Journal of Visualized Experiments

Analysis of Group IV viral SSHHPS Using In vitro and In silico Methods

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE60421R2
Full Title:	Analysis of Group IV viral SSHHPS Using In vitro and In silico Methods
Section/Category:	JoVE Immunology and Infection
Keywords:	Enzyme, Assay, FRET, nonstructural protein, alphavirus, Group IV, protease, SSHHPS, gel assay, in vitro, docking
Corresponding Author:	Patricia Legler NRL Washington, DC UNITED STATES
Corresponding Author's Institution:	NRL
Corresponding Author E-Mail:	patricia.legler@nrl.navy.mil
Order of Authors:	Xin Hu
	Jaimee R. Compton
	Patricia Legler
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Washington, DC, US

TITLE:

Analysis of Group IV Viral SSHHPS Using In Vitro and In Silico Methods

AUTHORS AND AFFILIATIONS:

Xin Hu[1], Jaimee R. Compton[2], Patricia M. Legler[2]

[1]National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD

[2]United States Naval Research Laboratory, Washington, D.C.

Email addresses of co-authors:

Xin Hu (xin.hu@nih.gov)

Jaimee R. Compton (jaimee.compton@nrl.navy.mil)

Corresponding author:

Patricia M. Legler (patricia.legler@nrl.navy.mil)

KEYWORDS:

enzyme, assay, FRET, nonstructural protein, alphavirus, Group IV, protease, SSHHPS, gel assay, in vitro, docking

SUMMARY:

We present a general protocol for identifying short stretches of homologous host-pathogen protein sequences (SSHHPS) embedded in the viral polyprotein. SSHHPS are recognized by viral proteases and direct the targeted destruction of specific host proteins by several Group IV viruses.

ABSTRACT:

Alphaviral enzymes are synthesized in a single polypeptide. The nonstructural polyprotein (nsP) is processed by its nsP2 cysteine protease to produce active enzymes essential for viral replication. Viral proteases are highly specific and recognize conserved cleavage site motif sequences (~6-8 amino acids). In several Group IV viruses, the nsP protease(s) cleavage site motif sequences can be found in specific host proteins involved in generating the innate immune responses and, in some cases, the targeted proteins appear to be linked to the virus-induced phenotype. These viruses utilize short stretches of homologous host-pathogen protein sequences (SSHHPS) for targeted destruction of host proteins. To identify SSHHPS the viral protease cleavage site motif sequences can be inputted into BLAST and the host genome(s) can be searched. Cleavage initially can be tested using the purified nsP viral protease and fluorescence resonance energy transfer (FRET) substrates made in *E. coli*. The FRET substrates contain cyan and yellow fluorescent protein and the cleavage site sequence (CFP-sequence-YFP). This protease assay can be used continuously in a plate reader or discontinuously in SDS-PAGE gels. Models of the bound peptide substrates can be generated in silico to guide substrate selection and mutagenesis studies. CFP/YFP substrates have also been utilized to identify protease inhibitors. These in vitro and in silico methods can be used in combination with cell-

based assays to determine if the targeted host protein affects viral replication.

INTRODUCTION:

Evidence of horizontal gene transfer from virus to host, or host to virus, can be found in a variety of genomes[1-4]. Examples of viral endogenization are the CRISPR spacer sequences found in bacterial host genomes[4]. Recently, we have found evidence of host protein sequences embedded in the nonstructural polyproteins of (+)ssRNA Group IV viruses. These sequences within the coding regions of the viral genome can be propagated generationally. The short stretches of homologous host-pathogen protein sequences (SSHHPS) are found in the virus and host[5,6]. SSHHPS are the conserved cleavage site motif sequences recognized by viral proteases that have homology to specific host proteins. These sequences direct the destruction of specific host proteins.

In our previous publication[6], we compiled a list of all of the host proteins that were targeted by viral proteases and found that the list of targets was non-random (**Table 1**). Two trends were apparent. First, the majority of the viral proteases that cut host proteins belonged to Group IV viruses (24 of 25 cases involved Group IV viral proteases), and one protease belonged to the (+)ssRNA Group VI retroviruses (HIV, human immunodeficiency virus)[7]. Second, the host protein targets being cut by the viral proteases were generally involved in generating the innate immune responses suggesting that the cleavages were intended to antagonize the host's immune responses. Half of the host proteins targeted by the viral proteases were known components of signaling cascades that generate interferon (IFN) and proinflammatory cytokines (**Table 1**). Others were involved in host cell transcription[8-10] or translation[11]. Interestingly, Shmakov et al.[4] have shown that many CRISPR protospacer sequences correspond to genes involved in plasmid conjugation or replication[4].

Group IV includes, among others, *Flaviviridae, Picornaviridae, Coronaviridae, Calciviridae,* and *Togaviridae*. Several new and emerging pathogens belong to Group IV such as the Zika virus (ZIKV), West Nile (WNV), Chikungunya (CHIKV), severe acute respiratory syndrome virus (SARS) and Middle East respiratory syndrome virus (MERS). The (+)ssRNA genome is essentially a piece of mRNA. To produce the enzymes necessary for genome replication, the (+)ssRNA genome first must be translated. In alphaviruses and other Group IV viruses, the enzymes necessary for replication are produced in a single polyprotein (i.e., nsP1234 for VEEV). The nonstructural polyprotein (nsP) is proteolytically processed (nsP1234 $\rightarrow$ nsP1, nsP2, nsP3, nsP4) by the nsP2 protease to produce active enzymes[12] (**Figure 1**). Cleavage of the polyprotein by the nsP2 protease is essential for viral replication; this has been demonstrated by deletion and site-directed mutagenesis of the active site cysteine of the nsP2 protease[13,14]. Notably, the translation of viral proteins precedes genome replication events. For example, nsP4 contains the RNA-dependent RNA polymerase needed to replicate the (+)ssRNA genome. Genome replication can produce dsRNA intermediates; these intermediates can trigger the host's innate immune responses. Thus, these viruses may cleave host innate immune response proteins early in infection in order to suppress their effects[15-17].

Silencing can occur at the level of DNA, RNA, and protein. What is common to each of the

silencing mechanisms shown in **Figure 1** is that short foreign DNA, RNA, or protein sequences are used to guide the destruction of specific targets to antagonize their function. The silencing mechanisms are analogous to "search and delete" programs that have been written in three different languages. The short cleavage site sequence is analogous to a "keyword". Each program has an enzyme that recognizes the match between the short sequence (the "keyword") and a word in the "file" that is to be deleted. Once a match is found, the enzyme cuts ("deletes") the larger target sequence. The three mechanisms shown in **Figure 1** are used to defend the host from viruses, or to defend a virus from a host's immune system.

Viral proteases recognize short cleavage site motif sequences between ~2-11 amino acids; in nucleotides, this would correspond to 6-33 bases. For comparison, CRISPR spacer sequences are ~26-72 nucleotides and RNAi are ~20-22 nucleotides[18,19]. While these sequences are relatively short, they can be recognized specifically. Given the higher diversity of amino acids, the probability of a random cleavage event is relatively low for a viral protease recognizing protein sequences of 6-8 amino acids or longer. The prediction of SSHHPS in host proteins will largely depend upon the specificity of the viral protease being examined. If the protease has strict sequence specificity requirements the chance of finding a cleavage site sequence is $1/20^6 = 1$ in 64 million or $1/20^8 = 1$ in 25.6 billion; however, most proteases have variable subsite tolerances (e.g., R or K may be tolerated at the S1 site). Consequently, there is no requirement for sequence identity between the sequences found in the host versus the virus. For viral proteases that have looser sequence requirements (such as those belonging to *Picornaviridae)* the probability of finding a cleavage site in a host protein may be higher. Many of the entries in **Table 1** are from the *Picornaviridae* family.

Schechter & Berger notation[20] is commonly used to describe the residues in a protease substrate and the subsites to which they bind, we utilize this notation throughout. The residues in the substrate that are N-terminal of the scissile bond are denoted as P3-P2-P1 while those that are C-terminal are denoted as P1'-P2'-P3'. The corresponding subsites in the protease that bind these amino acid residues are S3-S2-S1 and S1'-S2'-S3', respectively.

To determine which host proteins are being targeted, we can identify SSHHPS in the viral polyprotein cleavage sites and search for the host proteins that contain them. Herein, we outline procedures for identifying SSHHPS using known viral protease cleavage site sequences. The bioinformatic methods, protease assays, and in silico methods described are intended to be used in conjunction with cell-based assays.

Sequence alignments of the host proteins targeted by viral proteases have revealed species-specific differences within these short cleavage site sequences. For example, the Venezuelan equine encephalitis virus (VEEV) nsP2 protease was found to cut human TRIM14, a tripartite motif (TRIM) protein[6]. Some TRIM proteins are viral restriction factors (e.g., TRIM5α[21]), most are thought to be ubiquitin E3 ligases. TRIM14 lacks a RING (really interesting new gene) domain and is not thought to be an E3 ligase[22]. TRIM14 has been proposed to be an adaptor in the mitochondrial antiviral signalosome (MAVS)[22], but may have other antiviral functions[23]. Alignment of TRIM14 sequences from various species shows that equine lack the cleavage site

and harbor a truncated version of TRIM14 that is missing the C-terminal PRY/SPRY domain. This domain contains a polyubiquitination site (**Figure 2**). In equine, these viruses are highly lethal (~20-80% mortality) whereas in humans only ~1% die from VEEV infections[24]. Cleavage of the PRY/SPRY domain may transiently short circuit the MAVS signaling cascade. This cascade can be triggered by dsRNA and leads to the production of interferon and pro-inflammatory cytokines. Thus, the presence of the SSHHPS may be useful for predicting which species have defense systems against specific Group IV viruses.

In Group IV viruses, IFN antagonism mechanisms are thought to be multiply redundant[25]. Host protein cleavage may be transient during infection and concentrations may recover over time. We found in cells that TRIM14 cleavage products could be detected very early after transfection (6 h) with a plasmid encoding the protease (cytomegalovirus promoter). However, at longer periods, the cleavage products were not detected. In virus-infected cells, the kinetics were different and cleavage products could be detected between 6-48 h[6]. Others have reported the appearance of host protein cleavage products as early as 3-6 h post infection[9,11].

Proteolytic activity in cells is often difficult to catch; the cleavage products can vary in their solubility, concentration, stability, and lifetime. In cell-based assays, it cannot be assumed that cleavage products will accumulate in a cell or that the band intensities of cut and uncut protein will show compensatory increases and decreases as the cut protein may be degraded very quickly and may not be detectable in a Western blot at an expected molecular weight (MW) (e.g., the region containing the epitope could be cleaved by other host proteases or could be ubiquitinated). If the substrate of the viral protease is an innate immune response protein, its concentration may vary during infection. For example, some innate immune response proteins are present prior to viral infection and are induced further by interferon[26]. The concentration of the target protein may therefore fluctuate during infection and comparison of uninfected vs. infected cell lysates may be difficult to interpret. Additionally, all cells may not be uniformly transfected or infected. In vitro protease assays using purified proteins from *E. coli* on the other hand have fewer variables for which to control and such assays can be done using SDS-PAGE rather than immunoblots. Contaminating proteases can be inhibited in the early steps of the protein purification of the CFP/YFP substrate, and mutated viral proteases can be purified and tested as controls to determine if the cleavage is due to the viral protease or a contaminating bacterial protease.

One limitation of in vitro protease assays is that they lack the complexity of a mammalian cell. For an enzyme to cut its substrate, the two must be co-localized. Group IV viral proteases differ in structure and localization. For example, the ZIKV protease is embedded in the endoplasmic reticulum (ER) membrane and faces the cytosol, whereas the VEEV nsP2 protease is a soluble protein in the cytoplasm and nucleus[27]. Some of the cleavage site sequences found in the ZIKV SSHHPS analysis were in signal peptides suggesting that cleavage might occur co-translationally for some targets. Thus, the location of the protease and the substrate in the cell also needs to be considered in these analyses.

Cell-based assays can be valuable for establishing a role for the identified host protein(s) in

infection. Methods that aim to halt viral protease cleavage of host proteins such as the addition of a protease inhibitor[6] or a mutation in the host target[16] can be used to examine their effects on viral replication. Overexpression of the targeted protein also may affect viral replication[28]. Plaque assays or other methods can be used to quantify viral replication.

PROTOCOL:

1. Bioinformatics: identification of SSHHPS in the host genome using BLAST

NOTE: Protein BLAST can be found at blast.ncbi.nlm.nih.gov/Blast.cgi.

1.1. Input ~20 amino acids surrounding the scissile bond in the viral polyprotein. Select **non-redundant protein sequences** and type in the host genome to be searched (e.g., *Homo sapiens*).

1.1.1. If needed, select PHI-BLAST. Type in a pattern sequence (e.g., for the 25 residues of V12 shown below enter the pattern "AG" without quotes).

VEEV nsP1/2 V12 VEEPTLEADVDLML**Q**E**A**G**A**↓**G**SVETP
VEEV nsP2/3 V23 LSSTLTNIYTGSRL**H**E**A**G**C**↓**A**PSYHV
VEEV nsP3/4 V34 TREEFEAFVAQQQR**F**D**A**G**A**↓**Y**IFSSD

NOTE: In PHI-BLAST, square brackets [XY] indicate that amino acid X or Y can be at the subsite position (e.g., AG[AC][GAY]).

1.1.2. Inspect the BLAST results and identify the hits that have high sequence identity to residues that are conserved in the polyprotein cleavage sites (e.g., tripartite motif protein 14) (**Figure 3**).

NOTE: For serine proteases higher conservation of the P1 residue is expected, while for cysteine proteases higher conservation of the P2 residue is expected.

1.1.3. Color the residues that are identical to a cleavage site sequence and are in sequential order (no gaps). Color the residues tolerated at the subsite, but present in a different cleavage site in a second color.

NOTE: Residues that represent conservative substitutions (e.g., Leu vs. Val) that are not present in a viral cleavage site also may be found and may or may not be recognized by the viral protease.

1.1.4. Rank order the BLAST hits based upon the number of consecutive identical or tolerated residues that match a cleavage site sequence. From the list, select the proteins containing ≥6 identical or similar residues for analysis in protease assays.

1.1.5. Repeat the procedure for the other cleavage sites (nsP2/3, nsP3/4, etc.) and gradually strengthen the prediction by adding more highly conserved residues to the PHI-BLAST pattern.

2. In vitro assays: designing and preparing protease substrates

2.1. Construct a plasmid encoding the cyan fluorescent protein (CFP), ≤25 amino acids of the cleavage site sequence, followed by the yellow fluorescent protein (YFP, also known as Venus[29]).

NOTE: The plasmid can be constructed using sequence and ligation independent cloning (SLIC)[30] or commercial gene synthesis. A pet15b plasmid containing the sequence shown in **Figure 4** was synthesized commercially and was used here.

2.1.1. To optimize the substrate length, construct additional variable length FRET substrates containing 12-25 amino acids of the natural viral polyprotein cleavage site sequences using a 2-fragment SLIC reaction. Analyze cleavage using the SDS-PAGE gel-based assay or by measuring steady state kinetic parameters using the methods below.

NOTE: In some cases, cleavage sites can be identified by homology to known cleavage sites[31]. If cleavage of the substrates containing the polyprotein junction sequences is not observed, there may be a requirement for additional residues or a structural motif (e.g., an alpha helix[32]). Alternatively, the purified viral protease may be inactive. Confirm cleavage of the viral polyprotein sequences before pursuing SSHHPS analysis. The number of residues in the substrate was optimized for the VEEV protease using variable length substrates (12 to 25 amino acids) followed by analysis of V_{max} and K_m[32,33]. The Zika viral ns2B/nsB protease cleavage sites used in the examples have been published[34,35].

2.2. Prepare the CFP/YFP substrates by freshly transforming 8-20 µL of BL-21(DE3) *E. coli* competent cells with the CFP-V12-YFP plasmid according to manufacturer's directions and plate on Luria Bertani (LB) agar plates containing 50 µg/mL Ampicillin (37 °C).

2.2.1. Autoclave four 4 L flasks containing LB media (1.5 L media per flask) and 100 mL of LB in a 250 mL flask. Cap each flask with aluminum foil.

2.2.2. Inoculate the 100 mL culture with a colony of the freshly transformed bacteria and grow at 37 °C with shaking (200 rpm) overnight.

2.2.3. To make the CFP/YFP substrate, inoculate four 4 L flasks with 25 mL of an overnight culture. Begin shaking the cultures at 37 °C and monitor growth by UV-vis spectroscopy at 600 nm hourly.

2.2.4. When the bacteria reach an absorbance of ~1.0 at 600 nm (approximately ~3-4 h of growth) induce protein expression by adding 0.5 mL of 1 M isopropyl-β-D-thiogalactoside (IPTG) per flask. After adding IPTG, lower the temperature of the shaking incubator to 17 °C and allow expression to continue overnight for 17-20 h.

2.2.5. Pellet the bacteria using a high-speed centrifuge at 7,000 x *g* for 10 min (4 °C) and retain the pellets. Remove and discard liquid media. Store the pellets at -80 °C or lyse immediately.

2.2.6. Prepare 100 mL of lysis buffer containing 50 mM Tris pH 7.6, 500 mM NaCl, 35 mL of bacterial protein extraction reagent, 30 mg of lysozyme, 25 U of DNase, and 1 protease inhibitor tablet. Resuspend the pellets in lysis buffer with a pipette and transfer ~25-35 mL into 50 mL disposable conical tubes.

2.2.7. Place the tubes in a plastic beaker containing ice water. Insert the sonicator tip into the tubes so that the tip is ~1 cm from the bottom of the tube and sonicate the lysates 10-20 times on level 5 for 15 s intervals until the lysate becomes fluid and liquefied.

NOTE: Use hearing protection during sonication.

2.2.8. Transfer the lysate to high speed centrifuge tubes and centrifuge at 20,500 x *g* for 30 min at 4 °C. After the spin, retain the supernatant (~100 mL) and transfer it to a clean bottle. Discard the pellets.

2.2.9. Prepare 1 L of Buffer A (50 mM Tris pH 7.6, 500 mM NaCl). Prepare 300 mL of Buffer B (50 mM Tris pH 7.6, 500 mM NaCl, 300 mM Imidazole).

2.2.10. Equilibrate a 100 mL nickel column using 3 column volumes of Buffer A and a flow rate of 5 mL/min.

2.2.11. Load the lysate onto the nickel column using a flow rate of 2 to 5 mL/min. Wash the column with 2 column volumes of Buffer A, followed by ~5 column volumes of 20% Buffer B. During the 20% Buffer B wash, the absorbance at 280 nm (A_{280}) will increase as contaminants elute from the column during the wash. Continue washing the column until the A_{280} of the eluate has returned to baseline values.

2.2.12. Elute the protein with 2-3 column volumes of 100% Buffer B using a flow rate of 2-5 mL/min and collect 10 mL fractions. Measure the A_{280} of each fraction.

2.2.13. Combine and concentrate fractions containing A_{280} > 0.1 using a 15 mL centrifugal ultrafiltration unit. Spin the ultrafiltration units at 5,000 x *g* for 15 min and continue to add fractions until the volume has been reduced to ~50-75 mL.

2.2.14. Cut a 14 inch piece of dialysis tubing with a molecular weight cut-off (MWCO) of 6-8 kDa. Hydrate the dialysis tubing by boiling it fully submerged in 300 mL of water for 10 min. Tie a secure knot at one end of the membrane. Fill the bag with dialysis buffer to ensure that no cracks or leaks are present. Remove the buffer from the bag and keep the bag submerged in the dialysis buffer.

2.2.15. Transfer the concentrated protein from 2.2.13 into the dialysis bag with a plastic pipette. Remove any air bubbles from the bag. Close the bag with a second knot or a dialysis clip. Dialyze the protein against 500 mL of 50 mM Tris pH 7.6, 5 mM EDTA (ethylenediaminetetraacetic acid), 250 mM NaCl in a 500 mL graduated cylinder overnight at 4 °C.

2.2.16. Dialyze the protein a second time against 500 mL of 50 mM Tris pH 7.6 at 4 °C for 2 h.

2.3. For the anion exchange column prepare 500 mL of Buffer A (50 mM Tris pH 7.6) and 500 mL Buffer B (50 mM Tris pH 7.6, 1.0 M NaCl). Equilibrate a 30 mL anion exchange column with 3 column volumes of Buffer A (2-5 mL/min).

2.3.1. Remove the protein from the dialysis bag and transfer to a bottle. Keep the bottle on ice. Load the dialyzed protein onto the column (2-5 mL/min).

NOTE: The CFP/YFP protein will bind the column and will be yellow in appearance.

2.3.2. Wash the column with Buffer A until the A_{280} returns to baseline (5 mL/min). Elute the protein using a gradient (0-50% Buffer B, 100 mL) and collect 10 mL fractions.

2.3.3. Inspect the column fractions using SDS-PAGE. Combine those that are >95% pure.

2.3.4. Concentrate the protein to an A_{280} ~10-20 using a 15 mL centrifugal ultrafiltration unit. Spin the concentrator at 4,500 x *g* for 10 min at 4 °C and continue to add protein until all of the protein-containing fractions have been combined.

2.4. Carefully remove the protein from the concentrator with a pipette. Aliquot the protein into 1.5 mL microcentrifuge tubes and flash freeze in liquid nitrogen for long term storage at -80 °C. Buffer exchange the protein at room temperature using a PD-10 column equilibrated with the appropriate assay buffer prior to use.

2.5. Using Beer's law calculate the protein concentration using the A_{280} and a calculated extinction coefficient (e.g., for the V12 substrate the ε = 47,790 M^{-1} cm^{-1}).

NOTE: The extinction coefficient (ε) can be calculated from the protein sequence in **Figure 4** using the Expasy ProtParam program (https://web.expasy.org/protparam/).

3. Preparation of the alphaviral nsP2 cysteine protease

3.1. Design and construct a plasmid encoding the protease. For cysteine proteases, use the pet32 plasmid to construct a thioredoxin (Trx) fusion protein.

NOTE: The pet32 plasmid encodes a thrombin cleavage site (LVPR↓GS) for removal of the thioredoxin and His-tag (**Figure 5**). Thioredoxin will help maintain the active site cysteine in a reduced state during expression. For serine proteases, the thioredoxin is not needed and steps involving its removal by thrombin can be omitted. The VEEV nsP2 protease sequence was incorporated into a pet32b plasmid that was prepared commercially to avoid handling Select agents.

3.1.1. Freshly transform the plasmid DNA into BL21(DE3)pLysS *E. coli* according to manufacturer's directions. Plate the bacteria on LB agar plates containing Ampicillin.

NOTE: Chloramphenicol is only used for *E. coli* strains carrying the pLysS plasmid and is omitted if BL21(DE3) cells are used. It is not necessary to include chloramphenicol on the LB agar plate in this step.

3.1.2. Autoclave four 4 L flasks of 1.5 L of LB media (6 L total volume) and 100 mL of LB in a 250 mL flask. Cap each flask with aluminum foil.

3.1.3. Inoculate a 100 mL overnight culture of LB/Ampicillin with a colony from the plate and grow in a shaking incubator (200 rpm) at 37 °C.

3.1.4. Inoculate the 4 L flasks with 25 mL of the overnight culture and add the appropriate antibiotics.

NOTE: The media for the BL21(DE3) pLysS cells carrying the pet32 plasmid should have final concentrations of 25 µg/mL chloramphenicol and 50 µg/mL Ampicillin.

3.1.5. Induce protein expression by adding 0.5 mL of IPTG to the culture when the absorbance at 600 nm reaches 1.0. Lower the temperature of the shaking incubator to 17 °C. Allow expression to continue overnight (~17 h).

3.1.6. Pellet the cells by centrifugation (7,000 x *g* for 10 min at 4 °C). Remove and discard the liquid media.

NOTE: The pellets can be stored at -80 °C for months or lysed immediately.

3.1.7. Prepare 100 mL of lysis buffer (50 mM Tris pH 7.6, 500 mL NaCl, 2 mM beta mercaptoethanol (BME), 30 mg lysozyme, 5% glycerol, 25 U DNase, 35 mL bacterial protein extraction reagent). Open bottles of BME in a chemical hood when adding. Keep the bacterial lysate on ice or at 4 °C for this and all subsequent steps.

NOTE: For cysteine proteases, 2 mM BME is included to keep the nucleophilic cysteine reduced. The columns can be run at room temperature using chilled buffers. Buffers should be made with cold deionized water cooled to 4 °C.

3.1.8. Resuspend the bacterial pellets in ~25 mL of lysis buffer and transfer ~25 mL of the lysate into 4 x 50 mL disposable conical tubes. Place the tubes into plastic beakers containing ice water. Sonicate the lysate 10 times on level 5 for 15 s intervals.

3.1.9. Transfer the lysate into high speed centrifuge tubes. Clarify the lysate by centrifugation (30 min, 20,500 x *g* at 4 °C).

3.1.10. Prepare 0.5 L of Buffer A (50 mM Tris pH 7.6, 500 mM NaCl, 5% glycerol, 2 mM BME) and chill to 4 °C.

3.1.11. Prepare 250 mL of Buffer B (50 mM Tris pH 7.6, 500 mM NaCl, 5% glycerol, 2 mM BME, 300 mM imidazole) and chill to 4 °C.

3.1.12. Equilibrate a 50 mL nickel column with 3 column volumes of Buffer A. Load the clarified lysate onto the column at 2-5 mL/min and discard the pellets.

3.1.13. Wash the column (2-5 mL/min) with 2 column volumes of Buffer A followed by 5 column volumes of Buffer A containing 20% Buffer B (60 mM Imidazole). Elute the protein (5 mL/min) with 100% Buffer B and collect 10 mL fractions.

3.1.14. Combine and concentrate fractions containing the protease that have $A_{280} \geq 0.1$ using a 15 mL centrifugal ultrafiltration unit and 15 min spins at 5,000 x *g* at 4 °C. After the volume has been reduced to ~ 5 mL, buffer exchange the protein in the concentration unit by adding fresh dialysis buffer to the protein (50 mM Tris pH 7.6, 250 mM NaCl, 5 mM dithiothreitol (DTT), 1 mM EDTA, 5% Glycerol). Spin again at 5,000 x *g* at 4 °C for 15 min; repeat the buffer exchange step 2-3 times. Add thrombin to the protein (20 µL of 1 unit/µL) prior to dialysis to remove the thioredoxin and His-tag.

3.1.15. Transfer the protein into a dialysis bag and dialyze against 500 mL of the dialysis buffer (4 °C) in a 500 mL graduated cylinder overnight.

NOTE: The FPLC (fast protein liquid chromatography) system and the nickel column should be thoroughly cleaned with stripping buffer (2 M NaCl, 50 mM EDTA) before proceeding to the anion exchange column. Any residual nickel in the FPLC lines will turn the buffer solutions containing DTT brown when mixed. Wash the nickel column and FPLC system with 4 column volumes of water. Pump wash the FPLC system thoroughly with water. The nickel column can be regenerated by flowing 2 column volumes of 0.2 M nickel sulfate over the resin for subsequent purifications.

3.2. For the anion exchange column prepare 1 L of Buffer A (50 mM Tris pH 7.6, 5 mM DTT, 5% glycerol).

3.2.1. Prepare 0.5 L of Buffer B (50 mM Tris pH 7.6, 5 mM DTT, 5% glycerol, 1.25 M NaCl).

3.2.2. Equilibrate a 30 mL anion exchange column with Buffer A (3 column volumes, 2-5 mL/min). Place the tubes in the fraction collector for collection of the flow through.

NOTE: The VEEV protease has a calculated isoelectric point (pI) of 8.7 and will bind cation-exchange columns but will flow through anion exchange columns. The pI can be calculated from the protein sequence using the Expasy ProtParam program (https://web.expasy.org/protparam/).

3.2.3. Dilute the dialyzed protein 1:3 with Buffer A, and then load the protein (5 mL/min). Collect the flow-through in 10 mL fractions.

3.3. Remove the anion exchange column from the FPLC system. Connect a cation exchange column to the FPLC system. Equilibrate a 30 mL cation exchange column with 3 column volumes of Buffer A (5 mL/min).

3.3.1. Load the flow through of the anion exchange column onto the cation exchange column at 2-5 mL/min. Wash the column with Buffer A until the A_{280} returns to baseline level. Elute the protein with a 100 mL gradient (0-50% Buffer B) and collect 10 mL fractions.

NOTE: The VEEV protease will elute at around 0.6 M NaCl.

3.3.2. Inspect the column fractions using SDS-PAGE. Combine fractions that are >95% pure and concentrate to an $A_{280} \approx 2$ using 15 mL centrifugal ultrafiltration units. The enzyme can be flash-frozen in liquid nitrogen and stored at -80 °C.

4. Assaying the enzyme continuously using a plate reader

4.1. Prepare 50 mL of assay buffer (50 mM HEPES pH 7.0).

4.1.1. As the alphaviral proteases have relatively low k_{cat} values, dilute the enzyme in the assay buffer to 4.7 µM (this will roughly correspond to an $A_{280} = 0.2$ for the VEEV protease without Trx).

4.1.2. To measure the activity of the enzyme, prepare a stock of substrate in the assay buffer with a concentration of 185 µM; this will roughly correspond to an $A_{280} = 9$. In 8 microcentrifuge tubes, prepare the reaction mixes shown in **Table 2** by combining the appropriate volumes of the 185 µM substrate stock and buffer. In a black half-area 96-well plate pipet 45 µL of the reaction mixes into 3 wells (columns 1, 2, 3). Row A should contain the [S] = 5 µM reaction mix, and Row H should contain the [S] = 140 µM reaction mix.

4.1.3. Set the plate reader to detect simultaneously fluorescence at two wavelengths with a fixed photomultiplier tube (PMT) setting (e.g., low):

Wavelength 1	excitation = 434 nm,	emission = 527 nm
Wavelength 2	excitation = 434 nm,	emission = 470 nm

4.1.4. Set the read time to 20 min (measuring 1 read per minute) and select the wells to be read. Insert the plate into the plate reader and measure the spontaneous rate of hydrolysis for 20 min. Monitor the emission ratios (emission at 527/emission at 470) over time.

4.1.5. Run an endpoint read of the plate containing the “UNCUT” substrate.

NOTE: These values will be used in subsequent data calculations. The average of the emission ratios from 3 wells will be the values of the “UNCUT” substrate at t=0 in **Table 3**.

4.1.6. Remove the plate and pipet 5 µL of enzyme into each well. Read the plate again for 20 min with 1 read per minute. Set the plate reader to output absolute values.

NOTE: For this assay, the slopes will be negative. Each well will contain a total volume of 50 µL.

4.1.7. At the end of the read, seal the plate with film to prevent evaporation. Leave the plate at room temperature overnight to allow the enzyme to cut the substrate completely.

4.1.8. After ~24 h, remove the sealing film and perform an endpoint read of the plate using the same PMT as in the prior plates. Average these emission ratios and input into **Table 3** under "CUT". Confirm the cleavage of the substrate using the SDS-PAGE discontinuous assay described below (Step 5.1.).

4.1.9. Export the data to a spreadsheet. Output the fluorescence units at each time point for the 2 wavelengths (**Table 4**).

4.1.10. Calculate the nmol of substrate that have been cut at time t using equation (1) where X is the emission ratio (527 nm/470 nm) at a given time point, *neg* is the emission ratio of the "UNCUT" substrate at t=0, and *pos* is the emission ratio of the completely "CUT" substrate measured after 24 h of cutting (**Table 3**).

$$\frac{X-neg}{(pos-neg)} * nmol = nmols\ of\ S\ cut\ at\ time\ t \qquad (1)$$

NOTE: Representative fluorescence data are shown for one well (well E7) containing 80 µM substrate (4 nmol of S per well) in **Table 4**. The calculations were performed for each well in the plate.

4.1.11. For each well, plot nmol vs. time (min) and obtain the initial velocities (slopes) by fitting the data to y = mx + b. For the data collected in 4.1.5, plot nmol vs. time (min) for each well. The slope will equal the nmol product produced per minute. Subtract the spontaneous rates of hydrolysis measured in 4.1.5 from the enzyme-catalyzed reaction rates (**Table 5**).

NOTE: The first read can be clipped from the data if it is artifactually high due to movement of the plate into the plate reader.

4.1.12. Calculate the amount of enzyme in mg that was added to each well (e.g., 0.0009 mg). A unit is defined as a µmol of product produced per minute (µmol/min). Divide the nmol/min by the mg of enzyme present in the well to obtain mU/mg; divide by 1000 to obtain U/mg.

4.1.13. Plot [S] µM on the x-axis and U/mg on the y-axis and fit the data to the Michaelis-Menten equation to obtain V_{max} and K_m. This can be done in the software (e.g., GraFit).

5. Assaying the enzyme discontinuously using SDS-PAGE analysis

5.1. Prepare a 50 µL reaction containing 10 µM substrate and buffer in place of enzyme and label as "UNCUT."

NOTE: The volumes of substrate and buffer are shown in **Table 2**. If the continuous assay has been run, the samples can be used directly from the 96-well plate.

5.1.1. Prepare a 50 µL reaction containing 10 µM substrate and 5 µL enzyme and label as "CUT". Start the timer when the enzyme is added to the substrate.

NOTE: Inhibitors can be added to additional tubes containing enzyme and substrate. Adjust the volume of added buffer to compensate for the added volume of inhibitor. Concentrations of DMSO should not exceed 2%.

5.1.2. Incubate the reactions for ~15-24 h at room temperature (22 ± 3 °C). Stop the reactions by adding 50 µL of 2x Laemelli buffer. After stopping the reaction boil, each tube for 3-10 min.

5.1.3. Assemble the gel tank according to the manufacturer's directions. Insert a 17-well pre-cast 12% polyacrylamide gel cassette and a buffer dam on the other side. Fill the interior reservoir of the cell with 1x SDS running buffer until the buffer reaches the top of the cassette. Fill the external reservoir half-full with the same buffer.

5.1.4. To analyze cleavage using the discontinuous assay, load 5 µL of each reaction mixture into a lane of an SDS-PAGE gel beginning with the "UNCUT" reaction. Include a molecular weight marker in the first or last lane.

5.1.5. Attach the electrodes of the gel tank to the power supply and separate the products at 110 V for 60 min. Remove the gel from the cassette by inserting the cracking tool in between the plates. Place the gel in a plastic tray and submerge the gel in 5-10 mL of gel staining solution; bands will be visible within 30 min. After 1-24 hours remove the excess stain, submerge the gel in water and use a gel imager to take a picture of the gel.

6. Docking substrate peptides to the VEEV-nsP2 cysteine protease

6.1. Download the coordinate file for the VEEV cysteine protease from the PDB (https://www.rcsb.org/). The PDB code is 2HWK. Save the file as 2HWK.pdb.

6.1.1. Prepare the protein structure using MOE (https://www.chemcomp.com/). Load the protein PDB file into MOE. Click the **Select** and **Solvent** on the right hand side bar and delete the solvent.

6.1.2. Open the **Structure Preparation** panel from the top menu bar **Protein**. Automatically correct all structural items by clicking on **Correct** and protonate the structure by clicking on

Protonate3D. Add partial charges to the protein by opening **Partial Charges** panel and selecting **Amber 99** and **Adjust hydrogens and Lone pairs as required**. Finally, save the structure file as "2HWK_dock.pdb".

6.2. Build the structure for the substrate peptides (nsP12, nsP23, nsP34) and TRIM14 using MOE. Open the **Protein Builder** panel, enter the substrate sequence, set the **Geometry** as **Extended**, and click on **Build**. The structure will be shown in MOE window.

6.2.1. Minimize the peptide structure by clicking **Minimize** on the panel. Save the structure as a PDB file (**Figure 6**).

6.3. Dock the substrate peptides to VEEV-nsP2 using PyRx/AutoDock 4.2 (http://autodock.scripps.edu/). Open the PyRx Tool, edit the preference setting, inactivate all torsions for **Ligand Preparation**. Load the substrate molecule, right click the molecule name on the Navigator panel, select **Make ligand** to prepare the ligand docking file. Load the protein 2HWK_clean.pdb, and select **Make macromolecule** to prepare the pdbqt docking file (**Figure 7**).

6.3.1. Start the **AutoDock Wizard** on the docking panel at the bottom. Select the prepared ligand and protein files. Define the protein binding pocket by manually adjusting the grid dimension which is centered at the catalytic residue Cys-477. Using the default spacing parameter 0.375 Å. Click on **Run AutoGrid** to generate grid maps.

6.3.2. Run **AutoDock** and select the **Lamarckian Genetic Algorithm (LGA) method**. Click on the **Docking Parameters** and set the **Number of GA runs** to 50. Use the default parameters for others. Click on **Forward** to start the docking run.

6.3.3. Open the **Analyze Results** panel. Inspect all predicted binding poses. Select the best model with the lowest predicted binding energy and reasonable binding interactions between the Cys-477 and substrate on the cleavage site. Save the binding model as PDB file for further MD simulations.

7. MD simulations of docked VEEV-substrate complexes

7.1. Prepare the input files using Amber (http://ambermd.org/). Following the standard protocol, MD simulations are performed for the predicted substrate binding models using the AMBER package and the ff99SB force field.

NOTE: The solvated systems are subjected to a thorough energy minimization prior to MD simulations. Periodic boundary conditions are applied to simulate a continuous system. The particle mesh Ewald (PME) method was employed to calculate the long-range electrostatic interactions. The simulated system was first subjected to a gradual temperature increase from 0 K to 300 K over 100 ps, and then equilibrated for 500 ps at 300 K, followed by production runs of 2-ns length in total.

7.1.1. Run the simulation job at a high performance computing facility. Our simulations were run on the Biowulf cluster (https://hpc.nih.gov/) (**Figure 8**).

7.1.2. Visualize the trajectory output using the VMD program (www.ks.uiuc.edu/Research/vmd/). Analyze the binding interactions and conformational changes of the substrates and TRIM14 within the active site of nsP2 (**Figure 9**).

REPRESENTATIVE RESULTS:

SSHHPS analysis of the ZIKV ns2B/3 protease identified 4 host protein targets: FOXG1, SFRP1, a G_s alpha subunit from a retinal cDNA library, and the NT5M mitochondrial 5',3'-nucleotidase (**Figure 10**)[6]. Notably, no other method predicted these proteins as potential targets of the ZIKV protease. Mutations in the FOXG1 gene have been linked to a congenital syndrome characterized by impaired development and structural brain abnormalities such as microcephaly. SFRP1 is a secreted frizzled-related protein (SFRP); these soluble receptors can competitively bind Wnt ligands to antagonize and inhibit Wnt signaling. The Wnt signaling pathway is involved in the regulation of the IFN response during Flavivirus infection[36]. The cleavage of SFRP1 would be expected to enhance flavivirus replication. SFRP1 is also involved in Th17-cell differentiation[37]. Sequence alignments of the SSHHPS showed species-specific differences in the cleavage site sequences (**Figure 10D**). The cleavage site sequence in SFRP1 was identical in humans and chickens; ZIKV can induce mortality and microcephaly in chicken embryos[38]. In rodents, the highly conserved P1 residue (K/R)R↓G is substituted by a glycine (RGG). Immunocompetent strains of mice are generally resistant to ZIKV infection and disease[39].

Steady state kinetic parameters and inhibition constants can be measured for the viral polyprotein sequences and for the host protein sequences using the continuous assay in a plate reader[31,40,41] (**Figure 11A**). For qualitative cleavage information, such as cleavage of a particular sequence or the inhibition of the protease by various compounds, the discontinuous assay can be used (**Figure 11B**).

Optimization of the number of residues in between CFP and YFP may be required. A substrate-bound model can be made using the in silico methods. A representative docked model of the nsP1/nsP2 junction is shown in **Figure 9**. For the VEEV nsP2 protease, cleavage of a 12-amino acid Semliki Forest Virus (SFV) sequence had been reported (K_m = 0.58 mM[33]). Lengthening the substrate sequence to 19, 22, and 25 residues and reducing the ionic strength of the buffer led to a significant reduction in K_m. Examination of the VEEV nsP2 crystal structure and crystal packing also showed that a portion of one of the junctions was packed against the protease domain and was helical. Thus, the longer VEEV substrates may bind better due to the recognition of a secondary structural motif.

For TRIM14, we obtained a K_m = 21 µM[6,33]. The K_m for the substrate carrying the host protein sequence was comparable to the K_m values of the substrates containing the viral polyprotein cleavage site sequences (K_m(V12) = 12 µM and K_m(V34) = 21 µM). The cleavage site sequences at the nsP1/nsP2, nsP2/nsP3, and nsP3/nsP4 junctions were cut with different efficiencies. In the cell, this is thought to allow for sequential cleavage of the polyprotein[42].

Caution should be taken in interpreting negative results. If no cleavage occurs, the cleavage site may be too short or the purified protease may be inactive. For substrates that are cut, additional experiments are needed to confirm cleavage of the full length protein or cleavage in virus-infected cells. Appropriate follow-on experiments should be chosen. The effects of overexpression or silencing of the target protein on viral replication also can be tested.

FIGURE AND TABLE LEGENDS:

Figure 1. Three mechanisms of silencing. Silencing can occur at the level of DNA, RNA, or protein. These "search and delete" algorithms each use a "keyword" to direct the cleavage of a file containing the word. This figure has been modified from Morazzani et al.[32] and the references therein.

Figure 2. Species-specific differences in cleavage site sequences. The C-terminal PRY/SPRY domains of TRIM14 homologues are shown in the alignment. The PRY/SPRY domain can be identified by the conserved motifs highlighted in gray. Human TRIM14 is cut at QEAGA↓G by the VEEV nsP2 cysteine protease. The SSHHP sequence is shown in color. The residue in green is the P1' residue; in blue is the P4 residue, and in red are other conserved residues within the cleavage site motif sequence. Equine harbor a truncated version of TRIM14 lacking the PRY/SPRY domain. The lysine highlighted in cyan is poly-ubiquitinated and is important for the assembly of the MAVS signalosome. The C-terminal PRY/SPRY domain may be transiently cut by the nsP2 protease to impair the host's antiviral response intracellularly during an acute viral infection. In equine, this domain is always absent. This suggests that the PRY/SPRY domain of TRIM14 may have a protective function against VEEV infections. This figure has been reproduced from Morazanni et al.[6]

Figure 3. SSHHPS identification using BLAST. The cleavage site motif sequence at the VEEV nsP1/nsP2 junction is aligned with the SSHHP sequence in the host protein TRIM14. The residue colored in green is the P1' residue; in blue is the P4 residue and in red are other conserved residues of the cleavage site motif sequence. Most alignments contained homology to regions outside of the conserved cleavage site motif or did not include the P1/P1' scissile bond residues. TRIM14 showed a match to 6 residues in sequential order that included P1 and P1'.

Figure 4. Protein and DNA sequences of the CFP-V12-YFP substrate for the VEEV nsP2 cysteine protease. The NdeI (CATATG) and XhoI (CTCGAG) restriction sites are shown in capital letters. In red is the cleavage site sequence from the viral polyprotein that is in between nsP1 and nsP2. The residue in green is the P1' residue and in blue is the P4 residue of the cleavage site.

Figure 5. Protein sequence of the Trx-VEEV-nsP2 cysteine protease construct. Thioredoxin (Trx) is shown in yellow. The thrombin cleavage site and His-tag are shown in cyan. The Cys-His dyad are labeled in red.

Figure 6. Peptide structures in MOE.

Figure 7. Docking of substrate peptide using PyRx/AutoDock.

Figure 8. Jobs running on the Biowulf cluster.

Figure 9. Model of the VEEV P12 substrate containing the cleavage site sequence at the nsP1/nsP2 junction. The Cys-477/His-546 catalytic dyad is shown in blue. Figure was made using Pymol (https://pymol.org)**.**

Figure 10. SSHHPS Analysis of the Zika virus ns2B/ns3 protease. (**A**) Predicted host protein targets of the ZIKV ns2B/ns3 protease. Residues in red match a single cleavage site sequence. Residues in green are tolerated at the subsite and match residues at other cleavage sites. SFRP1 had the highest number of identical residues in consecutive order. (**B**) CFP-substrate-YFP proteins (~50-60 kDa) were expressed and purified containing the predicted SSHHP sequence from each host protein (human). The ZIKV protease cut human FOXG1, SFRP1, NT5M and a $G_s$alpha subunit isolated from a retinal cDNA library. The cleavage products are approximately 28-30 kDa. The substrate sequences are available in Morazzani et al.[6] (**C**) While the ns2B/ns3 protease cut SFRP1, it did not cut its homologues (SFRP2 and SFRP4). (**D**) Alignment of the cleavage sites from different animal species may be useful in selecting an animal model for a Group IV virus. Note that the conserved R↓G sequence differs between humans and rodents in SFRP1. Figure reproduced from Morazzani et al.[6]

Figure 11. Steady state kinetic analysis using the continuous and discontinuous assays. (**A**) The kinetic data shown in Table 5 was plotted in GraFit. The inset shows the Lineweaver-Burk plot. (**B**) SDS-PAGE gel showing the cleavage products of the CFP-V12-YFP substrate. In lane 1 is the "UNCUT" substrate (48 kDa). In lane 2 is the "CUT" substrate (31 kDa and 27 kDa). In lanes 3-9 different compounds were included to test their inhibitory activity. Lane 4 contains the E64d covalent inhibitor. These reactions were run overnight for ~17 h at room temperature. Boiling of the samples was required to achieve the sharp banding pattern. The nsP2 protease is visible (56 kDa) in the reactions containing enzyme, but not in lane 1. Lane 1 is the "no enzyme" control.

DISCUSSION:

Sequence-specific destruction of a protein or a nucleic acid guided by a foreign sequence is only seen in a few cases in biology. The mechanisms shown in **Figure 1** are defensive mechanisms that protect a host from a virus, or a virus from a host.

Using bioinformatic methods we can identify the targets that are destroyed by these systems. In our analyses of SSHHP sequences, we discovered that many of them could be found in proteins needed to generate innate immune responses. Some had obvious roles such as MAVS and TRIF (TIR-domain-containing adapter-inducing interferon-β), while others were related to immunity though more complex mechanisms (e.g., Histone H3, SFRP1, FOXG1)[8,9]. The target information stored in the SSHHP sequence has the potential to identify pathways that have antiviral effects against these viruses. Antiviral responses in vivo are often virus-specific[26,43]. For example, subsets of TRIM proteins have antiviral effects on different viruses[43-45], some are viral restriction factors

(e.g., HIV and TRIM5α). The specificity of TRIM proteins (~70 have been identified) still is being examined[44,45]. The information within SSHHPS may contribute to our understanding of how these viruses evade the innate immune responses. Other patterns and correlations may be uncovered as more SSHHPS are examined.

Species-specific differences were apparent in our analyses (**Figure 2, Figure 10**). These viruses are known to affect some species more than others. Information about host range, host susceptibility, and host defenses may be present within SSHHPS. For example, equine, the most susceptible species to equine encephalitis viruses, lacked the region of human TRIM14 that was transiently cut by the VEEV nsP2 protease. Humans rarely die from VEEV infections but can be infected[24]. The human TRIM14 protein carried an nsP2 protease cleavage sequence[6]. The presence of the cleavage site suggest that humans have a defense mechanism against these viruses. Birds have been thought to be potential reservoirs of these viruses[46]. The corresponding SSHHP sequence in the TRIM14 protein from chickens differed from the sequences found in humans and other species. Subtle differences like these may make a target host protein uncleavable or more readily cleaved. Aguirre et al.[16] showed that an uncleavable mutated STRING protein induced higher levels of IFN after Dengue virus infection and that mice naturally carry a version of STING that is not cut by the Dengue ns2B3 protease. The murine STING protein was not cut by the ZIKV protease[47]. In our SSHHPS analysis, we also observed differences in the ZIKV protease cleavage site sequences when we compared the human proteins with those of rodents [6] (**Figure 10D**). Reproducing the species-specific proteolytic cleavages of host proteins may be important in animal models used for Group IV viruses. The inhibition of host protein cleavage also has implications with regards to the development of Group IV protease inhibitors. In our previous publication, we showed that we could inhibit TRIM14 cleavage by the VEEV nsP2 protease using CA074 methyl ester[6]. This result suggests that small molecule inhibitors of these proteases may be able to modulate the innate immune responses that are capable of suppressing the infection[6,31].

Genetic variation within a species also has the potential to produce differences in proteolytic cleavage. Subtle differences in codon usage could affect ribosome pausing[48]. Since some Group IV viral proteases are embedded in the ER membrane, differences in these pauses could affect cleavage of a target if cleavage occurs co-translationally. Some of the cleavage sites that we identified were in predicted signal peptide sequences (e.g., SFRP1) while others were internal.

SSHHPS analysis can produce information that differs from other methods of host protein analyses. SSHHPS analysis was inexpensive and easy to employ. The use of a bacterial expression system allowed testing of short segments (~25 amino acids) of mammalian sequences without the use of mammalian cell culture. We found that the CFP-YFP substrates were able to tolerate all of the tested human protein sequences; however, yields varied. In similar assays, substrates containing human protein sequences as long as 63 amino acids were successfully expressed, purified, and utilized for kinetic analyses and inhibitor screening[49-51]. Since only small amounts of the substrate are needed for the discontinuous assay, a large number of targets can be explored. One advantage of the system is that the CFP/YFP substrates can be used for SDS-PAGE analyses and for more elaborate kinetic analyses (i.e., IC_{50}, K_i, K_m, V_{max}). For drug discovery, inhibitory

compounds can produce artifacts in fluorescent assays. Thus, the discontinuous assay in combination with a continuous assay allows one to confirm cleavage or inhibition of cleavage. The samples for the discontinuous SDS-PAGE assay can be taken directly out of the 96-well plates. CFP/YFP substrates have been used for compound library screening[52]. However, additional analyses are required to determine if a substrate is suitable for high throughput screening such as the calculation of a Z-factor[53].

One challenge in designing a substrate is identifying the region around the scissile bond that is bound and recognized by the protease. In the examples shown here, we began with 12 residue sequences that were centered around the scissile bond. After analyzing sequence alignments of the cleavage sites homology to the residues N-terminal of the scissile bond was found for the VEEV protease, whereas for the ZIKV protease homology to several of the C-terminal residues was found. An in silico model of the docked substrate can be used to design site-directed mutagenesis experiments that probe the binding sites of the substrate. Since the substrate and enzyme sequences are on plasmids, either can be mutated to test the in silico models or subsite tolerances. This can be advantageous if a crystal structure of the bound substrate(s) is not available.

SSHHPS analysis may also yield new information about the mechanisms by which virus-induced phenotypes are produced by viral enzymes. One of the ZIKV targets, SFRP1, is part of the Wnt signaling pathway and has roles in both brain and eye development and in immune responses[36,37,54-57]. We found that the other protein sequences that could be cut by the ZIKV ns2B/ns3 protease were also in proteins involved in brain and eye development; abnormalities in both have been observed in congenital Zika syndrome and are thought to be part of the virus-induced phenotype[58].

The predictability of host-pathogen interactions could be exploited for a variety of applications: target-specific oncolytic viral therapies; de-risking live virus vaccines; refinement, prediction or selection of animal models; prediction of host-range or susceptibility; prediction of zoonotic events; and prediction of host-defenses. Since the methods described are sequence-based, they may be of value to incorporate into software in the future.

ACKNOWLEDGMENTS:

This work was supported by Defense Threat Reduction Agency (DTRA) project numbers CB-SEED-SEED09-2-0061 and CBCall4-CBM-05-2-0019.

DISCLOSURES:

The opinions expressed here are those of the authors and do not represent those of the U. S. Navy, U.S. Army, U. S. Department of Defense, or the U. S. government.

REFERENCES:

1 Liu, H. et al. Widespread Horizontal Gene Transfer from Double-Stranded RNA Viruses to Eukaryotic Nuclear Genomes. *Journal of Virology.* **84** (22), 11876-11887 (2010).

2 Hagai, T., Azia, A., Babu, M. M., Andino, R. Use of host-like peptide motifs in viral proteins

836 is a prevalent strategy in host-virus interactions. *Cell Reports.* **7** (5), 1729-1739 (2014).
837 3 Gorbalenya, A. E. Host-related sequences in RNA viral genomes. *Seminars in Virology.* **3**
838 359-371 (1992).
839 4 Shmakov, S. A. et al. The CRISPR Spacer Space Is Dominated by Sequences from Species-
840 Specific Mobilomes. *MBio.* **8** (5), 1-18 (2017).
841 5 Legler, P. M., Morazzani, E., Glass, P.J., Compton, J.R. Proteome Editing System and A
842 Biomarker of Veev Infection. United States patent application (2018).
843 6 Morazzani, E. M. et al. Proteolytic cleavage of host proteins by the Group IV viral
844 proteases of Venezuelan equine encephalitis virus and Zika virus. *Antiviral Research.* **164**
845 106-122 (2019).
846 7 Alvarez, E., Castello, A., Menendez-Arias, L., Carrasco, L. HIV protease cleaves poly(A)-
847 binding protein. *Biochemical Journal.* **396** (2), 219-226 (2006).
848 8 Falk, M. M. et al. Foot-and-mouth disease virus protease 3C induces specific proteolytic
849 cleavage of host cell histone H3. *Journal of Virology.* **64** (2), 748-756 (1990).
850 9 Grigera, P. R., Tisminetzky, S. G. Histone H3 modification in BHK cells infected with foot-
851 and-mouth disease virus. *Virology.* **136** (1), 10-19 (1984).
852 10 Li, W., Ross-Smith, N., Proud, C. G., Belsham, G. J. Cleavage of translation initiation factor
853 4AI (eIF4AI) but not eIF4AII by foot-and-mouth disease virus 3C protease: identification
854 of the eIF4AI cleavage site. *FEBS Letters.* **507** (1), 1-5 (2001).
855 11 Kuyumcu-Martinez, M. et al. Calicivirus 3C-like proteinase inhibits cellular translation by
856 cleavage of poly(A)-binding protein. *Journal of Virology.* **78** (15), 8172-8182 (2004).
857 12 Pietila, M. K., Hellstrom, K., Ahola, T. Alphavirus polymerase and RNA replication. *Virus*
858 *Research.* **234** 44-57 (2017).
859 13 Hardy, W. R., Strauss, J. H. Processing the nonstructural polyproteins of sindbis virus:
860 nonstructural proteinase is in the C-terminal half of nsP2 and functions both in cis and in
861 trans. *Journal of Virology.* **63** (11), 4653-4664 (1989).
862 14 Strauss, E. G., De Groot, R. J., Levinson, R., Strauss, J. H. Identification of the active site
863 residues in the nsP2 proteinase of Sindbis virus. *Virology.* **191** (2), 932-940 (1992).
864 15 Wang, D. et al. Foot-and-mouth disease virus 3C protease cleaves NEMO to impair innate
865 immune signaling. *Journal of Virology.* **86** (17), 9311-9322 (2012).
866 16 Aguirre, S. et al. DENV inhibits type I IFN production in infected cells by cleaving human
867 STING. *PLoS Pathogens.* **8** (10), e1002934 (2012).
868 17 Barral, P. M., Sarkar, D., Fisher, P. B., Racaniello, V. R. RIG-I is cleaved during picornavirus
869 infection. *Virology.* **391** (2), 171-176 (2009).
870 18 Elbashir, S. M., Lendeckel, W., Tuschl, T. RNA interference is mediated by 21- and 22-
871 nucleotide RNAs. *Genes & Development.* **15** (2), 188-200 (2001).
872 19 Deveau, H., Garneau, J. E., Moineau, S. CRISPR/Cas system and its role in phage-bacteria
873 interactions. *Annual Review of Microbiology.* **64** 475-493 (2010).
874 20 Schechter, I., Berger, A. On the size of the active site in proteases. I. Papain. *Biochemical*
875 *and Biophysical Research Communications.* **27** (2), 157-162 (1967).
876 21 Bieniasz, P. D. Intrinsic immunity: a front-line defense against viral attack. *Nature*
877 *Immunology.* **5** (11), 1109-1115 (2004).
878 22 Zhou, Z. et al. TRIM14 is a mitochondrial adaptor that facilitates retinoic acid-inducible
879 gene-I-like receptor-mediated innate immune response. *Proceedings of the National*

Academy of Sciences of the U S A. **111** (2), E245-254 (2014).
23 Wang, S. et al. TRIM14 inhibits hepatitis C virus infection by SPRY domain-dependent targeted degradation of the viral NS5A protein. *Scientific Reports.* **6** 32336 (2016).
24 Zacks, M. A., Paessler, S. Encephalitic alphaviruses. *Veterinary Microbiology.* **140** (3-4), 281-286 (2010).
25 Hollidge, B. S., Weiss, S. R., Soldan, S. S. The role of interferon antagonist, non-structural proteins in the pathogenesis and emergence of arboviruses. *Viruses.* **3** (6), 629-658 (2011).
26 Carthagena, L. et al. Human TRIM gene expression in response to interferons. *PLoS One.* **4** (3), e4894 (2009).
27 Montgomery, S. A., Johnston, R. E. Nuclear import and export of Venezuelan equine encephalitis virus nonstructural protein 2. *Journal of Virology.* **81** (19), 10268-10279 (2007).
28 Nenasheva, V. V. et al. Enhanced expression of trim14 gene suppressed Sindbis virus reproduction and modulated the transcription of a large number of genes of innate immunity. *Immunologic Research.* **62** (3), 255-262 (2015).
29 Nagai, T. et al. A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology.* **20** (1), 87-90 (2002).
30 Li, M. Z., Elledge, S. J. SLIC: a method for sequence- and ligation-independent cloning. *Methods in Molecular Biology.* **852** 51-59 (2012).
31 Hu, X. et al. Kinetic, Mutational, and Structural Studies of the Venezuelan Equine Encephalitis Virus Nonstructural Protein 2 Cysteine Protease. *Biochemistry.* **55** (21), 3007-3019 (2016).
32 Morazzani, E. M. et al. Proteolytic cleavage of host proteins by the Group IV viral proteases of Venezuelan equine encephalitis virus and Zika virus. *Antiviral Research.* **164** 106-122 (2019).
33 Zhang, D., Tozser, J., Waugh, D. S. Molecular cloning, overproduction, purification and biochemical characterization of the p39 nsp2 protease domains encoded by three alphaviruses. *Protein Expression and Purification.* **64** (1), 89-97 (2009).
34 Lei, J. et al. Crystal structure of Zika virus NS2B-NS3 protease in complex with a boronate inhibitor. *Science.* **353** (6298), 503-505 (2016).
35 Shiryaev, S. A. et al. Characterization of the Zika virus two-component NS2B-NS3 protease and structure-assisted identification of allosteric small-molecule antagonists. *Antiviral Research.* **143** 218-229 (2017).
36 Smith, J. L., Jeng, S., McWeeney, S. K., Hirsch, A. J. A MicroRNA Screen Identifies the Wnt Signaling Pathway as a Regulator of the Interferon Response during Flavivirus Infection. *Journal of Virology.* **91** (8) (2017).
37 Lee, Y. S. et al. The Wnt inhibitor secreted Frizzled-Related Protein 1 (sFRP1) promotes human Th17 differentiation. *European Journal of Immunology.* **42** (10), 2564-2573 (2012).
38 Goodfellow, F. T. et al. Zika Virus Induced Mortality and Microcephaly in Chicken Embryos. *Stem Cells and Development.* **25** (22), 1691-1697 (2016).
39 Morrison, T. E., Diamond, M. S. Animal Models of Zika Virus Infection, Pathogenesis, and Immunity. *Journal of Virology.* **91** (8) (2017).
40 Morazzani, E. M. et al. in *Books of Abstracts, 254th American Chemical Society National*

Meeting, Washington, D.C. BIOL-20 (2017).
41 Compton, J. R., Mickey, M. J., Hu, X., Marugan, J. J., Legler, P. M. Mutation of Asn-475 in the Venezuelan Equine Encephalitis Virus nsP2 Cysteine Protease Leads to a Self-Inhibited State. *Biochemistry.* **56** (47), 6221-6230 (2017).
42 Vasiljeva, L. et al. Regulation of the sequential processing of Semliki Forest virus replicase polyprotein. *Journal of Biological Chemistry.* **278** (43), 41636-41645 (2003).
43 Uchil, P. D., Quinlan, B. D., Chan, W. T., Luna, J. M., Mothes, W. TRIM E3 ligases interfere with early and late stages of the retroviral life cycle. *PLoS Pathogens.* **4** (2), e16 (2008).
44 Ozato, K., Shin, D. M., Chang, T. H., Morse, H. C., 3rd. TRIM family proteins and their emerging roles in innate immunity. *Nature Reviews Immunology.* **8** (11), 849-860 (2008).
45 van Tol, S., Hage, A., Giraldo, M. I., Bharaj, P., Rajsbaum, R. The TRIMendous Role of TRIMs in Virus-Host Interactions. *Vaccines (Basel).* **5** (3) (2017).
46 Molaei, G. et al. Dynamics of Vector-Host Interactions in Avian Communities in Four Eastern Equine Encephalitis Virus Foci in the Northeastern U.S. *PLoS Neglected Tropical Diseases.* **10** (1), e0004347 (2016).
47 Ding, Q. et al. Species-specific disruption of STING-dependent antiviral cellular defenses by the Zika virus NS2B3 protease. *Proceedings of the National Academy of Sciences of the U S A.* **115** (27), E6310-E6318 (2018).
48 Angov, E., Legler, P. M., Mease, R. M. Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding. *Methods in Molecular Biology.* **705** 1-13 (2011).
49 Ruge, D. R. et al. Detection of six serotypes of botulinum neurotoxin using fluorogenic reporters. *Analytical Biochemistry.* **411** (2), 200-209 (2011).
50 Hu, X. et al. Structural insight into exosite binding and discovery of novel exosite inhibitors of botulinum neurotoxin serotype A through in silico screening. *Journal of Computer-Aided Molecular Design.* **28** (7), 765-778 (2014).
51 Dunning, F. M. et al. Detection of botulinum neurotoxin serotype A, B, and F proteolytic activity in complex matrices with picomolar to femtomolar sensitivity. *Applied and Environmental Microbiology.* **78** (21), 7687-7697 (2012).
52 Nguyen, T. G. et al. Development of fluorescent substrates and assays for the key autophagy-related cysteine protease enzyme, ATG4B. *Assay and Drug Development Technologies.* **12** (3), 176-189 (2014).
53 Zhang, J. H., Chung, T. D., Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *Journal of Biomolecular Screening.* **4** (2), 67-73 (1999).
54 Bovolenta, P., Esteve, P., Ruiz, J. M., Cisneros, E., Lopez-Rios, J. Beyond Wnt inhibition: new functions of secreted Frizzled-related proteins in development and disease. *Journal of Cell Science.* **121** (Pt 6), 737-746 (2008).
55 Esteve, P. et al. SFRPs act as negative modulators of ADAM10 to regulate retinal neurogenesis. *Nature Neuroscience.* **14** (5), 562-569 (2011).
56 Garcia-Hoyos, M. et al. Evaluation of SFRP1 as a candidate for human retinal dystrophies. *Molecular Vision.* **10** 426-431 (2004).
57 Marcos, S. et al. Secreted frizzled related proteins modulate pathfinding and fasciculation of mouse retina ganglion cell axons by direct and indirect mechanisms. *Journal of*

Neuroscience. **35** (11), 4729-4740 (2015).

58 Moore, C. A. et al. Characterizing the Pattern of Anomalies in Congenital Zika Syndrome for Pediatric Clinicians. *JAMA Pediatrics.* **171** (3), 288-295 (2017).

"Keyword"

Enzyme searches for a match between the keyword and words in file and cuts ("deletes").

File to delete

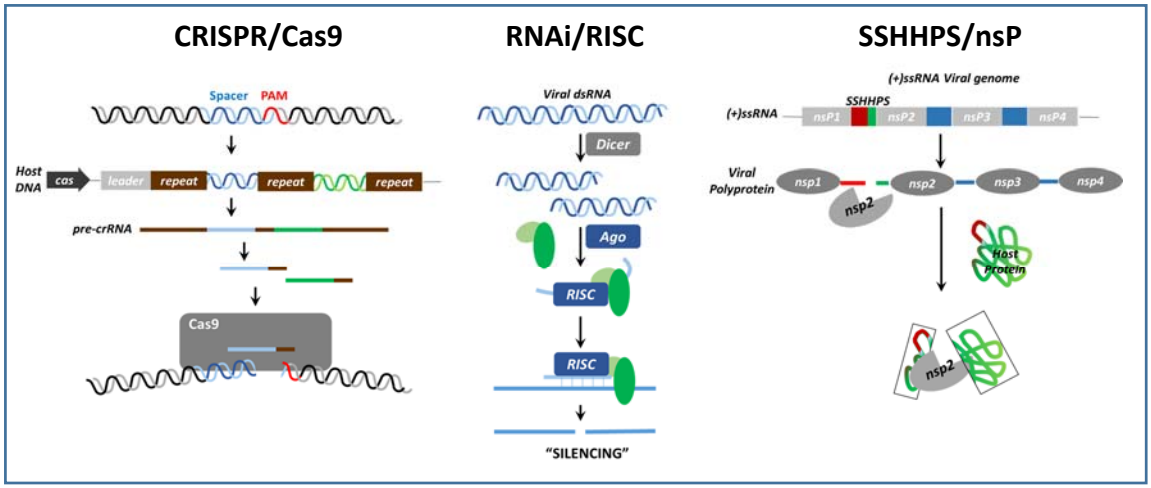


Figure2 Click here to access/download;Figure;Figure2.pdf

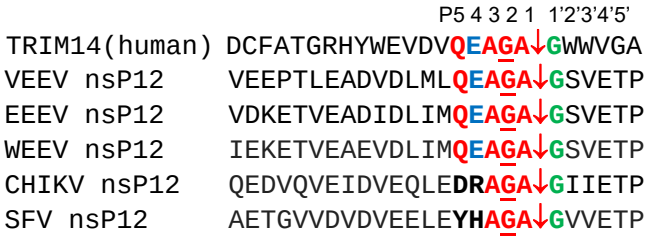
```
Human           SPSPERSLLLKYARTPTLDPDTMHARLRLSADRLTVRCGLLGSLGPVPVLRFDALWQVLA
Rat             SPSPERALFLKYARTPTLDPDTMHARLRLSTDGLTVRCSLLGRLGPRPAPRFDSLRQVLG
Mouse           SPSPERALFLKYARTPTLDPDTMHARLRLSPDGLTVRCSLLGRLGPRPAPRFDALRQVLG
Rabbit          SPSPERSLLLKYARTPTLDPDTMHARLRLSADRLTVRCALLGRLGQPPATRFDELWQVLG
dog             SPSPERSLFLKYARTPTLEPDTMHARLRLSADRLTVRCALVGRLGPAPALRFDKLWQVLG
Rhesus          SPSPERSLFLKYARTPTLDPDTMHARLRLSADRLTVRCGLLGSLGPVPALRFDALWQVLA
Pig             SPSPQRSLFLKYARTPTLEPDTMHARLRLSADRLTVRCGLFGRLGPTPTLCFDSLWQVLG
Cow             SPSPERSLFLKFFFR-EQAPDSLRAHCRSPGGGAGR---VLGRLGPTAALRFDALWQVLS
Sheep           SPSPERSLFLKYARTPCLEPDTMHARLRLSADLLTVRCGLLGRLGPTAALRFDALWQVLS
goat            SPSPERSLFLKYARTPCLEPDTMHARLRLSADLLTVRCGLLGRLGPTAALRFDALWQVLS
Chicken         MSPVDRMLFLKHARSPTWEYDSLHPRLKLSDDRLVVSCNWRRIFYPCGPQRFDKLWQVLS
Horse           SPSPERSLFLKCKTP---------------------------------------------
Donkey          SPSPERSLFLKCKTP---------------------------------------------
                 .. :* *:**
                          SSHHPS

Human           RDCFATGRHYWEVDVQEAGAGWWVGAAYASLRRRGASAAARLGCNRQSWCLKRYDLEYWA
Rat             RDGFAAGRHYWEVDVQEAGVGWWVGAAYPSLRRRGTSTAARLGCNRESWCVKRYDMEYWA
Mouse           RDGFAAGRHYWEVDVQEAGVGWWVGAAYPSLRRRGASAAARLGCNRESWCVKRYDLEYWA
Rabbit          RDCFAAGRHYWEVDVQEAGAGWWVGAAYGSLWRRGASAAARLGCNRQSWCLKRYDLEYWA
dog             RDCYAAGRHYWEVDVQEAGVGWWVGAAYASLRRCGASAAARLGCNRQSWCLKRYDLEYWA
Rhesus          RDCFAAGRHYWEVDVQEAGAGWWVGAAYASLRRRGASAAARLGCNRQSWCLKRYDLEYWA
Pig             RDCFAAGRHYWEVDVQEAGIGWWVGAAYGSLRRRGASAAARLGCNRQSWCLKRYDLEYWA
Cow             RDCFAAGRHYWEVDVQEAGVGWWVGAAYGSLRRHGASDAARLGCNRQSWCLKRYDLEYWA
Sheep           RDCFAAGRHYWEVDVQEAGVGWWVGAAYGSLRRHGASDAARLGCNRQSWCLKRYDLEYWA
goat            RDCFAAGRHYWEVDVQEAGVGWWVGAAYGSLRRHGASDAARLGCNRQSWCLKRYDLEYWA
Chicken         RDAFLSGSHYWEVDLLHAGAGWWIGAAYPSIGRKGDSESCRLGWNRASWCIKKFDLEYWA
Horse           ---RAEGGVHSGGRVTEMEAPLFFLREPTWVF----------------------------
Donkey          ---RAEGRVHSGGRVTEMEAPLFFLREPTWLRRRGASAAARLGCNRQSWCLKRYDLEYWA
                      *  :    : .     :.      :

Human           FHDGQRSRLRPRDDLDRLGVFLDYEAGVLAFYDVTGGMSHLHTFRATFQEPLYPALRLWE
Rat             FHDCQRSRLRLRRDPHRLGVFLDYEAGILTFYDVAGGMSHLHTFYAVFQEPLYPALRLWE
Mouse           FHDGQRSRLRPRRDPHRLGVFLDYEAGILAFYDVAGGMSHLHTFHAAFQEPLYPALRLWE
Rabbit          FHDGQRSRLRPRDDPDRVGVFLDYEAGVLAFYDVTGGMSHLHTFRSTFQEPLYPALRLWE
dog             FHDGQRSRLRPRDDPDRLGVFLDYEAGVLAFYDVSGGMSHLHTFRASFQEPLYPALRLWE
Rhesus          FHDGQRSRLRPRDDLDRLGVFLDYEAGVLAFYDVTGSMSHLHTFRATFQEPVYPALRLWE
Pig             FHDGQRSRLRLRDDPDRLGVFLDYEAGVLAFYDVSGGMSHLHTFRAAFQEPLYPALRLWE
Cow             FHDCQRSRLRPRHDPDRLGVFLDYEAGVLAFYDVTGGMSHLHTFRAAFQEPLYPALRLWE
Sheep           FHDCQRSRLRPRHDPDRLGVFLDYEAGVLAFYDVTDGMSHLHTFRAAFQEPLYPALRLWE
goat            FHDCQRSRLRPRHDPDRLGVFLDYEAGVLAFYDVTDGMSHLHTFRAAFQEPLYPALRLWE
Chicken         FHKGERIPLLIEDDPDRIGVFLDYEAGILSFYNVTDGMTHLHTFRCKFTEPVYPALRLWE
Horse           ------------------------------------------------------------
Donkey          FHDGQRSRLRPRGDPERLGVFLDYEAGVLAFYDVTGGMSHLHTFRAAFQEPLYPALRLWE
```

Figure3

Click here to access/download;Figure;Figure3.pdf



Most alignments:

Query	2	EEP-TL-EA-DVDLMLQEAGAG	20
		EEP TL E D DL LQE G G	
Sbjct	275	EEPQTLPETQDGDLHLQEQGSG	296

TRIM14 alignment:

Query	7	EADVDLMLQEAGAG	20
		E DV QEAGAG	
Sbjct	325	EVDV----QEAGAG	334

Figure4

Click here to access/download;Figure;Figure4.pdf

Figure5

Click here to access/download;Figure;Figure5.pdf

Figure6

Click here to access/download;Figure;Figure6.pdf

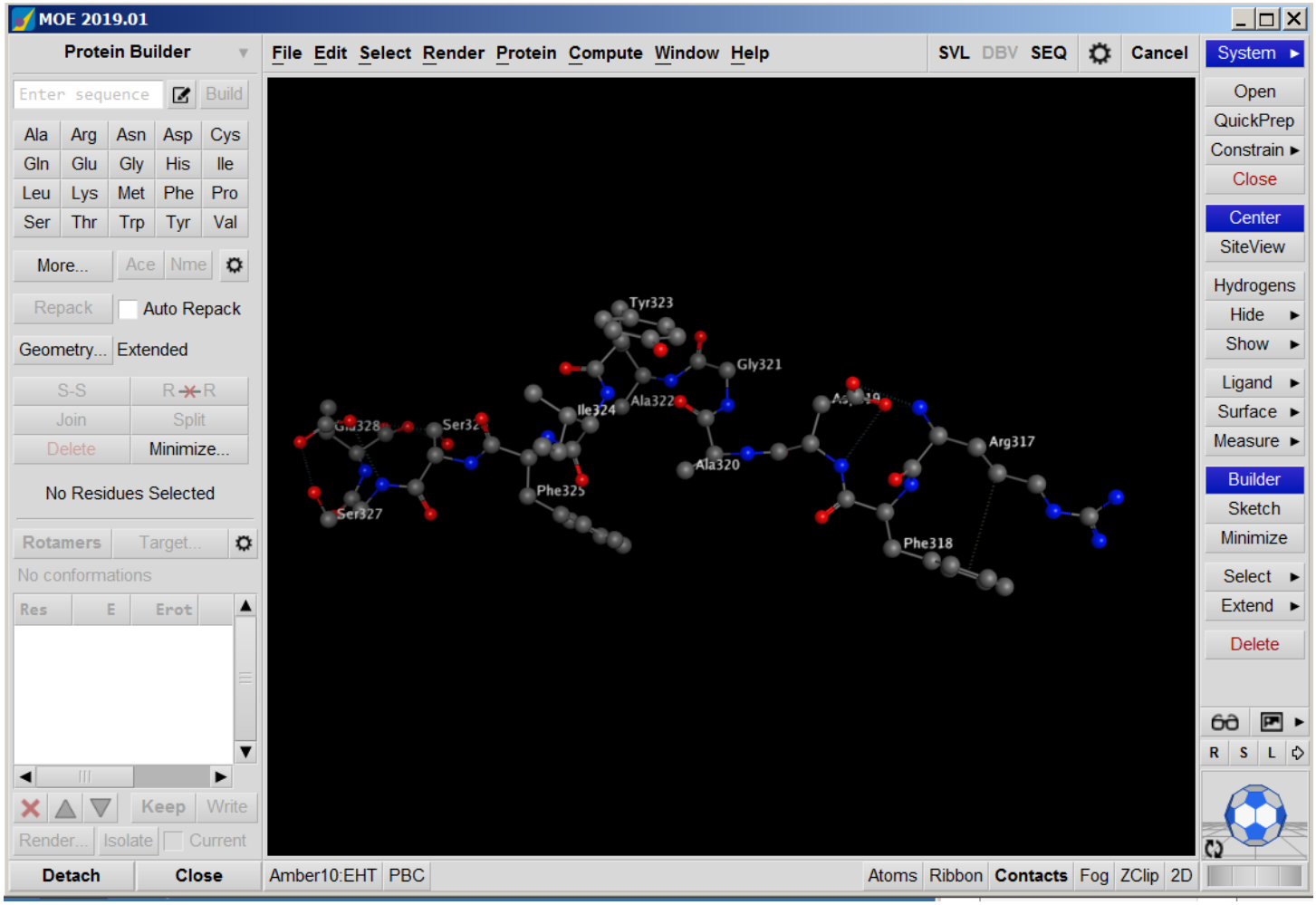


Figure7

Click here to access/download;Figure;Figure7.pdf

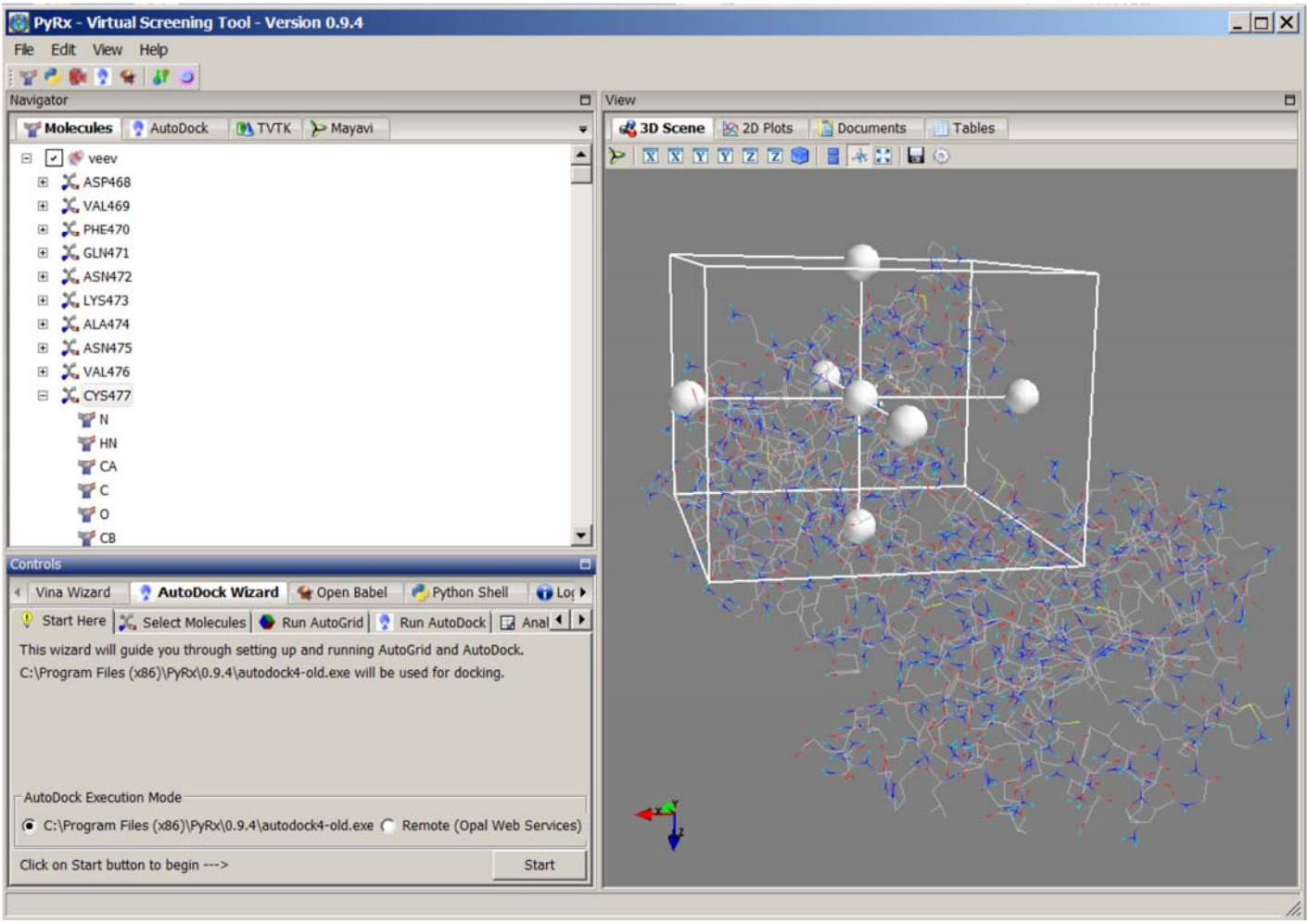


Figure8

Click here to access/download;Figure;Figure8.pdf

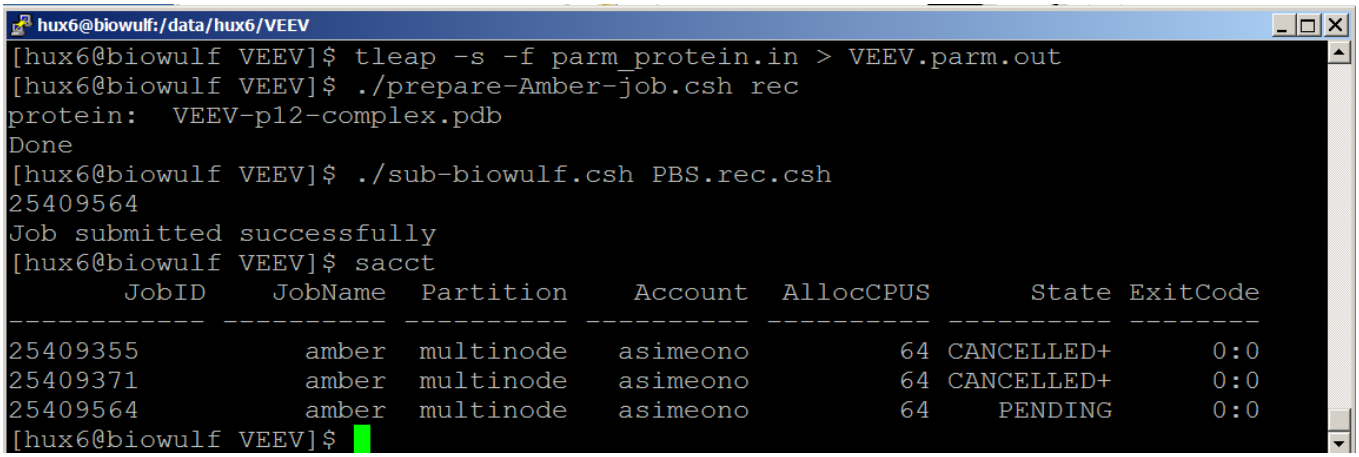
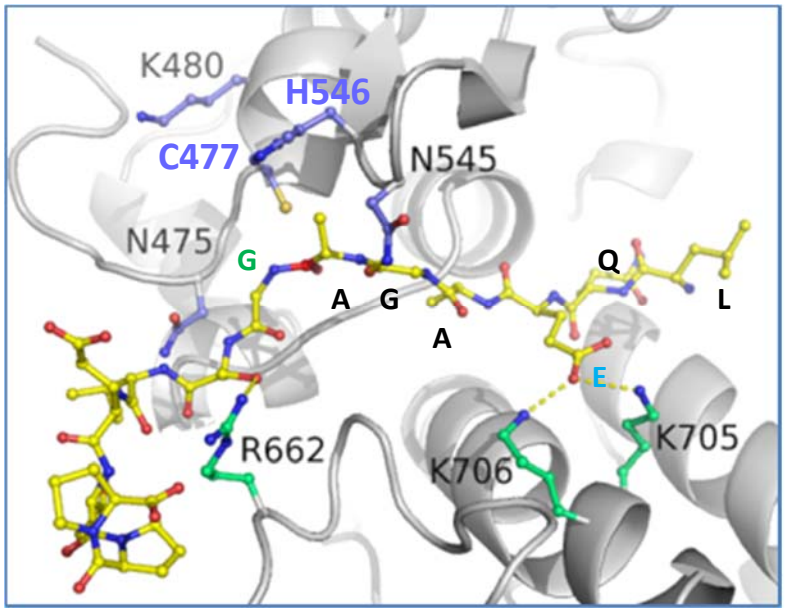


Figure9

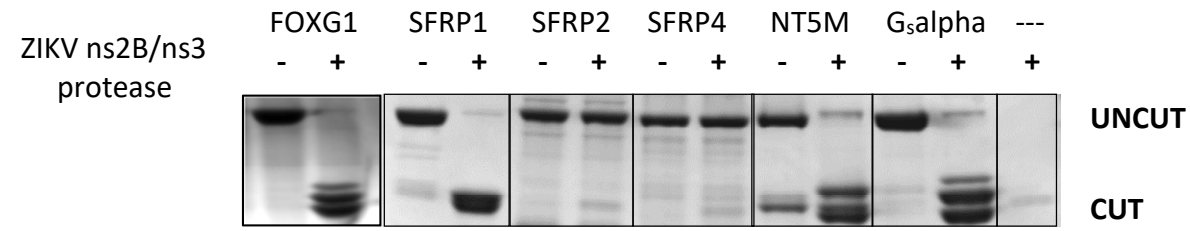
Click here to access/download;Figure;Figure9.pdf



A

Substrate	Cleavage Site Sequence	Accession No.
ZIKV C↓Ci	KER**KRR**↓**GADT**SIGI	
ZIKV ns2A↓ ns2b	TRS**GKR**↓**SWPP**SEVL	
ZIKV ns2B↓ns3	VKT**GKR**↓**SGAL**WDVP	
ZIKV ns3↓ns4A	FAA**GKR**↓**GAAL**GVME	
ZIKV ns4B↓ns5	GLV**KRR**↓**GGGT**GETL	
Secreted Frizzled-related protein 1 (SFRP1)	SEGGRR↓GAALGVLL	NP_003003.3, AAB70793.1
Gs Alpha subunit	QVAGRR|GAALPCSL	CAA39484.1
NT5M, 5'(3')-deoxy-ribonucleotidase, mitochondrial isoform X3	VPAGRR|GAAGGLGL	XP_011522268.1
Transcription factor HFK1 (Forkhead box protein G1)	KLAFKR|GARLTSTG	CAA52239.1
Ankyrin repeat domain-containing protein 65 isoform 1	AAAAGR|GAALRFLL	NP_001138682.1
Mitochondrial dynamics protein MID51 isoform 1	SHSGKR|SWEEPNWM	NP_061881.2
Voltage-dependent T-type calcium channel subunit alpha-1H isoform X7	GPPARR|SWPPSPQR	XP_016879309.1
Plexin-B3 isoform 1 and 2 precursor & plexin-B3	FAAGPR|GTQAALCA	NP_005384.2
PRRC2A	GRGDKR|SWPSPKNR	AQY77357.1
Leucine-rich repeat and calponin homology domain-containing protein 4 isoform 2	TEAGQR|GSALGDLA	NP_001276863.1
Protein phosphatase 1B (PPM1B)	RAAGKS|GSALELSV	AAY89640.1
FCRLB protein	GPREAR|GAALGGVV	AAH38564.2

B



C

```
SFRP1                MGIGRSEGGRRGAALGVLLALGAAL
SFRP2       MLQGPGSLLLLFLASHCCLGSARGLFLFGQPDFS
SFRP4          MFLSILVALCLWLHLALGVRGAPCEAVRIPMCRH
```

D

```
                                  SFRP1                                FOXG1                NT5M
                                    ↓                                    ↓                   ↓
ZIKV ns2B|ns3                    VKTGKRSGALWDVP
ZIKV ns3|ns4A                    FAAGKRGAALGVME
Homo sapiens           MGIGRSEGGRRGAALGVLLALGAAL   RSTTSRAKLAFKRGARLTSTG  CSAAVPAGRRGAAGGLGLAGG
Macaca mulatta         MGSGRSAGGCRGAALGVLLALGAAL   RSTTSRAKLAFKRGARLTSTG  CSAAVPAGRRGAAGGQGLAGG
Gorilla gorilla        MGSGRSVGGRRGAVLGVLLALGAAL   RSTTSRAKLAFKRGARLTSTG  CSAAVPAGRRGAAGGPGLAGG
Pongo abelii           MGSGRSVGGRRGAALGVLLALGAAL   RSTTSRAKLAFKRGARLTSTG  CSAAVPVGRRGAAGGPGLAGD
Pan troglodytes        MGSGRSVGGRRGAALGVLLALGAAL   RSTTSRAKLAFKRGARLTSTG  CSAAVPAGRRGAAGGPGLAGG
Myotis lucifugus       MGSGRGAGGRRGAAGGVLLALAAGL     (seq unavailable)    ----------------------
Tupaia chinensis       MGSGRGAGGRRGAAAGVLLALAAGL   RSTTSRAKLAFKRGARLTSTG  ----------------------
Gallus gallus          MGVGRSEGGRRGAALGVLLALGVAL   RSTTSRAKLAFKRGARLTSTG  HCGPLAGLGRGSCPTAGSRRA
Equus caballus         MGSGRGAGGRRGAAGGVLLALAAGL   RSTTSRAKLAFKRGARLTSTG  RGAAGPAGRRWAAGGPAGRAG
Xenopus laevis         -----MNGENGIWPLLLFWVTPGIL   RSTTSRAKLAFKRGARLTSTG  MAFLPSLLRRGNMLSPRLQNS
Bos taurus             ------MGGGRWAAAGALLALAAGL   RSTTSRAKLAFKRGARLTSTG  RGAAGPAGRRWASGGPAGRA-
Sus scrofa             ------MGGGRWAAAGVLLALAAGL   RSTTSRAKLAFKRGARLTSTG  TRPAGPAGRRWASGGPGGRA-
Camelus dromedarius    ------MGGGRWAAAGVLLALAAGL   RSTTSRAKLAFKRGARLTSTG  -MAMAMAARHTRP--------
Rattus norvegicus      MGVGRNARGRGGAASGVLLALAAAL   RSTTSRAKLAFKRGARLTSTG  RLRGCCARPRGAPLRAER---
Mus musculus           MGVGRSARGRGGAASGVLLALAAAL   RSTTSRAKLAFKRGARLTSTG  RLRGCCARPRGAPLRAERSR-
Salmo salar            ----MRSIERLCGWRIIPLALTMAV   RSTTSRAKLAFKRGARLTSTG  GRILLHDQFKCISAKMSSSSG
C. elegans             ----------MMLFPVLYILFAFSV   RPSS------------LSRAR     (No homologue)
```

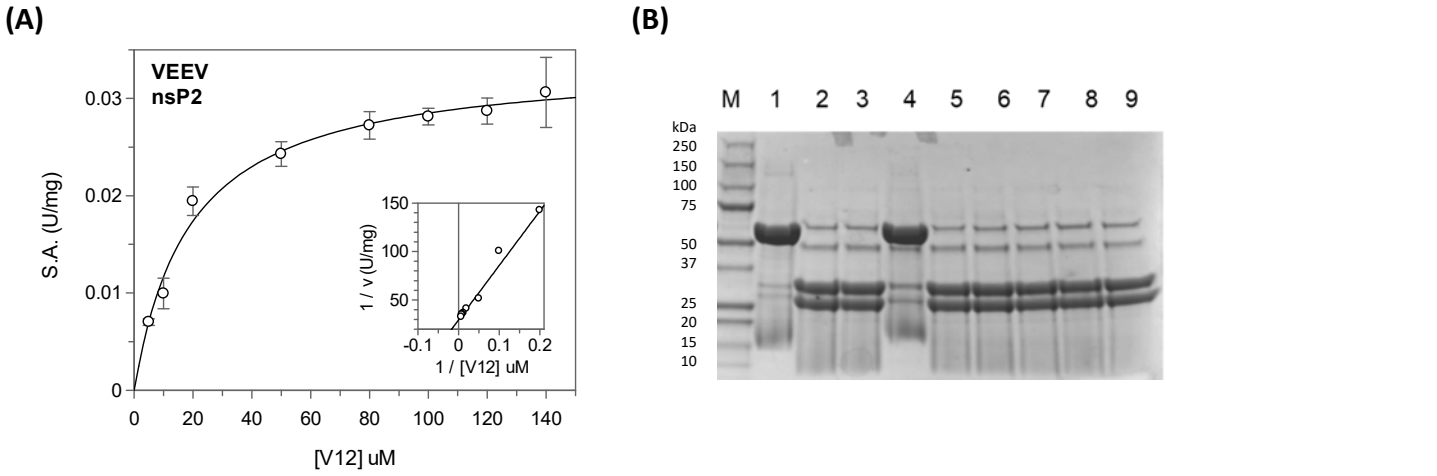


Table 1. List of host proteins cleaved by viral proteases. This table has been adapted from Morazanni, et al.

Virus	Family	Viral Protease	Viral Protease Cleavage Site Motif	Host Protein Substrate	Cleavage site in Host protein	PMID
Poliovirus	*Picornaviridae*	3Cpro	(QE)↓(LIGS)	**RIG-1**	LKKFPQ↓GQKGKV	19628239
				TATA-binding Protein	QGLASPQ↓GAMTPG	8388502
				TATA-binding Protein	AAAVQQ↓STSQQA	16014932
				Poly(A)-binding protein (PABP)	VHVQ↓GQ	14749392
				eIF5B	VMEQ↓G	18572216
		2Apro		TATA-binding Protein	MMPY↓GTGLTP	9261414
Rhinovirus type 1a		3Cpro	(AV)XXQ↓G	**NF-κB**	LLNQ↓GIP	15845545
				eIF5B	VMEQ↓G	18572216
Echovirus type 1				**NF-κB**	LLNQ↓GIP	15845545
Coxsackie B virus				eIF5B	VMEQ↓G	18572216
Foot and Mouth disease Virus (FMDV)		3Cpro	(QE)↓(LIGS)	**NEMO**	LALPSQ↓RRSPPE	22718831
				eIF4A	TNVRAE↓VQKLQM	11682048
				Histone H3	PRKQL↓ATKAA	2153239
		Leader protease		eIF4G	SFANLG↓RTTLST	15885108
Hepatitis A Virus		3Cpro	(LVI)X(TSA)(QEX)↓XXXX	**NEMO**	PVLKAQ↓ADIYK	24920812
		3ABC		**MAVS**	LASQ↓VDSP	17438296
		3CD		**TRIF**	DWSQ↓GCSL	21931545
				TRIF	IREQSQ↓HLDG	21931545
Norovirus MD145-12	*Calciviridae*	3CLpro	(YILA)(EKST)(LTF)(QE)↓(GA)(PKG)(ETKPD)	**PABP**	VHVQ↓GQN	15254188
					AIPQ↓TQE	15254188
Feline Calicivirus (Vesivirus)			(IFP)(REKA)(ASE)E↓(ADNS)(CVGK)(PAHG)	**PABP**	WTAQ↓GAR	15254188
Dengue	*Flaviviridae*	ns2B/ns3	QKKKQR↓SGVLWD	**STING (MITA)**	VRACLGCPLRR↓GALLLLSIY	22761576
West Nile Virus				**STING (MITA)**		29915078
Japanese Encephlitis Virus				**STING (MITA)**		29915078

Zika Virus		ns2B/ns3	(KG)(KR)R↓(SG)	**STING (MITA)**	HIHSRYR↓GSYWRTVR	29915078
				SFRP1	SEGGRR↓GAALGVLL	30742841
				Gs alpha	QVAGRR↓GAALPCSL	30742841
				NT5M	VPAGRR↓GAAGGLGL	30742841
				FOXG1	KLAFKR↓GARLTSTG	30742841
Hepatitis C Virus		ns3/4A	C↓(SA)	**MAVS**	EREVPC↓HRPS	20044805, 16177806
				TRIF	PPPPPSSTPC↓SAHLTPSSLE	15710891
VEEV	*Togaviridae*	nsP2	AG(ACR)↓(GAY)	**TRIM14**	DCFATGRHYWEVDVQEAGA↓GWWVGA	30742841

Table 2. Reaction Mixes for K_m and V_{max} Measurements in a 96-well plate. The table shown below can be made in Excel. The A280 values in blue are inputted. The amounts shown (3 wells x 50 µL = 150 µL) are sufficient to fill 3 wells in the 96-well plate so that data can be measured in triplicate and standard deviations can be calculated.

Stock	A280	mg/mL	µM
Substrate (V12)	8.85	10.5	185.2
Enzyme (WT)	0.2	0.18	4.7

$[S]_{final}$ µM	Substrate (nmol)	Substrate (µL)	Enz (µL)	50 mM HEPES pH 7.0 Buffer (µL)	Total Volume (µL)
5	0.25	4.1	15	131	**150**
10	0.5	8.1	15	126.9	**150**
20	1	16.2	15	118.8	**150**
50	2.5	40.5	15	94.5	**150**
80	4	64.8	15	70.2	**150**
100	5	81	15	54	**150**
120	6	97.2	15	37.8	**150**
140	7	113.4	15	21.6	**150**

Table 3. Emission ratios (527 nm/470 nm) of the "uncut" and "cut" CFP-V12-YFP substrate. The data for the wells containing 4 nmols of substrate (80 μM) are shown in Table 4.

V12		
Uncut (*neg*)	Cut (*pos*)	Substrate (nmols)
0.88	0.58	0.25
0.94	0.64	0.5
1.04	0.73	1
1.25	0.94	2.5
1.38	1.07	4
1.44	1.12	5
1.5	1.17	6
1.54	1.22	7

Table 4. Representative fluorescence data for the VEEV nsP2 cysteine protease and CFP-V12-YFP substrate. Data for well E7 in the 96-well plate are shown. This well contains the reaction mixture that has a substrate concentration of 80 µM. The well contains 50 µL of the reaction mix or 4 nmols of substrate. The percentage of the substrate that has been cut (% Cut) can be calculated using the values in Table 3 (*e.g* . Fraction Cut = (X-neg)/(pos-neg)). The nmols of Substrate that has been cut can be calculated by multiplying the Fraction Cut value by 4 nmols in this example.

	Ex/Em = 434/527	Ex/Em= 434/470	[S] = 80 µM		
Well	E7	E7	E7	E7	E7
t (Minutes)	Fluorescence Units	Fluorescence Units	X = FU (434/527) / FU (434/470)	Fraction Cut	nmols
0:00	22296	16535	1.3484	0.03	0.1238
1:00	22689	16742	1.3552	0.01	0.0357
2:00	22650	16770	1.3506	0.02	0.0957
3:00	22485	16686	1.3475	0.03	0.1352
4:00	22549	16750	1.3462	0.04	0.1523
5:00	22641	16807	1.3471	0.04	0.1408
6:00	22592	16794	1.3453	0.04	0.1639
7:00	22458	16701	1.3447	0.04	0.1716
8:00	22414	16715	1.3409	0.06	0.2201
9:00	22460	16768	1.3394	0.06	0.2395
10:00	22513	16806	1.3396	0.06	0.2373
11:00	22465	16800	1.3372	0.07	0.268
12:00	22365	16762	1.3343	0.08	0.3059
13:00	22413	16780	1.3357	0.07	0.2881
14:00	22372	16802	1.3316	0.09	0.3411
15:00	22391	16760	1.336	0.07	0.284
16:00	22341	16811	1.329	0.09	0.3741
17:00	22301	16816	1.3261	0.1	0.4111
18:00	22235	16722	1.3296	0.09	0.3659
19:00	22271	16816	1.3244	0.11	0.4334
20:00	22192	16767	1.3236	0.11	0.4444

Table 5. Calculation of K_m and V_{max}. The spontaneous rate of substrate hydrolysis is shown in column 2. The rate of enzyme-catalyzed substrate hydrolysis is shown in column 3. The difference between these rates are the initial velocities; these values are shown in column 4 (column 4 = column 3 - column 2). In column 5 the values in column 4 have been divided by the mg of enzyme in the well. The mg of enzyme can be calculated from Table 2 (e.g. 5 µL of 0.18 mg/mL of enzyme).

	Slope	Slope	nmol/min	µmol/ min*mg
[S] µM	Spont	w/Enz	DIFF (v_o)	U/mg
5	-0.0002	0.006	0.0062	0.007
10	-0.0004	0.0084	0.0088	0.0099
20	-0.0009	0.0163	0.0172	0.0194
50	-0.0025	0.019	0.0215	0.0243
80	-0.0046	0.0195	0.0241	0.0272
100	-0.0062	0.0187	0.0249	0.0281
120	-0.0071	0.0183	0.0254	0.0287
140	-0.0089	0.0182	0.0271	0.0306

Table of Materials

Click here to access/download;Table of Materials;JoVE_Table_of_Materials FINAL_AZ.xlsx

Name of Material/Equipment	Company	Catalog Number	Comments/Description
250 mL Erlenmeyer Flask	VWR	89000-362	
2-mercaptoethanol	Acros Organics (Fisher)	125472500	Danger: Acutely Toxic. Open bottle in hood to avoid inhaling the fumes.
4L Pyrex wide-mouth graduated Erlenmeyer flask with screw-cap	Millipore Sigma	CLS49954L-1EA	
AKTA Prime Plus	GE Healthcare	17-0729-01	
AKTA XK 16/20 Column	GE Healthcare	28988937	
Amicon Ultra-0.5 Centrifugal Filter Unit	Millipore Sigma	UFC501096	
Amicon Ultra-15 Centrifugal Filter Unit	Millipore Sigma	UFC901096	
Amicon Ultra-4 Centrifugal Filter Unit	Millipore Sigma	UFC801024	
Ampicillin	Sigma	A0166	Danger: Allergic reactions (skin or breathing).
Chelating Sepharose Fast Flow	GE Healthcare	17-0575-02	Once the resin is equilibrated with 0.2 M Nickel Sulfate it is refered to as a Nickel Column in the text. Column will have a green color after washing with water. The column will have a blue color after equilibrating with buffer.
Chloramphenicol	RPI	C61000	Danger: May cause cancer.
Corning 50 mL centrifuge tubes	Corning	430828	Suggestion: Polypropylene tubes are less likely to crack during sonication than Polyethylene tubes
Corning 96 Well Half-Area Microplate, Non-Binding Surface	Corning	3993	
Dialysis Tubing Clips	Fisher Scientific	PI68011	
Disposable PD-10 Desalting Column	GE Healthcare	17-0851-01	
DNAse	Sigma	DN25-1G	
DTT (DL-Dithiothreitol)	RPI	D11000-50.0	Warning: Acute Oral Toxicity; skin and eye irritation
EDTA	Fisher Scientific	S311-500	
Fisherbrand Petri Dishes with Clear Lid	Fisher Scientific	FB0875712	
Glycerol	Acros Organics (Fisher)	15892-0010	
HEPES	Millipore Sigma	H4034-1KG	
Imidazole	Acros Organics	301870010	Danger: Toxic, Irritant
IPTG (Isopropyl β-D-thiogalactopyranoside)	Calbiochem (Millipore Sigma)	420291	Do not breathe dust. Avoid contact with eyes and skin.
Laemmli Sample Buffer	BIO-RAD	1610737	
Luria Bertani Agar	Fluka (Millipore Sigma)	L3027-1KG	Suggestion: Autoclave with magnetic stirrer in the liquid, and stir while cooling. Wait to add antibiotic until you can hold your hands on the bottle without pain for 30 seconds.
Luria Bertani Media	Fisher Bioreagents	BP1426-2	
Lysozyme	Sigma	L4919-5g	
Mini-PROTEAN Tetra Vertical Electrophoresis CellGel Box	BIO-RAD		
Nalgene Oak Ridge High-Speed PPCO Centrifuge Tubes	Nalgene (Thermo Scientific)	3119-0050	
Nanodrop	Thermo Fisher		
New Brunswick Innova 42R Shaker Incubator	Eppendorf	M1335	
Nickel Sulfate Hexahydrate (Crystalline/Certified ACS), Fisher Chemical	Fisher Scientific	N73-500	Danger: Harmful if swallowed or inhaled, skin and eye irritation,
One Shot BL21(DE3) Chemically Competent *E. coli*	Invitrogen (Thermo Fisher)	C600003	May be harmful if inhaled or swallowed. May cause skin and eye irritation with susceptible people.
One Shot BL21(DE3) pLysS Chemically Competent E. coli	Invitrogen (Thermo Fisher)	C606003	May be harmful if inhaled or swallowed. May cause skin and eye irritation with susceptible people.

pet15b plasmid DNA	Novagen (Millipore Sigma)	69661	GenScript Inc. was used for commerical DNA synthesis. The pet15b plasmid was used for the CFP/YFP substrates.
pet32b	Novagen (Millipore Sigma)	69016-3	The pet32b plasmid was used for the cysteine protease construct.
Pierce Protease Inhibitor Mini Tablets, EDTA-free	Thermo Fisher	A32955	Warning: Skin corrosion/irriation; eye damage
Plate Reader	Molecular Devices	Model M5	
Precision Plus Protein All Blue Prestained Protein Standard	BIO-RAD	161-0373	
Protein Extraction Reagent	Novagen (Millipore Sigma)	70584-4	BugBuster or Bper (Catalog # 78248, ThermoFisher)
Q-Sepharose Fast Flow	G.E. Healthcare	17-0510-01	Anion exchange resin
RunBlue (12%) 17-well PAGE gels	Expedeon	BCG01227	Any 12% pre-cast polyacrylamide gel can be used
RunBlue 20x SDS Running Buffer	Expedeon	NXB50500	Dilute 50 mL with 950 mL deionized water to obtain 1x
RunBlue Instant Blue Gel Stain	Expedeon	ISB1L	Do not dilute, use as directed
Sodium Chloride	Fisher Chemical	S271-10	
Sonifier Cell Disrupter 450 Sonicator	Branson Ultrasonics (VWR)	Model No. 101-063-346R	Sonicator was used on level 5
Spectra/Por 6-8 kD MWCO	Spectrum Labs	132645T	Dialysis Tubing
SP-Sepharose Fast Flow	G.E. Healthcare	17-0729-01	Cation exchange resin
Thrombin from bovine plasma	Sigma	T6634-500UN	
Tris Base	Fisher Scientific	BP152-500	Caution: Eye/Skin Irritant

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: Analysis of GroupIV viral SSHHPSSequences Using In vitro and Insilico Methods

Author(s): Xin Hu, Jaimee R. Compton, Patricia M. Legler

Item 1: The Author elects to have the Materials be made available (as described at http://www.jove.com/publish) via:

☒ Standard Access ☐ Open Access

Item 2: Please select one of the following items: (All Authors are gov. employees)

☐ The Author is **NOT** a United States government employee.

☒ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. **Defined Terms.** As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of The Journal of Visualized Experiments; "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. **Background.** The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. **Grant of Rights in Article.** In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and(c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. **Retention of Rights in Article.** Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. **Grant of Rights in Video – Standard Access.** This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. **Grant of Rights in Video – Open Access.** This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this **Section 6** is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. **Government Employees.** If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. **Protection of the Work.** The Author(s) authorize JoVE to take steps in the Author(s) name and on their behalf if JoVE believes some third party could be infringing or might infringe the copyright of either the Author's Article and/or Video.

9. **Likeness, Privacy, Personality.** The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

10. **Author Warranties.** The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

11. **JoVE Discretion.** If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

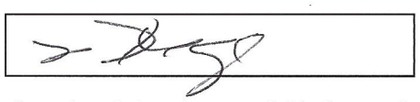
12. **Indemnification.** The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in JoVE or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

13. **Fees.** To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

14. **Transfer, Governing Law.** This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement is required per submission.

CORRESPONDING AUTHOR

Name:	Patricia Legler
Department:	CBMSE
Institution:	Naval Research Laboratory
Title:	Ph.D
Signature:	
Date:	06/14/2019

Please submit a **signed** and **dated** copy of this license by one of the following three methods:

1. Upload an electronic version on the JoVE submission site
2. Fax the document to +1.866.381.2236
3. Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02140

Editorial comments:

1. Please note that the editor has formatted the manuscript to match the journal's style. Please retain the same. The updated manuscript is attached and please use this version to incorporate the changes that are requested.
As requested, we have used the provided word.doc

2. Please adjust the numbering of the Protocol to follow the JoVE Instructions for Authors. Step 1 followed by 1.1, followed by 1.1.1, etc. Use subheadings and substeps for clarity if there are discrete stages in the protocol.
As requested, we have renumbered sections.

3. Please revise the Protocol to contain only action items that direct the reader to do something (e.g., “Do this,” “Ensure that,” etc.). The actions should be described in the imperative tense in complete sentences wherever possible. Avoid usage of phrases such as “could be,” “should be,” and “would be” throughout the Protocol. Any text that cannot be written in the imperative tense may be added as a “NOTE.” Please include all safety procedures and use of hoods, etc. However, notes should be used sparingly and actions should be described in the imperative tense wherever possible. Please move the discussion about the protocol to the Discussion.
As requested, we have revised the protocols and moved some text to the Discussion.

4. Please note that the highlighted protocol text will be used to generate the script for the video and must contain everything that you would like shown in the video. Software must have a GUI (graphical user interface) and software steps must be more explicitly explained ('click', 'select', etc.). Please add more specific details (e.g. button clicks or menu selections for software actions, numerical values for settings, etc.). There should be enough detail in each step to supplement the actions seen in the video so that viewers can easily replicate the protocol. See specific comments marked in the attached manuscript.
As requested, we have responded to the specific comments.

5. Please revise the Protocol steps so that individual steps contain only 2-3 actions per step and a maximum of 4 sentences per step. Use sub-steps as necessary.
As requested, we have limited each step to 4 sentences and reduced the number of actions.

6. Please address specific comments marked in the attached manuscript.
As requested, we have responded to the specific comments.

7. After you have made all the recommended changes to your protocol section (listed above), please highlight in yellow up to 2.75 pages (no less than 1 page) of protocol text (including headers and spacing) to be featured in the video. Bear in mind the goal of the protocol and highlight the critical steps to be filmed. Our scriptwriters will derive the video script directly from the highlighted text.
As requested, we have highlighted the regions to demonstrate.

8. Please highlight complete sentences (not parts of sentences). Please ensure that the highlighted steps form a cohesive narrative with a logical flow from one highlighted step to the next. The highlighted text must include at least one action that is written in the imperative voice per step. Notes cannot usually be filmed and should be excluded from the highlighting.
As requested, we have highlighted the regions to demonstrate.

9. Please include all relevant details that are required to perform the step in the highlighting. For example: If step 2.5 is highlighted for filming and the details of how to perform the step are given in steps 2.5.1 and 2.5.2, then the sub-steps where the details are provided must be highlighted.
As requested, we have highlighted the regions to demonstrate.

10. Please upload each Figure individually to your Editorial Manager account as a .png, .tiff, .pdf, .svg, .eps, .psd, or .ai file.
As requested, we have loaded individual files.

11. Please upload each Table individually to your Editorial Manager account as an .xlsx file. Avoid any coloring or formatting in the tables.
As requested, we have loaded individual files. Table 1 does have color in it, however coloring the residues is one of the steps in the protocol. Is it okay to use color here?

12. Please remove the titles and figure legends from the uploaded figures. Please include all the figure Legends together at the end of the Representative Results in the manuscript text.
As requested, we have removed the figure legends.

13. References: Please do not abbreviate journal titles; use full journal name.
As requested, we have included full names.

TITLE:

Analysis of Group IV viral SSHHPS Using *In vitro* and *In silico* Methods

AUTHORS AND AFFILIATIONS:

Xin Hu[1], Jaimee R. Compton[2], Patricia M. Legler*[2]

[1]National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD

[2] United States Naval Research Laboratory, 4555 Overlook Ave., Washington, D.C.

Email addresses of co-authors:

Xin Hu (xin.hu@nih.gov)

Jaimee R. Compton (jaimee.compton@nrl.navy.mil)

Corresponding author:

Patricia M. Legler (patricia.legler@nrl.navy.mil)

KEYWORDS:

Enzyme, Assay, FRET, nonstructural protein, alphavirus, Group IV, protease, SSHHPS, gel assay, *in vitro*, docking

SUMMARY:

We present a general protocol for identifying short stretches of homologous host-pathogen protein sequences (SSHHPS) embedded in the viral polyprotein. SSHHPS are recognized by viral proteases and direct the targeted destruction of specific host proteins by several Group IV viruses.

ABSTRACT:

Alphaviral enzymes are synthesized in a single polypeptide. The nonstructural polyprotein (nsP) is processed by its nsP2 cysteine protease to produce active enzymes essential for viral replication. Viral proteases are highly specific and recognize conserved cleavage site motif sequences (~6-8 amino acids). In several Group IV viruses, the nsP protease(s) cleavage site motif sequences can be found in specific host proteins involved in generating the innate immune responses~~,~~ and, in some cases, the targeted proteins appear to be linked to the virus-induced phenotype. These viruses utilize short stretches of homologous host-pathogen protein sequences (SSHHPS) for targeted destruction of host proteins. To identify SSHHPS the viral protease cleavage site motif sequences can be inputted into BLAST and the host genome(s) can be searched. Cleavage ~~can~~ initially can be tested using the purified nsP viral protease and fluorescence resonance energy transfer (FRET) substrates made in *E. coli*. The FRET substrates contain cyan and yellow fluorescent protein and the cleavage site sequence (CFP-sequence-YFP). This protease assay can be used continuously in a plate reader or discontinuously in SDS-PAGE gels. Models of the bound peptide substrates can be generated *in silico* to guide substrate selection and mutagenesis studies. CFP/YFP substrates have also been utilized to identify protease inhibitors. These *in vitro* and *in silico* methods can be used in combination with cell-based assays to determine if the targeted host protein affects viral replication.

INTRODUCTION:

Evidence of horizontal gene transfer from virus to host, or host to virus can be found in a variety of genomes [1-4]. Examples of viral endogenization are the CRISPR spacer sequences found in bacterial host genomes[4]. Recently, we have found evidence of host protein sequences embedded in the nonstructural polyproteins of (+)ssRNA Group IV viruses. These sequences within the coding regions of the viral genome can be propagated generationally. The short stretches of homologous host-pathogen protein sequences (SSHHPS) are found in the virus and host [5,6]. SSHHPS are the conserved cleavage site motif sequences recognized by viral proteases that have homology to specific host proteins. These sequences direct the destruction of specific host proteins.

In our previous publication[6], we compiled a list of all of the host proteins that were targeted by viral proteases and found that the list of targets was non-random (**Table 1**). Two trends were apparent. First, the majority of the viral proteases that cut host proteins belonged to Group IV viruses (24~~31~~ of 25~~32~~ cases involved Group IV viral proteases), and ~~.~~ ~~O~~one protease belonged to the (+)ssRNA Group VI retroviruses (HIV, human immunodeficiency virus) [7]. Second, the host protein targets ~~that were~~ being cut by the viral proteases were generally involved in generating the innate immune responses suggesting that the cleavages were intended to antagonize the host's immune responses. Half of the host proteins targeted by the viral proteases wea~~we~~re known components of signaling cascades that generate interferon (IFN) and proinflammatory cytokines (**Table 1**). Others were involved in host cell transcription[8-10] or translation[11]. Interestingly, Shmakov, et al,[4] have shown that many CRISPR protospacer sequences correspond to genes involved in plasmid conjugation or replication[4]. ~~.~~

Group IV includes, among others, *Flaviviridae, Picornaviridae, Coronaviridae, Calciviridae,* and *Togaviridae* ~~among others~~. Several new and emerging pathogens belong to Group IV such as the Zika virus (ZIKV), West Nile (WNV), Chikungunya (CHIKV), severe acute respiratory syndrome virus (SARS) and Middle East respiratory syndrome virus (MERS). The (+)ssRNA genome is essentially a piece of mRNA. To produce the enzymes necessary for genome replication the (+)ssRNA genome ~~must~~ first must be translated. In alphaviruses and other Group IV viruses the enzymes necessary for replication are produced in a single polyprotein (*i.e.* nsP1234 for VEEV). The nonstructural polyprotein (nsP) is proteolytically processed (nsP1234 → nsP1, nsP2, nsP3, nsP4) by the nsP2 protease to produce active enzymes[12] (**Figure 1**). Cleavage of the polyprotein by the nsP2 protease is essential for viral replication~~,~~ this has been demonstrated by deletion and site-directed mutagenesis of the active site cysteine of the nsP2 protease [13,14]. Notably, the translation of viral proteins precedes genome replication events. For example, nsP4 contains the RNA-dependent RNA polymerase needed to replicate the (+)ssRNA genome. Genome replication can produce dsRNA intermediates; these intermediates can trigger the host's innate immune responses. Thus, these viruses may cleave host innate immune response proteins early in infection in order to suppress their effects.[15-17]

Silencing can occur at the level of DNA, RNA, and protein. What is common to each of the silencing mechanisms shown in **Figure 1** is that short foreign DNA, RNA, or protein sequences are

used to guide the destruction of specific targets to antagonize their function. The silencing mechanisms are analogous to "search and delete" programs that have been written in three different languages. The short cleavage site sequence is analogous to a "keyword". Each program has an enzyme that recognizes the match between the short sequence (the "keyword") and a word in the "file" that is to be deleted. Once a match is found, the enzyme cuts ("deletes") the larger target sequence. The three mechanisms shown in **Figure 1** are used to defend the host from viruses, or to defend a virus from a host's immune system.

Viral proteases recognize short cleavage site motif sequences between ~2-11 amino acids, in nucleotides this would correspond to 6-33 bases. For comparison, CRISPR spacer sequences are ~26-72 nucleotides and RNAi are ~20-22 nucleotides [18,19]. While these sequences are relatively short, they can be recognized specifically. Given the higher diversity of amino acids, the probability of a random cleavage event is relatively low for a viral protease recognizing protein sequences of 6-8 amino acids or longer. The prediction of SSHHPS ~~sequences~~ in host proteins will largely depend upon the specificity of the viral protease being examined. If the protease has strict sequence specificity requirements the chance of finding a cleavage site sequence is $1/20^6 = 1$ in 64 million or $1/20^8 = 1$ in 25.6 billion; however, most proteases have variable subsite tolerances (e.g., R or K may be tolerated at the S1 site). Consequently, there is no requirement for sequence identity between the sequences found in the host versus the virus. For viral proteases that have looser sequence requirements (such as those belonging to *Picornaviridae*) the probability of finding a cleavage site in a host protein may be higher. Many of the entries in **Table 1** are from the *Picornaviridae* family.

Schechter & Berger notation [20] is commonly used to describe the residues in a protease substrate and the subsites to which they bind, we utilize. ~~t~~This notation ~~will be used~~ throughout. The residues in the substrate that are N-terminal of the scissile bond are denoted as P3-P2-P1 while those that are C-terminal are denoted as P1'-P2'-P3'. The corresponding subsites in the protease that bind these amino acid residues are S3-S2-S1 and S1'-S2'-S3', respectively.

To determine which ~~at~~ host proteins are being targeted we can identify SSHHPS in the viral polyprotein cleavage sites and search for the host proteins that contain them. Herein, we outline procedures for identifying SSHHPS ~~sequences~~ using known viral protease cleavage site sequences. The bioinformatic methods, protease assays, and *in silico* methods described are intended to be used in conjunction with cell-based assays.

Sequence alignments of the host proteins targeted by viral proteases have revealed species-specific differences within these short cleavage site sequences. For example, the Venezuelan equine encephalitis virus (VEEV) nsP2 protease was found to cut human TRIM14, a tripartite motif (TRIM) protein ~~.~~[6]. Some TRIM proteins are ~~species-specific~~ viral restriction factors (e.g., TRIM5α[21]), most are thought to be ubiquitin E3 ligases. ~~Sets of TRIMs have been associated with different viruses [22,23].~~ TRIM14 lacks a RING (really interesting new gene) domain and is not thought to be an E3 ligase[24]. TRIM14 has been proposed to be an adaptor in the mitochondrial antiviral signalosome (MAVS) [24], but may have other antiviral functions [25]. Alignment of TRIM14 sequences from various species shows that equine lack the cleavage site and harbor a truncated version of TRIM14 that is missing the C-terminal PRY/SPRY domain. This domain contains a polyubiquitination site (**Figure 2**). In equine, these viruses are highly lethal (~20-80% mortality) whereas in humans only ~1% die from VEEV infections [26]. Cleavage of the PRY/SPRY domain may transiently short circuit the MAVS signaling cascade. This cascade can be triggered by dsRNA and leads to the production of interferon ~~(IFN)~~ and pro-inflammatory cytokines ~~and can be triggered by dsRNA~~. Thus, the presence of the SSHHPS may be useful for predicting which species have defense systems against specific Group IV viruses.

In Group IV viruses IFN antagonism mechanisms are thought to be multiply redundant [27]. Host protein cleavage may be transient during infection and concentrations may ~~can~~ recover over time. We found in cells that TRIM14 cleavage products could be detected very early after transfection (6 h) with a plasmid encoding the protease (cytomegalovirus ~~CMV~~ promoter). However, ~~but~~ at longer periods, the cleavage products were not detected. In virus-infected cells, the kinetics were different and cleavage products could be detected between 6-48 h[6]. Others have reported the appearance of host protein cleavage products as early as 3-6 h post infection [9,11].

Proteolytic activity in cells is often difficult to catch; ~~t~~ as ~~t~~he cleavage products can vary in their solubility, concentration, stability, and lifetime. In cell-based assays it cannot be assumed that cleavage products will accumulate in a cell or that the band intensities of cut and uncut protein will show compensatory increases and decreases as the cut protein may be degraded very quickly and may not be detectable in a Western blot at an expected molecular weight (MW) (e.g., the region containing the epitope could be cleaved by other host proteases or could be ubiquitinated). If the substrate of the viral protease is an innate immune response protein, its concentration may vary during infection. For example, some innate immune response proteins are present prior to viral infection and are ~~further~~ induced further by interferon[28]. The concentration of the target protein may therefore fluctuate during infection and comparison of uninfected vs. infected cell lysates may be difficult to interpret. Additionally, all cells may not be uniformly transfected or infected. *In vitro* protease assays using purified proteins from *E.coli* on the other hand have fewer variables for which to control ~~for~~ and such assays can be done using SDS-PAGE rather than ~~instead of~~ immunoblots. Contaminating proteases can be inhibited in the early steps of the protein purification of the CFP/YFP substrate, and mutated viral proteases can be purified and tested as controls to determine if the cleavage is due to the viral protease or ~~vs.~~ a contaminating bacterial protease.

One limitation of *in vitro* protease assays is that they lack the complexity of a mammalian cell. For an enzyme to cut its substrate, the two must be co-localized. Group IV viral proteases differ in structure and localization. For example, the ZIKV protease is embedded in the endoplasmic reticulum (ER) membrane and faces the cytosol, whereas the VEEV nsP2 protease is a soluble protein in the cytoplasm and nucleus [29]. Some of the cleavage site sequences that we found in our ZIKV SSHHPS analysis were in signal peptides suggesting that cleavage might occur co-translationally for some targets. Thus, the location of the protease and the substrate in the cell also needs to be considered in these analyses.

Cell-based assays can be valuable for establishing a role for the identified host protein(s) in infection. Methods that aim to halt viral protease cleavage of host proteins such as the addition of a protease inhibitor[6] or a mutation in the host target[16] can be used to examine their effects on viral replication. Overexpression of the targeted protein ~~may~~ also may affect viral replication [30]. Plaque assays or other methods can be used to quantify viral replication.

PROTOCOL:

1. Bioinformatics - Identification of SSHHPS in the Host G~~g~~enome U~~u~~sing BLAST

NOTE: Protein BLAST can be found at blast.ncbi.nlm.nih.gov/Blast.cgi.

1.1. Input ~20 amino acids surrounding the scissile bond in the viral polyprotein. Select "non-redundant protein sequences" and type in the host genome to be searched (e.g., *Homo sapiens*).

1.1.1. An additional tool that can be selected is PHI-BLAST. Type in a pattern sequence (e.g., for the 25 residues of V12 shown below enter the pattern "AG" without quotes).

```
VEEV nsP1/2 V12 VEEPTLEADVDLMLQEAGA↓GSVETP
VEEV nsP2/3 V23 LSSTLTNIYTGSRLHEAGC↓APSYHV
VEEV nsP3/4 V34 TREEFEAFVAQQQRFDAGA↓YIFSSD
```

NOTE: ~~An additional tool that can be selected is PHI-BLAST.~~ In PHI-BLAST, square brackets [XY] indicate that amino acid X or Y can be at the subsite position (e.g., AG[AC][GAY]).

Commented [A1]: The editor rephrased these lines according to JoVE guidelines. Please review for accuracy.

Commented [A2R1]: Step 1.1.1. is for Phi-BLAST (but not for regular blast)

1.1.2. Inspect the BLAST results and identify the hits that have high sequence identity to residues that are conserved in the polyprotein cleavage sites (e.g., tripartite motif protein 14) (**Figure 3**).

NOTE: For serine proteases higher conservation of the P1 residue is expected, while for cysteine proteases higher conservation of the P2 residue is expected.

1.1.3. Color the residues that are identical to a cleavage site sequence and are in sequential order (no gaps). Color the residues tolerated at the subsite, but present in a different cleavage site in a second color.

NOTE: Residues that represent conservative substitutions (e.g., Leu vs. Val) that are not present in a viral cleavage site ~~may~~ also may be found and may or may not be recognized by the viral protease.

1.1.4. Rank -order the BLAST hits based upon the number of consecutive identical or tolerated residues that match a cleavage site sequence. From the list select the proteins containing ≥6 identical or similar residues for analysis in protease assays.

1.1.5. Repeat the procedure for the other cleavage sites (nsP2/3, nsP3/4, etc.) and gradually strengthen the prediction by adding more highly conserved residues to the PHI-BLAST pattern.

2. *In vitro* assays - designing and preparing protease substrates

2.1. Construct a plasmid encoding the cyan fluorescent protein (CFP), ≤25 amino acids of the cleavage site sequence, followed by the yellow fluorescent protein (YFP, also known as Venus[31]).

NOTE: The plasmid can be constructed using sequence and ligation independent cloning (SLIC)[32] or commercial gene synthesis. A pet15b plasmid containing the sequence shown in **Figure 4** was synthesized commercially and was used here.

2.1.1. To optimize the substrate length, construct additional variable length FRET substrates containing 12-25 amino acids of the natural viral polyprotein cleavage site sequences using a 2-fragment SLIC reaction. Analyze cleavage using the SDS-PAGE gel-based assay or by measuring steady state kinetic parameters using the methods below.

Commented [A3]: This step seems to be quite different from substeps 2.2.1-2.2.17.
Please consider numbering the substeps as parallel steps, i.e., 2.3-2.18.

NOTE: In some cases, cleavage sites can be identified by homology to known cleavage sites [33]. If cleavage of the substrates containing the polyprotein junction sequences is not observed, there may be a requirement for additional residues or a structural motif (e.g., an alpha helix [34]). Alternatively, the purified viral protease may be inactive. Confirm cleavage of the viral polyprotein sequences before pursuing SSHHPS analysis. The number of residues in the substrate was optimized for the VEEV protease using variable length substrates (12 to 25 amino acids) followed by analysis of V_{max} and K_m [34,35]. The Zika viral ns2B/nsB protease cleavage sites used in the examples have been published [36,37].

2.2.~~1.~~ Prepare the CFP/YFP substrates by f~~F~~reshly transforming 8-20 µL of BL-21(DE3) *E.coli* competent cells with the CFP-V12-YFP plasmid according to manufacturer's directions and plate on Luria Bertani (LB) agar plates containing 50 µg/mL Ampicillin (37 °C).

2.2.~~1~~2. Autoclave four 4 L flasks containing LB media (1.5 L media per flask) and 100 mL LB in a 250 mL flask. Cap each flask with aluminum foil.

2.2.~~2~~3. Inoculate the 100 mL culture with a colony of the freshly transformed bacteria and grow at 37 °C with shaking (200 rpm) overnight.

2.2.~~3~~4. To make the CFP/YFP substrate, inoculate four 4 L flasks with 25 mL of an overnight culture. Begin shaking the cultures at 37 °C and monitor growth by UV-vis spectroscopy at 600 nm hourly.

2.2.~~4~~5. When the bacteria reach an absorbance of ~1.0 at 600 nm (approximately ~3-4 h of growth) induce protein expression by adding 0.5 mL of 1 M isopropyl-β-D-thiogalactoside (IPTG) per flask. After adding IPTG, lower the temperature of the shaking incubator to 17 °C and allow expression to continue overnight for 17-20 h.

Commented [A4]: Please spell it out.

2.2.5~~6~~. Pellet the bacteria using a high-speed centrifuge at 7,000 x g for 10 min (4 °C) and retain the pellets. Remove and discard liquid media. Store the pellets at -80 °C or lyse immediately.

2.2.6~~7~~. Prepare 100 mL of lysis buffer containing 50 mM Tris pH 7.6, 500 mM NaCl, 35 mL of bacterial protein extraction reagent, 30 mg of lysozyme, 25 U of DNase, and 1 protease inhibitor tablet. Resuspend the pellets in lysis buffer with a pipet and transfer ~25-35 mL into 50 mL disposable conical tubes.

2.2.7~~8~~. Place the tubes in a plastic beaker containing ice water. Insert the sonicator tip into the tubes so that the tip is ~1 cm from the bottom of the tube and sonicate the lysates 10-20 times on level 5 for 15 second intervals until the lysate becomes fluid and liquefied.

NOTE: Use hearing protection during sonication.

2.2.8~~9~~. Transfer the lysate to high speed centrifuge tubes and centrifuge at 20,500 x g for 30 min at 4 °C. After the spin, retain the ~~approximately 100 mL of~~ supernatant (~100 mL) and transfer it to a clean bottle. Discard the pellets.

2.2.9~~10~~. Prepare 1 L of Buffer A (50 mM Tris pH 7.6, 500 mM NaCl). Prepare 300 mL of Buffer B (50 mM Tris pH 7.6, 500 mM NaCl, 300 mM Imidazole).

2.2.10~~1~~. Equilibrate a 100 mL nickel column using 3 column volumes of Buffer A and a flow rate of 5 mL/min.

2.2.11~~2~~. Load the lysate onto the nickel column using a flow rate of 2 to 5 mL/min. Wash the column with 2 column volumes of Buffer A, followed by ~5 column volumes of 20% Buffer B. During the 20% Buffer B wash the absorbance at 280 nm (A_{280}) will increase as ~~A~~ contaminants ~~peak will~~ elute from the column during the ~~is~~ wash. Continue washing the column until the ~~absorbance at 280 nm~~ (A_{280} of the eluate) ~~has~~ has returned to baseline values.

Commented [A5]: Observed by what?

Commented [A6]: Absorbance of what, the eluate? When is the absorbance measured? It is not mentioned.

2.2.12~~3~~. Elute the protein with 2-3 column volumes of 100% Buffer B using a flow rate of 2-5 mL/min and collect 10 mL fractions. Measure the A_{280} of each fraction.

2.2.13~~4~~. Combine and concentrate fractions containing A_{280} > 0.1 using a 15 mL centrifugal ultrafiltration unit. Spin the ultrafiltration units at 5,000 x g for 15 min and continue to add fractions until the volume has been reduced to ~50-75 mL.

2.2.14~~5~~. Cut a 14 inch piece of dialysis tubing with a molecular weight cut-off (MWCO) of 6-8 kDa. Hydrate the dialysis tubing by boiling it fully submerged in 300 mL of water for 10 min. Tie a secure knot at one end of the membrane. Fill the bag with dialysis buffer to ensure that no cracks or leaks are present. Remove the buffer from the bag and keep the bag submerged in the dialysis buffer.

2.2.15~~6~~. Transfer the concentrated protein from 2.2.13 into the dialysis bag with a plastic pipet. Remove any air bubbles from the bag. Close the bag with a second knot or a dialysis clip. Dialyze the protein against 500 mL of 50 mM Tris pH 7.6, 5 mM EDTA (ethylenediaminetetraacetic acid), 250 mM NaCl in a 500 mL graduated cylinder overnight at 4 °C.

Commented [A7]: Do you mean fractions from step 2.2.14?

2.2.16. ~~2.2.17.~~ Dialyze the protein a second time against 500 mL of 50 mM Tris pH 7.6 at 4 °C for 2 h.

2.3. For the anion exchange column p~~P~~repare 500 mL of Buffer A (50 mM Tris pH 7.6) and 500 mL Buffer B (50 mM Tris pH 7.6, 1.0 M NaCl). Equilibrate a 30 mL anion exchange column with 3 column volumes of Buffer A (2-5 mL/min).

2.3.1. Remove the protein from the dialysis bag and transfer to a bottle. Keep the bottle on ice. Load the dialyzed protein onto the column (2-5 mL/min).

NOTE: The CFP/YFP protein will bind the column and will be yellow in appearance.

2.3.2. Wash the column with Buffer A until the A_{280} returns to baseline (5 mL/min). Elute the protein using a gradient (0-50% Buffer B, 100 mL) and collect 10 mL fractions.

2.3.3. Inspect the column fractions using SDS-PAGE. Combine those that are >95% pure.

2.3.4. Concentrate the protein to an A_{280} ~10-20 using a 15 mL centrifugal ultrafiltration unit. Spin the concentrator at 4,500 x g for 10 min at 4 °C and continue to add protein until all of the protein-containing fractions have been combined.

2.4. Carefully remove the protein from the concentrator with a pipet. Aliquot the protein into 1.5 mL microcentrifuge tubes and flash freeze in liquid nitrogen for long term storage at -80 °C. Buffer exchange the protein at room temperature using a PD-10 column equilibrated with the appropriate assay buffer prior to use.

2.5. Using Beer's law c~~C~~alculate the protein concentration using the A_{280} and a calculated extinction coefficient (e.g., for the V12 substrate the ε = 47,790 M^{-1} cm^{-1}).

NOTE: The extinction coefficient (ϵ) can be calculated from the protein sequence in **Figure 4** using the Expasy ProtParam program (https://web.expasy.org/protparam/).

3. Preparation of the Alphaviral nsP2 Cysteine P~~p~~rotease

Commented [A8]: Please change the numbering in this section to 3.1-3.23.

3.1.

Design and construct a plasmid encoding the protease. For cysteine proteases, use the pet32 plasmid to construct a thioredoxin (Trx) fusion protein.

NOTE: The pet32 plasmid encodes a thrombin cleavage site (LVPR↓GS) for removal of the thioredoxin and His-tag (**Figure 5**). Thioredoxin will help maintain the active site cysteine in a reduced state during expression. For serine proteases, the thioredoxin is not needed and steps involving its removal by thrombin can be omitted. The VEEV nsP2 protease sequence was incorporated into a pet32b plasmid that was prepared commercially to avoid handling Sselect agents.

3.1.1.2 Freshly transform the plasmid DNA into BL21(DE3)pLysS *E.coli* according to manufacturer's directions. Plate the bacteria on LB agar plates containing Ampicillin.

NOTE: Chloramphenicol is only used for *E. coli* strains carrying the pLysS plasmid and is omitted if BL21(DE3) cells are used. It is not necessary to include chloramphenicol on the LB agar plate in this step.

2.0.0.3.1.1. Autoclave four 4 L flasks of 1.5 L of LB media (6 L total volume) and 100 mL LB in a 250 mL flask. Cap each flask with aluminum foil.

2.0.0.3.1.2. Inoculate a 100 mL overnight culture of LB/Ampicillin with a colony from the plate and grow in a shaking incubator (200 rpm) at 37 °C.

2.0.0.3.1.3. Inoculate the 4 L flasks with 25 mL of the overnight culture and add the appropriate antibiotics.

NOTE: The media for the BL21(DE3) pLysS cells carrying the pet32 plasmid should have final concentrations of 25 µg/mL chloramphenicol and 50 µg/mL Ampicillin.

2.0.0.3.1.4. Induce protein expression by adding 0.5 mL of IPTG to the culture when the absorbance at 600 nm reaches 1.0. Lower the temperature of the shaking incubator to 17 °C. Allow expression to continue overnight (~17 h).

2.0.0.3.1.5. Pellet the cells by centrifugation (7,000 x g for 10 min at 4 °C). Remove and discard the liquid media.

NOTE: The pellets can be stored at -80 °C for months or lysed immediately.

2.0.0.3.1.6. Prepare 100 mL of lysis buffer (50 mM Tris pH 7.6, 500 mL NaCl, 2 mM beta mercaptoethanol (BME), 30 mg lysozyme, 5% glycerol, 25 U DNase, 35 mL bacterial protein extraction reagent). Open bottles of BME in a chemical hood when adding. Keep the bacterial lysate on ice or at 4 °C for this and all subsequent steps.

NOTE: For cysteine proteases, 2 mM ~~beta mercaptoethanol (BME)~~BME is included to keep the nucleophilic cysteine reduced. The columns can be run at room temperature using chilled buffers. Buffers should be made with cold deionized water cooled to 4 °C.

2.0.0.3.1.7. Resuspend the bacterial pellets in ~25 mL of lysis buffer and transfer ~25 mL of the lysate into 4 x 50 mL disposable conical tubes. Place the tubes into plastic beakers containing ice water. Sonicate the lysate 10 times on level 5 for 15 second intervals.

2.0.0.3.1.8. Transfer the lysate into high speed centrifuge tubes. Clarify the lysate by centrifugation (30 min, 20,500 x g at 4 °C).

2.0.0.3.1.9. Prepare 0.5 L of Buffer A (50 mM Tris pH 7.6, 500 mM NaCl, 5% glycerol, 2 mM BME) and chill to 4 °C.

2.0.0.3.1.10. Prepare 250 mL of Buffer B (50 mM Tris pH 7.6, 500 mM NaCl, 5% glycerol, 2 mM BME, 300 mM Imidazole) and chill to 4 °C.

2.0.0.3.1.11. Equilibrate a 50 mL nickel column with 3 column volumes of Buffer A. Load the clarified lysate onto the column at 2-5 mL/min and discard the pellets.

2.0.0.3.1.12. Wash the column (2-5 mL/min) with 2 column volumes of Buffer A followed by 5 column volumes of Buffer A containing 20% Buffer B (60 mM iimidazole). Elute the protein (5 mL/min) with 100% Buffer B and collect 10 mL fractions.

2.0.0.3.1.13. Combine and concentrate fractions containing the protease that have $A_{280} \geq 0.1$ using a 15 mL centrifugal ultrafiltration unit and 15 min spins at 5,000 x g at 4 °C. After the volume has been reduced to ~ 5 mL buffer exchange the protein in the concentration unit by adding fresh dialysis buffer to the protein (50 mM Tris pH 7.6, 250 mM NaCl, 5 mM dithiothreitol (DTT), 1 mM EDTA, 5% Glycerol). Spin again at 5,000 x g at 4 °C for 15 min; repeat the buffer exchange step 2-3 times. Add thrombin to the protein (20 µL of 1 unit/µL) prior to dialysis to remove the thioredoxin and His-tag.

2.0.0.3.1.14. Transfer the protein into a dialysis bag and dialyze against 500 mL of the dialysis buffer (4 °C) in a 500 mL graduated cylinder overnight.

NOTE: The FPLC (fast protein liquid chromatography) system and the nickel column should be thoroughly cleaned with stripping buffer (2 M NaCl, 50 mM EDTA) before proceeding to the anion exchange column. Any residual nickel in the FPLC lines will turn the buffer solutions containing DTT brown when mixed. Wash the nickel column and FPLC system with 4 column volumes of water. Pump wWash the FPLC system thoroughly with water. The nickel column can be regenerated by flowing 2 column volumes of 0.2 M nNickel sSulfate over the resin for subsequent purifications.

2.0.3.2. For the anion exchange column pPrepare 1 L of Buffer A (50 mM Tris pH 7.6, 5 mM DTT, 5% glycerol).

2.0.0.3.2.1. Prepare 0.5 L of Buffer B (50 mM Tris pH 7.6, 5 mM DTT, 5% glycerol, 1.25 M NaCl).

~~3.1.19.~~3.2.2. Equilibrate a 30 mL anion exchange column with Buffer A (3 column volumes, 2-5 mL/min). Place the tubes in the fraction collector for collection of the flow through.

NOTE: The VEEV protease has a calculated isoelectric point (pI) of 8.7 and will bind cation-exchange columns but will flow through anion exchange columns. The pI can be calculated from the protein sequence using the Expasy ProtParam program (https://web.expasy.org/protparam/).

~~3.1.20.~~3.2.3. Dilute the dialyzed protein 1:3 with Buffer A, then load the protein (5 mL/min). Collect the flow-through in 10 mL fractions.

~~3.1.~~3.3. Remove the anion exchange column from the FPLC system. Connect a cation exchange column to the FPLC system. Equilibrate a 30 mL cation exchange column with 3 column volumes of Buffer A (5 mL/min).

~~3.1.22.~~3.3.1. Load the flow through of the anion exchange column onto the cation exchange column at 2-5 mL/min. Wash the column with Buffer A until the A_{280} returns to baseline level. Elute the protein with a 100 mL gradient (0-50% Buffer B) and collect 10 mL fractions.

NOTE: The VEEV protease will elute at around 0.6 M NaCl.

~~3.1.23.~~3.3.2. Inspect the column fractions using SDS-PAGE. Combine fractions that are >95% pure and concentrate to an $A_{280} \approx 2$ using 15 mL centrifugal ultrafiltration units. The enzyme can be flash-frozen in liquid nitrogen and stored at -80 °C.

~~3.~~4. Assaying the Enzyme Continuously Using a Plate Reader

Commented [A9]: Please change the numbering in this section to 4.1-4.15.

4.1. Prepare 50 mL of assay buffer (50 mM HEPES pH 7.0~~, 150 mM NaCl, 5 mM DTT~~).

~~3.0.0.~~4.1.1. The alphaviral proteases have relatively low k_{cat} values. Dilute the enzyme in the buffer ~~to be assayed to~~to 4.7 µM (this will roughly correspond to an A_{280} = 0.2 for the VEEV protease without Trx).

Commented [A10]: With what?

4.1.~~2.~~3 To measure the activity of the enzyme prepare a stock of substrate in the assay buffer with a concentration of 185 µM, this will roughly correspond to an A_{280} = 9. In 8 microcentrifuge tubes prepare the reaction mixes shown in **Table 2** by combining the appropriate volumes of the 185 µM substrate stock and buffer. In a black half-area 96-well plate pipet 45 µL of the reaction mixes into 3 wells (columns 1, 2, 3). Row A should contain the [S] = 5 µM reaction mix, and Row H should contain the [S] = 140 µM reaction mix.

4.1.~~3.~~4 Set the plate reader to ~~simultaneously~~ detect simultaneously fluorescence at two wavelengths with a fixed photomultiplier tube (PMT) setting (e.g., ~~l~~Low):

Wavelength 1	excitation = 434 nm, emission = 527 nm
Wavelength 2	excitation = 434~~——~~ nm, emission = 470 nm

~~3.0.3.~~4.1.4. Set the read time to 20 min (measuring 1 read per minute) and select the wells to be read. Insert the plate into the plate reader and measure the spontaneous rate of hydrolysis for 20 min. Monitor the emission ratios (Emission at 527/Emission at 470) over time.

~~3.0.3.~~4.1.5. Next, run an e~~E~~ndpoint read of the plate containing the "UNCUT~~ncut~~" substrate.

NOTE: These values will be used in subsequent data calculations. The average of the emission ratios from 3 wells will be the values of the "UNCUT~~ncut~~" substrate at t=0 in **Table 3**.

4.1.~~6.~~7 Remove the plate and pipet 5 µL of enzyme into each well. Read the plate again for 20 min with 1 read per minute. Set the plate reader to output a~~A~~bsolute values.

NOTE: For this assay, the slopes will be negative. Each well will contain a total volume of 50 µL.

~~3.0.6.~~4.1.7. At the end of the read, seal the plate with film to prevent evaporation. Leave the plate at room temperature overnight to allow the enzyme to cut the substrate completely.

Commented [A11]: What does this mean?

~~3.0.6.~~4.1.8. After ~24 h, remove the sealing film and perform an e~~E~~ndpoint read of the plate using the same PMT as in the prior plates. These emission ratios should be averaged and inputted into **Table 3** under "CUT~~ut~~". The cleavage of the substrate can be confirmed using the SDS-PAGE discontinuous assay described below (Step 5.1.~~34~~).

Commented [A12]: Please specify the specific steps.

Formatted: Highlight

~~3.0.6.~~4.1.9. Export the data to Excel. The fluorescence units at each time point should be outputted for the 2 wavelengths (**Table 4**).

~~3.0.6.~~4.1.10. ~~To c~~Calculate the nmols ~~of product~~ substrate that haves been cut at time t usinge equation (1) where X is the emission ratio (527 nm/470 nm) at a given time point, *neg* is the emission ratio of the "UNCUT~~ncut~~" substrate at t=0, and *pos* is the emission ratio of the completely "CUT~~ut~~" substrate measured after 24 h of cutting (**Table 3**).

$$\frac{X-neg}{(pos-neg)} * nmol = nmols\ of\ S\ cut\ at\ time\ t \qquad (1)$$

NOTE: ~~The nmols of substrate cut at time = t can be obtained from equation 1 and the values in Table 3.~~ Representative fluorescence data arei~~s~~ shown for one well (well E7) containing 80 µM substrate (4 nmols of S per well) ~~is~~ in **Table 4**. The calculations were performed for each well in the plate.

~~3.0.6.~~4.1.11. For each well plot nmols vs. time (min) and~~to~~ obtain the initial velocities (slopes) by fitting the data to y = mx + b~~. This should be done. This can be done in~~ using GraFit (Erithacus Software Limited). For the data collected in 4.1.5. plot nmols vs. time (min) for ~~the Spontaneous Hydrolysis data as well~~each well. The slope will equal the nmols product produced per minute. Subtract t~~T~~he spontaneous rates of hydrolysis measured in 4.1.5. ~~can be subtracted~~ from the

enzyme-catalyzed reaction rates (**Table 5**).

NOTE: The first read can be clipped from the data if it is artifactually high due to movement of the plate into the plate reader.

~~3.0.6.~~4.1.12. Calculate the amount of enzyme in mg that was added to each well (e.g., 0.0009 mg). A unit is defined as a µmol of product produced per minute (µmol/min). Divide the nmol/min by the mg of enzyme present in the well to obtain mU/mg; divide by 1000 to obtain U/mg.

~~3.0.6.~~4.1.13. Plot [S] µM on the x-axis and U/mg on the y-axis and fit the data to the Michaelis-Menten equation to obtain V_{max} and K_m. This can be done in GraFit (Erithacus Software Limited).

5. Assaying the Enzyme Discontinuously Using SDS-PAGE Analysis

~~4.0.~~5.1. Prepare a 50 µL reaction containing 10 µM substrate and buffer in place of enzyme label as "UNCUT.".

NOTE: The volumes of substrate and buffer are shown in **Table 2**. If the continuous assay has been run, the samples can be used directly from the 96-well plate.

~~4.0.0.~~5.1.1. Prepare a 50 µL reaction containing 10 µM substrate and 5 µL enzyme and label as "CUT". Start the timer when the enzyme is added to the substrate.

NOTE: Inhibitors can be added to additional tubes containing enzyme and substrate. Adjust the volume of added buffer to compensate for the added volume of inhibitor. Concentrations of DMSO should not exceed 2%.

5.1.2. Incubate the reactions for ~15-24 h at room temperature (22 ± 3 °C). Stop the reactions by adding 50 µL 2x Laemelli buffer. After stopping the reaction boil each tube for 3-10 min.

~~4.0.0.~~5.1.3. Assemble the gel tank according to the manufacturer's directions. Insert~~and place~~ well pre-cast 12% polyacrylamide gel cassette ~~into the gel tank, place~~and a buffer dam on the other side. Fill the interior reservoir of the cell with 1x SDS running buffer until the buffer reaches the top of the cassette. Fill the external reservoir half-full with the same buffer.

~~4.0.0.~~5.1.4. To analyze cleavage using the discontinuous assay load 5 µL of each reaction a lane of~~in~~ a ~~12%~~ SDS-PAGE gel beginning with the "UNCUT" reaction. Include a molecular weight marker in the first or last lane.

~~4.0.0.~~5.1.5. Attach the electrodes of the ~~power supply to the lid of the~~ gel tank to the power and sSeparate the products at 110 V for 60 min. Remove the gel from the cassette by inserting the cracking tool in between the plates. Place the gel in a plastic tray and submerge the gel in 5-10 mL of gel sStaining solution;~~:~~ ~~the gel until~~ bBbands will be~~are~~ visible within ~~(>~~30 min~~)~~. After 1-24 hours rRemove the excess stain, ~~and place~~ submerge the gel in water and use a gel ~~-~~imager to take a picture of the gel.

6. Docking substrate peptides to the VEEV-nsP2 cysteine protease

6.1. Download the coordinate file for the VEEV cysteine protease from the PDB (https://www.rcsb.org/). The PDB code is 2HWK. Save the file as 2HWK.pdb.

6.1.1. ~~6.2~~ Prepare the protein structure using MOE (https://www.chemcomp.com/). Load the protein PDB file into MOE. Click the "Select" and "Solvent" on the right hand side bar and delete the solvent.

6.1.2. Open the "Structure Preparation" panel from the top menu bar "Protein". Automatically correct all structural items by clicking on "Correct" and protonate the structure by clicking on "Protonate3D". Add partial charges to the protein by opening "Partial Charges" panel and selecting "Amber 99" and "Adjust hydrogens and Lone pairs as required". Finally, save the structure file as "2HWK_dock.pdb".

~~Prepare the protein structure using MOE (https://www.chemcomp.com/). Remove the solvent and correct the catalytic residue Cys-477 (change CSO to CYS and delete the alternative conformation). Protonate the structure and save the coordinates using new name for the pdb file.~~

~~6.2.1 Generate the coordinate (.pdb) files for the substrate peptides (nsP12, nsP23, nsP34) and TRIM14 using MOE. Minimize the generated peptide structures in an unfolded conformation (**Figure 6**).~~

6.2. ~~6.3~~ Build the structure for the substrate peptides (nsP12, nsP23, nsP34) and TRIM14 using MOE. Open the Protein Builder panel, enter the substrate sequence, set the "Geometry" as "Extended", and click on~~the~~ "Build". The structure will be shown in MOE window.

6.2.1. Minimize the peptide structure by clicking "Minimize" on the panel. Save the structure as a PDB file (**Figure 6**). ~~Dock the substrate peptides to the VEEV-nsP2 protein target using PyRx/AutoDock 4.2 (http://autodock.scripps.edu/) (**Figure 7**). The amide bond of the substrate peptide should be constrained in the docking search.~~

6.3. Dock the substrate peptides to VEEV-nsP2 using PyRx/AutoDock 4.2 (http://autodock.scripps.edu/). Open the PyRx Tool, edit the preference setting, inactivate all torsions for "Ligand Preparation". Load the substrate molecule, right click the molecule name on the Navigator panel, select "Make ligand" to prepare the ligand docking file. Load the protein 2HWK_clean.pdb, select "Make macromolecule" to prepare the pdbqt docking file (**Figure 7**).

~~6.3.1 Define the binding pocket centered at the catalytic residue Cys-477. Run AutoGrid to generate grid maps. Run AutoDock using the Lamarckian Genetic Algorithm (LGA). Generate 50 docking poses.~~

6.3.1. ~~Inspect and select the best binding model based on 1) key interactions between the~~

Commented [A13]: Please write the text in the imperative tense.

Commented [A14]: With what and how?

Commented [A15]: What is done after this?

Commented [A16]: Software steps must be more explicitly explained ('click', 'select', etc.). Please add more specific details (e.g., button clicks or menu selections for software actions, numerical values for settings, etc.).

Commented [A17]: Please describe how.

Commented [A18]: How?

Formatted: List Paragraph

Commented [A19]: How?

Commented [A20]: How?

Commented [A21]: How?

Formatted: No bullets or numbering

Formatted: List Paragraph

~~substrate and Cys-477 at the cleavage site, 2) the predicted binding energy.~~Start the "AutoDock Wizard" on the docking panel at the bottom. Select the prepared ligand and protein files. Define the protein binding pocket by manually adjusting the grid dimension which is centered at the catalytic residue Cys-477. Using the default spacing parameter 0.375 Å. Click on "Run AutoGrid" to generate grid maps.

6.3.2. Run AutoDock and ~~.-s~~Select the Lamarckian Genetic Algorithm (LGA) method. Click on the "Docking Parameters" and set the "Number of GA runs" to 50. Use the default parameters for others. Click on "Forward" to start the docking run.

6.3.3. ~~-~~Open the "Analyze Results" panel. Inspect all predicted binding poses. Select the best model with the lowest predicted binding energy and reasonable binding interactions between the Cys-477 and substrate on the cleavage site. Save the binding model as PDB file for further MD simulations.

~~6.3.2~~

7. MD simulations of D~~d~~ocked VEEV-substrate C~~c~~omplex~~e~~s

7.1. Prepare the input files using Amber (http://ambermd.org/). Following the standard protocol, MD simulations are performed for the predicted substrate binding models using the AMBER package and the ff99SB force field.

7.1.1. The solvated systems are subjected to a thorough energy minimization prior to MD simulations. Periodic boundary conditions are applied to simulate a continuous system. The particle mesh Ewald (PME) method was employed to calculate the long-range electrostatic interactions. The simulated system was first subjected to a gradual temperature increase from 0 K to 300 K over 100 ps, and then equilibrated for 500 ps at 300 K, followed by production runs of 2-ns length in total.

7.1.2~~3~~. Run the simulation job at a high performance computing facility, our simulations were run on the Biowulf cluster (https://hpc.nih.gov/) (**Figure 8**).

7.1.3~~4~~. Visualize the trajectory output using the VMD program (www.ks.uiuc.edu/Research/vmd/). Analyze the binding interactions and conformational changes of the substrates and TRIM14 within the active site of nsP2 (**Figure 9**).

~~7.1. Following the standard protocol, MD simulations wereas performed to refine the predicted substrate binding models using the AMBER package and the ff99SB force field (http://ambermd.org/).~~

~~7.1.1 Prepare the topology and coordinate input files using the program tleap in AMBER, which solvate the protein complex, add ions (Cl-), build and save the AMBER parameter files.~~

~~7.1.2 Prepare job scripts for MD simulations. The job consists of the following steps: 1) energy minimization of the water molecules while holding the solute frozen (1,000 steps using the steepest descent algorithm), followed by 5,000 steps of conjugate gradient minimization of the whole system; 2) heat up the system to a gradual temperature increased from 0 K to 300 K over 100 ps; 3) Equilibrate the system for 500 ps at 300 K; 4) production run for 10 ns.~~

~~7.1.3 Submit the simulation job to the NIH high performance computing Biowulf Cluster (https://hpc.nih.gov/) (**Figure 8**).~~

~~7.2 Visualize the trajectory output using the VMD program (www.ks.uiuc.edu/Research/vmd/). Analyze the binding interactions and conformational changes of the substrates and TRIM14 within the active site of nsP2 (**Figure 9**).~~

~~7.1.1 Prepare the input files using Amber (http://ambermd.org/). Following the standard protocol, MD simulations are performed for the predicted substrate binding models using the AMBER package and the ff99SB force field.~~

~~7.1.2 The solvated systems are subjected to a thorough energy minimization prior to MD simulations. Periodic boundary conditions are applied to simulate a continuous system. The particle mesh Ewald (PME) method was employed to calculate the long-range electrostatic interactions. The simulated system was first subjected to a gradual temperature increase from 0 K to 300 K over 100 ps, and then equilibrated for 500 ps at 300 K, followed by production runs of 2-ns length in total.~~

~~7.1.3 Run the simulation job at a high performance computing facility, our simulations were run on the Biowulf cluster (https://hpc.nih.gov/) (**Figure 8**).~~

~~7.1.4 Visualize the trajectory output using the VMD program (www.ks.uiuc.edu/Research/vmd/). Analyze the binding interactions and conformational changes of the substrates and TRIM14 within the active site of nsP2 (**Figure 9**).~~

REPRESENTATIVE RESULTS:

SSHHPS analysis of the ZIKV ns2B/3 protease identified 4 host protein targets: FOXG1, SFRP1, a G_s alpha subunit from a retinal cDNA library, and the NT5M mitochondrial 5',3'-nucleotidase (**Figure 10**).[6] Notably, no other method predicted these proteins as potential targets of the ZIKV protease. Mutations in the FOXG1 gene have been linked to a congenital syndrome characterized by impaired development and structural brain abnormalities such as microcephaly. SFRP1 is an s~~S~~ecreted frizzled-related protein~~s~~ (SFRP); these ~~are~~ soluble receptors ~~that~~ can competitively bind Wnt ligands ~~(competitively)~~ to antagonize and inhibit Wnt signaling. ~~SFRP1 is also involved in T-cell differentiation[38].~~ The Wnt signaling pathway is ~~was previously~~involved ~~linked~~ ~~into~~ the regulation of the IFN response during Flavivirus infection~~; microRNAs that repressed Wnt/β-catenin signaling had strong anti-flaviviral effects~~.[39]. T~~hus, t~~he ~~proteolytic~~ cleavage of SFRP1 would be expected to enhance flavivirus replication. SFRP1 is also involved in Th17-cell differentiation[38]. Sequence alignments of the SSHHPS showed species-specific differences in the cleavage site sequences (**Figure 10D**). The cleavage site sequence in SFRP1 was identical in humans and chickens; ZIKV can induce mortality and microcephaly in chicken embryos [40]. In

Formatted: Indent: Left: 0.5", No bullets or numbering

Formatted: No bullets or numbering

Formatted: Font: 12 pt

Formatted: Indent: Left: 0", First line: 0", Outline numbered + Level: 2 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Indent: Left: 0", First line: 0", Outline numbered + Level: 3 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0" + Indent at: 0.5"

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Commented [A22]: Please revise these lines to avoid previously published text.

Commented [A23]: Please revise these lines to avoid previously published text.

rodents, the highly conserved P1 residue (K/R)R↓G is substituted by a glycine (RGG). Immunocompetent strains of mice are generally resistant to ZIKV infection and disease [41].

Steady state kinetic parameters and inhibition constants can be measured for the viral polyprotein sequences and for the host protein sequences using the continuous assay in a plate reader [33,42,43] (**Figure 11A**). For qualitative cleavage information, such as cleavage of a particular sequence or the inhibition of the protease by various compounds, the discontinuous assay can be used (**Figure 11B**).

Optimization of the number of residues in between CFP and YFP may be required. A substrate-bound model can be made using the *in silico* methods. A representative docked model of the nsP1/nsP2 junction is shown in **Figure 9**. For the VEEV nsP2 protease, cleavage of a 12-amino acid Semliki Forest Virus (SFV) sequence had been reported (K_m = 0.58 mM [35]). Lengthening the substrate sequence to 19, 22, and 25 residues and reducing the ionic strength of the buffer led to a significant reduction in K_m. Examination of the VEEV nsP2 crystal structure and crystal packing also showed that a portion of one of the junctions was packed against the protease domain and was helical. Thus, the longer VEEV substrates may bind better due to the recognition of a secondary structural motif.

For TRIM14, we obtained a K_m = 21 μM [6,35]. The K_m for the substrate carrying the host protein sequence was comparable to the K_m values of the substrates containing the viral polyprotein cleavage site sequences (K_m(V12) = 12 μM and K_m(V34) = 21 μM). The cleavage site sequences at the nsP1/nsP2, nsP2/nsP3, and nsP3/nsP4 junctions were cut with different efficiencies. In the cell, this is thought to allow for sequential cleavage of the polyprotein [44].

Caution should be taken in interpreting negative results. If no cleavage occurs, the cleavage site may be too short, or the purified protease may be inactive. For substrates that are cut, additional experiments are needed to confirm cleavage of the full length protein or cleavage in virus-infected cells. Appropriate follow-on experiments should be chosen. The effects of overexpression or silencing of the target protein on viral replication ~~can~~ also can be tested.

FIGURE AND TABLE LEGENDS:

Figure 1. Three mechanisms of silencing. Silencing can occur at the level of DNA, RNA, or protein. These "search and delete" algorithms each use a "keyword" to direct the cleavage of a file containing the word. This figure has been modified from ~~-~~Morazzani, et al.[34] and the references therein.

Figure 2. Species ~~-~~specific differences in cleavage site sequences. The C-terminal PRY/SPRY domains of TRIM14 homologues are shown in the alignment. The PRY/SPRY domain can be identified by the conserved motifs highlighted in gray. Human TRIM14 is cut at QEAGA↓G by the VEEV nsP2 cysteine protease. The SSHHP sequence is shown in color. The residue in green is the P1′ residue; in blue is the P4 residue, and in red are other conserved residues within the cleavage site motif sequence. Equine harbor a truncated version of TRIM14 lacking the PRY/SPRY domain. The lysine highlighted in cyan is poly-ubiquitinated and is important for the assembly of the MAVS signalosome. The C-terminal PRY/SPRY domain may be transiently cut by the nsP2 protease to impair the host's antiviral response intracellularly during an acute viral infection. In equine, this domain is always absent. This suggests that the PRY/SPRY domain of TRIM14 may have a protective function against VEEV infections. This figure has been reproduced from Morazanni, et al.[6] ~~The TRIM14 protein sequences from different species are aligned. The C-terminal PRY/SPRY domain is shown. Human TRIM14 is cut at QEAGA↓G by the VEEV nsP2 cysteine protease. Equine harbor a truncated version of TRIM14 lacking the PRY/SPRY domain. The lysine highlighted in cyan is poly-ubiquitinated and this is important for the assembly of the MAVS signalosome. The C-terminal PRY/SPRY domain may be transiently cut by the nsP2 protease to impair the host's antiviral response intracellularly during an acute viral infection. In equine, this domain is always absent. This suggests that the PRY/SPRY domain of TRIM14 may have a protective function against VEEV infections.~~

Figure 3. SSHHPS identification using BLAST. The cleavage site motif sequence at the VEEV nsP1/nsP2 junction is aligned with the SSHHP sequence in the host protein TRIM14. The residue colored in green is the P1′ residue; in blue is the P4 residue and in red are other conserved residues of the cleavage site motif sequence. Most alignments contained homology to regions outside of the conserved cleavage site motif or did not include the P1/P1′ scissile bond residues. TRIM14 showed a match to 6 residues in sequential order that included P1 and P1′.

Figure 4. Protein and DNA sequences of the CFP-V12-YFP substrate for the VEEV nsP2 cysteine protease. The NdeI (CATATG) and XhoI (CTCGAG) restriction sites are shown in capital letters. In red is the cleavage site sequence from the viral polyprotein that is in between nsP1 and nsP2. The residue in green is the P1′ residue and in blue is the P4 residue of the cleavage site. ~~The NdeI (CATATG) and XhoI (CTCGAG) are shown in capital letters. In red is the cleavage site sequence motif from the viral polyprotein that is in between nsP1 and nsP2.~~

Figure 5. Protein sequence of the Trx-VEEV-nsP2 cysteine protease construct. Thioredoxin (Trx) is shown in yellow. The thrombin cleavage site and His-tag are shown in cyan. The Cys-His dyad are labeled in red.

Figure 6. Peptide structures in MOE.

Figure 7. Docking of substrate peptide using PyRx/AutoDock.

Figure 8. Jobs running on the Biowulf cluster.

Figure 9. Model of the VEEV P12 substrate containing the cleavage site sequence at the nsP1/nsP2 junction. The Cys-477/His-546 catalytic dyad is shown in blue. Figure was made using Pymol (https://pymol.org).

Figure 10. SSHHPS Analysis of the Zika virus ns2B/ns3 protease. (A) Predicted host protein targets of the ZIKV ns2B/ns3 protease. Residues in red match a single cleavage site sequence.

Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Font: 12 pt
Formatted: Font: 12 pt
Formatted: Font: Not Bold

Residues in green are tolerated at the subsite and match residues at other cleavage sites. SFRP1 had the highest number of identical residues in consecutive order. **(B)** CFP-substrate-YFP proteins (~50-60 kDa) were expressed and purified containing the predicted SSHHP sequence from each host protein (human). The ZIKV protease cut human FOXG1, SFRP1, NT5M and a G_{α}alpha subunit isolated from a retinal cDNA library. The cleavage products are approximately 28-30 kDa. The substrate sequences are available in Morazzani, et al.[6] **(C)** While the ns2B/ns3 protease cut SFRP1, it did not cut its homologues (SFRP2 and SFRP4). **(D)** Alignment of the cleavage sites from different animal species may be useful in selecting an animal model for a Group IV virus. Note that the conserved R↓G sequence differs between humans and rodents in SFRP1. Figure reproduced from Morazzani, et al.[6]~~A) Predicted host protein targets of the ZIKV ns2B/ns3 protease. Residues in red match a single cleavage site sequence. Residues in green are tolerated at the subsite and match residues at other cleavage sites. SFRP1 had the highest number of identical residues in consecutive order. (B) CFP-substrate-YFP proteins were expressed and purified and contained the predicted SSHHPS from each host protein (human). The ZIKV protease cut FOXG1, SFRP1, NT5M and a G$_{\alpha}$alpha subunit from a retinal cDNA library. (C) While the ns2B/ns3 protease cut SFRP1, it did not cut its homologues (SFRP2 and SFRP4). (D) Alignment of the cleavage sites from different animal species may be useful in selecting an animal model for a Group IV virus. Note that the conserved R↓G sequence differs between humans and rodents in SFRP1.~~

Figure 11. Steady state kinetic analysis using the continuous and discontinuous assays. (A) The kinetic data shown in Table 5 was plotted in GraFit. The inset shows the Lineweaver-Burk plot. **(B)** SDS-PAGE gel showing the cleavage products of the CFP-V12-YFP substrate. In lane 1 is the "UNCUT" substrate (48 kDa). In lane 2 is the "CUT" substrate (31 kDa and 27 kDa). In lanes 3-9 different compounds were included to test their inhibitory activity. Lane 4 contains the E64d covalent inhibitor. These reactions were run overnight for ~17 h at room temperature. Boiling of the samples was required to achieve the sharp banding pattern. The nsP2 protease is visible (56 kDa) in the reactions containing enzyme, but not in lane 1. Lane 1 is the "no enzyme" control.

DISCUSSION:

Sequence-specific destruction of a protein or a nucleic acid guided by a foreign sequence is only seen in a few cases in biology. The mechanisms shown in **Figure 1** are defensive mechanisms that protect a host from a virus, or a virus from a host.

Using bioinformatic methods we can identify the targets that are destroyed by these systems. In our analyses of SSHHP sequences, we ~~found~~ discovered that many of them could be found in proteins needed to generate innate immune responses. Some had obvious roles such as MAVS and TRIF (TIR-domain-containing adapter-inducing interferon-β), while others were related to immunity though more complex mechanisms (e.g., Histone H3, SFRP1, FOXG1)[8,9]. The target information stored in the SSHHP sequence has the potential to identify pathways that have antiviral effects against these viruses. Antiviral responses *in vivo* are often virus-specific [23,28]. For example, subsets of TRIM proteins have antiviral effects on different viruses [22,23,45], some are viral restriction factors (e.g., HIV and TRIM5α). The specificity of TRIM proteins (~70 have been identified) ~~is~~ still is being examined [22,45]. The information within ~~these sequences~~SSHHPS may contribute to our understanding of how these viruses evade the innate immune responses. Other patterns and correlations may be uncovered as more SSHHPS are examined.

Species-specific differences were apparent in our analyses (**Figures 2, 10**). These viruses are known to affect some species more than others. Information about host range, host susceptibility, and host defenses may be present within SSHHPS. For example, equine, the most susceptible species to equine encephalitis viruses, lacked the region of human TRIM14 that was transiently cut by the VEEV nsP2 protease. Humans rarely die from VEEV infections but can be infected[26]. The human TRIM14 protein carried an nsP2 protease cleavage sequence.[6] The presence of the cleavage site suggest that humans have a defense mechanism against these viruses. Birds have been thought to be potential reservoirs of these viruses[46]. The corresponding SSHHP sequence in the TRIM14 protein from chickens differed from the sequences found in humans and other species. Subtle differences like these may make a target host protein uncleavable or more readily cleaved. Aguirre, et al.[16] showed that an uncleavable mutated STRING protein induced higher levels of IFN after Dengue virus infection and that mice naturally carry a version of STING that is not cut by the Dengue ns2B3 protease. The murine STING protein ~~was also~~ was not cut by the ZIKV protease[47]. In our SSHHPS analysis, we also observed differences in the ZIKV protease cleavage site sequences when we compared the human proteins with those of rodents [6] (**Figure 10D**). Reproducing the species-specific proteolytic cleavages of host proteins may be important in animal models used for Group IV viruses. The inhibition of host protein cleavage also has implications with regards to the development of Group IV protease inhibitors. In our previous publication, we showed that we could inhibit TRIM14 cleavage by the VEEV nsP2 protease using CA074 methyl ester [6]. This result suggests that small molecule inhibitors of these proteases may be able to modulate the innate immune responses that are capable of suppressing the infection.[6,33]

Genetic variation within a species also has the potential to produce differences in proteolytic cleavage. Subtle differences in codon usage could affect ribosome pausing [48]. Since some Group IV viral proteases are embedded in the ER membrane, differences in these pauses could affect cleavage of a target if cleavage occurs co-translationally. Some of the cleavage sites that we identified were in predicted signal peptide sequences (e.g., SFRP1) while others were internal.

SSHHPS analysis can produce information that differs from other methods of host protein analyses. SSHHPS analysis was inexpensive and easy to employ. The use of a bacterial expression system allowed testing of short segments (~25 amino acids) of mammalian sequences without the use of mammalian cell culture. We found that the CFP-YFP substrates were able to tolerate all of the tested human protein sequences; however, yields varied. In similar assays, substrates containing human protein sequences as long as 63 amino acids were successfully expressed, purified, and utilized for kinetic analyses and inhibitor screening [49-51]. Since only small amounts of the substrate are needed for the discontinuous assay, a large number of targets can be explored. One advantage of the system is that the CFP/YFP substrates can be used for SDS-PAGE analyses and for more elaborate kinetic analyses (*i.e.* IC_{50}, K_i, K_m, V_{max}). For drug discovery, ~~i~~Inhibitory compounds can produce artifacts in fluorescent assays. ~~T~~thus, the discontinuous

Formatted: Font: Not Bold
Formatted: Font: Not Bold, Subscript
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Font: Not Bold
Formatted: Font: Not Bold

assay in combination with continuous assay allows one to confirm cleavage or inhibition of cleavage. The samples for the discontinuous SDS-PAGE assay can be taken directly out of the 96-well plates. CFP/YFP substrates have been used for compound library screening [52]. However, additional analyses are required to determine if a substrate is suitable for high throughput screening such as the calculation of a Z-factor [53].

One challenge in designing a substrate is identifying the region around the scissile bond that is bound and recognized by the protease. In the examples shown here, we began with 12 residue sequences that were centered around the scissile bond. After analyzing sequence alignments of the cleavage sites homology to the residues N-terminal of the scissile bond was found for the VEEV protease, whereas for the ZIKV protease homology to several of the C-terminal residues was found. An *in silico* model of the docked substrate can be used to design site-directed mutagenesis experiments that probe the binding sites of the substrate. Since the substrate and enzyme sequences are on plasmids, either can be mutated to test the *in silico* models or subsite tolerances. This can be advantageous if a crystal structure of the bound substrate(s) is not available.

SSHHPS analysis may also yield new information about the mechanisms by which virus-induced phenotypes are produced by viral enzymes. One of the ZIKV targets, SFRP1, is part of the Wnt signaling pathway and has roles in both brain and eye development and in immune responses [38,39,54-57]. We found that the other protein sequences that could be cut by the ZIKV ns2B/ns3 protease were also in proteins involved in brain and eye development; abnormalities in both have been observed in congenital Zika syndrome and are thought to be part of the virus-induced phenotype [58].

The predictability of host-pathogen interactions could be exploited for a variety of applications: target-specific oncolytic viral therapies; de-risking live virus vaccines; refinement, prediction or selection of animal models; prediction of host-range or susceptibility; prediction of zoonotic events; and prediction of host-defenses. Since the methods described are sequence-based, they may be of value to incorporate into software in the future.

ACKNOWLEDGMENTS:

This work was supported by Defense Threat Reduction Agency (DTRA) project numbers CB-SEED-SEED09-2-0061 and CBCall4-CBM-05-2-0019.

DISCLOSURES:

The opinions expressed here are those of the authors and do not represent those of the U. S. Navy, U.S. Army, U. S. Department of Defense, or the U. S. government. ~~The authors have nothing to disclose.~~

REFERENCES:

Commented [A24]: References: Please do not abbreviate journal titles; use full journal name.

1 Liu, H. *et al.* Widespread Horizontal Gene Transfer from Double-Stranded RNA Viruses to Eukaryotic Nuclear Genomes. *Journal of Virology.* **84** (22), 11876-11887, (2010).

2 Hagai, T., Azia, A., Babu, M. M. & Andino, R. Use of host-like peptide motifs in viral proteins is a prevalent strategy in host-virus interactions. *Cell Reports.* **7** (5), 1729-1739, (2014).

3 Gorbalenya, A. E. Host-related sequences in RNA viral genomes. *Seminars in Virology.* **3** 359-371, (1992).

4 Shmakov, S. A. *et al.* The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes. *MBio.* **8** (5), 1-18, (2017).

5 Legler, P. M., Morazzani, E., Glass, P.J., Compton, J.R. Proteome Editing System and A Biomarker of Veev Infection. United States patent (2018).

6 Morazzani, E. M. *et al.* Proteolytic cleavage of host proteins by the Group IV viral proteases of Venezuelan equine encephalitis virus and Zika virus. *Antiviral Research.* **164** 106-122, (2019).

7 Alvarez, E., Castello, A., Menendez-Arias, L. & Carrasco, L. HIV protease cleaves poly(A)-binding protein. *Biochemical Journal.* **396** (2), 219-226, (2006).

8 Falk, M. M. *et al.* Foot-and-mouth disease virus protease 3C induces specific proteolytic cleavage of host cell histone H3. *Journal of Virology.* **64** (2), 748-756, (1990).

9 Grigera, P. R. & Tisminetzky, S. G. Histone H3 modification in BHK cells infected with foot-and-mouth disease virus. *Virology.* **136** (1), 10-19, (1984).

10 Li, W., Ross-Smith, N., Proud, C. G. & Belsham, G. J. Cleavage of translation initiation factor 4AI (eIF4AI) but not eIF4AII by foot-and-mouth disease virus 3C protease: identification of the eIF4AI cleavage site. *FEBS Letters.* **507** (1), 1-5, (2001).

11 Kuyumcu-Martinez, M. *et al.* Calicivirus 3C-like proteinase inhibits cellular translation by cleavage of poly(A)-binding protein. *Journal of Virology.* **78** (15), 8172-8182, (2004).

12 Pietila, M. K., Hellstrom, K. & Ahola, T. Alphavirus polymerase and RNA replication. *Virus Research.* **234** 44-57, (2017).

13 Hardy, W. R. & Strauss, J. H. Processing the nonstructural polyproteins of sindbis virus: nonstructural proteinase is in the C-terminal half of nsP2 and functions both in cis and in trans. *Journal of Virology.* **63** (11), 4653-4664, (1989).

14 Strauss, E. G., De Groot, R. J., Levinson, R. & Strauss, J. H. Identification of the active site residues in the nsP2 proteinase of Sindbis virus. *Virology.* **191** (2), 932-940, (1992).

15 Wang, D. *et al.* Foot-and-mouth disease virus 3C protease cleaves NEMO to impair innate immune signaling. *Journal of Virology.* **86** (17), 9311-9322, (2012).

16 Aguirre, S. *et al.* DENV inhibits type I IFN production in infected cells by cleaving human STING. *PLoS Pathogens.* **8** (10), e1002934, (2012).

17 Barral, P. M., Sarkar, D., Fisher, P. B. & Racaniello, V. R. RIG-I is cleaved during picornavirus infection. *Virology.* **391** (2), 171-176, (2009).

18 Elbashir, S. M., Lendeckel, W. & Tuschl, T. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes & Development.* **15** (2), 188-200, (2001).

19 Deveau, H., Garneau, J. E. & Moineau, S. CRISPR/Cas system and its role in phage-bacteria interactions. *Annual Review of Microbiology.* **64** 475-493, (2010).

20 Schechter, I. & Berger, A. On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications.* **27** (2), 157-162, (1967).

21 Bieniasz, P. D. Intrinsic immunity: a front-line defense against viral attack. *Nature Immunology.* **5** (11), 1109-1115, (2004).

22 van Tol, S., Hage, A., Giraldo, M. I., Bharaj, P. & Rajsbaum, R. The TRIMendous Role of

TRIMs in Virus-Host Interactions. *Vaccines (Basel).* **5** (3), (2017).
23 Uchil, P. D., Quinlan, B. D., Chan, W. T., Luna, J. M. & Mothes, W. TRIM E3 ligases interfere with early and late stages of the retroviral life cycle. *PLoS Pathogens.* **4** (2), e16, (2008).
24 Zhou, Z. *et al.* TRIM14 is a mitochondrial adaptor that facilitates retinoic acid-inducible gene-I-like receptor-mediated innate immune response. *Proceedings of the National Academy of Sciences of the U S A.* **111** (2), E245-254, (2014).
25 Wang, S. *et al.* TRIM14 inhibits hepatitis C virus infection by SPRY domain-dependent targeted degradation of the viral NS5A protein. *Scientific Reports.* **6** 32336, (2016).
26 Zacks, M. A. & Paessler, S. Encephalitic alphaviruses. *Vet Microbiol.* **140** (3-4), 281-286, (2010).
27 Hollidge, B. S., Weiss, S. R. & Soldan, S. S. The role of interferon antagonist, non-structural proteins in the pathogenesis and emergence of arboviruses. *Viruses.* **3** (6), 629-658, (2011).
28 Carthagena, L. *et al.* Human TRIM gene expression in response to interferons. *PLoS One.* **4** (3), e4894, (2009).
29 Montgomery, S. A. & Johnston, R. E. Nuclear import and export of Venezuelan equine encephalitis virus nonstructural protein 2. *Journal of Virology.* **81** (19), 10268-10279, (2007).
30 Nenasheva, V. V. *et al.* Enhanced expression of trim14 gene suppressed Sindbis virus reproduction and modulated the transcription of a large number of genes of innate immunity. *Immunologic Research.* **62** (3), 255-262, (2015).
31 Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnology.* **20** (1), 87-90, (2002).
32 Li, M. Z. & Elledge, S. J. SLIC: a method for sequence- and ligation-independent cloning. *Methods in Molecular Biology*. **852** 51-59, (2012).
33 Hu, X. *et al.* Kinetic, Mutational, and Structural Studies of the Venezuelan Equine Encephalitis Virus Nonstructural Protein 2 Cysteine Protease. *Biochemistry.* **55** (21), 3007-3019, (2016).
34 Morazzani, E. M. *et al.* Proteolytic cleavage of host proteins by the Group IV viral proteases of Venezuelan equine encephalitis virus and Zika virus. *Antiviral Research.* **164** 106-122, (2019).
35 Zhang, D., Tozser, J. & Waugh, D. S. Molecular cloning, overproduction, purification and biochemical characterization of the p39 nsp2 protease domains encoded by three alphaviruses. *Protein Expression and Purification.* **64** (1), 89-97, (2009).
36 Lei, J. *et al.* Crystal structure of Zika virus NS2B-NS3 protease in complex with a boronate inhibitor. *Science.* **353** (6298), 503-505, (2016).
37 Shiryaev, S. A. *et al.* Characterization of the Zika virus two-component NS2B-NS3 protease and structure-assisted identification of allosteric small-molecule antagonists. *Antiviral Research.* **143** 218-229, (2017).
38 Lee, Y. S. *et al.* The Wnt inhibitor secreted Frizzled-Related Protein 1 (sFRP1) promotes human Th17 differentiation. *European Journal of Immunology.* **42** (10), 2564-2573, (2012).
39 Smith, J. L., Jeng, S., McWeeney, S. K. & Hirsch, A. J. A MicroRNA Screen Identifies the Wnt Signaling Pathway as a Regulator of the Interferon Response during Flavivirus Infection. *Journal of Virology.* **91** (8), (2017).
40 Goodfellow, F. T. *et al.* Zika Virus Induced Mortality and Microcephaly in Chicken Embryos. *Stem Cells and Development.* **25** (22), 1691-1697, (2016).
41 Morrison, T. E. & Diamond, M. S. Animal Models of Zika Virus Infection, Pathogenesis, and Immunity. *Journal of Virology.* **91** (8), (2017).
42 Morazzani, E. M. *et al.* in *Books of Abstracts, 254th American Chemical Society National Meeting, Washington, D.C.* BIOL-20, (2017).
43 Compton, J. R., Mickey, M. J., Hu, X., Marugan, J. J. & Legler, P. M. Mutation of Asn-475 in the Venezuelan Equine Encephalitis Virus nsP2 Cysteine Protease Leads to a Self-Inhibited State. *Biochemistry.* **56** (47), 6221-6230, (2017).
44 Vasiljeva, L. *et al.* Regulation of the sequential processing of Semliki Forest virus replicase polyprotein. *Journal of Biological Chemistry.* **278** (43), 41636-41645, (2003).
45 Ozato, K., Shin, D. M., Chang, T. H. & Morse, H. C., 3rd. TRIM family proteins and their emerging roles in innate immunity. *Nature Reviews Immunology.* **8** (11), 849-860, (2008).
46 Molaei, G. *et al.* Dynamics of Vector-Host Interactions in Avian Communities in Four Eastern Equine Encephalitis Virus Foci in the Northeastern U.S. *PLoS Neglected Tropical Diseases.* **10** (1), e0004347, (2016).
47 Ding, Q. *et al.* Species-specific disruption of STING-dependent antiviral cellular defenses by the Zika virus NS2B3 protease. *Proceedings of the National Academy of Sciences of the U S A.* **115** (27), E6310-E6318, (2018).
48 Angov, E., Legler, P. M. & Mease, R. M. Adjustment of codon usage frequencies by codon harmonization improves protein expression and folding. *Methods in Molecular Biology.* **705** 1-13, (2011).
49 Ruge, D. R. *et al.* Detection of six serotypes of botulinum neurotoxin using fluorogenic reporters. *Analytical Biochemistry.* **411** (2), 200-209, (2011).
50 Hu, X. *et al.* Structural insight into exosite binding and discovery of novel exosite inhibitors of botulinum neurotoxin serotype A through in silico screening. *Journal of Computer-Aided Molecular Design.* **28** (7), 765-778, (2014).
51 Dunning, F. M. *et al.* Detection of botulinum neurotoxin serotype A, B, and F proteolytic activity in complex matrices with picomolar to femtomolar sensitivity. *Applied and Environmental Microbiology.* **78** (21), 7687-7697, (2012).
52 Nguyen, T. G. *et al.* Development of fluorescent substrates and assays for the key autophagy-related cysteine protease enzyme, ATG4B. *Assay and Drug Development Technologies.* **12** (3), 176-189, (2014).
53 Zhang, J. H., Chung, T. D. & Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *Journal of Biomolecular Screening.* **4** (2), 67-73, (1999).
54 Bovolenta, P., Esteve, P., Ruiz, J. M., Cisneros, E. & Lopez-Rios, J. Beyond Wnt inhibition: new functions of secreted Frizzled-related proteins in development and disease. *Journal of Cell Science.* **121** (Pt 6), 737-746, (2008).
55 Esteve, P. *et al.* SFRPs act as negative modulators of ADAM10 to regulate retinal neurogenesis. *Nat Neurosci.* **14** (5), 562-569, (2011).
56 Garcia-Hoyos, M. *et al.* Evaluation of SFRP1 as a candidate for human retinal dystrophies. *Mol Vis.* **10** 426-431, (2004).

57 Marcos, S. *et al.* Secreted frizzled related proteins modulate pathfinding and fasciculation of mouse retina ganglion cell axons by direct and indirect mechanisms. *J Neurosci.* **35** (11), 4729-4740, (2015).
58 Moore, C. A. *et al.* Characterizing the Pattern of Anomalies in Congenital Zika Syndrome for Pediatric Clinicians. *JAMA Pediatrics.* **171** (3), 288-295, (2017).

CC: "Xin Hu" xin.hu@nih.gov, "Jaimee R. Compton" jaimee.compton@nrl.navy.mil

Dear Dr. Legler,

Your manuscript, JoVE60421 "Analysis of Group IV viral SSHHPS Using In vitro and In silico Methods," has been editorially and peer reviewed, and the following comments need to be addressed. Note that editorial comments address both requirements for video production and formatting of the article for publication. Please track the changes within the manuscript to identify all of the edits.

After revising and uploading your submission, please also upload a separate rebuttal document that addresses each of the editorial and peer review comments individually. Please submit each figure as a vector image file to ensure high resolution throughout production: (.psd, ai, .eps., .svg). Please ensure that the image is 1920 x 1080 pixels or 300 dpi. Additionally, please upload tables as .xlsx files.

Your revision is due by Aug 05, 2019.

To submit a revision, go to the JoVE submission site <http://www.editorialmanager.com/jove> and log in as an author. You will find your submission under the heading "Submission Needing Revision". Please note that the corresponding author in Editorial Manager refers to the point of contact during the review and production of the video article.

Best,

Peer Review,
Peer Review
JoVE <http://www.jove.com/>
617.674.1888
Follow us: Facebook <https://www.facebook.com/JOVEjournal> | Twitter <https://twitter.com/jovejournal> | LinkedIn <https://www.linkedin.com/company/312490>
About JoVE <http://www.jove.com/about>

_____

Editorial comments:

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues. The JoVE editor will not copy-edit your manuscript and any errors in the submitted revision may be present in the published version.

As requested, we have proof-read the manuscript.

2. Please obtain explicit copyright permission to reuse any figures from a previous publication. Explicit permission can be expressed in the form of a letter from the editor or a link to the editorial policy that allows re-prints. Please upload this information as a .doc or .docx file to your Editorial Manager account. The Figure must be cited appropriately in the Figure Legend, i.e. "This figure has been modified from [citation]."

As requested, we have received permission from Antiviral research (See Attached Email).

3. Please revise lines 103-105, 597-605, 650-651, 652-655 to avoid textual overlap with previously published work.

As requested, we have edited these sentences.

4. Please remove the embedded figure(s) from the manuscript. All figures should be uploaded separately to your Editorial Manager account.

As requested we have moved these.

5. Please remove the embedded table(s) from the manuscript. All tables should be uploaded separately to your Editorial Manager account in the form of an .xls or .xlsx file.

As requested we have moved these.

6. Please adjust the numbering of the Protocol to follow the JoVE Instructions for Authors. Step 1 followed by 1.1, followed by 1.1.1, etc. Each step should include 1−2 actions and contain 2−3 sentences. Use subheadings and substeps for clarity if there are discrete stages in the protocol. Please refrain from using bullets, dashes, or indentations.

As requested we have reformatted the methods.

7. JoVE cannot publish manuscripts containing commercial language. This includes trademark symbols (™), registered symbols (®), and company names before an instrument or reagent. Please remove all commercial language from your manuscript and use generic terms instead. All commercial products should be sufficiently referenced in the Table of Materials. You may use the generic term followed by "(Table of Materials)" to draw the readers' attention to specific commercial names. Examples of commercial sounding language in your manuscript are: BugBuster, Sepharose, GenScript Inc., etc.

As requested, we have removed these and changed the names.

8. Please revise the Protocol to contain only action items that direct the reader to do something (e.g., "Do this," "Ensure that," etc.). The actions should be described in the imperative tense in complete sentences wherever possible. Avoid usage of phrases such as "could be," "should be," and "would be" throughout the Protocol. Any text that cannot be written in the imperative tense may be added as a "NOTE." Please include all safety procedures and use of hoods, etc. However, notes should be used

sparingly and actions should be described in the imperative tense wherever possible. Please move the discussion about the protocol to the Discussion.

9. Lines 332-333: What happens after centrifugation? Discard the supernatant?
As requested, we have added step 2.2.7

10. Line 343: What is the molecular-weight cut-off of the dialysis bag?
We have added the MWCO (6-8 kDa).

11. Line 403: What volume of lysis buffer is used to resuspend the pellet?
Approximately 25 mL.

12. Please combine some of the shorter Protocol steps so that individual steps contain 2-3 actions and maximum of 4 sentences per step.
As requested, we combined several steps.

13. Please include single line spacing between each numbered step or note in the protocol.
As requested, we have added these.

14. After you have made all the recommended changes to your protocol section (listed above), please highlight in yellow up to 2.75 pages (no less than 1 page) of protocol text (including headers and spacing) to be featured in the video. Bear in mind the goal of the protocol and highlight the critical steps to be filmed. Our scriptwriters will derive the video script directly from the highlighted text.
As requested, we have highlighted sections.

15. Please highlight complete sentences (not parts of sentences). Please ensure that the highlighted steps form a cohesive narrative with a logical flow from one highlighted step to the next. The highlighted text must include at least one action that is written in the imperative voice per step. Notes cannot usually be filmed and should be excluded from the highlighting.

16. Please include all relevant details that are required to perform the step in the highlighting. For example: If step 2.5 is highlighted for filming and the details of how to perform the step are given in steps 2.5.1 and 2.5.2, then the sub-steps where the details are provided must be highlighted.

17. Figure 11: Please use the micro symbol µ instead of u. Please mark the fragment sizes in panel B.
As requested, we have corrected these.

18. Table of Materials: Please remove any ™/®/© symbols. Please sort the materials alphabetically by material name.
As requested, we have corrected these.

19. Please ensure that the references appear as the following: [Lastname, F.I., LastName, F.I., LastName, F.I. Article Title. Source. Volume (Issue), FirstPage – LastPage (YEAR).] For more than 6 authors, list only the first author then et al. Please do not abbreviate journal titles. See the example below:
Bedford, C.D., Harris, R.N., Howd, R.A., Goff, D.A., Koolpe, G.A. Quaternary salts of 2-[(hydroxyimino)methyl]imidazole. Journal of Medicinal Chemistry. 32 (2), 493-503 (1998).
We have checked the references.

Reviewers' comments:
Reviewer #1:
Manuscript Summary:
The manuscript describes a protocol for the in vitro and in silico studies on short stretches of homologous host-pathogen protein sequences. The method is respresented on the example of alphaviral proteases. The manuscript content is of interest to the readership of JoVE, and may be potentially adapted to study other viral proteases.

Major Concerns:
I do not have any major concern. The protocol seems to be very robust and carefully optimised.

Minor Concerns:
The minor concerns are listed below. The title is concise, the abstract is appropriate for this method article. Overall, the protocol manuscript is well written, I think language editing is not necessary, grammatical errors will be corrected by the help of the copy-editing service of JoVE.

We thank Reviewer #1 for reading and reviewing the manuscript. We greatly appreciate your comments and have made the requested revisions.

Comments regarding protocol:

Step 1.1 - For BLAST analysis, please provide examples for possible tools or specify the one which was used in the presented protocol.
As requested, we have added information about PHI-BLAST which is a tool within BLAST that can help refine the search.

Step 1.2 and 1.3 - Is there any automated method for filtering and rank-ordering? If yes, indicating those may aid to carry out the protocol.
Unfortunately, there is not. We are applying for funding to make a computer program that will do this using a BLAST output file as input.

Step 2 and 3
- Indicate the use of AKTA Prime Plus system in every case, show the applicable flow rates (mL/min) and the steps where different flow rate need to be applied.
We agree, we have added the flow rates.

- Amicon Ultra-15, -4, and 0.5 centrifugal filter units are listed in the tables of materials, but it is not indicated in the written protocol which types of units need to be applied in the different steps. Complete the protocol accordingly every case. However, three types of units are listed, the use of only two is written in the protocol. Complete the description of the given steps by the applicable temperature of buffer exchange.
We have removed the Amicon Ultra-4 unit. We thank the Reviewer for catching this.

Step 2.1 and 2.2 - To avoid misunderstandings, specify the plasmid names in every case, or complete table of materials and indicate in the Comments/Description that plasmids having pet15b core were used for CFP-V12-YFP substrate expression, while pet32 plasmid for VEEV nsP2 protease. Define plasmid names clearly and uniformly throughout the manuscript, e.g. pet15b (step 2.1) and CFP-V12-YFP (step 2.2) name is also used for the same plasmid.

We agree, we have added to the notes in the Materials table and have added a sentence to section 3.

Step 2.13 - Indicate temperature of dialysis.
We have added the temperature.

Step 4.2 - You suggest dilution of enzyme to 4.7 µM concentration. Is it necessary to determine the percentage of properly folded proteins which possess activity? Does the indicated concentration correspond to the total protein concentration or to the concentration of the active enzyme?
An active site titration could be done with a covalent inhibitor, however, we have found that the viral proteases have very different inhibitor specificities than the human enzymes. Most commercially available cysteine protease inhibitors are designed for human enzymes. Here we omitted the titration since it may or may not be possible to perform. We added a sentence to 4.1.2.

Step 4.11 - Use "room temperature" instead of "R.T".
As requested, we have changed the abbreviation.

Comments regarding tables:

Table 1.
- It should be noted in the table legend that the table has been prepared by the modification of the table published previously (Morazzani et al. 2019). Due to this, some references included in the table are not listed in the reference list of the manuscript (e.g. Barral et al., Das et al., etc), which may be disturbing for the readers. The table should be understandable alone, without the original article. I suggest the correction, for example indicating PMIDs may be acceptable.
As requested, we have made changes.

Table 2.
- I think that the Table should be subdivided to part A and B, because part A shows no reaction mixtures but reaction components.
- In the upper panel "substrate (V12)" and "enzyme (wt)" names are suggested, for use for clarity.
- In the upper panel showing mM (3rd column) is not necessary, mM is not shown neither in step 4.2 nor in 4.3.
- In the first column of the lower panel "[S]f" is shown. What does "f" mean? This label for substrate concentration is not used elsewhere. Is it a typo?
- It is unclear what kind of values are shown in the 2nd column of the lower panel. Amount of which reaction component is shown here? How these values have been calculated?
This was an Excel calculator, the top part shows where one can input the concentrations (in A280). We have edited the column labels to make them more clear.

Table 3.
- Why 5th row is highlighted by grey background? Whether this value is used in the presented calculations? Please explain and complete table legend accordingly.
The fifth row corresponds to the raw data in Table 4. We have expanded the legends.

Table 4.
- In the first row (title line), use "µ" symbol instead of "u" to show [S].
- Please, define "X" an "%" in the table. Write "emission ratio (527 nm/470 nm)" instead of using only "X", and define how percentage can be obtained.

As requested, we have added the information.

Table 5.
- In the first column, use "μ" symbol instead of "u" to show [S].
- In the second column (slope) what does the "Spont" mean? Please define the abbreviation in the table legend or in the text.
- In the second column (slope) what does the "w/Enz" mean?
As requested, we have expanded the legend.

Comments regarding figures:

Figure 1 and 2.
- Figure 1 and 2 are slightly modified versions of figures from Ref.3. This need to be indicated in the figure legends, e.g. as it is shown in case of figure 10 ("Figure reproduced from Morazzani et al. 3").
As requested, we have added the information.

Figure 2, 3 and 4.
- In these figures (and in lane 204-206, as well), the "QEAGAG" sequence is colored, but the color code is not defined in the figure legend. Please, complete figure legends in order to indicate why E and G residues are colored by blue and green, respectively. If coloring is not important, do not highlight residues.
As requested, we have edited the legends.

Figure 3.
- Lower panel showing "Most alignments" is hardly visible, the grey color need to be changed in order to increase contrast, black color is suggested.
As request, we have made it black.

Figure 6.
- Labels of residue 319 and 328 are not visible in the figure. Please indicate the sequence in the figure legend.
Unfortunately, the labels were added by the program. We were not able to move them.

Figure 9.
- I suggest to label names of all residues which are represented by sticks, both in the case of enzyme and substrate residues. For substrate, the substrate sequence needs to be shown at least in figure legend. Showing catalytic residue and cleavage position would also help understanding.
As requested, we have labeled the figure.

Figure 10.
- It is necessary to complete part B of the figure - showing SDS-PAGE gels - by labeling the molecular weights of the standard proteins.
- In figure part B sample labeling does not fits to the corresponding lanes (lane 693-694). Neither protein names nor the labels showing presence ("+") or absence ("-") of the protease are positioned properly, they are not above the corresponding lanes in every case. Spacing between the labels need to be checked.

These were done on 2 different gels so the MW markers are not shown. We have labeled the cut and uncut and have added the approximate MW's to the legend.

Figure 11.
- Indicate in figure legend sample number and whether error bars indicate SD or SEM (figure part A).
- Define "S.A." abbreviation in the column title.
- In figure legend change "Lineweaver-Burke" to "Lineweaver-Burk" (lane759 and 818).
- It is hard to identify bands based on molecular weights without labeling the molecular weight standards. It is necessary to complete figure part B - showing the SDS-PAGE gel - with the molecular weights of the marker ladder proteins.

As requested, we have added the labels to this figure and the MW markers.

Once again, we would like to thank Reviewer #1 for their careful read of the manuscript. Your suggestions were very helpful.

Reviewer #2:
Manuscript Summary:
The manuscript by Hu et al., describes a series of bioinformatic and in vitro approaches to identify and characterize the presence of SSHHPS. Briefly, the authors describe the use of degenerate BLAST searching to identify potential cleavage sites within host genes, and use comparative genomics to determine the level of conservation across multiple host species. They then describe the utilization of an in vitro cleavage system based on FRET to assess cleavage efficiency in a reductionist recombinant system. The protocols described, in my estimation, represent a complete description of the necessary steps to identify and test novel SSHHPs. On the whole, the manuscript is fairly well written, and the proposed methodology easy to understand and reproduce for other systems. The text in several areas seems to lack a cohesive focus, and could be improved upon.

Moreover, the methodology described represents a superficial identification of host / pathogen interactions, and the discussion does not adequately address the limitations of the assays.

I must unfortunately, recommend the rejection of this manuscript.

This is a methods paper for Morazzani, et al. Morazzani, et al was published in a peer reviewed journal in 2019. The Morazzani, et al manuscript contains the virology and the cleavage of the full length protein. The article is available online: https://www.ncbi.nlm.nih.gov/pubmed/30742841.

Table 1 has a list of reference citations (peer-reviewed journals) for 24 other cases of host protein cleavage by viral proteases. The references listed in Table 1 are virology papers. These papers are from other labs and were published over a 30 year period.

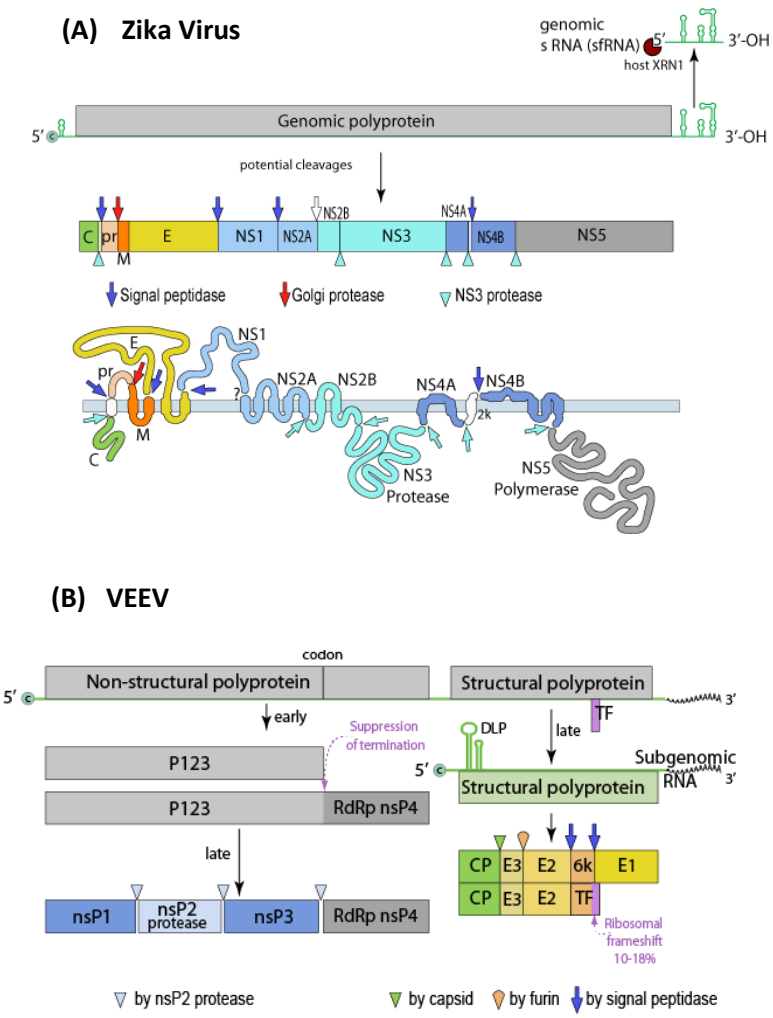
Major Concerns:

1) The cliche real estate saying "Location, location, location" applies here; however, a better example would be "Context, context, context". The methodology described in the paper will result in the identification of putative cleavage sites, and the use of the FRET method will determine their relative cleavage efficiency. But, the capacity of a host factor to be cleaved (and thus it's relative importance) is highly dependent on the context in which the target peptide is found. As such, there are several significant components of this that are not discussed in this

manuscript- What if the peptide is buried in the protein structure?, What if the protein and protease are in different subcellular compartments?, What if the protease is inactivated during viral replication?, What if rate limiting amounts of protease are generated during infection?

We have added sentences to the Intro.

This is virus-dependent. For flaviviruses the ns2B/ns3 protease is embedded in the ER and faces the cytosol; the signal peptidase faces the ER lumen. In the case of Zika virus, we found some of the cleavage sites in signal peptides suggesting that the proteolytic cleavage may alter the localization of the protein in the cell. It also suggests that cleavage may occur co-translationally (as the polypeptide is coming off of the ribosome). Several of the reasons you stated are why we and others perform cell-based assays using live virus after the protease assay. These assays determine if the cleavage can occur in the context of an infection and if the cleavage has any effect on viral replication. This manuscript is focused on the bioinformatic methods, protease assay, and *in silico* modeling. These methods are geared at delineating the cleavage motif sequence and the tolerances to substitutions.



For VEEV (an alphavirus) the protease is soluble and can be found in both the cytoplasm an in the nucleus (the nucleus is contiguous with the ER).

We found that most of the cleavages occurred within the 1st day of infection and appear to be most relevant to the early establishment of infection.

Without addressing / discussing these concerns the utility of the manuscript can be distilled to "identify sites that look like known viral cleavage motifs, clone them into a (admittedly clever) FRET system, and see if the viral protease can cleave a motif that looks like the one it should cleave".

If I was reviewing a manuscript that used this approach I would require the demonstration of cleavage in a native context, and would be unwilling to accept the described methodology at face value.

Cleavage of the natural sites in the polyprotein was published for VEEV (PMID 30742841, PMID: 19013248) and for Zika (PMID: 27386922, PMID: 28461069). Most of the viral cleavage site sequences are known. The native context may not be relevant to co-translational cleavage.

The bioinformatics and FRET assay are the first steps of a long process and simply determine if proteolytic cleavage of a given sequence is physically possible. Expression and purification of full-length mammalian proteins for protease assays is very time consuming (>7 days) and the stability of the cleavage products can be poor and difficult to detect in cells. The in vitro assays here can be done quickly and with many target protein sequences using *E.coli* (1.5 days). This method identifies the potential host-targets so that those proteins can be detected in Western blots or by mass spectrometry in cell-based assays.

We agree that these are good points to discuss. We have included more information in the Protocol and Discussion about how we optimized the substrates.

2) At certain points in the introduction statements are made that at face value are confusing. For instance the statement on lines 48-49 inferring that SSHHPs are vertically transferred lacks substantiation, and is contrary to the "molecular arms race" models put forth by others.

Micro RNA also follow the logic. They are stored keywords that direct the destruction of a foreign target. RISC is a keyword generator that can generate keywords on the fly as invading viral RNA enters the cell.

CRISPR/Cas9 and RNAi/RISC are strategies to protect a host from a virus, SSHHPS is just the opposite. It's a strategy to protect a virus from a host. The algorithm is the same (e.g. "search & delete" program using a keyword).

Here "vertically transferred" means that alphaviruses pass the information stored in the cleavage site sequence to their viral progeny (mother to child); we have revised this.

How horizontal transmission mechanisms occur is more complicated, they may be due to recombination events since innate immune response genes are expressed during an infection or they be due to coincidental assimilation (ref. Hagai, 2014).

Minor Concerns:
1) The introduction of other restriction pathways (ie. CRISPR), is distracting from the purpose of the manuscript.

These 3 mechanisms are all "search and delete" algorithms ("programs") that utilize a "keyword". The "keyword" is used to delete a larger target sequence (a "file" containing the keyword). SSHHPS/nsP follow the logic. These 3 mechanisms perform sequence-specific destruction of a larger target (they cut or "delete" a file). Silencing can therefore occur at the level of DNA, RNA, or protein. The protein mechanism we describe here follows the logic. We have revised Fig. 1.

2) There are grammatical errors in the document.

We agree and have corrected them. We thank Reviewer 2 for their suggestions and comments.

Reviewer #3:
Major Concerns:
The manuscript describes a platform for studying the proteolytic activity of an alphaviral protease, combining in vitro and in silico methods. The overall concept is well described and technical details are given with sufficient details in order to make the protocol useful. However, the manuscript would benefit from adding more representative results to demonstrate the capacity of the method to identify eg. host proteins cleaved by the viral protease.

Regarding the applicability of the methods, the authors state in the abstract and in the body text that the assays can be repurposed for drug discovery. It is true that in general, microplate-based fluorescent assays are more suitable for screening applications than more laborious techniques, but not all microplate-based bioassays can be converted into screening assays. The authors provide no data on assay robustness (eg. in terms of its signal window, Z' factor or other assay quality parameters) nor do they present any data on the use of the assay for small molecule inhibitor studies. If the authors wish to claim their method to be useful for drug discovery purposes they must add data to demonstrate this. Otherwise it is recommended to remove this statement.

We agree, and have added a sentence about the Z-value in the Discussion. Here we meant that one can use this assay to look at inhibition in a gel or with a plate reader (2 orthogonal methods). Typically, for drug discovery a primary and a secondary assay is used followed by a counterscreen. We have already used these assays for *in silico* drug discovery on a small scale (500 compounds were tested and Ki's were measured) (manuscript in preparation). A similar substrate has been used for large library screening by another group; BoTest is a CFP/YFP construct for botulinum neurotoxin zinc metalloprotease. BioSentinel did successfully use this type of substrate for library screening (https://biosentinelpharma.com/home/product_information). We cited their paper (Ruge, et al) and received advice from Dr. Ward Tucker. We also found another reference where a similar substrate had been used for screening.

Fig 11B is a qualitative small molecule inhibitor assay. One can measure Ki by measuring Km and Vmax in the presence of varying concentrations of inhibitor. Here, we did not have enough space (there's a 10 page limit) to explain the Ki measurements and calculations, we may be able to write another methods paper on just inhibition kinetics later. This paper is primarily focused on how to identify a host protein substrate (as opposed to a viral protease inhibitor).

In addition to the various figures and tables related to the methodology, the manuscript contains an introductory figure (Fig 1) which has little value for the paper. To keep the manuscript more focused, this figure could be omitted.
We have revised Fig 1 to focus on the logic of the algorithms.

We thank Reviewer #3 for their helpful advice and comments.

Reviewer #4:
Manuscript Summary:
The article describes how SSHHPS can be investigated, with specific protocols for in vitro and in silico assays.

Major Concerns:
The protocols have sufficient detail, but are not presented in terms of a clearly stated background and research-related objectives. There needs to be more careful attention to citation. It is not clear about how well the protocol works across a range of scenarios, but enough evidence is given to show typical function. The lack of synthesis in the discussion and the haphazard aspects of the introduction mean that the article does not live up to its title as describing an effective method of analysis.

Abstract
Line 41: Drug discovery is mentioned, but indication of what a "repurposing" could be is needed, otherwise omit from abstract, because this appears to be too tangential/unfounded, and this point is not well-described in the rest of the manuscript (it appears in the discussion, but I do not understand what the authors are saying there).
As requested, we have revised it.

Introduction
The first two paragraphs are haphazard.

Lines 46-50 (paragraph 1): the first paragraph needs to be rewritten. In the first sentence, evolution is mentioned, with virus-to-host and host-to-virus sequence transfer as being "common in evolution", but there are many evolutionary scenarios that do not involve this sequence transfer. For instance, how does the evolutionary scenario of finch beaks in the Galapagos islands relate to this phenomenon that is "common in evolution"? If evolution is mentioned, be more specific as to what specific mechanisms or modalities have been inferred, and their importance to this particular study (coevolution, mutualistic, antagonistic, commensal, purifying, diversifying, drift, etc). The paragraph then jumps to bacterial CRISPR sequences and viral endogenization. If you are going to mention CRISPR's evolutionary origins as involving viral endogenization, make an appropriate citation (Shmakov et al 2017?) Then, the paragraph jumps to how host protein sequences are embedded into "viral polyproteins that can be vertically transferred". In a more direct sense, genomic nucleic acid is vertically transferred, and not protein sequences, and I would caution the authors about the need for accuracy. Finally, as I try and assess the overall meaning of what has been written, are the authors trying to state that the DNA for host protein sequences has been incorporated into the nucleic acid for viruses (instead of viral nucleic acid sequence being incorporated into host DNA). Has that been conclusively shown?

We have removed the word “evolution”. We have added a citation for Shmakov. As requested, we have revised the wording.

Yes, in the viral polyprotein there are sequences that can be found in host proteins. The cleavage site motif sequences correspond to specific host proteins. The host proteins are all generally involved in generating the innate immune responses, the “hit list” of host proteins shown in Table 1 is a non-random list of proteins. The MAVS and Toll-receptor pathways can generate interferon. Components of these pathways are being targeted by the viral proteases to short circuit the cascade (Fig 1 on the right). Table 1 has the 31 published cases of host protein cleavage by viral proteases that we could find (there was only 1 non-Group IV case, HIV protease, discussed it in the text to save a column). All of the other viral proteases that cut host proteins belong to Group IV we reported this correlation in Morazzani, et al earlier. All are (+)ssRNA viruses or retroviruses. To our knowledge, we are the only authors who have compiled a list of host protein cleavages by viral proteases and have identified trends in the list. The data in Table 1 comes from multiple labs over a 30+ year period.

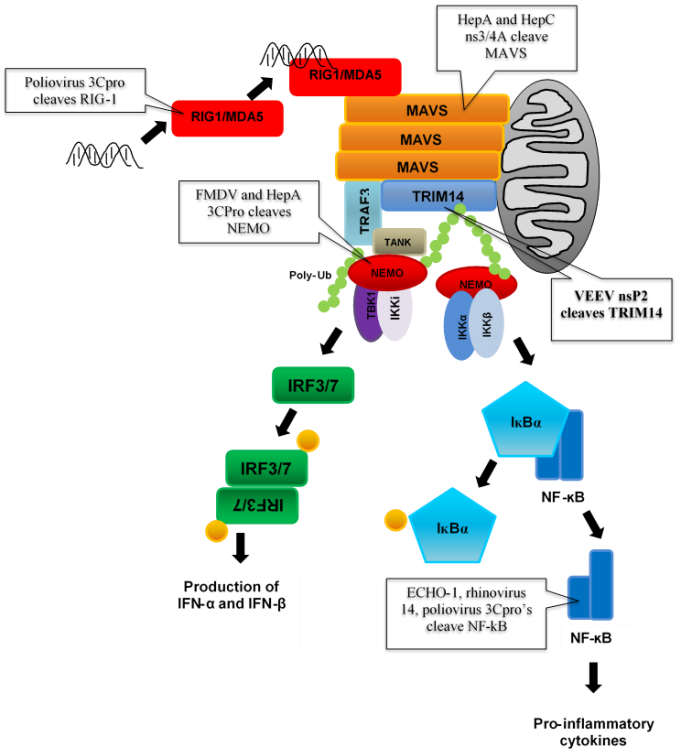


Figure 1. The MAVS Signalosome is targeted by multiple viral proteases (from Morazzani, et al).

Lines 51-59 (paragraph 2 of the introduction): "The information stored within the SSHHPS has only recently begun to emerge." This sentence is hard to interpret. Do you mean the dynamics surrounding SSHHPS are not yet fully understood? What part of viral phenotype can be expected to be recapitulated (host range, host susceptibility, host defenses)? And, is there an existing definition of recapitulate that supports your usage of this word in your intended sense? All of a sudden, targets are mentioned (what is the target, and what is targeting the targets?) and other methods are reported as not to be working (to do what, and is there a citation supporting this claim of other methods' failures, describing what these other methods are, and how these other methods failed?). Why say "SSHHP sequences" at the end of this paragraph, when you mainly say SSHHPS most everywhere else?

We currently do not know what portion of the phenotype is due to the viral proteases that are targeting host proteins for destruction. However, we did notice that the protein sequences cut by the ZIKV protease were in proteins involved in brain and eye development. Some had roles in both development and antiviral responses.

By recapitulate we mean that if the protease cuts a host protein(s) in humans, the same protein(s) should be cut in the animal model. In some cases an “uncleavable” homologue is found in rodents (Ref. Aguirre, et al 2012).

We revised the sentence, SSHHPS analysis produces different information from existing methods (proteomics).

It was difficult to distinguish between the plural and singular forms of SSHHPS's, where appropriate we spelled out "sequences."

Lines 60-71 (paragraph 3 of the introduction): We now jump from dsRNA (first citation in paragraph #1 about dsRNA viruses and eukaryotic cells) into ssRNA viruses. What does knowledge about dsRNA viruses impart into ssRNA viruses, where ssRNA viruses seem to constitute a major topical area for the rest of the manuscript? How do the prior two paragraphs relate to the rest of the introduction? The flow of writing does not seem to have been well-connected at this point. Remove this sentence: "In some cases the host proteins targeted by the viral proteases appear to be linked to the virus-induced phenotype." As relates to those SSHHPS that have been fully verified as being targets (and what fraction have an effect on the immune system), is it the strong majority of all SSHHPS or only a minor fraction, etc?

During (+)ssRNA replication, dsRNA intermediates are generated. The dsRNA is what triggers the MAVS cascade (it is sensed by RIG-I in the figure above).

As requested, we have revised the paragraph and have removed the sentence.

Remove "(n.b. nsP4 is the RNA-dependent RNA polymerase)." In terms of content, you seem to imply that viral translation is brought about by the RNA-dependent RNA polymerase, yet this is not the case. My own reaction in terms of style - why on earth use this rare abbreviation (n.b.) for nota bene (meaning "note well") when, in general, everything in a well-written scientific manuscript (that has an economy of prose) should be "noted well" by the reader?

As requested, we have removed the abbreviation. The RNA-dependent RNA polymerase is for replication of the (+)ssRNA genome. Translation precedes genome replication. Genome replication will produce dsRNA. The dsRNA will trigger MAVS. Thus, by short circuiting the MAVS pathway, the downstream effects of the MAVS pathway can be halted.

Lines 86-87: "Thus, these viruses may cleave host innate immune response proteins early in infection in order to suppress their effects." Is this offered as sheer speculation, or is there some prior report in the literature of this expectation and some supporting evidence?

Several authors reported cleavage products within the 3-5 hours post-infection [1, 2]. In the first report of a host protein cleavage by a viral protease (Histone H3) the cleavage was observed within 3 hours post-infection. We found this to be common. We have added references to support the statement. We also discuss the "uncleavable" case.

We also showed in Morazzani, et al that we could inhibit the VEEV protease and inhibit TRIM14 cleavage in infected cells. Another group showed that if you overexpressed TRIM14 that you could reduce alphaviral replication. The papers in Table 1 have similar evidence and conclusions.

Line 89: "CRISPR spacer sequences are ~26-72 nucleotides and RNAi are ~20-22 nucleotides" - does not seem relevant to the paragraph.

SSHHP sequences are 6-8 amino acids, in base pairs this would correspond to 18-24 bp. For ZIKV we found a protein with 11 amino acids in common, this would correspond to 33 bp. While these sequences may seem short, they are specifically recognized.

Lines 96-97: "Thus, for viral proteases that have loser sequence requirements such as those belonging to Picornaviridae the probability of finding a cleavage site in a host protein is likely higher." Did you mean "looser sequence requirements"? Also, is there a citation about the Picornaviridae claim?

Several of the entries in Table 1 are from *Piconaviridae*. From the papers there is very little sequence identity between the cleavage sequence in the virus vs. the host sequence (colored residues in column 6 of Table 1). In some cases, only 2 amino acids match. For the other viruses, the sequence identity or homology was higher (the ZIKV protease targets had 11 matching residues). The proteases are different, from the Table one can infer that the Picornaviral proteases must have very broad tolerances at each of their subsites since there is little sequence identity. Consequently, the host targets would be more difficult to predict using BLAST.

Figure 1 is yet another smattering mention of how nsP proteases compare to RNAi and CRISPR. If you are going to make this comparison, it would need to be consolidated into one well-focused paragraph (it is not otherwise a foundational theme to be continually referred to, as relates to the the title and abstract of this manuscript). Depending on goals with the associated video, the figure may or may not be relevant.
We have revised Figure 1. Here we want to show that SSHHPS follows the logic. These are algorithms or "programs". In particular, this is a search and delete program that utilizes a keyword. The methods described allow one to find the "keyword" and to find the file that may be "deleted" by the enzyme (protease or nuclease). The programs are written in 3 different languages (DNA, RNA, protein). SSHHPS is an example of post-translational silencing.

Lines 135-143: Is there a corresponding difference in human versus equine susceptibility, or does this analysis mainly relate to hypothesis generation?
Yes, as the name suggests equine encephalitis viruses kill horses and donkeys. The equine TRIM14 was missing the entire region that was transiently cut by the VEEV protease (and corresponded to a PRY/SPRY domain). The PRY/SPRY domain is a binding domain that recognizes a second protein in the MAVS signalosome.

Only 1% of humans die from VEEV infections. The species that were susceptible to infection (humans, rodents) had the cleavage sites suggesting that they had a defense mechanism (i.e. MAVS). The proposed reservoir (birds) had 2 amino acid differences in the cleavage site. The most susceptible species (horses) was missing the domain that was cut off. These alignments may be useful for host-reservoir analysis or for predicting which species will die from infection. For animal model selection, these differences may be important to examine. We discuss the species-specific differences in more detail in the Intro.

Line 170: A better topic sentence is needed. I am guessing that the authors mean to indicate that in vitro assays outperform limitations of cell-based assays (in which case, is there a citation to support this claim)?
We agree, we have revised this. Cleavage product analysis in cell-based assays is far more time-consuming than the *in vitro* protease assays. The cleavage is transient in cells and there are competing processes in mammalian cells (e.g. proteasome, proteases, etc.). It was very difficult to catch the cleavage of TRIM14 in cells. The amount of TRIM14 in cells is very low. We were not able to find a good reference for this.

Lines 193-195: "Cell-based assays should be used in combination with the in vitro and in silico methods described herein to confirm that sequence is a SSHHPS." Yet, prior content of the manuscript indicates that cell-based assays may sometimes be futile, and none of the protocol seems to show how a cell-based assay would be conducted.
Group IV is a large group. The proteases are not all the same. For instance, the flaviviruses have proteases that are embedded in the ER membrane. Some of the predicted cleavage sites were in signal peptides suggesting co-translational cleavage. Some of the reviewers suggested showing full-length protein cleavage, however, if cleavage is co-translational the cleavage of a fully folded native protein may not be relevant. For instance, if the sequence is buried, it may have been accessible during translation.

The alphaviral protease is in the cytoplasm and in the nucleus. The cleavage in cell-based assays is transient (first 24 hours) so one may not catch it in a cell-based assay, however, it does not necessarily mean that cleavage did not occur. Since the protease is essential for viral replication, we can't simply mutate the protease in a virus to test if cleavage is important.

Since the substrates of the proteases may all be very different, we decided to not address the cell-based assays since they may or may not apply to a particular case.

Lines 830-832: "The use of a bacterial expression system allows testing of short segments (~25 amino acids) of mammalian sequences without the use of cell culture." Specify as "mammalian cell culture".
As requested, we have inserted mammalian.

I do not find the discussion (Lines 828-881) to be an effective synthesis, and the writing needs improvement, on par with what has been described for other sections of the manuscript. What happened to objectives such as drug discovery, host range, host sensitivity, and host defenses?
We agree, we have revised it.

Minor Concerns:
Citation content for Citation #2, line 894 (Legler, 2018), seems abstruse - shouldn't the citation indicate more specifically that this citation relates to a patent?
This is a patent application, which is basically an abbreviated version of Morazzani, et al. It was technically published prior to Morazzani, et al and would be considered to be the first report.

We thank Reviewer #4 for their helpful comment and suggestions they have all greatly improved the quality of the manuscript.

---

[1] Graham, K. L., Gustin, K. E., Rivera, C., Kuyumcu-Martinez, N. M., Choe, S. S., Lloyd, R. E., Sarnow, P., and Utz, P. J. (2004) Proteolytic cleavage of the catalytic subunit of DNA-dependent protein kinase during poliovirus infection, *J Virol 78*, 6313-6321.

[2] Barral, P. M., Morrison, J. M., Drahos, J., Gupta, P., Sarkar, D., Fisher, P. B., and Racaniello, V. R. (2007) MDA-5 is cleaved in poliovirus-infected cells, *J Virol 81*, 3677-3684.

Patricia Legler

From: Vethakkan, Anita Mercy M. (ELS-CHN) <A.Vethakkan@elsevier.com>
Sent: Monday, July 29, 2019 3:35 AM
To: Patricia Legler
Subject: RE: Reproducing figures from another paper [190723-017799]

Dear Patricia Legler

Thank you for your email.

Please note that, as one of the Authors of this article, you retain the right to reuse portion or excerpts in a new work. You do not require formal permission to do so.

For full details of your rights as a Journal Author, please visit: https://www.elsevier.com/about/our-business/policies/copyright

I'd like to inform you that in your proposed publication please state that the figures are reprinted with permission from Elsevier and any readers of your publication would be required to contact Elsevier for permission to reuse them as Elsevier owns the copyright.

Please feel free to contact me if you have any queries.

Kind regards

Anita

Anita Mercy

Senior Copyrights Coordinator - Global Rights

Elsevier | Health Content Operations

(A division of Reed Elsevier India Pvt. Ltd.)

Ascendas International Tech Park, Crest - 12th Floor|Taramani, Chennai 600113 • India|

Tel: +91 42994696

a.vethakkan@elsevier.com <mailto:a.vethakkan@elsevier.com>

Permissions Helpdesk no - +1 215 239 3867

Join the Elsevier Connect Community

www.elsevier.com <http://www.elsevier.com/>

Twitter <https://twitter.com/ElsevierConnect> | Facebook <http://www.facebook.com/ElsevierConnect> | LinkedIn <http://www.linkedin.com/company/elsevier> | Google+ <https://plus.google.com/109387593599114360742>

From: Patricia Legler <patricia.legler@nrl.navy.mil>
Sent: Wednesday, July 24, 2019 4:59 PM
To: Rights and Permissions (ELS) <Permissions@elsevier.com>
Subject: FW: Reproducing figures from another paper [190723-017799]

*** External email: use caution ***

Hi,

I would like to re-publish one of my figures from my paper (which is open access) in another journal.

Can I get permission to do this?

Pat

From: Researcher Support [mailto:support@elsevier.com]
Sent: Wednesday, July 24, 2019 3:48 AM
To: Patricia Legler
Subject: Re: Reproducing figures from another paper [190723-017799]

How was our service today?
<https://service.elsevier.com/rd?1=APONxAqSDv8a~atpGm8a~yCgrOcqeC75Mv_K~zj~PP9D&2=7&6=1&7=27116982>
<https://service.elsevier.com/rd?1=APONxAqSDv8a~atpGm8a~yCgrOcqeC75Mv_K~zj~PP9D&2=6&6=1&7=27116982>

Dear Dr. Legler,

Article reference: AVR4465

Thank you for your enquiry.

I understand that you would like to reuse some of your figures.

I can confirm that that authors can use their articles, in full or in part, for a wide range of scholarly, non-commercial purposes one of which is inclusion in a thesis or dissertation. See the following link for further information on this https://www.elsevier.com/about/our-business/policies/copyright/personal-use <https://www.elsevier.com/about/our-business/policies/copyright/personal-use>

As you will be able to see from the above link, our policy is to allow authors to use their work in their thesis or dissertation (provided that this is not to be published commercially).

If you require further clarification please contact permissions@elsevier.com <mailto:permissions@elsevier.com> who is best positioned to assist you.

Kind regards,

Maria Jonnalyn L. Flaviano
Researcher Support
ELSEVIER

Find out some simple ways to share your research data <https://www.elsevier.com/authors/author-services/research-data?utm_campaign=STMJ_1527601374_AUTH_SERV_OTR&utm_medium=email&utm_source=Other&dgcid=STMJ_1527601374_AUTH_SERV_OTR> , including features that are directly available when you submit your research article to an Elsevier journal.

For assistance, please visit our Customer Support site <https://service.elsevier.com/app/home/supporthub/publishing/> where you can search for solutions on a range of topics and find answers to frequently asked questions.

_____

From: Administrator
Date: 23/07/2019 08.52 PM

Dear Customer,

Thank you for submitting your question. This is to confirm that we have received your request and we aim to respond within 24 hours.

For future correspondence about this question, please provide this reference number: [190723-017799]. Please do not change the subject line of this email when you reply.

You can reach our support center at: https://service.elsevier.com/app/home/supporthub/publishing <https://service.elsevier.com/app/home/supporthub/publishing/track/APONxAqSDv8a~atpGm8a~yCgrOcqeC75Mv_K~zj~PP9D>

Kind regards,
Elsevier Customer Service

_____

From: Patricia Legler
Date: 23/07/2019 08.52 PM

I am writing a methods paper for JOVE. I would like to reuse some of my figures. How do I obtain explicit permission to reuse the figures?

_____

[---001:002413:53857---]