immunity inducer proteins (Mukherjee et al., 2011), as well as receptor-interacting serine/threonine-protein kinase 3 (RIPK3), which is essential for necroptosis (Harris et al., 2015). It can therefore be hypothesized that this dual protease functionality must depend on cleavage motif homology between the viral and host proteomes, and that by examining these motifs within viral polyprotein cleavage sites, currently unknown host protein cleavage sites can be discovered proactively. However, a significant barrier to investigating this hypothesis in the context of human health is the size of the human proteome itself, which contains tens of thousands of different polypeptides, many of which could be false-positives due to restrictive structural contexts.

To establish accurate and precise parameters for the search for host proteins targeted for antagonization, it is possible to view this obstacle through the lens of evolution. It has shown that several primate species are resistant to infection by dengue virus, which is a member of the family *Flaviviridae*, closely related to *Picornaviridae*, and that this resistance is at least partially

attributable to a single amino acid difference in STING protein, an important inducer of interferon production (Stabell et al., 2018). It has also been shown that several primate species are resistant to infection by Hepatitis C virus (HCV), which is also a member of the family *Flaviviridae*, and that this resistance is partially attributable to a single amino acid difference in MAVS (Patel et al., 2012). In both of these independent previous studies, it was concluded that the genetic diversities of these single amino acid differences across species indicate that these are sites of rapid evolution among primates, which is a trend common among genes involved in the innate and adaptive immune systems. This suggests that the existence of antagonistic viral proteases is acting as a strong selection pressure to drive a host evolutionary trajectory towards amino acid motifs that are unrecognizable as substrates for viral proteases, and that this pressure can be mapped with signatures of rapid evolution.

In this context of an evolutionary arms race, we hypothesized that we could overlay rapid evolution and cleavage motifs to identify new targets within the human proteome under attack by Picornaviral and Flaviviral proteases. The computational pipeline that executes this overlay was critical in narrowing the search parameters, allowing for the prediction and discovery of novel cleavage sites in the human proteome and elucidating possible new mechanisms for infection and replication of Picornaviruses and Flaviviruses within primates and humans.

**Viral Polyprotein Cleavage Motif Computational Pipeline:**

The polyprotein translated by the viruses in the family *Picornaviridae* is known to contain eight cleavage sites recognized by that virus's own protease, which is called 3C. Crystallographic characterization of the viral cleavage sites has shown them to be disordered loops rather defined secondary structures. Two genera of interest to human health research within the family *Picornaviridae* are Enteroviruses, which includes Coxsackievirus strain B3 (CVB3), and

Cardioviruses, which includes encephalomyocarditis virus (EMCV). By compiling the cleavage motifs from the polyproteins CVB3 and EMCV, it was possible to develop a consensus sequence based on the relative abundances of amino acids within a window of eight peptides (Figure 1) (Brian Tsu, unpublished data). The window of eight amino acids for determining the consensus sequence is the standard in the field, and it is used to balance the necessities of capturing true-positive predictions while excluding false-positives; too wide and unspecific a window would capture more false-positives, while too narrow and specific a window would exclude more true-positives.

The consensus cleavage motif shows a very strong bias towards glutamine in the P1 position, a relatively slightly weaker bias towards glycine in the P1' position, and a moderate bias towards alanine in the P4 position, annotated as [A--Q/G---]. Although it is currently unknown the exact degree to which variation in the consensus cleavage motif affects cleavage efficiency for Picornaviruses, the aforementioned previous study with primate STING protein and dengue virus, a Flavivirus, implies that single amino acid differences could have significant consequences for cleavage efficiency for Picornaviruses, which are closely related to Flaviviruses.

The picornavirus consensus cleavage motif was used as a training set for a computational pipeline to predict cleavage sites within the human proteome. The produced list of potential-positives was further filtered by known gene ontological (GO) terms of positive selection in the predicted cleavage site, and of IFN-upregulation of gene expression, which is a common trend among genes related to the immune response (Ivashkiv et al., 2015). These filters were intended to narrow the search parameters towards human immune system proteins that could be potentially targeted for antagonization by viral proteases.

## METHODS:

### Reporter System and Cloning Cleavage Motifs:

This project necessitated the development of a novel reporter system, which is composed of three plasmid constructs that individually express a functional viral protease, a potential target motif, and eGFP-expressing transfection control, which has been previously shown to be an effective quantifier of transfection efficiency (Dandekar et al., 2005). These three plasmids were transiently transfected into Huh7 cell lines, a human hepatocyte-derived carcinomic cell line that has been previously established to be particularly vulnerable to infection by hepatitis C virus and other related viruses (Sainz et al., 2009). The molar ratios of the three plasmid constructs in the transfection protocol were 1:2.5:1.5, respectively. Huh7 cells were split onto 24-well plates for experimentation, transfected 24 hours post-splitting, and harvested 24 hours post-transfection.

The aforementioned potential target motif was not contained in its native protein. Although the native protein would provide important contextual information about the biology of a certain potential target motif, such as when the motif is buried within the structure of its protein, it is known that cloning whole proteins into vector plasmids is a slow and difficult process, which was a definitive weakness to that route of investigation when there were thousands of potential motifs that could be tested.

Therefore, in the interest of generating a high-throughput reporter system, the potential target motifs were instead cloned into a reporter construct composed of a destabilizing domain, a multiple cloning site containing the target motif itself, mCherry RFP, and FLAG epitope tags (Figure 2 and Figure 3). The multiple cloning site adopts a disordered loop structure, reflective of the viral cleavage motifs. The destabilizing domain constitutively targets the translated protein to the proteasome for degradation (Banaszynski et al., 2006), but in the event of cleavage of the target

motif in the multiple cloning site by viral protease, the mCherry and FLAG domains remain in the cytoplasm, providing routes for detection of cleavage events by microscopy, flow cytometry, and western blot. Additionally, the small synthetic molecule Shield-1 can be used to stabilize the destabilizing domain and prevent degradation of the reporter construct (Banaszynski et al., 2006), which could potentially allow a low-activity protease to accumulate sufficient levels of cleavage product to be visible in detection assays.

For testing potential cleavage sites, the reporter system requires an insertion of 24 base pairs into its multiple cloning site, which was accomplished with long oligo primers and polymerase chain reactions on the vector of the reporter system, the plasmid pDL2064. This allowed for rapid development and testing of cleavage of human peptides, and due to the various human motifs being contained within otherwise identical reporter constructs, cleavage events always produce cleavage products of identical molecular weight.

**Transient Transfection of Tissue Culture:**

As mentioned previously, all transfections were conducted using tissue cultures of Huh7 cell lines. It has been generally shown that transient transfection typically yields low efficiency of DNA uptake, and this would logically negatively affect protein expression levels in the reporter system. However, while it would have been possible to counteract this problem by generating stably transfected Huh7 cell lines, the selection process of stable transfection would have had to have been repeated for each potential target motif, and the resultant cell lines would have suffered from the toxicity of constitutively overexpressed reporter proteins. This route was deemed prohibitively labor-intensive and risky in what was intended to be a high-throughput and reliable reporter system, and also unnecessary as protein expression levels in the transient system were ultimately sufficient for detection of cleavage events.

**Cleavage Efficiency Analysis by Microscopy:**

The inclusion of paired RFP/GFP within the reporter system enables detection of cleavage events by fluorescence microscopy. This assay was not quantified, instead serving only to indicate a binary presence or absence of a cleavage event. In future investigations, the reporter system would permit quantification of cleavage events by fluorescence-activated cell sorting (FACS), as applied in flow cytometry, but such experimentation was beyond the scope of this project.

**Cleavage Efficiency Analysis of by Western Blot:**

Due to a cleavage event resulting in a significant change in the construct protein molecular weight, this reporter system also enables detection of cleavage events by western blot. Huh7 cells were harvested by trypsinization and the protein samples were prepared with 1X SDS solution boiled at 100 ºC for 10 minutes. All immunoblot signal was detected by anti-FLAG antibodies grown in mouse, and anti-GAPDH antibodies grown in rabbit for relative protein expression control.

**Whole Protein Cleavage Analysis:**

Due to the unavailability of many proteins in publicly accessible ORFeome libraries, it was necessary to obtain these from in-house cloning. All in-house cloning was performed with A549, an alveolar basal epithelium-derived adenocarcinomic cell line, induced with IFN-alpha. Messenger RNA from cultures of IFN-induced A549 cells was harvested for RT-PCR, and the generated cDNA templates were used for cloning PCR.

## RESULTS:

### The Reporter System:

The reporter system was intended to be a multifunction grouping of three plasmid constructs, the keystone of which expressed the novel reporter construct. This reporter construct was able to clearly detect the presence and absence of cleavage events, which was shown by the comparison of reporter construct with an unmodified multiple cloning site (hereafter referred to as the "vehicle") to CVB3-2C-3A, the viral polyprotein site with the highest predicted cleavability within the CVB3 computational pipeline. The cleavage efficiency of CVB3-2C-3A, as well as all other sites with significant cleavage, was approximated by comparing the strengths of visible signal within the uncleaved (52 kDa) and cleaved bands (32 kDa) on the western blot.

### Cleavage Efficiency Analysis by Microscopy:

At 24 hours post-transfection, Huh7 tissue cultures were analyzed by fluorescence microscopy. Expression levels of eGFP were sufficiently high to identify consistent transfection efficiencies, but expression levels of RFP were sufficiently low that it was not possible to make significant visual distinctions between most of the tested potential target motifs and their associated negative controls (which contained non-expressing vector instead of viral protease).

However, in some tested potential target motifs, cleavage efficiency was sufficiently high that RFP signal was significantly more visible, allowing for positive identification of cleavage. This successful experimentation also indicates that the reporter system is applicable for FACS in flow cytometry. Although that particular course was beyond the scope of this project, future experimentation could benefit greatly from the increased sensitivity of FACS-based analysis on this reporter construct, permitting finer precision detection of cleavage events.

**Cleavage Efficiency Analysis by Western Blot:**

The majority of this signal was visible as uncleaved protein that had not yet undergone proteasome degradation, possibly due to overload of the cellular machinery due to overexpression of protein. The detection of a true-positive cleavage event was considered to be the presence of a visible product of lower molecular weight than the uncleaved product, and the efficiency of the cleavage was approximated by comparing the intensities of the two bands (Figure 4).

It was found that the viral polyprotein cleavage sites exhibit a range of cleavage efficiencies (Figure 5). The majority of polyprotein cleavage sites producing no visible cleavage products, but a few, particularly CVB3-2C-3A and CVB3-3A-B, generated significant cleavage efficiencies. In the previously described viral polyprotein consensus cleavage motifs, CVB3 and EMCV exhibit similar characteristics in their consensus sequences, most notably the motif bias towards [A--Q/G--] (Figure 1). However, cross-species cleavage analysis has shown that the proteases of the two viruses are incapable of producing significant amounts of cleavage product in each other's polyprotein motifs.

To investigate the possibility of the structural context of the viral polyprotein affecting cleavage efficiency, the reporter systems for the viral polyprotein cleavage motifs were modified to include 50 additional amino acids (or 150 additional base pairs) from the viral genomes both upstream and downstream from the vector's multiple cloning site, a number judged sufficient to produce disordered loops or alpha-helical secondary structures within the reporter construct. It was found that the addition of structural context did not significantly impact the cleavage efficiencies of the viral polyprotein cleavage motifs (Figure 6).

After these initial findings were established, the progressing steps in the project were to generate and test human peptides within the reporter system. All of the following reported cleavage

efficiencies were approximated by comparing the visual intensities of the cleaved and uncleaved products (Figure 7 and Table 1). RIPK3, which has been mentioned previously, was cleaved by CVB3 by 40% and by EMCV by 0%. CD74, which plays a critical role in innate immunity by serving as a cell surface receptor (Gil-Yarom et al., 2016), was cleaved by CVB3 by 45% and by EMCV by 0%. PPP1R15A, which facilitates cell stress recovery by reversing protein synthesis shutoff (Clavarino et al., 2012), was cleaved by CVB3 by 5% and by EMCV by 0%. NCOA7, which enhances the transcriptional activities of several nuclear receptors (Herold, 2019), was cleaved by CVB3 by 5% and by EMCV by 0%. OAS3, which is a known antiviral enzyme with activity against dsRNA and ssRNA viruses (Ibsen et al., 2014), was cleaved by CVB3 by 30% and by EMCV by 0%. SAMD9, which is associated with antiviral pathways (Nounamo et al., 2017), was cleaved by CVB3 by 45% and by EMCV by 0%. ASB11 (isoform 3), which mediates ubiquitination and subsequent proteasomal degradation of target proteins (Andresen et al., 2013), was cleaved by CVB3 by 10% and by EMCV by 0%. HES4, which is a DNA-binding transcriptional repressor (Zhou et al., 2012), was cleaved by CVB3 by 20% and by EMCV by 0%. EHD3, which is involved in endocytic tubulation and transport (Galperin et al., 2002), was cleaved by CVB3 by 60% and by EMCV by 0%. UNC93B1, which plays a critical role in innate immunity by regulating TLR-signaling (Pohar et al., 2012), was cleaved by CVB3 by 5% and by EMCV by 0%. BATF2, which is a transcription factor that controls differentiation of cell lineages in the immune system (Chang et al., 2018), was cleaved by CVB3 by 30% and by EMCV by 0%. CDHR4, which is a cell adhesion protein (Sotomayor et al., 2014), was cleaved by CVB3 by 10% and by EMCV by 0%. C1RL (variant 3), which mediates cleavage of haptoglobin in the cell ER (Wicher & Fries, 2004), was cleaved by CVB3 by 40% and by EMCV by 0%. IFI44, which appears to have a variety of functions, including some antiviral activity against HIV (Huang et al., 2018;

Power et al., 2015), was cleaved by CVB3 by 20% and by EMCV by 5%. OAS1, which, like OAS3, is a known antiviral protein with activity against ssRNA viruses (Eskildsen, 2003); while the wild-type OAS1 is predicted to be uncleavable, there exists a single-nucleotide polymorphism in OAS1 in the human population which is predicted to be cleavable. This mutant OAS1 was cleaved by CVB3 by 40% and by EMCV by 0%. CARD6, which is a regulator of NF-kappaB activation (Dufner et al., 2006), was cleaved by CVB3 by 0% and by EMCV by 0%. FOXP3, which is a transcriptional regulator that controls the development of T-cells in the immune system, and has antitumor functions (Larsen et al., 2013; Roncador et al., 2005), was cleaved by CVB3 by 0% and by EMCV by 0%. APOBEC3A, which is a DNA deaminase that is known to restrict viruses and retrotransposons by targeting foreign DNA, was cleaved by CVB3 by 0% and by EMCV by 0%.

The human proteins tested for cleavage in their native forms as well as in the reporter system were MX1, NINL, and GBP5 (Figure 8 and Table 1). MX1 (interferon-induced GTP-binding protein) is known to possess antiviral activity against a wide range of RNA viruses and some DNA viruses (Verhelst et al., 2013). NINL (ninein-like protein) is involved in microtubule organization and cargo transportation, with possible interactions with viral factors (Bachmann-Gagescu et al., 2015; Kowanda et al., 2016; Chris Beierschmitt, unpublished data). GBP5 (Guanylate-binding protein 5) is a GTPase known to activate the NLRP3 inflammasome and have antiviral functions (Fujiwara et al., 2016; Hotter et al., 2016).

Cleavage motif analysis of MX1 within the reporter system context showed that the predicted cleavage motif is cleaved by CVB3 by 50% and by EMCV by 0%, and this was capitulated in the native protein context (Figure 8A).

Cleavage motif analysis of NINL within the reporter system context showed an efficiency gradient between the three predicted cleavage motifs when the reporter system was exposed to CVB3 proteases, but nothing when exposed to EMCV protease (Figure 8B). This efficiency gradient was accurately predicted by the computational pipeline, which had assigned each site with different cleavability scores. However, cleavage motif analysis of NINL within the native protein context showed only two cleavage events.

Cleavage motif analysis of GBP5 within the reporter system context showed that the predicted cleavage motif is cleaved by CVB3 by 40% and by EMCV by 0% (Figure 8C). However, cleavage motif analysis of GBP5 within the native protein context showed that the target is not cleaved by CVB3 protease, but instead by EMCV protease. The cleavage event by EMCV protease also produces a cleavage product of a different size than would be expected from the CVB3 protease, indicating that a different motif is being targeted for antagonization.

## DISCUSSION:

The viability of the novel reporter system is repeatedly demonstrated throughout the testing of the viral polyprotein motifs and the potential human motifs. Although its utility for fluorescence microscopy and FACS has yet to be fully explored, its potency for western blotting is appreciable. The primary limitation of the system is the necessity of cloning a new reporter construct for each individual potential motif, but the high-throughput nature of the system is a significant counterbalance. With regards to the insensitivity of the microscopy assay, it is possible that the visual signal of the RFP fluorescence is not sufficiently intense to counterbalance the constitutive degradation of the reporter construct by its destabilizing domain, or to match the visual signal of the western blot antibody signal. Future iterations of this reporter system could possibly see a more active fluorophore than mCherry being used to increase the visual signal.

It was established throughout the viral polyprotein testing that the viral 3C proteases has varying degrees of cleavage efficiency within the CVB3 and EMCV polyproteins, ranging from a very high ~50% efficiency to undetectably low, and also that the two viral proteases largely did not cleave across species, despite the similarities between the two viruses' consensus cleavage motifs. More broadly, the overall experimentation has shown that the reporter system, more specifically a simple motif of eight amino acids, was sufficient to produce efficient cleavage of an artificial protein construct. Furthermore, the testing of additional structural context upon CVB3 polyprotein motifs shows that in cases where the cleavage efficiency is already undetectably low in the 8meric model, the addition of flanking secondary structures does not rescue a poor cleavage phenotype. However, it is still possible that a wider window than 100 amino acids may produce secondary or tertiary structures that could increase cleavage efficiency.

In addition to validating the reporter system as a potent tool for analyzing cleavage events, these findings suggest a hierarchy of affinities for cleavage among the polyprotein cleavage sites, possibly enabling serial cleavage events at different stages of the viral replication cycle, as opposed to simultaneous cleavage events across the entire polyprotein. The significance of *trans* interaction of protease and target in Huh7 cytoplasm as opposed to *cis* interaction in the native polyprotein context is unknown, though the overexpression of the protease and target in their constructed vectors may possibly counteract the limitation of a *trans* context. Also, of significance is the notably lower activity exhibited by EMCV protease as compared to CVB3 protease, even upon the cleavage sites from the EMCV polyprotein. There are a multitude of possible theories for the dissonant behaviors of the two proteases: EMCV protease may simply be a less catalytically active protease; it may have higher requirements for *cis* interactions; it may be that structural context is more important for EMCV protease than it is for CVB3 protease; or it may be that the variations beyond [A--Q/G---] within the window of eight amino acids hold some presently unknown significance.

When evaluating the predicted human protein cleavage motifs, the reporter system validated known targets of antagonization like RIPK3, but novel targets were found as well. CD74, OAS3, SAMD9, EHD3, BATF2, C1RL, and mutant OAS1 all featured cleavage efficiencies by CVB3 of over 30%, comparable to some of the viral polyprotein cleavage motifs. The inclusion of mutant OAS1 in this project shows that single nucleotide polymorphisms (SNPs) may confer species- or individual-specific resistance to viral pathogenesis, which is a strong example of the evolutionary arms race in action. The only cleavage event ever exhibited by EMCV in the reporter system for human motifs was for IFI44, and that was still extremely low. This lends additional credit to the possibility that the consensus cleavage motifs, and the computational pipeline built

upon them, require additional refinement for more stringent proteases like EMCV 3C, but overall, the predictive model is very promising for more promiscuous proteases like CVB3 3C.

The screening of human motifs within a whole protein context also revealed valuable information about the predictive model, namely that the cleavage of the 8meric motif can be considered a good indication of cleavage within the whole protein. The cleavage pattern of whole MX1 protein is the most straightforward example of this phenomenon, in which a strong cleavage product in the 8meric form is echoed by a strong cleavage product in the whole protein form. This is also capitulated by whole NINL protein, which contained three predicted cleavage motifs.

The cleavage pattern of whole NINL protein suggests that the gradient of cleavage efficiency between its three motifs manifests as the different sites exhibiting different affinities as ligands to CVB3 protease. Additional investigation (Chris Beierschmitt, unpublished data) has shown that directed mutagenesis of the higher affinity cleavage sites also results in single cleavage events that produce differently sized products, indicating that the CVB3 protease targets only the site with the highest available affinity. It is currently unknown how these cleavage events affect the functionality of NINL protein, and if the various sites produce the same effect on cleaved NINL. Overall, the cleavage pattern of whole MX1 and whole NINL indicates that the computational pipeline can accurately predict multiple cleavage sites within a single protein.

These results are contrasted by the cleavage pattern of whole GBP5, which showed a cleavage event by EMCV instead of by CVB3, indicates that the computational pipeline did not accurately predict all of the cleavage sites within GBP5 protein, but the reporter system was still capable of serendipitously discovering novel cleavage motifs. Further refinements are necessary for the pipeline, such as peptide variation within the window of eight amino acids around the cleavage sites, to more accurately catch such proteins, but more broadly, the analysis of all three

human proteins, from their 8meric forms to their whole forms, has shown that this evolution-guided reporter system is capable of discovering new host-pathogen biology.

## FIGURES

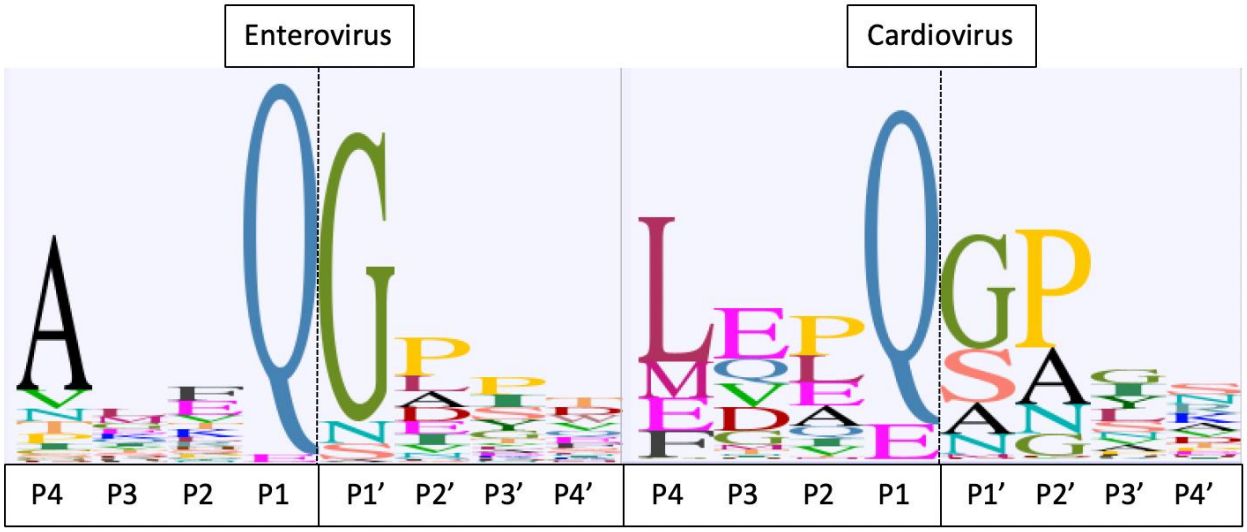



**Figure 1: Viral Polyprotein Consensus Cleavage Motifs:** These amino acid alignment compilations show the relative abundance of amino acids within a window of eight amino acids surrounding known cleavage sites for two genera within the family *Picornaviridae*: Enterovirus and Cardiovirus. The Enterovirus data was compiled from 6000 known polyprotein cleavage events and the Cardiovirus data was compiled from 300 known polyprotein cleavage events.

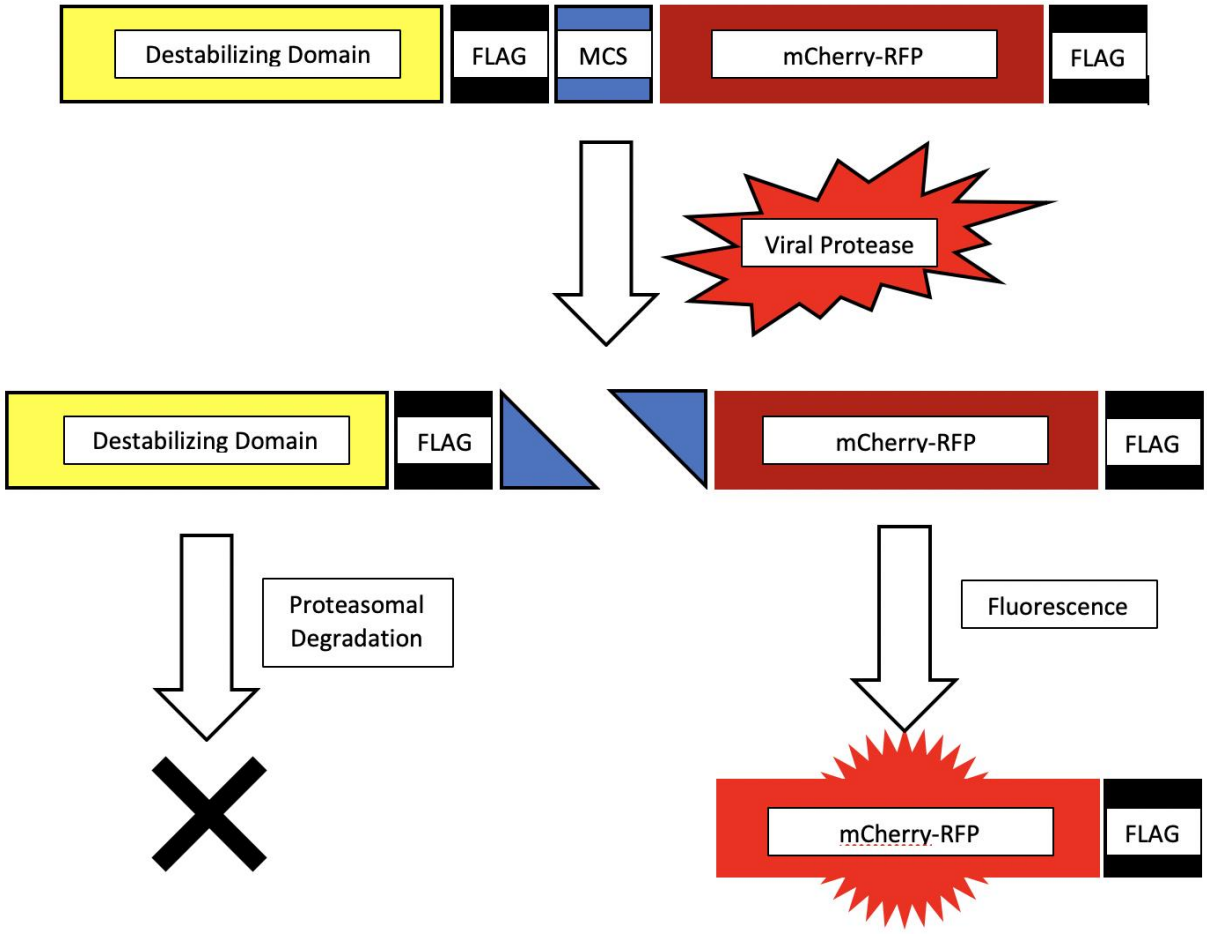



**Figure 2: Reporter Construct Schematic:** The destabilizing domain constitutively targets the construct to the proteasome for degradation. The mCherry-RFP constitutively fluoresces, allowing detection by microscopy. The FLAG epitope tags allow detection by western blotting. The multiple cloning site (MCS) contains multiple restriction enzyme sites, allowing insertion of a potential cleavage motif. The proteolytic cleavage of the MCS results in the destablizing domain being degraded while the RFP-FLAG domains remain in the cytoplasm for detection by microscopy and/or western blotting.

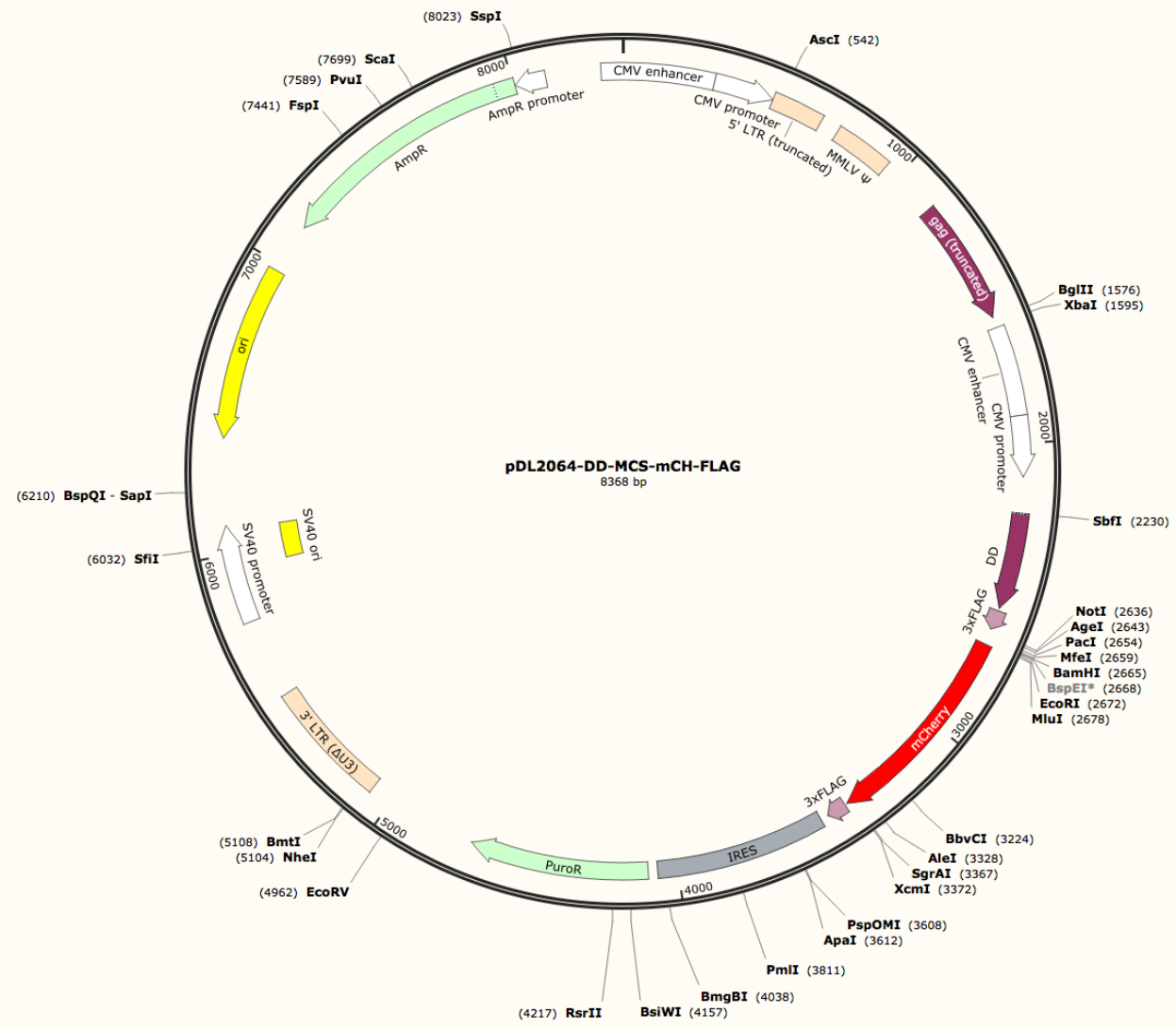



**Figure 3: Reporter System Vector Plasmid:** The plasmid construct is 8368 bp in size and contains a gene for ampicillin resistance (AmpR), allowing for reliable selection during the cloning process. The entire plasmid is expressed constutitively using a universal CMV promoter region.

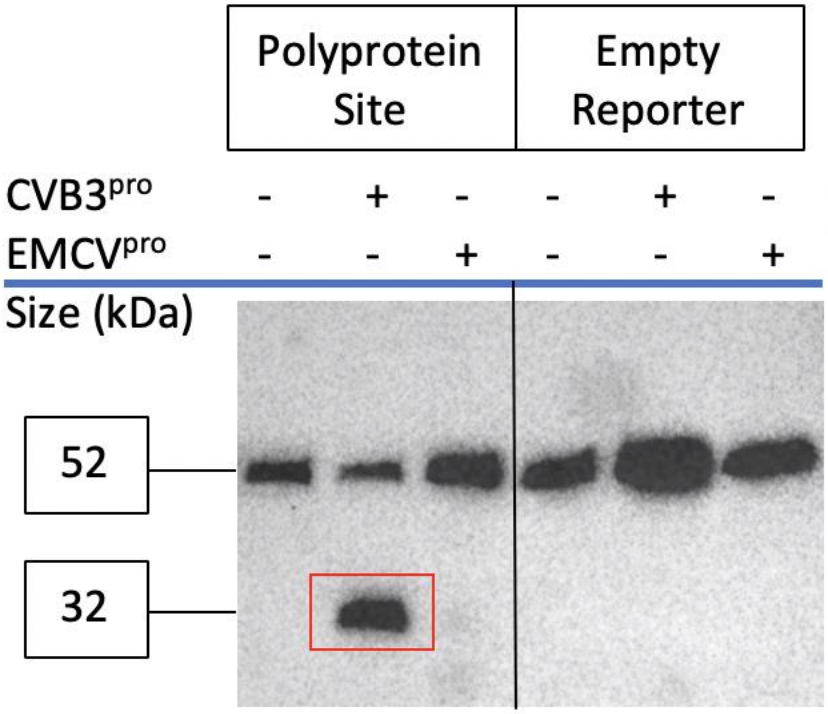



**Figure 4: Demonstration of Reporter System Cleavage Detection:** The red box indicates a significant cleavage product, a direct indication of a cleavage event within the reporter system. The polyprotein site was the viral polyprotein motif CVB3-2C-3A.

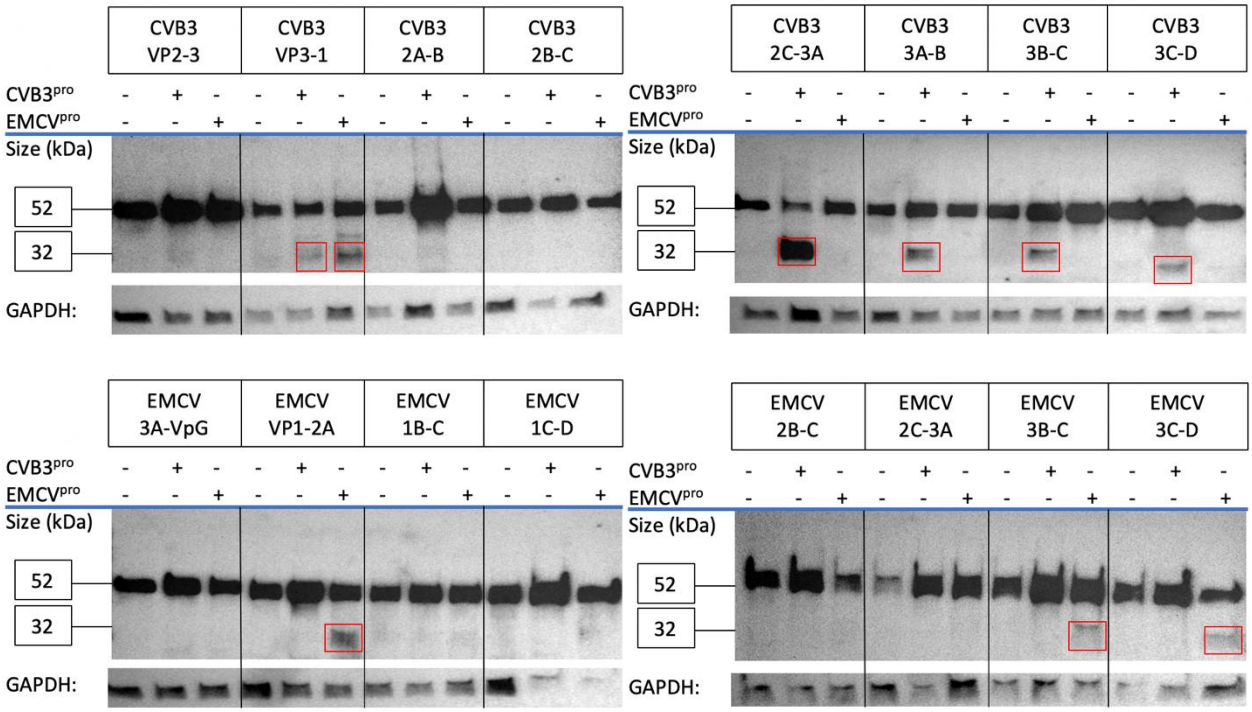



**Figure 5: Viral Polyprotein Motif Cleavage Efficiency:** Western blots showing a range of cleavage efficiencies for the two sets of eight viral polyprotein cleavage motifs, and also that the two species of viral protease will largely not cleave across species. Since CVB3 3C protease was able to cleave five of its own motifs, while EMCV was only able to cleave three of its own motifs and one of CVB3's motifs, it was determined that EMCV 3C protease may be less catalytically active or may have more stringent requirements for a motif to be vulnerable to being cleaved.

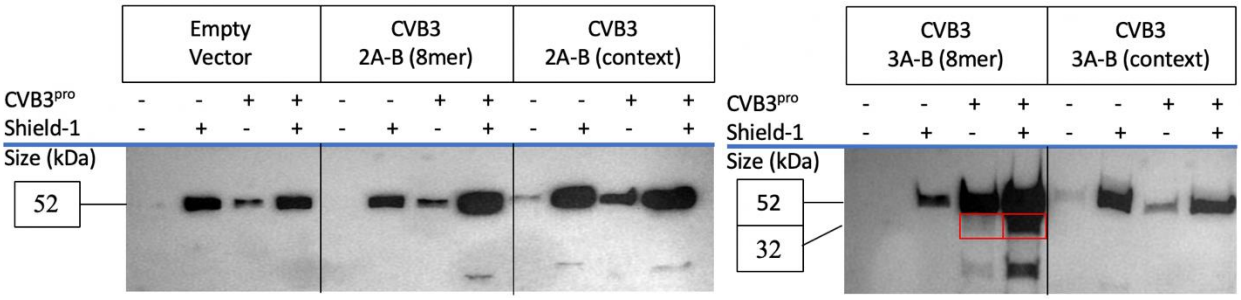



**Figure 6: Effects of Structural Context and Shield-1 Degradation Protection on Viral Polyprotein Cleavage Efficiency:** Western blots showing that structural context was added to two different CVB3 motifs, 2A-B and 3A-B. As in Figure 5, the CVB3-2A-B 8mer shows no cleavage and the CVB3-3A-B 8mer shows some cleavage. The addition of structural context does not significantly improve the cleavage efficiencies of either of these two motifs, even in the presence of Shield-1, which should be amplifying the visibility of low-intensity cleavage products. The red boxes significant cleavage products. The unboxed low-weight bands are spurious degradation products due to overabundance of protein in cellular cytoplasm due to Shield-1.

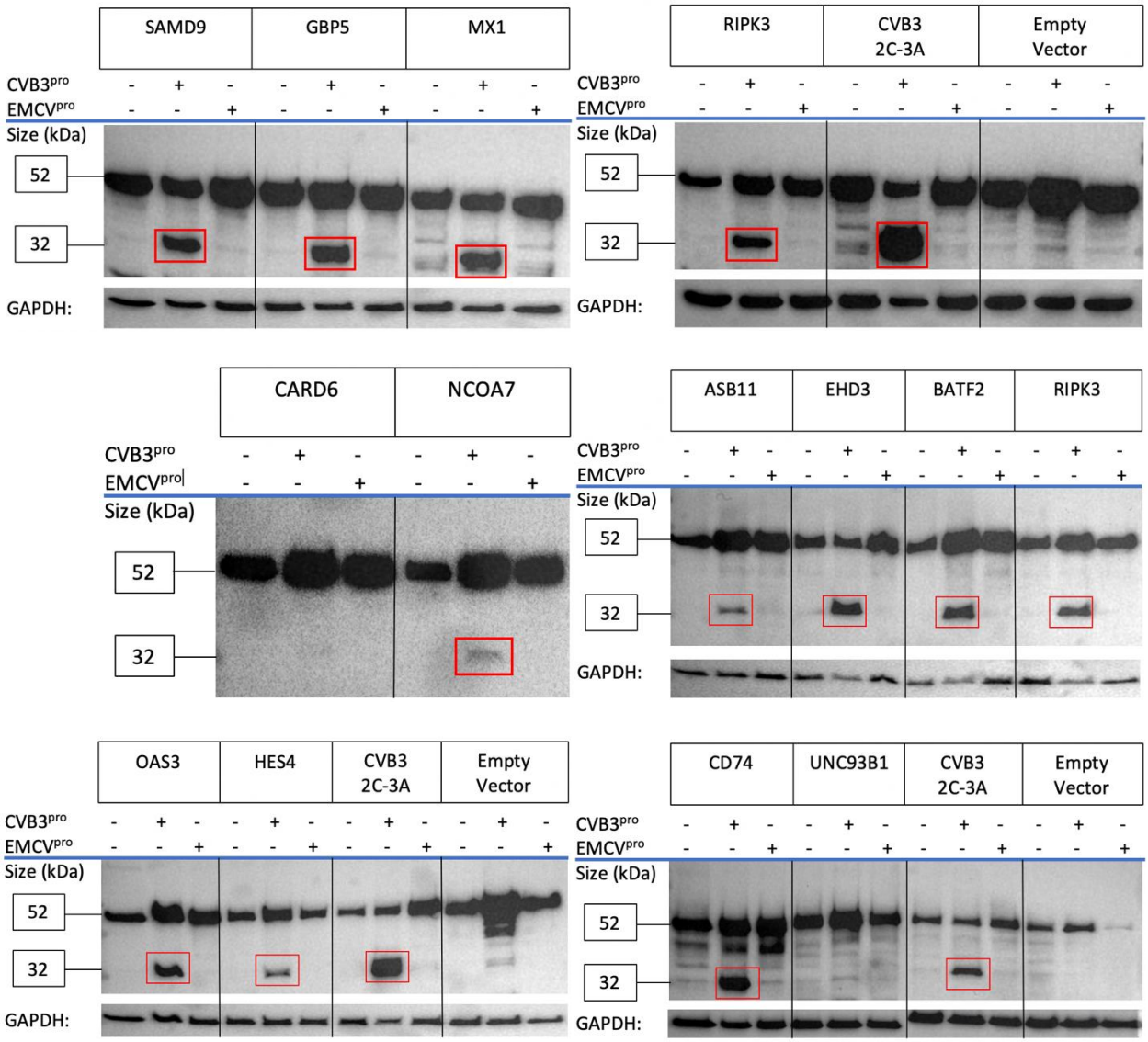
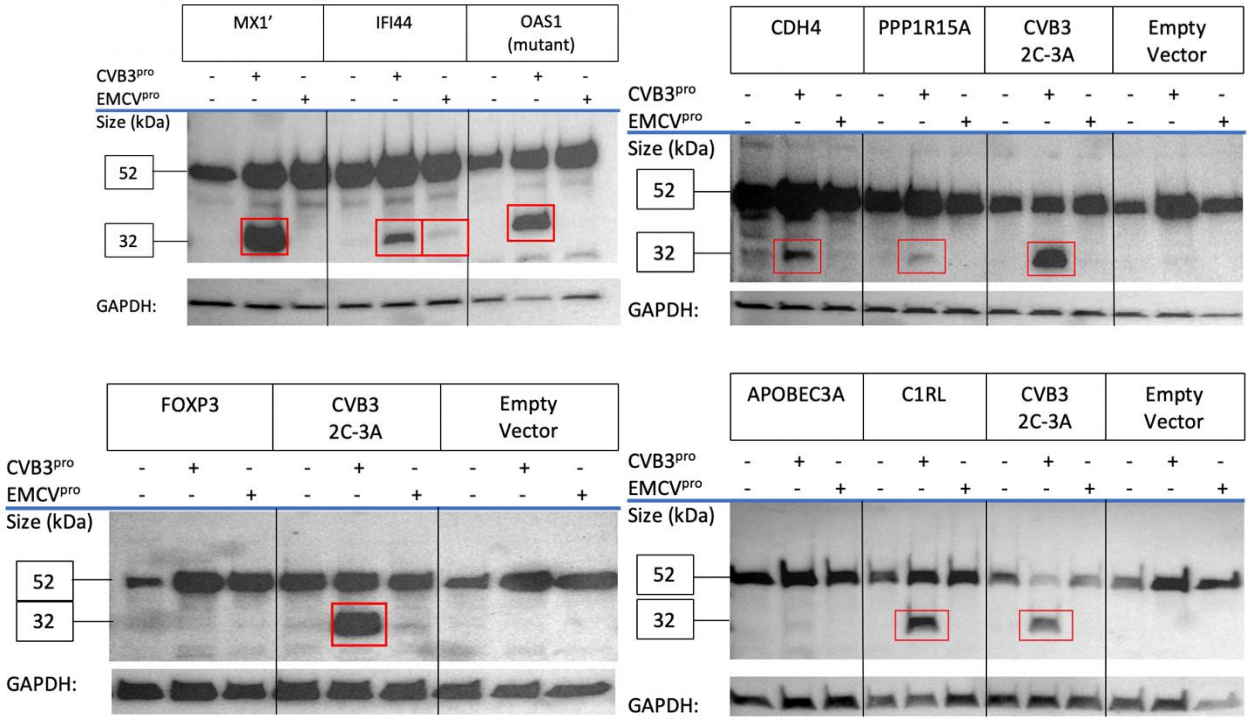



**Figure 7: Human Protein Motif Cleavage Efficiency:** The red boxes indicate significant cleavage products. Western blots showing that, similarly to the viral polyprotein motifs, the human protein motifs exhibit a range of cleavage efficiencies by mostly CVB3 3C protease, with EMCV 3C protease showing barely any activity at all. The presence of some human motifs that are not vulnerable to cleavage indicates that the predictive pipeline is not completely accurate, and further investigation is necessary for refinement.

**(Figure 7, continued from page 22)**



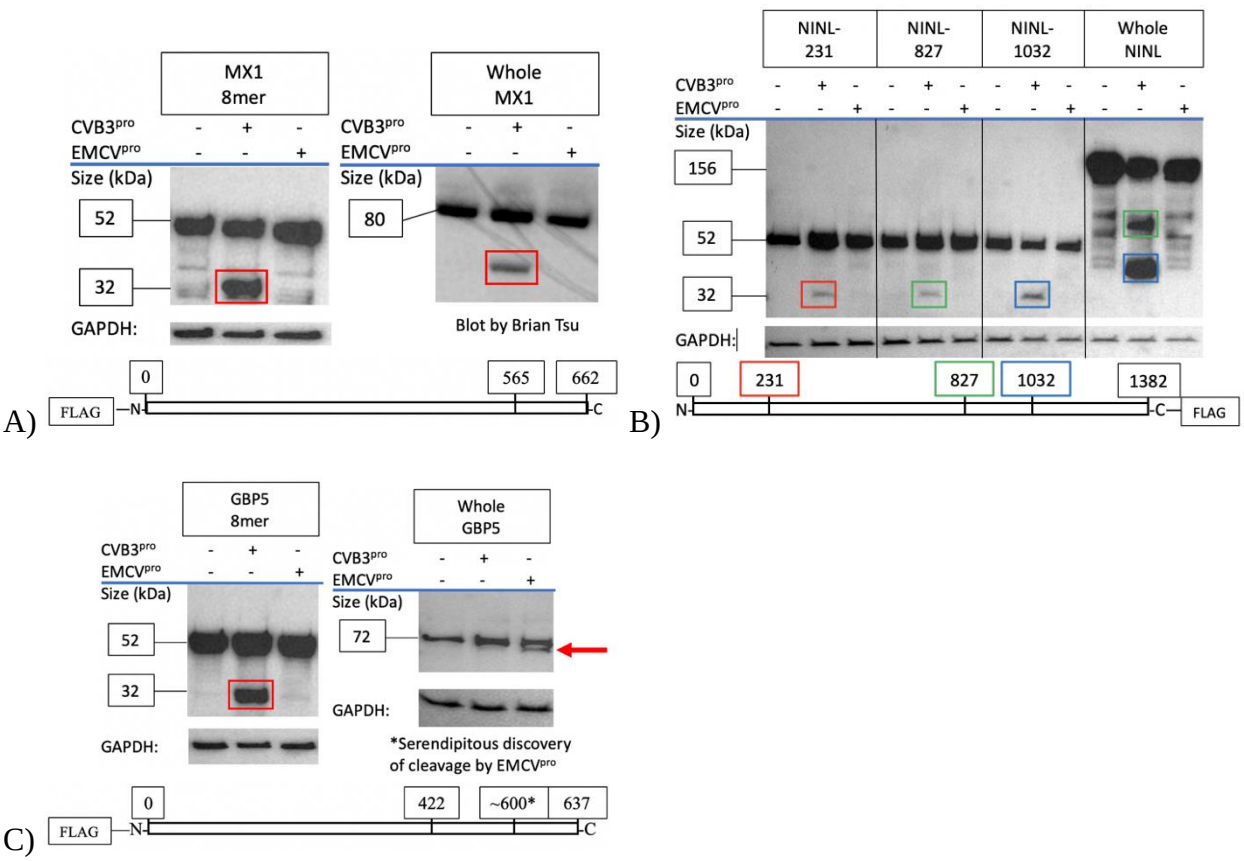



**Figure 8: Whole Human Protein Motif Cleavage Efficiency:** Images and western blots showing the cleavage site locations within the whole protein sequence, and the cleavage efficiencies of these sites, respectively. **A)** MX1 shows a strong cleavage efficiency in both its predicted 8meric form and in its predicted whole protein form, validating the reporter system for predicting cleavage in whole proteins. **B)** NINL shows a range of cleavage efficiencies between its three predicted 8meric motifs, but only two cleavage products are shown for the whole protein form of NINL. The colors of the boxes indicate which motifs are being cleaved in the whole protein form. **C)** The 8meric form of GBP5 is cleaved by CVB3 but not EMCV, while the whole protein form of GBP5 is cleaved by EMCV but not CVB3 (as indicated by the red arrow), and in a different location than it was predicted to be cleaved by CVB3. (*) Indicates the approximate location of the serendipitous cleavage motif in the whole GBP5 protein.

**Table 1: Approximate Quantification of Human Protein Cleavage Efficiency:** The *italicized* data represents the human motifs that were tested within the reporter system. The **bolded** data represents the human motifs that were tested within the native protein context.

| Name | Length (bp) | Size (kDa) | Motif | Cleavage by CVB3 | Cleavage by EMCV |
|---|---|---|---|---|---|
| *RIPK3* | *518* | *56.8* | *AERQGMNW* | *40%* | *0%* |
| *CD74* | *296* | *33.5* | *ALPQGPMQ* | *45%* | *0%* |
| *PPP1R15A* | *674* | *73.5* | *AARQGPWE* | *5%* | *0%* |
| *NCOA7* | *942* | *106.2* | *ARVQGYPW* | *5%* | *0%* |
| *OAS3* | *1,087* | *121.2* | *AWEQGCRQ* | *30%* | *0%* |
| *SAMD9* | *1,589* | *184.3* | *AGYQGEIE* | *45%* | *0%* |
| *ASB11 (isoform 3)* | *302* | *33* | *HTFQGDCW* | *10%* | *0%* |
| *HES4* | *221* | *23.5* | *AGPQGPGG* | *20%* | *0%* |
| *EHD3* | *535* | *60.9* | *AVMQGDME* | *60%* | *0%* |
| *UNC93B1* | *597* | *66.6* | *AGPQGDED* | *5%* | *0%* |
| *BATF2* | *274* | *29.4* | *ATWQGLVV* | *35%* | *0%* |
| *CDHR4* | *788* | *85.8* | *AFEQGQLW* | *10%* | *0%* |
| *C1RL (variant 3)* | *314* | *33.7* | *VQPQGPGK* | *40%* | *0%* |
| *IFI44* | *444* | *50.5* | *CCNQGPTL* | *20%* | *5%* |
| *OAS1 (mutant)* | *400* | *46* | *AWEQGSMK* | *40%* | *0%* |
| *CARD6* | *1,037* | *116.5* | *ASQQGVQM* | *0%* | *0%* |
| *FOXP3* | *431* | *47.2* | *AGSQGPVV* | *0%* | *0%* |
| *APOBEC3A* | *199* | *23.0* | *VDHQGCPF* | *0%* | *0%* |
| *MX1* | *662* | *75.5* | *ALMQGEET* | *50%* | *0%* |
| *NINL-1032* | *1382* | *156.3* | *ADPQGSWQ* | *15%* | *0%* |
| *NINL-827* | *1382* | *156.3* | *AEMQALPK* | *5%* | *0%* |
| *NINL-228* | *1382* | *156.3* | *VGLQGLEK* | *10%* | *0%* |
| **NINL** | **1382** | **156.3** | **ADPQGSWQ** | **15%** | **0%** |
| **GBP5** | **637** | **72.3** | **AVKQGIYS** | **40%** | **5%** |
| **MX1** | **662** | **75.5** | **GAFQSSSA** | **50%** | **0%** |

## REFERENCES:


Andresen, C. A., Smedegaard, S., Sylvestersen, K. B., Svensson, C., Iglesias-Gato, D., Cazzamali, G., Nielsen, T. K., & Flores-Morales, A. (2013). Protein Interaction Screening for the Ankyrin Repeats and Suppressor of Cytokine Signaling (SOCS) Box (ASB) Family Identify Asb11 as a Novel Endoplasmic Reticulum Resident Ubiquitin Ligase. *Journal of Biological Chemistry*, *289*(4), 2043–2054. doi: 10.1074/jbc.m113.534602

Bachmann-Gagescu, R., Dona, M., Hetterschijt, L., Tonnaer, E., Peters, T., Vrieze, E. D., Mans, D. A., Beersum, S. E., Phelps, I. G., Arts, H. H., Keunen, J. E., Ueffing, M., Roepman, R., Boldt, K., Doherty, D., Moens, C. B., Neuhauss, S. C., Kremer, H., & Wijk, E. V. (2015). The Ciliopathy Protein CC2D2A Associates with NINL and Functions in RAB8-MICAL3-Regulated Vesicle Trafficking. *PLOS Genetics*, *11*(10). doi: 10.1371/journal.pgen.1005575

Banaszynski, L. A., Chen, L.-C., Maynard-Smith, L. A., Ooi, A. G. L., & Wandless, T. J. (2006). A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules. *Cell*, *126*(5), 995–1004. doi: 10.1016/j.cell.2006.07.025

Chang, Y. K., Zuo, Z., & Stormo, G. D. (2018). Quantitative profiling of BATF family proteins/JUNB/IRF hetero-trimers using Spec-seq. *BMC Molecular Biology*, *19*(1). doi: 10.1186/s12867-018-0106-7

Chaplin, D. D. (2010). Overview of the immune response. *Journal of Allergy and Clinical Immunology*, *125*(2). doi: 10.1016/j.jaci.2009.12.980

Clavarino, G., Claudio, N., Dalet, A., Terawaki, S., Couderc, T., Chasson, L., Ceppi, M., Schmidt, E. K., Lecuit, M., Gatti, E., & Pierre, P. (2012). Protein phosphatase 1 subunit Ppp1r15a/GADD34 regulates cytokine production in polyinosinic:polycytidylic acid-stimulated dendritic cells. *Proceedings of the National Academy of Sciences*, *109*(8), 3006–3011. doi: 10.1073/pnas.1104491109

Dufner, A., Pownall, S., & Mak, T. W. (2006). Caspase recruitment domain protein 6 is a microtubule-interacting protein that positively modulates NF- B activation. *Proceedings of the National Academy of Sciences*, *103*(4), 988–993. doi: 10.1073/pnas.0510380103

Eskildsen, S. (2003). Characterization of the 2-5-oligoadenylate synthetase ubiquitin-like family. *Nucleic Acids Research*, *31*(12), 3166–3173. doi: 10.1093/nar/gkg427

Fujiwara, Y., Hizukuri, Y., Yamashiro, K., Makita, N., Ohnishi, K., Takeya, M., Komohara, Y., & Hayashi, Y. (2016). Guanylate-binding protein 5 is a marker of interferon-γ-induced classically activated macrophages. *Clinical & Translational Immunology*, *5*(11). doi: 10.1038/cti.2016.59

Galperin, E., Benjamin, S., Rapaport, D., Rotem-Yehudar, R., Tolchinsky, S., & Horowitz, M. (2002). EHD3: A Protein That Resides in Recycling Tubular and Vesicular Membrane Structures and Interacts with EHD1. *Traffic*, *3*(8), 575–589. doi: 10.1034/j.1600-0854.2002.30807.x

Gelderblom, H. R. (1996). Chapter 41 Structure and Classification of Viruses. In *Medical Microbiology. 4th edition*. S.l.: University of Texas Medical Branch at Galveston.

Gil-Yarom, N., Radomir, L., Sever, L., Kramer, M. P., Lewinsky, H., Bornstein, C., Blecher-Gonen, R., Barnett-Itzhaki, Z., Mirkin, V., Friedlander, G., Shvidel, L., Herishanu, Y., Lolis, E. J., Becker-Herman, S., Amit, I., & Shachar, I. (2016). CD74 is a novel transcription regulator. *Proceedings of the National Academy of Sciences*, *114*(3), 562–567. doi: 10.1073/pnas.1612195114

Harris, K. G., Morosky, S. A., Drummond, C. G., Patel, M., Kim, C., Stolz, D. B., Bergelson, J. M., Cherry, S., & Coyne, C. B. (2015). RIP3 Regulates Autophagy and Promotes Coxsackievirus B3 Infection of Intestinal Epithelial Cells. *Cell Host & Microbe*, *18*(2), 221–232. doi: 10.1016/j.chom.2015.07.007

Herold, N. (2019). Overexpression of the Interferon-Inducible Isoform 4 of NCOA7 Dissects the Entry Route of Enveloped Viruses and Demonstrates that HIV Enters Cells via Fusion at the Plasma Membrane. *Viruses*, *11*(2), 121. doi: 10.3390/v11020121

Hotter, D., Sauter, D., & Kirchhoff, F. (2016). Guanylate binding protein 5: Impairing virion infectivity by targeting retroviral envelope glycoproteins. *Small GTPases*, *8*(1), 31–37. doi: 10.1080/21541248.2016.1189990

Huang, W.-C., Tung, S.-L., Chen, Y.-L., Chen, P.-M., & Chu, P.-Y. (2018). IFI44L is a novel tumor suppressor in human hepatocellular carcinoma affecting cancer stemness, metastasis, and drug resistance via regulating met/Src signaling pathway. *BMC Cancer*, *18*(1). doi: 10.1186/s12885-018-4529-9

Ibsen, M. S., Gad, H. H., Thavachelvam, K., Boesen, T., Despres, P., & Hartmann, R. (2014). The 2′-5′-Oligoadenylate Synthetase 3 Enzyme Potently Synthesizes the 2′-5′-Oligoadenylates Required for RNase L Activation. *Journal of Virology*, *88*(24), 14222–14231. doi: 10.1128/JVI.01763-14

Ivashkiv, L. B., & Donlin, L. T. (2015). Regulation of type I interferon responses. *Nature Reviews Immunology*, *14*(1), 36–49. doi: 10.1038/nri3581

Kikkert, M. (2019). Innate Immune Evasion by Human Respiratory RNA Viruses. *Journal of Innate Immunity*, *12*(1), 4–20. doi: 10.1159/000503030

Kowanda, M., Bergalet, J., Wieczorek, M., Brouhard, G., Lécuyer, É., & Lasko, P. (2016). Loss of function of theDrosophilaNinein-related centrosomal protein Bsg25D causes mitotic defects and impairs embryonic development. *Biology Open*, *5*(8), 1040–1051. doi: 10.1242/bio.019638

Larsen, S. K., Munir, S., Woetmann, A., Frøsig, T. M., Odum, N., Svane, I. M., Becker, J. C., & Andersen, M. H. (2013). Functional characterization of Foxp3-specific spontaneous immune responses. *Leukemia*, *27*(12), 2332–2340. doi: 10.1038/leu.2013.196

Mukherjee, A., Morosky, S. A., Delorme-Axford, E., Dybdahl-Sissoko, N., Oberste, M. S., Wang, T., & Coyne, C. B. (2011). The Coxsackievirus B 3Cpro Protease Cleaves MAVS and TRIF to Attenuate Host Type I Interferon and Apoptotic Signaling. *PLoS Pathogens*, *7*(3). doi: 10.1371/journal.ppat.1001311

Narayanan, A., Wu, X., & Yang, Z. R. (2002). Mining viral protease data to extract cleavage knowledge. *Bioinformatics*, *18*(Suppl 1). doi: 10.1093/bioinformatics/18.suppl_1.s5

Nounamo, B., Li, Y., O'Byrne, P., Kearney, A. M., Khan, A., & Liu, J. (2017). An interaction domain in human SAMD9 is essential for myxoma virus host-range determinant M062 antagonism of host anti-viral function. *Virology*, *503*, 94–102. doi: 10.1016/j.virol.2017.01.004

Patel, M. R., Loo, Y.-M., Horner, S. M., Gale, M., & Malik, H. S. (2012). Convergent Evolution of Escape from Hepaciviral Antagonism in Primates. *PLoS Biology*, *10*(3). doi: 10.1371/journal.pbio.1001282

Pohar, J., Pirher, N., Benčina, M., Manček-Keber, M., & Jerala, R. (2012). The Role of UNC93B1 Protein in Surface Localization of TLR3 Receptor and in Cell Priming to Nucleic Acid Agonists. *Journal of Biological Chemistry*, *288*(1), 442–454. doi: 10.1074/jbc.m112.413922

Power, D., Santoso, N., Dieringer, M., Yu, J., Huang, H., Simpson, S., Seth, I., Gao, G., Miao, H., Elledge, S., & Zhu, J. (2015). IFI44 suppresses HIV-1 LTR promoter activity and facilitates its latency. *Virology*, *481*, 142–150. doi: 10.1016/j.virol.2015.02.046

Roncador, G., Brown, P. J., Maestre, L., Hue, S., Martínez-Torrecuadrada, J. L., Ling, K.-L., Pratap, S., Toms, C., Fox, B. C., Cerundolo, V., Powrie, F., & Banham, A. H. (2005). Analysis of FOXP3 protein expression in human CD4 CD25 regulatory T cells at the single-cell level. *European Journal of Immunology*, *35*(6), 1681–1691. doi: 10.1002/eji.200526189

Rouzine, I. M., & Rozhnova, G. (2018). Antigenic evolution of viruses in host populations. *PLOS Pathogens*, *14*(9). doi: 10.1371/journal.ppat.1007291

Rustagi, A., & Gale, M. (2014). Innate Antiviral Immune Signaling, Viral Evasion and Modulation by HIV-1. *Journal of Molecular Biology*, *426*(6), 1161–1177. doi: 10.1016/j.jmb.2013.12.003

Sainz, B., Barretto, N., & Uprichard, S. L. (2009). Hepatitis C Virus Infection in Phenotypically Distinct Huh7 Cell Lines. *PLoS ONE*, *4*(8). doi: 10.1371/journal.pone.0006561

Sotomayor, M., Gaudet, R., & Corey, D. P. (2014). Sorting out a promiscuous superfamily: towards cadherin connectomics. *Trends in Cell Biology*, *24*(9), 524–536. doi: 10.1016/j.tcb.2014.03.007

Stabell, A. C., Meyerson, N. R., Gullberg, R. C., Gilchrist, A. R., Webb, K. J., Old, W. M., Perera, R., & Sawyer, S. L. (2018). Dengue viruses cleave STING in humans but not in nonhuman primates, their presumed natural reservoir. *ELife*, *7*. doi: 10.7554/elife.31919

Stuchbury, G., & Münch, G. (2010). Optimizing the generation of stable neuronal cell lines via pre-transfection restriction enzyme digestion of plasmid DNA. *Cytotechnology*, *62*(3), 189–194. doi: 10.1007/s10616-010-9273-1

Verhelst, J., Hulpiau, P., & Saelens, X. (2013). Mx Proteins: Antiviral Gatekeepers That Restrain the Uninvited. *Microbiology and Molecular Biology Reviews*, *77*(4), 551–566. doi: 10.1128/mmbr.00024-13

Voskarides, K., Christaki, E., & Nikolopoulos, G. K. (2018). Influenza Virus—Host Co-evolution. A Predator-Prey Relationship? *Frontiers in Immunology*, *9*. doi: 10.3389/fimmu.2018.02017

Wicher, K. B., & Fries, E. (2004). Prohaptoglobin is proteolytically cleaved in the endoplasmic reticulum by the complement C1r-like protein. *Proceedings of the National Academy of Sciences*, *101*(40), 14390–14395. doi: 10.1073/pnas.0405692101

Zhou, M., Yan, J., Ma, Z., Zhou, Y., Abbood, N. N., Liu, J., Su, L., Jia, H., & Guo, A.-Y. (2012). Comparative and Evolutionary Analysis of the HES/HEY Gene Family Reveal Exon/Intron Loss and Teleost Specific Duplication Events. *PLoS ONE*, *7*(7). doi: 10.1371/journal.pone.0040649