

HHS Public Access

Author manuscript

Biochemistry. Author manuscript; available in PMC 2017 February 03.

Published in final edited form as:

Biochemistry. 2016 May 31; 55(21): 3007–3019. doi:10.1021/acs.biochem.5b00992.

Kinetic, Mutational, and Structural Studies of the Venezuelan Equine Encephalitis Virus Nonstructural Protein 2 Cysteine Protease

Xin Hu[†], Jaimee R. Compton^N, Dagmar H. Leary[‡], Mark A. Olson[∮], Michael S. Lee[∮], Jonah Cheung[§], Wenjuan Ye[†], Mark Ferrer[†], Noel Southall[†], Ajit Jadhav[†], Pamela J. Glass[∮], Juan Marugan[†], and Patricia M. Legler^{‡,*}

[†]NIH Chemical Genomics Center, National Center for Advancing Translational Sciences, Rockville, MD 20850

^NNova Research, Inc., Alexandria, VA 22308

§New York Structural Biology Center, New York, NY

∮United States Army Medical Research Institute of Infectious Diseases, Frederick, MD 21702

[‡]Center for Bio/molecular Science and Engineering, U.S. Naval Research Laboratory, Washington, DC 20375

Abstract

The Venezuelan equine encephalitis virus (VEEV) nonstructural protein 2 (nsP2) cysteine protease (EC 3.4.22.-) is essential for viral replication and is involved in the cytopathic effects (CPE) of the virus. The VEEV nsP2 protease is a member of MEROPS Clan CN and characteristically contains a papain-like protease linked to an S-adenosyl-L-methionine dependent RNA methyltransferase (SAM MTase) domain. The protease contains an alternative active site motif, ⁴⁷⁵NVCWAK⁴⁸⁰, which differs from papain's (CGS²⁵CWAFS), and the enzyme lacks a transition state (TS) stabilizing residue homologous to Q19 in papain. To understand the roles of conserved residues in catalysis we determined the structure of the free enzyme, and the first structure of an inhibitorbound alphaviral protease. The peptide-like E64d inhibitor was found to bind beneath a β-hairpin at the interface of the SAM MTase and protease domains. His-546 adopted a conformation that differed from that found in the free enzyme, each conformer may assist in leaving group departure of either the amine or Cys thiolate during the catalytic cycle. Interestingly, E64c (200 µM), the carboxylic acid form of the E64d ester, did not inhibit the nsP2 protease. To identify key residues involved in substrate binding, a number of mutants were analyzed. Mutation of the motif residue, N475A, led to a 24-fold reduction in k_{cat}/K_m, and the conformation of this residue did not change after inhibition. N475 forms a hydrogen bond with R662 in the SAM MTase domain, and the R662A and R662K mutations both led to 16-fold reductions in k_{cat}/K_m. N475 forms the base of the P1 binding site and likely orients the substrate for nucleophilic attack or plays a role in product

^{*}CORRESPONDING AUTHOR FOOTNOTE. Patricia.legler@nrl.navy.mil, (202) 404-6037, 4555 Overlook Ave., Washington, DC 20375.

SUPPORTING INFORMATION PARAGRAPH. Supplemental information is included and contains protein sequences of the FRET substrates and VEEV nsP2 thioredoxin fusion protein, sequence alignments, structural overlays, models, pH-rate profile, and mass spectra.

release. An Asn homologous to N475 is similarly found in coronaviral papain-like proteases (PLpro) of the Severe Acute Respiratory Syndrome (SARS) virus and Middle Eastern Respiratory virus (MERS). Mutation of another motif residue, K480A, led to a 9-fold decrease in k_{cat} and k_{cat}/K_m . K480 likely enhances the nucleophilicity of the Cys. Consistent with our substrate-bound models, the SAM MTase domain K706A mutation increased the K_m 4.5-fold to 500 μM . Within the β -hairpin, the N545A mutation slightly, but not significantly increased the k_{cat} and K_m . The structures and identified active site residues may facilitate the discovery of protease inhibitors with antiviral activity.

Keywords

VEEV; alphaviral; nsP2; cysteine protease; substrate specificity; E64d; antiviral drug target

The Venezuelan Equine Encephalitis Virus (VEEV) is a highly infectious New World alphavirus from the Togaviridae family. It is a positive-sense RNA virus (1). The zoonotic arthropod-borne virus (arbovirus) can be transmitted via mosquitos to birds, horses, rodents, humans and other vertebrate species (2;3). Equine are unusually susceptible to the virus, and fatality rates for horses are 20–80 % (3). VEEV is both neuroinvasive and neurotropic; the virus replicates in both lymphoid tissue and in the brain (3;4). VEEV causes acute viral infections which are typically resolved by innate and adaptive immune responses, however, neurological symptoms occur in approximately 14 % of infected humans and ~1 % of cases result in lethal encephalitis (3;5). The more deadly Eastern (EEEV, 36 % mortality (6)) and Western (WEEV, 10 % mortality (7)) equine encephalitis viruses share 53 % and 46 % sequence identity with VEEV, respectively. The related Old World alphaviruses such as Chikungunya (CHIKV), have only been associated with fever, arthralgia, and skin rashes (8;9). VEEV viral particles are unusual in that they are highly resistant to desiccation and can be stably lyophilized and aerosolized (3;10–12). Inhaled virus can pass into the brain via the olfactory neurons (4;13;14), which has implications for its use as a biological weapon.

Currently, there are no licensed vaccines or therapeutics for alphaviruses. However, two vaccines are used under Investigational New Drug (IND) status to protect laboratory personnel (15). The live-attenuated VEEV vaccine, TC-83, protects against both subcutaneous and aerosol exposure in animals, and ~82 % of vaccinated humans develop immunity after vaccination (15). A formalin-inactivated vaccine derived from TC-83 called C-84 (16) protected against subcutaneous infection, but failed to protect against aerosol exposure (17) suggesting poor induction of mucosal immunity. Some small molecule inhibitors targeting host or viral proteins have been identified (18–20), but none have been shown to directly inhibit the viral nsP2 cysteine protease or have significant effects on WEEV or EEEV. For viral encephalitis, poor neuronal and blood-brain barrier (BBB)-permeability limit the effect of most antivirals.

The nonstructural protein 2 (nsP2) cysteine protease plays an essential role in alphaviral replication (21;22). The viral RNA encoding the nonstructural (ns) proteins of VEEV is translated into a single polypeptide, nsP123 or nsP1234 (1). The nsP2 cysteine protease cleaves the nsP1234 polypeptide into four proteins (nsP1234 \rightarrow nsP1, nsP2, nsP3, nsP4)

(21;22) (Figure 1). The nsP2 polypeptide contains a short N-terminal region (23), a helicase domain, a cysteine protease domain, and a S-adenosyl-L-methionine-dependent RNA methyltransferase domain (SAM MTase) (Figure 1A). Alphaviruses cause a lytic infection of mammalian cells and the nsP2 protease-SAM MTase also plays a role in the cytopathic effects (CPE) of the virus (24–26). Temperature-sensitive mutations which lead to altered plaque phenotypes have been localized to the nsP2 protease-SAM MTase (26;27) and have been referred to as non-cytopathic effect variants (non-CPE). Some non-CPE mutations are near the predicted substrate binding site, while others are present on the SAM MTase domain distant from the active site (e.g. P713S). VEEV nsP2 localizes to both the cytoplasm and nucleus (28), and the high pI of the SAM MTase domain (pI = 9.45) is typical of a nucleic acid binding protein. These two domains are not only found linked together in alphaviruses, but also in Hepatitis E virus (HEV) (29). However, in HEV the SAM MTase and cysteine protease domains are permuted.

The alphaviral nsP2 proteases belong to Clan CN in the MEROPS peptidase database, which characteristically contain papain-like protease domains linked to SAM MTase domains. The role of the SAM MTase domain has been unclear. While three-dimensional structures of the VEEV and CHIKV nsP2 protease-SAM MTase domains (PDB 2HWK (30) and PDB 3TRK (31), respectively) have revealed details of the protease domain assembled with the SAM MTase domain (44–45), substrate or inhibitor bound structures have not been previously reported. Comparison of the VEEV nsP2 cysteine protease (PDB 2HWK) with other cysteine proteases suggests that the peptide substrate binds within a cleft formed by the protease and SAM methyltransferase domains (Figure 1B) (30;32). However, residues within the SAM MTase domain have not been formally shown to be required for proteolysis. Notably, the Clan CN nsP2 cysteine proteases of alphaviruses contain an alternative active site motif, ⁴⁷⁵NVCWAK⁴⁸⁰, which differs from papain's (CGS²⁵CWAFS), and lacks a key TS-stabilizing residue homologous to Q19 in papain.

Here we present the crystal structures of the nsP2 cysteine protease of VEEV in the free and E64d-bound states. This is the first report of an inhibitor bound alphaviral nsP2 protease structure. To investigate the catalytic mechanism of the enzyme, we performed kinetic and site-directed mutagenesis studies to identify key active site residues involved in substrate binding and catalysis. The structure of the peptide-like E64d adduct shows that the inhibitor binds beneath a β -hairpin at the interface of the SAM MTase and protease domains, and that His-546 adopts a different conformation from that in the apo state, which may be relevant to its role in leaving group departure in the catalytic cycle. A number of key residues in both the protease and SAM MTase domains were found to have significant effects on protease activity or substrate binding. The hydrogen bonding interactions between the inhibitor and enzyme may be beneficial for future inhibitor development.

MATERIALS & METHODS

Materials

All general chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Plasmid constructs were synthesized by Genscript USA, Inc. (Piscataway, NJ). BugBuster was purchased from Novagen, Inc (Billerica, MA). Column resins and PD-10 gel filtration

columns were purchased from G. E. Healthcare (Pittsburgh, PA). Roche EDTA-free Protease inhibitor tablets were from Sigma-Aldrich. Black half-area non-binding surface 96-well plates were from Corning Inc. (Corning, NY). Pierce Precise Tris-HEPES acrylamide gels (8–16 % gradient) and BupH Tris-HEPES SDS-PAGE running buffer were from Thermo Scientific (Rockford, IL). The JCSG+ screen was from Qiagen (Valencia, CA).

Plasmid Constructs of FRET Substrates

A pET15b plasmid (Ampicillin^R) encoding cyan fluorescent protein (CFP), an nsP2 protease cleavage site motif, $A\underline{G}(A/C)\downarrow(G/Y/A)$, and yellow fluorescent protein (YFP) in between the NdeI and XhoI cut sites was synthesized for each cleavage site. A total of five CFP-YFP constructs were made: (1) Semliki Forest Virus (SFV) P12 11:6, (2) SFV P12 19:6, (3) VEEV P12 19:6, (4) VEEV P23 19:6, and (5) VEEV P34 19:6. An N-terminal hexahistidine tag preceded a thrombin cleavage site. Protein sequences are included in the Supplemental Information. The cleavage site is denoted by a colon to demarcate the residues N-terminal of the scissile bond and residues that are C-terminal. The 25-residue substrate with the cleavage site between nsP1 and nsP2 is refered to as CFP-P1/2₂₅-YFP where the subscript refers to the length of the substrate.

Plasmid Constructs of VEEV nsP2 Protease

The VEEV nsP2 protease was fused to thioredoxin (Trx) and a hexa-histidine tag with a thrombin cleavage site in a pET32a vector (Novagen, Inc.). The His-tag was associated with thioredoxin. Site-directed mutagenesis was performed using QuikChange (Agilent Technologies).

Plasmid Construct of the CHIKV nsP2 Protease

Appropriate oligonucleotides were used to PCR amplify a DNA fragment coding for a construct of nsP2 (residue positions 1006 to 1326 of the non-structural polyprotein P1234) from a purchased synthetic DNA template (DNA 2.0). The fragment was inserted into a modified pMCSG9 vector (33) to fuse a decahistidine-tagged maltose-binding-protein at the N-terminus that could be cleaved using TEV protease.

Expression & Purification of the VEEV nsP2 Cysteine Protease

To ensure purification of the reduced state of the VEEV nsP2 cysteine protease (34–36), we constructed an nsP2-thioredoxin (Trx) fusion protein containing the protease and SAM MTase domains (residues 457–792) (Supplemental Information). BL-21(DE3) pLysS *E.coli* were transformed with the Trx-VEEV-nsP2 plasmid. Luria Bertani (LB) media (3–6 L) containing 50 µg/mL ampicillin (Amp) and 25 µg/mL chloramphenicol was inoculated and grown to an OD₆₀₀ of approximately 1.0 and induced with 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) overnight at 17 °C. Cells were pelleted and lysed with lysis buffer (50 mM Tris pH 7.6, 500 mM NaCl, 35 % BugBuster, 5 % glycerol, 2 mM β -mercaptoethanol (BME), 25 U of DNase, 0.3 mg/mL lysozyme) and sonicated ten times for 15 second intervals in an ice bath. Lysates were clarified by centrifugation at 20,000 x g for 30 minutes, and loaded onto a nickel column equilibrated with 50 mM Tris pH 7.6, 500 mM NaCl, 2 mM BME, 5 % glycerol. The column was washed with buffer containing 60 mM

imidazole. Protein was eluted using the same buffer containing 300 mM Imidazole. Protein was dialyzed overnight with bovine thrombin against 50 mM Tris pH 7.6, 250 mM NaCl, 5 mM dithiothreitol (DTT), 1 mM ethylenediamine tetraacetic acid (EDTA), 5 % glycerol, and then loaded onto an SP-Sepharose column equilibrated with 50 mM Tris pH 7.6, 5 % glycerol and 5 mM DTT. The cut tag-free nsP2 protease had a high pI and was eluted using a salt gradient (from 0 to 1.25 M NaCl). The protein was judged to be >95 % pure based upon SDS-PAGE gels. A C477A variant was constructed and purified by the same method to confirm that no other contaminating proteases had co-purified. Protein was concentrated, flash frozen in liquid nitrogen and stored at -80 °C. The thioredoxin-His-tag free enzymes were used for kinetic analysis. The buffer was exchanged to the corresponding assay buffer prior to all kinetic experiments using PD-10 columns. Protein concentration was determined from the A_{280} using a calculated extinction coefficient. $\varepsilon = 42,860 \, \text{M}^{-1} \, \text{cm}^{-1}$. The CHIKV nsP2 protease was expressed and purified by the same method, however, the His-tag and MBP were not removed.

Expression & Purification of FRET Protein Substrates

BL-21(DE3) E. coli were transformed with the plasmids encoding the substrates. LB/Amp (1.5 to 3.0 L) was inoculated and grown to an OD₆₀₀ of approximately 1.0 and induced with 0.3 mM IPTG overnight with shaking at 17 °C. Cells were pelleted by centrifugation, lysed with lysis buffer (50 mM Tris pH 7.6, 500 mM NaCl, 35 % BugBuster, 2 mM BME, 0.3 mg/mL lysozyme, 1 EDTA-free protease inhibitor tablet), and briefly sonicated for 1 minute in an ice bath. Lysates were clarified by centrifugation (20,500 x g for 30 minutes at 4 °C) and loaded onto a nickel column equilibrated with 50 mM Tris pH 7.6, 500 mM NaCl, 2 mM BME. The column was washed with the same buffer after loading, and with 10-20 column volumes of buffer containing 60 mM imidazole until the A₂₈₀ returned to baseline. The protein was eluted with the same buffer containing 300 mM imidazole. The protein was dialyzed against 50 mM Tris pH 7.6, 150 mM NaCl overnight at 4 °C. All substrates were produced in high yield (typical yields were 60-80 mg per liter of media) and could be readily concentrated to 9.0-10.5 mg/mL. The substrates were used for continuous and discontinuous assays. Similar substrates have been used to study other proteases (37). The substrate was confirmed to be folded at pH 6.0 at room temperature by circular dichroism spectroscopy. At pH values less than 5 the substrate was difficult to concentrate.

Continuous FRET Assay

For measurement of steady state kinetic parameters the method described by Ruge et al. was followed (38). Cleavage of the YFP/CFP FRET substrates was monitored continuously at room temperature (23 \pm 3 °C) using excitation/emission wavelengths of 434/470 nm and 434/527 nm and a SpectraMax M5 plate reader from Molecular Devices. The substrate was buffer-exchanged into 0.1 M sodium phosphate buffer pH 6.0 using PD-10 columns. Enzyme concentrations of 0.2 to 1 μM and a substrate concentration range of 10–140 μM (8 different concentrations) were used in assays; for high K_m values an additional point at 200 μM was used. Data were collected in triplicate (50 μL reaction volumes) in half-area black low binding surface 96-well plates from Corning, Inc. Plates were read for 10–120 minutes at 1 to 2 minute intervals. After the reads were completed the plates were sealed with film and allowed to digest overnight at room temperature 23 \pm 3 °C. For the variants tested

excess WT enzyme was added and the final emission ratios were read the next day. The time needed for complete cleavage of the substrate was confirmed at each pH using SDS-PAGE gels. The fraction of substrate cleaved, *f*, was calculated from the emission ratios at each time point using the following equation:

$$f = \frac{\left[\frac{\left(\frac{ex434}{em527}\right)}{\left(\frac{ex434}{em470}\right)} - r_{uncut}\right]}{\left(r_{cut} - r_{uncut}\right)}$$

The nmols of substrate cleaved at each time point was calculated by multiplying f by the nmols of substrate at t=0 ($\rm S_{o}$). The value of $\rm r_{\it uncut}$ corresponds to the emission ratio measured in the absence of enzyme, and the value of $\rm r_{\it cut}$ is the emission ratio measured when the substrate was fully cleaved. Initial velocities were calculated at each [$\rm S$] concentration from the linear range (f 20 %). Plots of time vs. nmols were linearly fit for each [$\rm S$] concentration, and $\rm v_{o}$ was obtained from the slopes of the lines. Rates of spontaneous hydrolysis were measured in the absence of enzyme, and were subtracted from the enzyme catalyzed rates. Data were fit to the Michaelis-Menten equation, $\rm v_{o}$ = ($\rm V_{max}*[\rm S]$)/($\rm K_{m}$ + [$\rm S$]) using Grafit (Erithricus Software Ltd.). For variants with $\rm K_{m}$ values greater than the highest substrate concentration tested the values of $\rm K_{m}$ were distantly extrapolated from Lineweaver-Burk plots of 1/ $\rm v_{o}$ vs. 1/[$\rm S$]. The $\rm k_{cat}/\rm K_{m}$ values were also calculated from the slope of the linear part of the Michaelis-Menten plot at substrate concentrations less than the $\rm K_{m}$ (39) for comparison with the $\rm k_{cat}/\rm K_{m}$ determined from the extrapolated values.

Discontinuous Gel-Based Assay

Reaction mixtures (2.6 μ M nsP2-Trx, 30 μ M FRET substrate, 1x PBS pH 7.4, 5 mM DTT) were incubated overnight (~18 hours) at room temperature (23 \pm 3 °C). The reactions were run until >90 % of the substrate was cleaved by the WT enzyme. Reactions were stopped by mixing with Laemelli buffer (1:1) and heating the samples for at least 3 minutes at 70 °C. Cleavage products (10 μ L) were separated by SDS-PAGE in 12-well 8–16 % gradient gels in BupH running buffer (100 mM Tris, 100 mM HEPES, 3 mM SDS, pH 8.0 \pm 0.5) at 110 V for 50 minutes. The calculated molecular weight of the uncut FRET substrate containing a 25 amino acid cleavage sequence was 58.3 kDa, and 30.9 kDa and 27.4 kDa for the cut CFP and YFP products. The molecular weight of the enzyme for the thioredoxin-His-tagged enzyme was 52.207 kDa, and 38.29 kDa for the Tag-free enzyme. The bands were well separated in 8–16 % gradient gels and boiling of the samples was required to achieve the sharp banding pattern. The BioRad Gel Dock Imager software (BioRad Inc.) was used for densitometry.

Time-Dependent Inhibition of VEEV nsP2 by E-64d

Progressive inhibition by E64d (CID 65663, Loxistatin) was measured in the absence of substrate with at least 5 different concentrations of inhibitor (in DMSO) essentially as described in (40). The K_p is the dissociation constant for the initial reversible complex. k_2 is the maximum rate constant for inhibition at saturating inhibitor concentration.

$$k_{obs} = k_2/(1 + (K_p/[I]))$$
 (1)

The apparent bimolecular rate constant for inhibition, k_i , describes the rate of formation of the covalent E-I complex from free enzyme and inhibitor in the absence of substrate, and was calculated according to equation 2.

$$k_i = k_2 / K_p$$
 (2)

Inhibition was carried out in 1x PBS pH 7.4 containing 0.08 % Tween-20 at room temperature (22 ± 3 °C). For inhibition experiments reducing agent was removed using a PD-10 column just prior to the experiment (41).

Crystallization of free enzyme and the E64d Adduct

Crystals were obtained in several conditions within the JCSG+ screen. The free enzyme was crystallized in Condition #15 (0.1 M Bicine pH 8.5, 20 % PEG 6000). The inhibitor-bound enzyme was crystallized in Condition #8, this condition produced the strongest density for the inhibitor. Enzyme in 50 mM Tris pH 7.6 was inhibited with E64d, excess inhibitor was removed using a PD-10 column. The inhibited enzyme (17 mg/mL) was mixed 1:1 with 0.2 M ammonium formate, 20 % PEG 3350 (condition #8 of the JCSG+ screen). Rod shaped crystals were obtained in hanging drops at 17 °C. Crystals were soaked in drop solution containing 30 % glycerol, looped, and flash frozen in liquid nitrogen. Diffraction data was collected at 150 K using a Bruker Micro-STAR rotating anode equipped with Helios optics and a Bruker Platinum 135 CCD area detector. The crystals of the free and bound enzyme diffracted out to 1.66 Å and 2.16 Å, respectively, on our home source. The space group, P2₁2₁2₁, was the same as for the previously reported structure of the free enzyme (PDB 2HWK, 2.45 Å resolution, pH 8.0 (30)), however the unit cell dimensions and crystal packing differed. The structure was solved by molecular replacement using PDB 2HWK (30) as the starting model and Amore (42). Model building was performed using Coot (43;44). Simulated annealing was carried out using CNS 1.1(45) and maximum likelihood refinement with Refmac5 (46).

Cell-based ELISA Assays

ATCC-Vero 76 cells were seeded at 40,000 cells/well in 96-black well plates and incubated at 37 °C with 5 % CO₂ with EMEM (Invitrogen) complete with 10 % FBS + 1% L-Glutamine and Pen/Strep (EMEM-C) for 24 hours. Compound was plated at varying dilutions (2x), then 50 μl of compound was transferred to cell plates 24 hours after seeding. Compound was added to three (3) cell plates, two (2) for virus infection evaluation, and one (1) for evaluation of cell viability via the Promega Cell Titer-Glo assay. At BSL-3, 50 μl of virus (VEEV Trinidad) diluted in MEM was added to corresponding infection plates. After adding virus, plates were incubated at 37 °C with 5 % CO₂ for 18–24 hours. After incubation, plates were fixed in 10 % buffered formalin and incubated at 4 °C for a

minimum of 24 hours before being removed from BSL-3 containment and transferred to BSL-2 for staining using an antibody cell-based ELISA detection assay (47).

To evaluate VEEV inhibition, plates were blocked for 1 hour at room temperature using a blocking buffer (1x PBS containing 3 % BSA). After 1 hour, plates were washed 3 times with PBS. Virus infection was detected in cells using the 1A4A monoclonal antibody (50 μ l of 1:4000 dilution of 1 mg/ml) added to each well and incubated at room temperature for 2 hours. Plates were washed 3 times with PBS; after washing, 65 μ l of a 1:10,000 dilution of HRP goat anti-mouse secondary was added per well and incubated for 1 hour at room temperature. After 1 hour, plates were washed 3 times with PBS and 100 μ l of Pierce HRP Pico Luminescent substrate (component A and B mixed) was added per well. Plates were examined for luminescence utilizing a Gemini EM from Molecular Devices. The VEEV assay described evaluates the amount of virus spread in the cell layer per well for which a percent inhibition can be calculated.

Modeling of substrate binding interactions

The binding models of substrates including VEEV P12, P23, and P34 were predicted with an ensemble-docking protocol using the AutoDock program (48). Multiple conformations of the VEEV nsP2 structure (PDB 2HWK) and the CHIKV nsP2 (PDB 3TRK) were obtained from MD simulations and cluster analysis. The active site of the protein was defined by a grid of 70 x 70 x 70 points with a grid spacing of 0.375 Å centered at the catalytic residue Cys-477. The Lamarckian Genetic Algorithm (LGA) was applied with 50 runs, and the best pose with the most favorable binding free energy was selected. MD simulations were performed for the predicted substrate binding models using the AMBER 14 package and the ff99SB force field (49). The solvated systems were subjected to a thorough energy minimization prior to MD simulations. Periodic boundary conditions were 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. The binding free energies were calculated using the MM-PBSA method. Decomposition of the calculated binding free energies was performed using the same MM-PBSA module in the AMBER 14 package.

RESULTS

Inhibition by a Peptide-like Inhibitor

E64d is a membrane permeable pro-drug for *in vivo* inhibition of calpains and Cathepsin B (50;51). In the cell the E64d ester is hydrolyzed by esterases to produce the negatively charged E64c acid (52). Many cysteine proteases are readily inhibited by E64d and E64c; papain is inhibited by nanomolar concentrations of E64c (53). E64c also inhibits Cathepsin B, H, and L (54–56). We evaluated the inhibitory activities of these two cysteine protease inhibitors against the VEEV nsP2 cysteine protease using both continuous and discontinuous assays and a cell-based ELISA assay. E64d weakly inhibited VEEV nsP2 (Figure 2A). Using the method of Kitz and Wilson (57) the maximal rate constant for inactivation $k_2 =$

 $0.05 \pm 0.02~\text{min}^{-1}$ and the dissociation constant, $K_p = 40 \pm 10~\mu\text{M}$, were measured for time-dependent inhibition in the absence of DTT. In discontinuous assays (SDS-PAGE gels) overnight incubation of E+I+S at room temperature (RT, $23 \pm 3~\text{°C}$) without DTT led to complete inhibition using 200 μM E64d, however, 200 μM E64c did not significantly inhibit the enzyme (Supplemental Fig. S1). In cell-based assays both E64d and E64c were ineffective at 50 μM , and no reduction in viral replication was observed (Figure 2B). The only difference between the two inhibitors is the carboxylic acid vs. the ester. Analysis of the E64d binding model of the nsP2 cysteine protease indicated that the negatively charged oxygen of the E64c acid would be near His-476-N81 and Cys-477-S⁻. Further comparison with four other cysteine proteases (papain, actinidin, Cathepsin L and Cathepsin K) showed that the carboxylic acid of the related E64 inhibitor was within hydrogen bonding distance of Gln-19 in these enzymes. A residue homologous to Gln-19 is notably absent in the VEEV nsP2 cysteine protease and may account for the inhibitor specificity. Thus, the negatively charged thiolate may make E64c binding unfavorable.

X-ray Crystal Structure of the E64d Adduct

To identify residues important for catalysis and substrate binding we determined the X-ray crystal structure of E64d bound to the protease (Table 1). The E64d inhibitor was found beneath the β -hairpin formed by residues R543-H546 (Figure 2D). The density was continuous with that of Cys-477-S γ and was evidence of covalent inhibition (Supplemental Fig. S1). The inhibitor was fitted with an occupancy of 0.75, and some additional density near a second conformer of Cys-477 was observed. Nucleophilic attack occurred at the C2 carbon of the epoxide, similar to what had been observed in papain (53;58). The backbone amide of His-510 (3.1 Å) and the backbone carbonyl group of Asn-545 (3.5 Å) were within hydrogen bonding distance to the inhibitor, however, no hydrogen bonding interactions involving side chains were observed in the E64d-bound structure.

Notably, a conformational change in residue His-546 was found when comparing the free and bound states (Figure 3). Crystal structures of papain suggested that His-159 rotates about its C_{β} - C_{γ} bond in order to donate its proton to the amide nitrogen and facilitate the collapse of the tetrahedral intermediate and release of the free amine (59). In overlays of the free and bound structures (PDB 5EZQ and 5EZS) one conformer of His-546-N81 was within hydrogen bonding distance to the ester oxygen which mimics the amine leaving group (Figure 3). In the E64d inhibited enzyme the His side chain was displaced by a water molecule, and could be refined with an occupancy of 0.8 to a position closer to the Cys-477-S γ . The density for the imidazole ring was poor suggesting static or dynamic disorder. The alternate conformer may allow the His to assist the thiolate leaving group. A third conformer of the His which differs from those observed here was identified in the previously reported crystal structure of the free enzyme (PDB 2HWK) (30).

Conformational changes in both the free and inhibited enzyme structures were observed in residues Asp-765 to Lys-775 which are distant from the active site (Supplemental Figure S1). His-771 rotated inwards towards the core of the SAM MTase, while Asn-772 rotated out. Residues within this region were involved in crystal packing interfaces, but are also likely involved in the nsP2/nsP3 interface. Overlays of the structures with the pre-cleavage

form of nsP23 from Sindbis virus (SINV) (PDB 4GUA) showed a similar conformation of the homologous His (His-1325) in the nsP23 complex (60). The significance of this conformational change is unclear, however, several mutations leading to altered plaque phenotypes have been identified near this region in the Old World alphavirus, SINV (26;61).

Since residues that stabilize the transition state have not been identified in the alphaviral nsP2 proteases we searched for hydrogen bonding donors near the carbonyl oxygen of E64d. The carbonyl oxygen of the E64d ester was 2.9 Å from the Cys-477-NH consistent with a role in TS stabilization. In the inhibitor-bound state no other backbone or side chain amine groups were directed towards the carbonyl oxygen. Only a water molecule was found 4.1 Å from the carbonyl oxygen. Substrate induced conformational changes to stabilize the TS in the VEEV nsP2 protease have been proposed by others (32), however, in the E64d structure no significant conformational changes occurred in residues of the active site with the exception of His-546.

Substrate Binding Interactions

We tested 17- and 25-residue substrates containing the SFV P1/2 cleavage site and measured the steady state kinetic parameters for each substrate (Figure 4C). A 12-residue substrate containing the SFV P1/2 cleavage site (EYHAGA → GVVETP) in between maltose binding protein (MBP) and NusG was previously characterized by Zhang et al. and produced a K_m = 0.58 mM and $k_{cat} = 0.96 \text{ min}^{-1}$ (62). Using a 17-residue (CFP-P1/2₁₇-YFP) and 25-residue (CFP-P1/2₂₅-YFP) SFV substrate (DVEELEYHA \underline{G} A \downarrow GVVETP and AETGVVDVDVEELEYHA $\underline{\mathbf{G}}$ A \downarrow GVVETP) we obtained a K_m of 600 ± 200 μ M and a k_{cat} of 3.0 \pm 1.0 min^{-1} for the shorter 17-residue substrate, and a K_m = 230 \pm 20 μM and k_{cat} of $2.0 \pm 0.1 \text{ min}^{-1}$ for the longer 25-residue substrate. The analogous VEEV substrates had K_m values ranging from 110–260 µM, indicating that residues beyond P6 or better accessibility to these residues, or both, are required for tight binding (Table 2). On the C-terminus of the substrate 6 residues were found to be sufficient, and no enhancement in cleavage was observed with longer sequences (data not shown). In fact, a variety of substitutions were tolerated within the P1'-P6' residues (Figure 4), suggesting that the viral protease may utilize backbone contacts as opposed to specific side chain interactions for substrate binding. The predominance of backbone interactions has been reported for other viral proteases (63).

To further probe the key residues important for substrate binding specificity we modeled the binding interactions of the protease with the P1/2, P2/3, and P3/4 substrates (Figure 4A, Supplemental Fig.S2). Previous studies of the substrate models with nsP2 were limited to the P1'-P4' site (32). We extended the substrates from P6 to P6' in the binding models as these residues are important for substrate recognition. In serine proteases the substrate specificity is largely determined by the P1 residue, whereas in cysteine proteases the P2 residue (underlined glycine) of the protease cleavage site motif, $A\underline{G}(A/C)\downarrow(G/Y/A)$, is highly conserved and the P1 residue is variable (64). In our models the glycine hydrogens of the P2 residue are directed towards the Trp-478 and Ile-542 side chains which may explain why larger amino acids are not accommodated at this position. All predicted models adopted favorable binding interactions at the C-terminal end of the substrate peptide (P1'-P6') that was pointed down towards the SAM domain (Figure 4A). Interestingly, the C-terminal end

of the substrate underwent significant conformational changes in the MD simulations and was also found to point upwards towards the protease domain (Figure 4B). The upward binding model appeared to be more stable with extensive interactions with residues surrounding the active site, placing the carbonyl oxygen of the scissile bond closer to the active site Cys. Based on the binding models, we identified several key residues important for substrate binding and performed site-directed mutagenesis and kinetic analysis as described below.

Effects of the K705A and K706A mutations on Substrate Binding

Residues K705 and K706 from the SAM domain showed remarkable binding interactions with the P4 residue of the substrates (Figure 4, Supplemental Figure S2). K706 is highly conserved in VEEV/EEEV/WEEV (Supplemental Figure S3), while K705 in VEEV is replaced by an Arg in EEEV and by a Cys in WEEV. The K706A and K705A/K706A mutations increased the K_m 4–5 fold; similarly mutation of K705 to glutamate also increased the K_m 3-fold (Table 3). K706 was partially substituted by Gln, K706Q, suggesting that the hydrogen bonding interactions of the side chain play an important role in substrate binding. This may also account for the broader substrate specificity of the VEEV nsP2 protease, which was able to cleave both the Old World SFV (P4 = His) and New World substrates (P4 = Asp/Glu).

While the VEEV nsP2 protease cleaved both the Old World SFV CFP-P1/2₂₅-YFP substrate and New World VEEV substrates, the Old World CHIKV nsP2 protease did not significantly cleave the New World VEEV CFP-P1/2₂₅-YFP, CFP-P2/3₂₅-YFP, and CFP-P3/4₂₅-YFP substrates (Supplemental Figure S4). The nsP2 protease of CHIKV (PDB 3TRK (31)) is structurally similar to the VEEV nsP2 protease (RMSD = 0.45 Å); however, two key residues in the VEEV nsP2 SAM MTase domain, R662 and K706, are replaced by N202 and D246 in the CHIKV nsP2 protease, respectively. As shown in the predicted binding model of the SFV P1/2 substrate with the CHIKV nsP2 protease (Supplemental Figure S4), the P4 residue (His or Arg) of the substrate may form a salt bridge interaction with D246. In contrast, the VEEV substrates possess an Asp or Glu at the P4 position that appears to make favorable interactions with the corresponding K705 and K706 residues in the VEEV nsP2 cysteine protease (32).

Effects of the N475A, R662A, R662K, and K480A mutations

VEEV contains a sequence motif, ⁴⁷⁵NVCWAK⁴⁸⁰, which differs from that of the papain-like cysteine proteases. An Asn is not present in the papain motif, but is present in the VEEV motif. The corona viral proteases, MERS PLpro (65) and SARS PLpro (66), also contain an Asn homologous to Asn-475. This Asn was proposed to stabilize the transition state (67). However, in our structure of the E64d-inhibited VEEV nsP2 cysteine protease the side chain of Asn-475 is directed towards the backbone carbonyl of Asp-507 on the SAM MTase domain (2.9 Å from the carbonyl oxygen) and no conformational changes are observed, suggesting that this residue likely does not stabilize the oxyanion via direct hydrogen bonding. In addition, the Asn-475 side chain is within hydrogen bonding distance of the Arg-662 guanidinium group, and this hydrogen bonding interaction remains intact in both the E64d structure and in our P3/4 substrate bound model. We mutated both Asn-475

and Arg-662 to examine their functional roles in substrate binding. The N475A mutation led to a 7-fold reduction in k_{cat} and a 4-fold increase in K_m , while mutation of Arg-662 to either Ala or Lys led to a 16-fold reduction in the k_{cat}/K_m and the reductions were predominantly due to reductions in k_{cat} (Table 3). Thus, consistent with the structural model, Asn-475 may play a role in positioning the substrate for nucleophilic attack, and the hydrogen bonding interaction between Asn-475 and Arg-662 may further stabilize substrate binding.

Lys-480 is located near the nucleophilic cysteine; its side chain amine is \sim 8 Å from the Cys-477-S γ . The motif residue is conserved among both Old and New World alphaviruses. In the free enzyme density for the side chain is observed, and the Lys-480 side chain is directed towards the oxidized (occupancy of 0.4) and reduced (occupancy of 0.6) Cys-477 conformers. In the E64d inhibited enzyme density for the side chain is poor suggesting static or dynamic disorder. Mutation of Lys-480 to alanine showed a 7-fold reduction in k_{cat} , but had no significant effect on the K_m of the substrate. The result indicated that, instead of participating in substrate binding, Lys-480 may enhance the nucleophilicity of the Cys-SH. The helix dipole has also been proposed to contribute to the microenvironment that stabilizes the thiolate, Cys-477 is found at the beginning of a helix and this structural feature may also contribute to the nucleophilicity of Cys-477 (59;68).

Effect of the N545A Mutation

MD simulations of the protein in the substrate-free state showed that the β -hairpin residues were rather flexible, but were significantly stabilized in the substrate binding complex due to binding interactions between N545 and the bound substrate. N545 was found to form binding interactions with the substrate in all predicted models (Figure 4, Supplemental Figure S2). The side chain NH of Asn-545 was within hydrogen bonding distance of the P3 carbonyl oxygen of the substrate (3.4 Å), and the backbone carbonyl oxygen of Asn-545 was within hydrogen bonding distance (2.8 Å) of the backbone NH of the P1 residue of the substrate. To examine the role of N545 we tested the N545A variant (Table 3). Mutation of N545 to alanine had minor effects on K_m and k_{cat} , indicating that N545 is involved in substrate binding mainly through backbone to backbone interactions rather than a side chain to backbone interactions. A hydrogen bond to the N545 backbone carbonyl oxygen was also observed in the E64d-bound complex (Figure 2). N545 is located at the end of the β -hairpin and precedes H546 which ion pairs with the C477 thiolate. N545 is more than 7.8 Å from H546 and does not appear to function as a catalytic triad residue.

DISCUSSION

A fundamental difference between cysteine and serine proteases was shown by Asboth et al. using thiono ester substrates where the C=O had been substituted by a C=S (69;70). Sulfur is a poorer hydrogen bond acceptor than oxygen. Papain was able to hydrolyze a C=S substrate analogue, whereas in a similar experiment carried out with serine hydrolases no detectable hydrolysis was observed (70). A second key difference noted by Asboth et al. is the separation of charges in the ground state vs. the TS (71). In serine hydrolases the ground state is neutral and a separation of charges occurs during the TS (i.e. the His acquires a proton, and the oxyanion forms) and the oxyanion must be stabilized by hydrogen bonds. In

cysteine proteases a negatively charged thiolate is present in the ground state and ion pairs with an imidazolium of a His; in the TS the negative charge of the Cys-S⁻ migrates to the carbonyl oxygen while the His remains protonated. Based upon their data Asboth et al. concluded that there was a minimal requirement for oxyanion hole residues in cysteine proteases (69–71).

The k_{cat}/K_m of the VEEV nsP2 cysteine protease is ~600-fold lower than that of papain. In vitro studies have shown that the papain like Cys protease of MERS also has a relatively low k_{cat}/K_m when compared with papain (72). This suggested the absence or substitution of an active site residue(s), or the absence of an intracellular activator (e.g. calpains are activated by calcium). Papain utilizes Gln-19 and the backbone NH of the nucleophilic Cys to stabilize its transition state (Figure 5A). Mutation of Gln-19 to serine (Q19S) reduces its k_{cat}/K_m 500-fold (73). The Clan CN VEEV cysteine protease lacks a residue homologous to the transition state stabilizing residue, Q19, in papain. Interestingly, the VEEV nsP2 cysteine protease is not the only viral cysteine protease lacking a residue homologous to Gln-19 in papain. Trp-107 in the SARS PLpro is thought to stabilize the oxyanion (72) in a manner similar to Gln-19 in papain. Trp-107 was mutated to Ala and no activity was observed (67) (Figure 5B). Lin et al. noted that in the MERS PLpro the Trp is replaced by a Leu, a residue which cannot function as a hydrogen bond donor (Figure 5C), and then showed that the mutation of the Leu to Trp actually increased the k_{cat}/K_m 23-fold (72). Consequently, the WT MERS PLpro has a lower k_{cat}/K_m than the SARS PLpro (72). MERS has a higher mortality rate than the SARS virus suggesting that the slower rate of turnover is not necessarily detrimental to the virus.

The low turnover numbers of these viral proteases may be of value to on-going drug discovery efforts (18;19) since the slowest step in replication governs the rate of the entire process, and the inhibition of a slow step may lead to a more apparent effect in a cell. The lack of TS stabilizing residues may not only impact turnover, but also inhibition of the enzyme since most cysteine protease inhibitors are covalent. For instance, the lack of inhibition of the VEEV nsP2 by 200 μ M E64c may be associated with the absence of a residue homologous to papain's Gln-19. Gln-19 is within hydrogen bonding distance to the carboxylic acid of E64, the compound from which E64d and E64c are derived (53).

Mutagenesis studies have been traditionally used to identify residues important to catalysis; however, the conformation of a residue is also important in the assignment of a proposed function. Asn-475 is conserved in both Old and New World alphaviruses, and was investigated here for a potential role in TS stabilization. In the structure of the free enzyme, the E64d adduct, and in one of our P34 models we found the Asn-475 side chain hydrogen bonded to the side chain of Arg-662 and to the backbone carbonyl oxygen of Asp-507 (Figure 5D). This residue did not rotate towards the carbonyl oxygen of the E64d inhibitor. In the Clan CN VEEV nsP2 cysteine protease only one backbone amide NH was within hydrogen bonding distance of the carbonyl oxygen of the E64d inhibitor consistent with minimal stabilization of the oxyanion in the TS (Figure 5). In the SARS PLpro Asn-110 which is homologous to Asn-475 had been proposed to play a possible role in TS stabilization (67), but similarly was not observed to rotate towards the carbonyl oxygen in the substrate bound C112S variant (PDB 4M0W) (74). While the N475A mutation affects

 k_{cat} and k_{cat}/K_m , the orientation of this side chain is not consistent with oxyanion stabilization by direct hydrogen bonding. Asn-475 forms the base of the P1 binding site and likely orients the substrate for nucleophilic attack or plays a role in product release. Notably, the corona virus motif (DNNCYL) of the Middle East respiratory syndrome (MERS-CoV PL^{pro}) and severe acute respiratory syndrome (SARS) papain like proteases (66;72) also contains an Asn at the n-2 position relative to the nucleophilic Cys, and this Asn aligns with Asn-475 in the Clan CN motif (475 NVCWAK 480) when the Cys-His ion pairs of these proteases are superposed (Figure 5).

The Clan CN proteases characteristically contain a papain-like cysteine protease domain and a SAM MTase domain. While peptide substrates have been previously modeled at the interface of these two domains, the substrate binding site had not been confirmed by structural or mutagenesis studies. Using a covalent peptide-like inhibitor and X-ray crystallography we determined the first inhibitor bound structure of a Clan CN protease and confirmed that the substrate binding site is complemented by the SAM MTase domain. We have shown that at least three residues, R662, K705, and K706, within the SAM MTase domain are used by the cysteine protease for substrate binding and recognition. Mutation of these residues led to significant reductions in k_{cat}/K_m (12 to 60-fold) and underscored the importance of the SAM MTase domain to protease activity. The structural and kinetic characterization of key active site residues within the conserved motif, not only provides insights into the catalytic mechanism of the nsP2 cysteine proteinase of VEEV, but should also assist in structure-based development of novel and potent antiviral drugs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Dr. Jim Schmidt (USAMRIID) for his assistance with the optimization of the substrates developed herein, and Drs. Ward Tucker (Bio Sentinel Pharmaceuticals, Inc.) and David Waugh (NCI) for helpful discussions. This work was supported by DTRA project number CB-SEED-SEED09-2-0061 and CBCall4-CBM-05-2-0019. The opinions expressed here are those of the authors and do not represent those of the U.S. Army, U.S. Navy, U.S. Department of Defense or the U.S. government.

References

- 1. Strauss JH, Strauss EG. The alphaviruses: gene expression, replication, and evolution. Microbiol Rev. 1994; 58:491–562. [PubMed: 7968923]
- 2. Deardorff ER, Forrester NL, Travassos-da-Rosa AP, Estrada-Franco JG, Navarro-Lopez R, Tesh RB, Weaver SC. Experimental infection of potential reservoir hosts with Venezuelan equine encephalitis virus, Mexico. Emerg Infect Dis. 2009; 15:519–525. [PubMed: 19331726]
- Zacks MA, Paessler S. Encephalitic alphaviruses. Vet Microbiol. 2010; 140:281–286. [PubMed: 19775836]
- Schafer A, Brooke CB, Whitmore AC, Johnston RE. The role of the blood-brain barrier during Venezuelan equine encephalitis virus infection. J Virol. 2011; 85:10682–10690. [PubMed: 21849461]
- Johnson KM, Martin DH. Venezuelan equine encephalitis. Adv Vet Sci Comp Med. 1974; 18:79– 116. [PubMed: 4609399]

 Deresiewicz RL, Thaler SJ, Hsu L, Zamani AA. Clinical and neuroradiographic manifestations of eastern equine encephalitis. N Engl J Med. 1997; 336:1867–1874. [PubMed: 9197215]

- 7. Encyclopedia of microbiology. Elsevier; New York: 2009.
- Suhrbier A, Jaffar-Bandjee MC, Gasque P. Arthritogenic alphaviruses--an overview. Nat Rev Rheumatol. 2012; 8:420–429. [PubMed: 22565316]
- 9. Ryman KD, Klimstra WB. Host responses to alphavirus infection. Immunol Rev. 2008; 225:27–45. [PubMed: 18837774]
- Steele, KE., Reed, D., Glass, PJ., Hart, MK., Ludwig, GV., Pratt, WD., Parker, MD., Smith, JF. Alphavirus Encephalitides. In: Dembek, ZF., editor. Medical Aspects of Biological Warfare. Office of the Surgeon General; Falls Church: 2007. p. 241-270.
- 11. Hanson RP, Sulkin SE, Beuscher EL, Hammon WM, MCKINNEY RW, Work TH. Arbovirus infections of laboratory workers. Extent of problem emphasizes the need for more effective measures to reduce hazards. Science. 1967; 158:1283–1286. [PubMed: 6058003]
- 12. Ehrlich R, Miller S. Effect of relative humidity and temperature on airborne Venezuelan equine encephalitis virus. Appl Microbiol. 1971; 22:194–199. [PubMed: 4999523]
- Steele KE, Twenhafel NA. REVIEW PAPER: pathology of animal models of alphavirus encephalitis. Vet Pathol. 2010; 47:790–805. [PubMed: 20551475]
- 14. Ryzhikov AB, Tkacheva NV, Sergeev AN, Ryabchikova EI. Venezuelan equine encephalitis virus propagation in the olfactory tract of normal and immunized mice. Biomed Sci. 1991; 2:607–614. [PubMed: 1841630]
- Pittman PR, Makuch RS, Mangiafico JA, Cannon TL, Gibbs PH, Peters CJ. Long-term duration of detectable neutralizing antibodies after administration of live-attenuated VEE vaccine and following booster vaccination with inactivated VEE vaccine. Vaccine. 1996; 14:337–343.
 [PubMed: 8744562]
- Cole FE Jr, May SW, Eddy GA. Inactivated Venezuelan equine encephalomyelitis vaccine prepared from attenuated (TC-83 strain) virus. Appl Microbiol. 1974; 27:150–153. [PubMed: 4809906]
- Jahrling PB, Stephenson EH. Protective efficacies of live attenuated and formaldehyde-inactivated Venezuelan equine encephalitis virus vaccines against aerosol challenge in hamsters. J Clin Microbiol. 1984; 19:429–431. [PubMed: 6715512]
- 18. Chung DH, Jonsson CB, Tower NA, Chu YK, Sahin E, Golden JE, Noah JW, Schroeder CE, Sotsky JB, Sosa MI, Cramer DE, McKellip SN, Rasmussen L, White EL, Schmaljohn CS, Julander JG, Smith JM, Filone CM, Connor JH, Sakurai Y, Davey RA. Discovery of a novel compound with anti-venezuelan equine encephalitis virus activity that targets the nonstructural protein 2. PLoS Pathog. 2014; 10:e1004213. [PubMed: 24967809]
- Kehn-Hall K, Narayanan A, Lundberg L, Sampey G, Pinkham C, Guendel I, Van DR, Senina S, Schultz KL, Stavale E, Aman MJ, Bailey C, Kashanchi F. Modulation of GSK-3beta activity in Venezuelan equine encephalitis virus infection. PLoS One. 2012; 7:e34761. [PubMed: 22496857]
- 20. Julander JG, Bowen RA, Rao JR, Day C, Shafer K, Smee DF, Morrey JD, Chu CK. Treatment of Venezuelan equine encephalitis virus infection with (–)-carbodine. Antiviral Res. 2008; 80:309–315. [PubMed: 18675850]
- 21. Ding MX, Schlesinger MJ. Evidence that Sindbis virus NSP2 is an autoprotease which processes the virus nonstructural polyprotein. Virology. 1989; 171:280–284. [PubMed: 2525839]
- 22. Hardy WR, Strauss JH. 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. J Virol. 1989; 63:4653–4664. [PubMed: 2529379]
- 23. Vasiljeva L, Merits A, Golubtsov A, Sizemskaja V, Kaariainen L, Ahola T. Regulation of the sequential processing of Semliki Forest virus replicase polyprotein. J Biol Chem. 2003; 278:41636–41645. [PubMed: 12917405]
- Garmashova N, Gorchakov R, Frolova E, Frolov I. Sindbis virus nonstructural protein nsP2 is cytotoxic and inhibits cellular transcription. J Virol. 2006; 80:5686–5696. [PubMed: 16731907]
- Garmashova N, Gorchakov R, Volkova E, Paessler S, Frolova E, Frolov I. The Old World and New World alphaviruses use different virus-specific proteins for induction of transcriptional shutoff. J Virol. 2007; 81:2472–2484. [PubMed: 17108023]

 Mayuri Geders TW, Smith JL, Kuhn RJ. Role for conserved residues of sindbis virus nonstructural protein 2 methyltransferase-like domain in regulation of minus-strand synthesis and development of cytopathic infection. J Virol. 2008; 82:7284–7297. [PubMed: 18495773]

- 27. Petrakova O, Volkova E, Gorchakov R, Paessler S, Kinney RM, Frolov I. Noncytopathic replication of Venezuelan equine encephalitis virus and eastern equine encephalitis virus replicons in Mammalian cells. J Virol. 2005; 79:7597–7608. [PubMed: 15919912]
- 28. Montgomery SA, Johnston RE. Nuclear import and export of Venezuelan equine encephalitis virus nonstructural protein 2. J Virol. 2007; 81:10268–10279. [PubMed: 17652399]
- 29. Karpe YA, Lole KS. Deubiquitination activity associated with hepatitis E virus putative papain-like cysteine protease. J Gen Virol. 2011; 92:2088–2092. [PubMed: 21653754]
- 30. Russo AT, White MA, Watowich SJ. The crystal structure of the Venezuelan equine encephalitis alphavirus nsP2 protease. Structure. 2006; 14:1449–1458. [PubMed: 16962975]
- 31. Cheung J, Franklin M, Mancia F, Rudolph M, Cassidy M, Gary E, Burshteyn F, Love J. Structure of the Chikungunya virus nsP2 protease. 2011
- 32. Russo AT, Malmstrom RD, White MA, Watowich SJ. Structural basis for substrate specificity of alphavirus nsP2 proteases. J Mol Graph Model. 2010; 29:46–53. [PubMed: 20483643]
- 33. Donnelly MI, Zhou M, Millard CS, Clancy S, Stols L, Eschenfeldt WH, Collart FR, Joachimiak A. An expression vector tailored for large-scale, high-throughput purification of recombinant proteins. Protein Expr Purif. 2006; 47:446–454. [PubMed: 16497515]
- 34. Legler PM, Cai M, Peterkofsky A, Clore GM. Three-dimensional solution structure of the cytoplasmic B domain of the mannitol transporter IImannitol of the Escherichia coli phosphotransferase system. J Biol Chem. 2004; 279:39115–39121. [PubMed: 15258141]
- 35. LaVallie ER, Lu Z, Blasio-Smith EA, Collins-Racie LA, McCoy JM. Thioredoxin as a fusion partner for production of soluble recombinant proteins in Escherichia coli. Methods Enzymol. 2000; 326:322–340. [PubMed: 11036651]
- LaVallie ER, Blasio-Smith EA, Collins-Racie LA, Lu Z, McCoy JM. Thioredoxin and related proteins as multifunctional fusion tags for soluble expression in E. coli. Methods Mol Biol. 2003; 205:119–140. [PubMed: 12491883]
- 37. Dong M, Tepp WH, Johnson EA, Chapman ER. Using fluorescent sensors to detect botulinum neurotoxin activity in vitro and in living cells. Proc Natl Acad Sci U S A. 2004; 101:14701–14706. [PubMed: 15465919]
- 38. Ruge DR, Dunning FM, Piazza TM, Molles BE, Adler M, Zeytin FN, Tucker WC. Detection of six serotypes of botulinum neurotoxin using fluorogenic reporters. Anal Biochem. 2011; 411:200–209. [PubMed: 21216216]
- 39. Steinkuhler C, Urbani A, Tomei L, Biasiol G, Sardana M, Bianchi E, Pessi A, De FR. Activity of purified hepatitis C virus protease NS3 on peptide substrates. J Virol. 1996; 70:6694–6700. [PubMed: 8794305]
- 40. Kitz R, Wilson IB. Esters of methanesulfonic acid as irreversible inhibitors of acetylcholinesterase. J Biol Chem. 1962; 237:3245–3249. [PubMed: 14033211]
- 41. Lee H, Torres J, Truong L, Chaudhuri R, Mittal A, Johnson ME. Reducing agents affect inhibitory activities of compounds: results from multiple drug targets. Anal Biochem. 2012; 423:46–53. [PubMed: 22310499]
- 42. Navaza J. Implementation of molecular replacement in AMoRe. Acta Crystallogr D Biol Crystallogr. 2001; 57:1367–1372. [PubMed: 11567147]
- 43. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
- 44. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010; 66:486–501. [PubMed: 20383002]
- 45. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr. 1998; 54:905–921. [PubMed: 9757107]
- 46. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr. 1994; 50:760–763. [PubMed: 15299374]

47. Spurgers KB, Hurt CR, Cohen JW, Eccelston LT, Lind CM, Lingappa VR, Glass PJ. Validation of a cell-based ELISA as a screening tool identifying anti-alphavirus small-molecule inhibitors. J Virol Methods. 2013; 193:226–231. [PubMed: 23764417]

- 48. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009; 30:2785–2791. [PubMed: 19399780]
- Paessler S, Yun NE, Judy BM, Dziuba N, Zacks MA, Grund AH, Frolov I, Campbell GA, Weaver SC, Estes DM. Alpha-beta T cells provide protection against lethal encephalitis in the murine model of VEEV infection. Virology. 2007; 367:307–323. [PubMed: 17610927]
- 50. Huang Z, McGowan EB, Detwiler TC. Ester and amide derivatives of E64c as inhibitors of platelet calpains. J Med Chem. 1992; 35:2048–2054. [PubMed: 1597856]
- 51. Tamai M, Matsumoto K, Omura S, Koyama I, Ozawa Y, Hanada K. In vitro and in vivo inhibition of cysteine proteinases by EST, a new analog of E-64. J Pharmacobiodyn. 1986; 9:672–677. [PubMed: 3023601]
- 52. Hook V, Hook G, Kindy M. Pharmacogenetic features of cathepsin B inhibitors that improve memory deficit and reduce beta-amyloid related to Alzheimer's disease. Biol Chem. 2010; 391:861–872. [PubMed: 20536395]
- Matsumoto K, Mizoue K, Kitamura K, Tse WC, Huber CP, Ishida T. Structural basis of inhibition of cysteine proteases by E-64 and its derivatives. Biopolymers. 1999; 51:99–107. [PubMed: 10380357]
- 54. Yamamoto A, Tomoo K, Matsugi K, Hara T, In Y, Murata M, Kitamura K, Ishida T. Structural basis for development of cathepsin B-specific noncovalent-type inhibitor: crystal structure of cathepsin B-E64c complex. Biochim Biophys Acta. 2002; 1597:244–251. [PubMed: 12044902]
- McDonald JK, Kadkhodayan S. Cathepsin L--a latent proteinase in guinea pig sperm. Biochem Biophys Res Commun. 1988; 151:827–835. [PubMed: 3348813]
- 56. Noda T, Isogai K, Katunuma N, Tarumoto Y, Ohzeki M. Effects of cathepsin B, H, and D in pectoral muscle of dystrophic chickens (line 413) of in vivo administration of E-64-c (N-[N-(L-3-transcarboxyoxirane-2-carbonyl)-L-leucyl]-3-methyl-butylamine). J Biochem. 1981; 90:893–896. [PubMed: 7309707]
- KITZ R, WILSON IB. Esters of methanesulfonic acid as irreversible inhibitors of acetylcholinesterase. J Biol Chem. 1962; 237:3245–3249. [PubMed: 14033211]
- 58. Yamamoto D, Matsumoto K, Ohishi H, Ishida T, Inoue M, Kitamura K, Mizuno H. Refined x-ray structure of papain. E-64-c complex at 2.1-A resolution. J Biol Chem. 1991; 266:14771–14777. [PubMed: 1860874]
- Storer AC, Menard R. Catalytic mechanism in papain family of cysteine peptidases. Methods Enzymol. 1994; 244:486–500. [PubMed: 7845227]
- Shin G, Yost SA, Miller MT, Elrod EJ, Grakoui A, Marcotrigiano J. Structural and functional insights into alphavirus polyprotein processing and pathogenesis. Proc Natl Acad Sci U S A. 2012; 109:16534–16539. [PubMed: 23010928]
- 61. Frolov I, Agapov E, Hoffman TA Jr, Pragai BM, Lippa M, Schlesinger S, Rice CM. Selection of RNA replicons capable of persistent noncytopathic replication in mammalian cells. J Virol. 1999; 73:3854–3865. [PubMed: 10196280]
- 62. Zhang D, Tozser J, Waugh DS. Molecular cloning, overproduction, purification and biochemical characterization of the p39 nsp2 protease domains encoded by three alphaviruses. Protein Expr Purif. 2009; 64:89–97. [PubMed: 19013248]
- 63. Tan, SL. Hepatitis C viruses genomes and molecular biology. Horizon Bioscience; Wymondham, Norfolk, U.K: 2006.
- 64. Smith, HJ., Simons, C. Enzymes and their inhibition drug development. CRC Press; Boca Raton: 2005.
- 65. Lei J, Mesters JR, Drosten C, Anemuller S, Ma Q, Hilgenfeld R. Crystal structure of the papainlike protease of MERS coronavirus reveals unusual, potentially druggable active-site features. Antiviral Res. 2014; 109:72–82. [PubMed: 24992731]

66. Ratia K, Kilianski A, Baez-Santos YM, Baker SC, Mesecar A. Structural Basis for the Ubiquitin-Linkage Specificity and delSGylating activity of SARS-CoV papain-like protease. PLoS Pathog. 2014; 10:e1004113. [PubMed: 24854014]

- 67. Ratia K, Saikatendu KS, Santarsiero BD, Barretto N, Baker SC, Stevens RC, Mesecar AD. Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme. Proc Natl Acad Sci U S A. 2006; 103:5717–5722. [PubMed: 16581910]
- 68. Harris TK, Turner GJ. Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. IUBMB Life. 2002; 53:85–98. [PubMed: 12049200]
- 69. Asboth B, Stokum E, Khan IU, Polgar L. Mechanism of action of cysteine proteinases: oxyanion binding site is not essential in the hydrolysis of specific substrates. Biochemistry. 1985; 24:606–609. [PubMed: 3888259]
- Asboth B, Polgar L. Transition-state stabilization at the oxyanion binding sites of serine and thiol
 proteinases: hydrolyses of thiono and oxygen esters. Biochemistry. 1983; 22:117–122. [PubMed:
 6338911]
- 71. Polgar L, Asboth B. The basic difference in catalyses by serine and cysteine proteinases resides in charge stabilization in the transition state. J Theor Biol. 1986; 121:323–326. [PubMed: 3540454]
- 72. Lin MH, Chuang SJ, Chen CC, Cheng SC, Cheng KW, Lin CH, Sun CY, Chou CY. Structural and functional characterization of MERS coronavirus papain-like protease. J Biomed Sci. 2014; 21:54. [PubMed: 24898546]
- 73. Carrière, J. Characterization of Oxyanion Hole Mutants of the Cysteine Proteases Papain and Cathepsin B. McGill University; 1992.
- 74. Chou CY, Lai HY, Chen HY, Cheng SC, Cheng KW, Chou YW. Structural basis for catalysis and ubiquitin recognition by the severe acute respiratory syndrome coronavirus papain-like protease. Acta Crystallogr D Biol Crystallogr. 2014; 70:572–581. [PubMed: 24531491]
- 75. Lei H, Santarsiero BD, Lee H, Johnson ME. Crystal structure of PLpro from Middle East Respiratory Syndrome (MERS) coronavirus, PDB 4PT5 ed. Protein Data Bank. 2014
- 76. Schechter I, Berger A. On the size of the active site in proteases. I. Papain. 1967. Biochem Biophys Res Commun. 2012; 425:497–502. [PubMed: 22925665]

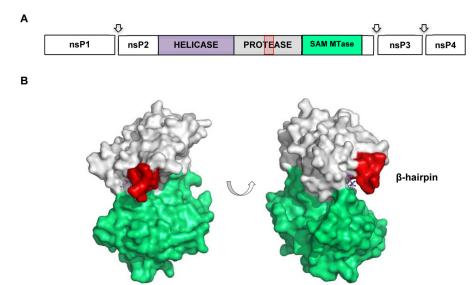


Figure 1. Organization of the VEEV nonstructural protein 2 (nsP2). (A) The nsP2 contains: an N-terminal region of unknown function, a helicase domain, a cysteine protease domain, and a SAM methyltransferase (MTase) domain. The nsP2 cysteine protease cleaves the nonstructural polypeptide into 4 segments (nsP1, nsP2, nsP3, nsP4). The cleavage is essential for viral replication. (B) The cysteine protease and SAM MTase domains are shown in white and green, respectively. The substrate is thought to bind beneath a β-hairpin (colored red) which forms a beak-like protrusion. The peptide-like E64d inhibitor was found beneath the β-hairpin and is shown in blue stick (PDB 5EZS).

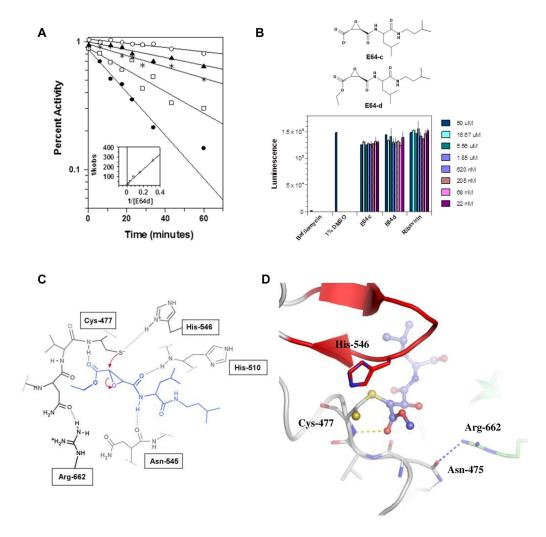


Figure 2.

(A) Time-dependent inhibition of the VEEV nsP2 protease by E64d. (B) Structure of the E64d prodrug. In cells esterases convert the E64d ester to the E64c acid. E64c is a prototypical cysteine protease inhibitor, but did not significantly inhibit the protease. E64d inhibited the enzyme *in vitro*, however, no inhibition of viral replication was observed with E64d or E64c in cell-based assays. (C) Proposed mechanism for covalent inhibition based upon the observed density in the E64d-inhibited enzyme structure. (D) Structure of the E64d VEEV nsP2 cysteine protease adduct (PDB 5EZS). The carbonyl oxygen of the ester was within hydrogen bonding distance of the Cys-477-NH (2.8 Å), this interaction likely stabilizes the TS in the enzyme-catalyzed proteolytic reaction. Other nearby NH groups were not directed towards the carbonyl oxygen. Asn-475 and Arg-662 were found hydrogen bonded in the adduct and no conformational change occurred in these residues upon inhibitor binding.

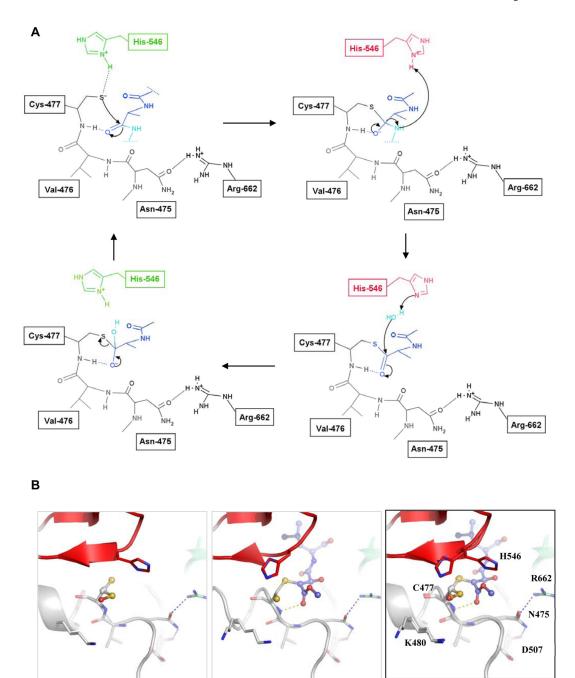


Figure 3. Proposed mechanism of the VEEV nsP2 Clan CN cysteine protease. (A) The mechanism is based upon the structures of the bound (PDB 5EZS) and free (PDB 5EZQ) enzyme described here. His-546 was found in two different conformations (lower panels) which may be relevant to different steps in the catalytic cycle. In one conformation His-546 could form an ion pair with the Cys-S γ thiolate. In the free enzyme the side chain of Lys-480 was also directed towards the Cys-S γ . During the catalytic cycle His-546 may rotate away from the Cys-S γ in order to donate its proton to the amide nitrogen and facilitate the collapse of the tetrahedral TS and release of the free amine. (B) In overlays, the second conformation of the

His-546-N81 was near the E64d ester oxygen and was well positioned to donate its hydrogen. Notably, the Asn-475 side chain did not rotate towards the carbonyl oxygen. Asn-475 forms the base of the P1 binding site and remained hydrogen bonded to Arg-662 and to the backbone carbonyl oxygen of Asp-507.

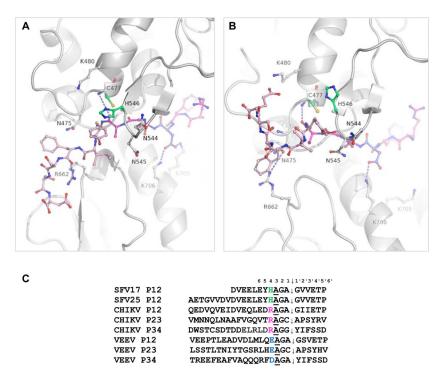


Figure 4.

Binding models of the VEEV nsP2 P34 substrates. (A) The C-terminal portion of the substrate (colored in light pink) was found directed downwards towards the SAM MTase domain, or (B) upwards towards the protease domain. The N-terminal portion of the substrate is colored in magenta. In both models the P4 residue (color in blue) interacted with K706. In the P34 model on the right the carbonyl of the scissile bond was closer to the Cys-Sγ. The hydrogen bond between R662 and N475 was maintained in the model on the right, and broken in the other. Other substrate bound models can be found in the Supplemental Information. (C) The 25-residue cleavage site sequences which were embedded between the YFP and CFP fluorescent proteins to make the substrates are shown for VEEV. The 17- and 25-residue SFV substrate sequences are also shown. The cleavage sequences for CHIKV are shown for comparison. The nomenclature of Berger and Schechter is used to identify residues on the amino (P1, P2, etc.) or carboxy (P1', P2', etc.) termini of the scissile bond. The P4 residue is colored, and the arrow indicates the location of the cleavage site. Cleavage sites in between the nonstructural proteins contain a common motif, $AG(A/C) \downarrow (G/Y/A)$. The P4 residue (colored) differs significantly between the Old and New World alphaviruses and is thought to be specifically recognized by residues in the SAM MTase domain (K705 and K706 in the VEEV nsP2).

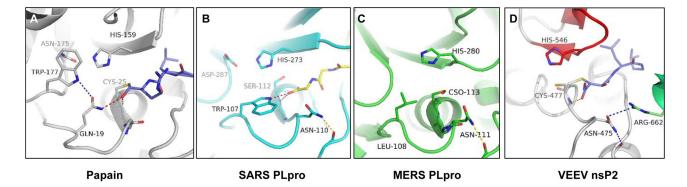


Figure 5.

Comparison of the active sites of (A) papain, the corona viral PLpro proteases from (B) SARS (C112S variant, PDB 4M0W) (74) and (C) MERS (PDB 4PT5) (75), and the (D) alphaviral nsP2 cysteine protease of VEEV (PDB 5EZS). Cysteine proteases utilize the backbone NH of the Cys and in some cases the side chains of various residues to stabilize the transition state. The architecture of the active site is distinctly different from that of a serine hydrolase which utilizes two to three backbone NH groups to form an oxyanion hole. Notably, the Asn-110 in the SARS protease, the Asn-111 in the MERS protease, and the Asn-475 in the VEEV protease are present at the n-2 position relative to the Cys and are

conserved. The Asn does not hydrogen bond to the carbonyl oxygen of the substrate or inhibitor in the SARS PLpro or in the VEEV nsP2 cysteine protease, respectively.

Hu et al.

Table 1

X-ray crystallography data collection and refinement statistics.

Page 25

	VEEV WT (PDB 5EZQ)	VEEV WT E64d (PDB 5EZS)
Space group	P2(1)2(1)2(1)	P2(1)2(1)2(1)
Unit Cell Dimensions (Å)	60.81, 63.72, 86.08	60.89, 63.63, 86.75
Wavelength (Å)	1.54	1.54
Resolution Range (Å) ^a	63.72-1.66 (1.76-1.66)	60.89-2.16 (2.26-2.16)
Unique Reflections	40,293 (6,260)	18,605 (2,273)
R _{sym} b	0.046 (0.348)	0.097 (0.266)
$I/\sigma I$	21.9 (3.2)	14.2 (4.9)
Completeness	99.6 (97.7)	99.2 (96.2)
Redundancy	6.01 (2.95)	7.3 (4.0)
Refinement Statistics:		
Resolution (Å)	51.21-1.66	51.31-2.16
No. of reflections	38,017	17,649
$R_{factor}^{\mathcal{C}}$	0.195	0.192
R_{free}^{d}	0.211 (5%)	0.227 (5%)
Number of Atoms:		
Protein	2567	2579
Solvent	218	208
Other	0	21
Average B-factors (Å ²)		
Protein	16.9	17.4
Solvent	24.6	23.1
R.m.s.d. from ideal geometry:		
Bond lengths (Å)	0.008	0.010
Bond angles (degrees)	1.22	1.383
Ramachandran plot		
Most favored regions (%)	91.8%	90.4%
Additional allowed regions (%)	8.2%	9.6%
Generously allowed regions (%)	0.0%	0.0%
Disallowed regions (%)	0.0%	0.0%

 $^{^{}a}$ Values in parentheses are for the outer most data shell

 $b_{\rm Rsym~was~calculated~from}\,R_{sym} {=}^{\sum |I_i - < I>} \big| \Big/_{\sum I_i}$

 $c_{\mbox{Rfactor} \mbox{ for working set of reflections was calculated using:}} R_{factor} = \sum \|F_o| - |F_c| / \sum |F_o|$

 $d_{\mbox{\sc Rfree}}$ for test set and size of test set as % total reflections in parentheses.

Hu et al.

Table 2

Steady state kinetic parameters for the tag-free WT VEEV nsP2 measured for the P12, P23, and P34 substrates in 0.1 M sodium phosphate buffer pH 6.0 at R.T. $(23 \pm 3 \, ^{\circ}\text{C})$.

Enzyme	V _{max} U/mg	${f k}_{ m cat}$ ${f min}^{-1}$	$K_{m}\\\mu M$	k_{cat}/K_m F $1/min*\mu M$	Fold Change in k _{cat}	Fold Change in K _m	Fold Change in $k_{\rm cat}$ – Fold Change in K_m – Fold Difference in $k_{\rm cat}/K_m$
P12	0.135 ± 0.008	5.2 ± 0.3	110 ± 10	0.135 ± 0.008 5.2 ± 0.3 110 ± 10 0.047 ± 0.005	-	-	
P23	0.016 ± 0.007 0.6 ± 0.3	0.6 ± 0.3	260 ± 150	$260 \pm 150 0.002 + 0.002$	9-fold	2-fold	24-fold
P34	0.081 ± 0.008	3.1 ± 0.3	180 ± 30	3.1 ± 0.3 180 ± 30 0.017 ± 0.003	2-fold	2-fold	3-fold

Page 26

Hu et al.

Page 27

Table 3

Steady State Kinetic parameters of the tag-free WT VEEV nsP2 and its variant and the P12 substrate. Parameters were measured in 0.1 M sodium phosphate pH 6.0 buffer at R.T. (23 \pm 3 °C).

Enzyme	V _{max} U/mg	$\mathbf{k}_{\mathrm{cat}}$ min $^{-1}$	К _т µМ	k_{cat}/K_{m} 1/min* μM	Fold Change in k _{cat}	Fold Change in K _m	Fold Change in $k_{\rm cat}$ – Fold Change in $K_{\rm m}$ – Fold Change in $k_{\rm cat}/K_{\rm m}$
WT	0.135 ± 0.008	5.2 ± 0.3	110 ± 10	0.047 ± 0.005		-	-
N475A	0.02 ± 0.01	0.8 ± 0.4	400 ± 200	0.002 ± 0.001	7-fold	4-fold	24-fold
K480A	0.019 ± 0.005	0.7 ± 0.2	140 ± 50	0.005 ± 0.002	7-fold	1	9-fold
N545A	0.18 ± 0.03	7 ± 1	170 ± 30	0.041 ± 0.009	!	1	-
R662A	0.019 ± 0.005	0.7 ± 0.2	220 ± 70	0.003 ± 0.001	7-fold	2-fold	16-fold
R662K	0.008 ± 0.002	0.3 ± 0.1	110 ± 40	0.003 ± 0.001	17-fold	1	16-fold
K705E	0.03 ± 0.01	1.1 ± 0.4	300 ± 100	0.004 ± 0.002	5-fold	3-fold	12-fold
K706A	0.03 ± 0.01	1.3 ± 0.4	500 ± 200	0.003 ± 0.002	5-fold	5-fold	16-fold
K706Q	0.011 ± 0.002	0.42 ± 0.08	100 ± 30	0.004 ± 0.001	12-fold	1	12-fold
K705A/K706A	0.008 ± 0.003	0.3 ± 0.1	400 ± 200	0.0008 ± 0.0005	17-fold	4-fold	ploJ-09
P713S	0.077 ± 0.005	2.9 ± 0.2	110 ± 10	0.026 ± 0.003	2-fold	1	2-fold