

Identification of novel proteolytically inactive mutations in coronavirus 3C-like protease using a combined approach

Junwei Zhou,^{*,†} Liurong Fang,^{*,†} Zhixiang Yang,^{*,†} Shangen Xu,^{*,†} Mengting Lv,^{*,†} Zheng Sun,^{*,†} Jiyao Chen,^{*,†} Dang Wang,^{*,†,1} Jun Gao,^{*,1,2} and Shaobo Xiao^{*,†,1,3}

^{*}State Key Laboratory of Agricultural Microbiology, College of Veterinary Medicine, and [†]Agricultural Bioinformatics Key Laboratory of Hubei Province, College of Informatics, Huazhong Agricultural University, Wuhan, China; and ¹Key Laboratory of Preventive Veterinary Medicine in Hubei Province, Cooperative Innovation Center for Sustainable Pig Production, Wuhan, China

ABSTRACT: Coronaviruses (CoVs) infect humans and multiple other animal species, causing highly prevalent and severe diseases. 3C-like proteases (3CL^{Pro}s) from CoVs (also called main proteases) are essential for viral replication and are also involved in polyprotein cleavage and immune regulation, making them attractive and effective targets for the development of antiviral drugs. Herein, the 3CL^{Pro} from the porcine epidemic diarrhea virus, an enteropathogenic CoV, was used as a model to identify novel crucial residues for enzyme activity. First, we established a rapid, sensitive, and efficient luciferase-based biosensor to monitor the activity of PDEV 3CL^{Pro} *in vivo*. Using this luciferase biosensor, along with confirming the well-known catalytic residues (His41 and Cys144), we identified 4 novel proteolytically inactivated mutants of PDEV 3CL^{Pro}, which was also confirmed in mammalian cells by biochemical experiments. Our molecular dynamics (MD) simulations showed that the hydrogen bonding interactions occurring within and outside of the protease's active site and the dynamic fluctuations of the substrate, especially the van der Waals contacts, were drastically altered, a situation related to the loss of 3CL^{Pro} activity. These data suggest that changing the intermolecular dynamics in protein-substrate complexes eliminates the mechanism underlying the protease activity. The discovery of novel crucial residues for enzyme activity in the binding pocket could potentially provide more druggable sites for the design of protease inhibitors. In addition, our in-depth study of the dynamic substrate's envelope model using MD simulations is an approach that could augment the discovery of new inhibitors against 3CL^{Pro} in CoVs and other viral 3C proteases.—Zhou, J., Fang, L., Yang, Z., Xu, S., Lv, M., Sun, Z., Chen, J., Wang, D., Gao, J., Xiao, S. Identification of novel proteolytically inactive mutations in coronavirus 3C-like protease using a combined approach. FASEB J. 33, 000–000 (2019). www.fasebj.org

KEY WORDS: biosensor · catalytic residue · molecular dynamics · molecular mechanism

Coronaviruses (CoVs) are important pathogens capable of causing severe, fatal, and highly prevalent diseases in humans and other animals (1, 2). Since the outbreak of severe

acute respiratory syndrome (SARS) CoV in 2003 (3) and the outbreak of Middle East respiratory syndrome CoV in 2012 (4, 5), CoVs have attracted more and more attention. CoVs are prone to genetic mutation, bringing about new variants and the reemergence of old ones. For example, porcine epidemic diarrhea virus (PEDV), a swine enteropathogenic CoV that causes lethal watery diarrhea in piglets, was first identified in the early 1970s (6). PEDV reemerged in 2010, with a large-scale outbreak in China that rapidly spread to the United States and other countries, resulting in enormous economic losses to the global pig farming industry (7). In addition, this emerging PEDV variant possesses the potential to infect humans, thereby posing a significant threat to public health (8). Although vaccines against PEDV have been developed, the continuous emergence of new serotypes and recombination events between field and vaccine strains mean that vaccination is only partially successful (9, 10).

CoVs 3C-like protease (3CL^{Pro}), which are also referred to as the main protease in these viruses, are encoded by

ABBREVIATIONS: 3CL^{Pro}, 3C-like protease; CoV, coronavirus; DnaE, catalytic subunit α of DNA polymerase III; HA, hemagglutinin; HCV, hepatitis C virus; HEK-293T, human embryonic kidney; MD, molecular dynamics; PEDV, porcine epidemic diarrhea virus; RMSD, root mean square deviation; SARS, severe acute respiratory syndrome; TEV, tobacco etch virus; vdW, van der Waals; WT, wild type

¹ These authors contributed equally to this work.

² Correspondence: Agricultural Bioinformatics Key Laboratory of Hubei Province, College of Informatics, Huazhong Agricultural University, 1 Shi-zi-shan St., Wuhan 430070, Hubei, P.R. China. E-mail: gaojun@mail.hzau.edu.cn

³ Correspondence: Laboratory of Animal Virology, College of Veterinary Medicine, Huazhong Agricultural University, 1 Shi-zi-shan St., Wuhan 430070, Hubei, P.R. China. E-mail: vet@mail.hzau.edu.cn

doi: 10.1096/fj.201901624RR

This article includes supplemental data. Please visit <http://www.fasebj.org> to obtain this information.

nonstructural protein 5 and are essential for viral replication. 3CL^{pro} in CoVs share highly conserved substrate recognition pockets, which are responsible for cleaving the viral polyprotein and the host factors involved in the innate immune response, including the signal transducer and activator of transcription 2 and the NF- κ B essential modulator signaling protein (11–14). Thus, targeting 3CL^{pro} serves as a 2-pronged attack on the virus by preventing viral maturation and restoring the natural immune response. One strategy used in the rational design of protease inhibitor drugs is to exploit the interactions occurring in the protease's active site, an approach mainly based on the in-depth study of the substrate–active site interaction. Furthermore, such inhibitors are often designed to be in close proximity with the catalytic residues in the protease active site to avoid drug resistance (14–16). Hence, the discovery of more crucial residues in protease active site could theoretically potentiate potential druggable sites. A series of inhibitors was reported to act against 3CL^{pro} from CoVs to prevent viral replication since the SARS outbreak in 2003 (17–22). Nevertheless, a theoretical understanding of proteases, substrate–active site interactions, and high-level resistance to protease inhibitors in viruses is not yet fully developed. Therapeutic options and treatment outcomes for patients infected with HIV or hepatitis C virus (HCV) have greatly benefited from structure and molecular dynamics (MD)–based drug design approaches, specifically with respect to viral protease inhibitor development. Moreover, the dynamic substrate envelope model with MD simulations has clearly explained the molecular mechanism of drug resistance in a clinically significant variant of the HCV and HIV proteases. Unlike the static information gained from crystal structures, the MD simulations in several studies permitted a detailed analysis of the interaction network, in terms of the direct interactions with substrate within the active site and the internal electrostatic network throughout the enzyme, both of which are reportedly critical requirements for tight substrate binding (14, 23–28).

When considering the development of protease inhibitors, the most important criterion is the ability to detect protease activity. However, the traditional methods often involve protein purification and enzyme activity inhibition experiments *in vitro*, which are inefficient and cannot meet the requirements of high-throughput screening *in vivo*. Therefore, there is an urgent need for a simple, efficient, and high-throughput method to detect protease activity at the cellular level in order to fully reflect the biologic characteristics of a protease. As a reporter protein, firefly luciferase is widely used to detect apoptosis and enzyme activity and is also used to screen for antiapoptotic drugs and identify enzyme recognition sequences. In theory, a firefly luciferase reporter–based approach would also allow for the identification and screening of the specific amino acids affecting the activity of a viral protease (29–32).

Consequently, in this study, we developed a combined strategy to identify novel proteolytically inactive mutants of a viral protease. Using PEDV 3CL^{pro} as the model, we established a luciferase-based biosensor to monitor protease activity in cells and identified 4 novel amino acids

essential for the activity of the PEDV 3CL^{pro} (Trp31, Phe139, Gly142, and His162). MD simulations were also performed on the wild-type (WT) or single-substitution variants of 3CL^{pro} to calculate the dynamic substrate envelopes. In agreement with the experimental loss in protease activity, the single-substitution mutants (W31A, F139A, G142A, and H162A) were seen to significantly disrupt the intermolecular hydrogen bonding network and intermolecular dynamic correlations for the active sites, thus affecting the intermolecular hydrogen bond network and the substrate binding affinity. Our results explain the potential molecular basis whereby the 3CL^{pro} mutants were proteolytically inactivated, thereby providing more potential target sites for drug design.

MATERIALS AND METHODS

Plasmids

The cDNA expression construct that encodes PEDV 3CL^{pro} and the luciferase reporter plasmids (233D and 358D) were previously described in refs. 13 and 33. The cDNA expression construct encoding PEDV 3CL^{pro} was PCR amplified and cloned into the C-terminal hemagglutinin (HA) tag–encoding pCAGGS-HA-C plasmid. First, secondary structures in CoV 3CL^{pro} were analyzed using ESPrpt (<http://esprpt.ibcp.fr/ESPrpt/ESPrpt/index.php>). Then, 7 aa sites were chosen based on the predicted amino acid interactions with an online method (<https://mistic2.leloir.org.ar/>). Mutagenesis of the PEDV 3CL^{pro} constructs (to produce W31A, C38A, H41A, F139A, G142A, C144A, G145A, Y160A, and H162A) was carried out by overlapping extension PCR using specific mutagenic primers. Luciferase reporter plasmids (233D and 358D), which contain oligonucleotides corresponding to ENLYFQ↓YS [cleaved by tobacco etch virus (TEV) 3C^{pro}], were used as the reporter controls (32). The construction strategy for the luciferase-based biosensor plasmids (233DP and 358DP) to monitor the activity of PEDV 3CL^{pro} was as follows. The DNA sequences encoding the N- and C-terminal halves of the catalytic subunit α of DNA polymerase III (DnaE), which encompass the protein's *trans*-splicing activity, were synthesized and cloned into the pCAGGS-multiple cloning site (MCS) vector to construct pCAGGS-DnaE. The sequences corresponding to the N-terminal fragments (aa 4–233) and the C-terminal fragments (aa 235–544) of firefly luciferase were PCR amplified from the firefly luciferase reporter vector pGL4.21luc2P/Puro (Promega, Madison, WI, USA). The sequences encoding aa 4–233 and 235–544, which were fused to the corresponding amino sequence YNSTLQ↓AGLRKM (the N-terminal auto-cleavage sequence in PEDV 3CL^{pro}), were cloned into pCAGGS-DnaE to create the 233DP reporter. The same construction strategy was used to generate the 358DP reporter (Fig. 1A). All the constructs were validated by DNA sequencing.

Luciferase reporter gene assays

Human embryonic kidney (HEK-293T) cells, obtained from the China Center for Type Culture Collection (Wuhan University, Wuhan, China), were cultured at 37°C in 5% CO₂ in DMEM (Thermo Fisher Scientific, Waltham, MA, USA) supplemented with 10% fetal bovine serum. The luciferase reporter constructs (233DP and 358DP) and their controls (233D and 358D, respectively) were used to detect PEDV-specific 3CL^{pro} activity. HEK-293T cells plated in 48-well plates were transfected with various 3CL^{pro} expression plasmids or the empty control plasmid, together with the luciferase reporter plasmid and pRL-TK (Promega), which was used as an

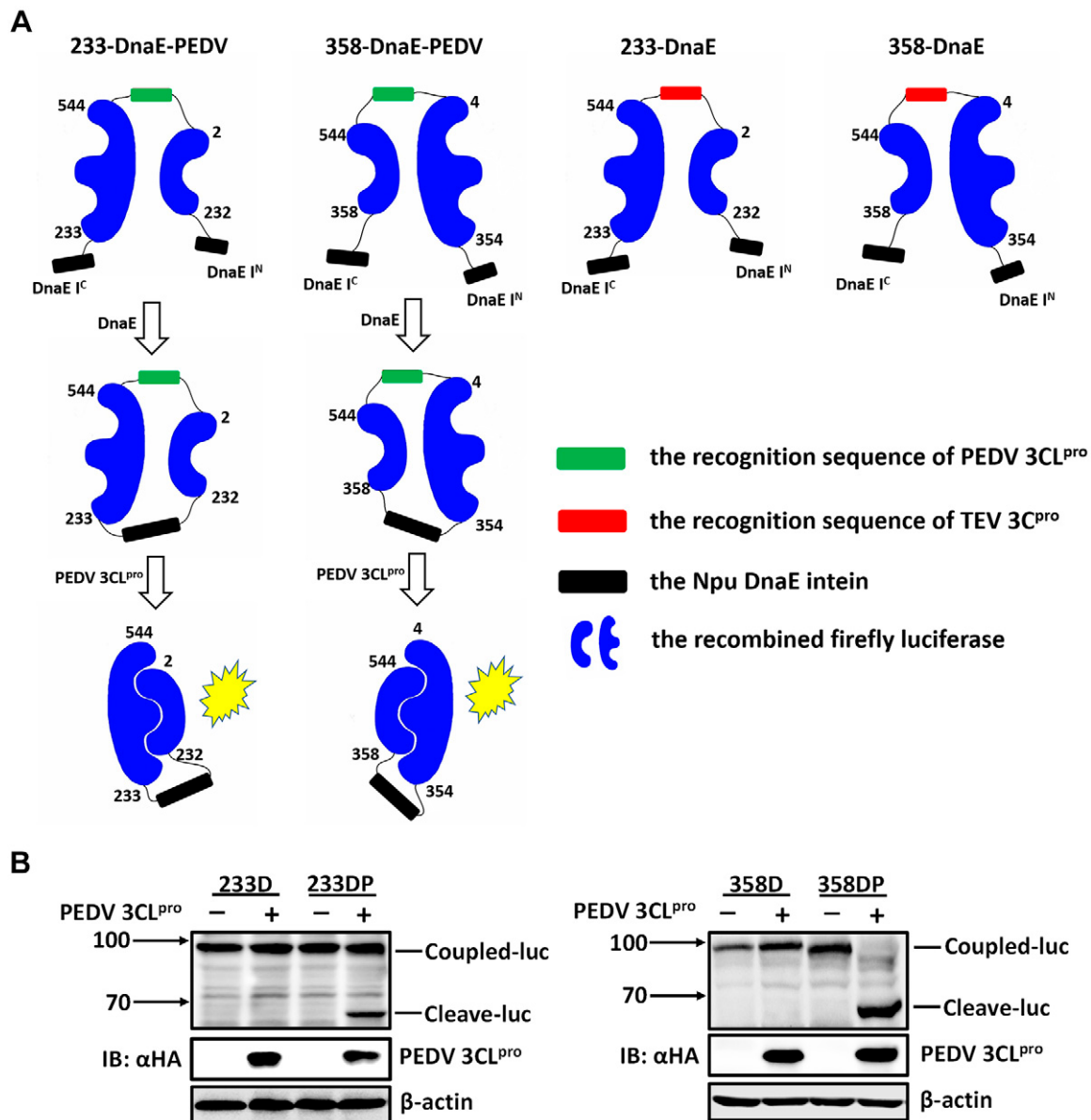


Figure 1. Exploiting the biosensor assay to evaluate PEDV 3CL^{pro} activity *in vivo*. **A)** Diagram showing the generation of 233DP and 358DP constructs and their controls (233D and 358D, respectively). The blue structure represents the peptide sequence of the recombinant firefly luciferase. The black rectangle represents the *Nostoc punctiforme* (Npu) DnaE intein (DnaE I) peptide sequence used to cyclize the protein. The green rectangle represents the YNSTLQLAGLRKM protease recognition sequence that was used to assess PEDV 3CL^{pro} activity, and the red rectangle represents the ENLYFQIYS protease recognition sequence for TEV 3C^{pro}, which was used as the control. **B)** HEK-293T cells were transfected with 233DP or 358DP, or the corresponding controls (233D and 358D, respectively) and the plasmid encoding PEDV 3CL^{pro}. After 30 h, cell lysates were prepared and analyzed by Western blotting. αHA, antihemagglutinin; IB, immunoblotting; luc, luciferase.

internal control to normalize the transfection efficiency. At 36 h post-transfection, the cells were lysed, and a luciferase reporter assay system (Promega) was utilized to determine the luciferase activities in the lysed cells. The activities were normalized to the corresponding *Renilla* luciferase activities.

Western blotting analysis

Briefly, HEK-293T cells cultured in 60-mm dishes were transfected with the various plasmids. After 30 h, the cells were harvested by adding lysis buffer, and the protein concentrations were measured in the whole cell extracts. The samples were

resolved by SDS-PAGE and then transferred to PVDF membranes (MilliporeSigma, Burlington, MA, USA) to determine the protein expression levels. The membranes were then incubated with antibodies and secondary antibodies. The overexpression of PEDV 3CL^{pro} WT and its distinct mutants was evaluated using an anti-HA antibody (Medical and Biological Laboratories, Nagoya, Japan). An anti-goat monoclonal secondary antibody (Promega) was used to analyze the expression level of each luciferase reporter gene. An anti-β-actin mouse monoclonal antibody (Beyotime, Shanghai, China) was utilized to monitor β-actin's expression level to confirm that the protein loading was equal for the samples. The lane with the 100- and 70-kDa molecular mass bands was revealed by protein markers (26616; Thermo Fisher Scientific).

MD simulation protocol

Because the dimer structure of 3CL^{Pro} from CoVs is necessary for enzyme activity (34–36), the protease model used in this study was built from the X-ray structure of a dimeric PEDV 3CL^{Pro} mutant (C144A) bound to a peptide substrate (Protein Data Bank ID: 4ZUH; <https://www.rcsb.org/>). The C-terminal deletion PEDV 3CL^{Pro} in 4ZUH was substituted, and the substrate in the complex was replaced with YNSTLQ↓AGLRKM (the N-terminal auto-cleavage sequence of PEDV 3CL^{Pro}) using the software in the SYBYL-X program (v.2.0; <https://omictools.com/sybyl-x-tool>). For consistency, this crystal structure was used as the template for constructing the single mutant complexes by SYBYL-X. All water molecules in the crystal structure were retained. The Leap module in Amber18 was used to add all of the missing hydrogen atoms (37). The ff14SB Amber force field was used to assign bonded and nonbonded parameters to the protein and its peptide substrate (38). Each system was solvated with a 12 Å shell of the transferable intermolecular potential with 3 points (also known as TIP3P) water in a truncated octahedron simulation box with periodic boundary conditions (39). Sodium (Na⁺) or chloride (Cl[−]) counterions were added to neutralize the overall charge of the system.

For each complex, the MD simulations, which were collected for 50 ns using the Amber ff14SB force field in Nanoscale Molecular Dynamics (v.2.13; <https://www.ks.uiuc.edu/Research/namd/>), were repeated 3 times (38, 40). To relieve bad contacts and to direct each system toward energetically favorable conformations, each system was minimized using a 2-step, extensive energy minimization process based on the steepest descent method followed by the conjugate gradient algorithm. First, water molecules and counterions were relaxed by restraining the complex with a harmonic constant of 100 kcal/mol·Å^{−2}. Second, the restraint was removed to allow all of the atoms to move freely. After minimization, each system was gently heated from 0 to 310 K in 500 ps at a constant volume and equilibrated at 310 K for another 2 ns. Finally, a 50 ns MD simulation without any restrictions was performed at constant pressure, and the coordinates of the atoms were saved every 5 ps. During the MD simulation, bonds involving hydrogens were constrained by the SHAKE algorithm, and a time step of 2 fs was adopted (41). The Langevin thermostat approach was employed to control the temperature with a collision frequency of 1.0 ps^{−1} (42). The particle mesh Ewald method was used to treat the long-range electrostatic interactions (43, 44), and the cutoff distances for the long-range electrostatic and van der Waals (vdW) interactions were set at 10 Å.

Analysis of the MD simulations

Root mean square deviation calculations

Root mean square deviation (RMSD) calculations were performed using the Visual Molecular Dynamics software package (45). The frames from each interval were aligned to the first frame of the trajectory, and the RMSD values were calculated using all of the backbone α carbon atoms.

vdW contact potential calculations

The vdW contact potential energy between the protease and its substrate was calculated over an MD trajectory and averaged using the molecular mechanics Poisson-Boltzmann surface area method. The values were averaged over 120 ns (*i.e.*, the last three 40 ns of each repetition system).

Hydrogen bond calculations

The percentage of time that a hydrogen bond existed during a trajectory was calculated using the HBonds Plugin from Visual Molecular Dynamics and averaged over 120 ns (*i.e.*, the last three 40 ns of each repetition system) (45). A hydrogen bond was defined as having a donor-acceptor distance of a maximum of 3.5 Å, where only the polar atoms (nitrogen, oxygen, sulfur, and fluorine) were involved. The donor-hydrogen acceptor angle was defined as being <40°. Hydrogen bonds were summed over each residue and substrate except when otherwise indicated.

Cross-correlation analysis

To explore the effect of residue mutation on the conformation and internal dynamics changes of the protease-substrate complex, the cross-correlation matrix elements C_{ij} , which reflect the fluctuation of coordinates of the C_α atoms relative to their mean positions, were calculated from the last 40 ns of the MD trajectory for each system using the following equation, where the angle brackets represent the mean times over the recorded snapshots:

$$CO_{ij} = \frac{\langle \Delta R_i \rangle \langle \Delta R_j \rangle}{\langle \langle \Delta R_i^2 \rangle \langle \Delta R_j^2 \rangle \rangle^{1/2}}$$

ΔR_i indicates the fluctuation in the position vector R of site i , and ΔR_j is the fluctuation in the position vector R of site j (46). A more positive C_{ij} value represents a stronger correlated atomic fluctuation in the i th and j th residues.

Statistical analysis

The results are presented as the means ± SD of at least 3 experiments. Significant differences were detected using Student's t test. Values of $P < 0.05$ were considered statistically significant.

RESULTS

Exploiting the biosensor assay to evaluate PEDV 3CL^{Pro} activity *in vivo*

To establish a firefly luciferase reporter system to monitor the activity of PEDV-3CL^{Pro} in mammalian cells, we used an inverted, cyclized recombinant firefly luciferase construct (pCAGGS-DnaE) separated by an engineered site corresponding to the N-terminal YNSTLQ↓AGLRKM auto-cleavage sequence in PEDV 3CL^{Pro}. DnaE is widely used in protein cyclization because it improves the sensitivity of luciferase detection without affecting the luciferase's activity (31, 47, 48). As shown in Fig. 1A, the expressed N- and C-terminal fragments (233DP and 358DP) were cyclized to restrict the movement of the 2 domains in the presence of DnaE, which locked the enzyme into a more inactive form. Upon cleavage by PEDV 3CL^{Pro}, which recognizes the engineered cleavage site, the 2 firefly luciferase domains could theoretically interact freely and change into an active form of the luciferase. To detect any nonspecific cleavage by PEDV 3CL^{Pro}, the 233D and 358D systems were fused to the ENLYFQ↓YS sequence, which is recognized by TEV 3C^{Pro}, and used as the controls for the corresponding proteins (Fig. 1A).

To further determine whether the system fused to the N-terminal auto-cleavage sequence of PEDV 3CL^{pro} was successfully recognized and cleaved by PEDV 3CL^{pro}, HEK-293T cells were transfected with the PEDV 3CL^{pro} expression plasmid, together with the reporter 233DP or 358DP constructs, or the corresponding controls. Western blotting analyses showed that the protein bands from the cells cotransfected with PEDV 3CL^{pro} and 233DP or 358DP migrated fastest. On account of the cyclization conferred by DnaE, 233DP and 358DP were linearized after cleavage, thereby possessing greater mobility and resulting in a slightly smaller sized product than the cyclized protein on a Western blot (Fig. 1B). No cleavage activity was detected in the cells transfected with the 233D or 358D systems fused to the recognition sequence of TEV 3C^{pro} (Fig. 1B). These results confirm that the recombinant luciferase constructs fused to the N-terminal auto-cleavage sequences of PEDV 3CL^{pro} are specifically recognized and cleaved by PEDV 3CL^{pro}, suggesting their potential utility in assessing the activity of PEDV 3CL^{pro} in HEK-293T cells.

Reliability of the cyclized luciferase-based biosensor (233DP) at detecting PEDV 3CL^{pro} activity in mammalian cells

To evaluate the function of the reporter in the luciferase activity assay, the PEDV 3CL^{pro} expression plasmid, in

addition to each of the reporters or their respective controls and *Renilla* luciferase plasmids, was transfected into HEK-293T cells. The cells were lysed at 36 h post-transfection, and a dual-luciferase assay was performed on the lysates. As shown in Fig. 2A, the activity of 233DP was markedly induced by PEDV 3CL^{pro}, whereas that of the control reporter 233D remained low. Nevertheless, the background activity of 358DP was higher than that of 358D to some extent without the expression of PEDV 3CL^{pro} (Fig. 2A), suggesting that the increased activities of the reporter luciferase might be nonspecific. These results show that the 233DP reporter is a more sensitive and reliable biosensor assay for evaluating PEDV 3CL^{pro} activity. To further verify the effect of 233DP in the luciferase activity assay, HEK-293T cells were transfected with different amounts of the PEDV 3CL^{pro} expression plasmid and the 233D or 233DP reporter. As shown in Fig. 2B, a dose-dependent response was evident, with increasing amounts of protease expression leading to higher luciferase activity levels. Western blotting also revealed that PEDV 3CL^{pro} was able to cleave the recombinant firefly luciferase in a dose-dependent fashion, producing a faster migrating protein band (Fig. 2C). The consistency of the cleavage and the fold induction confirms that a correlation exists between the luciferase activity assay and reporter construct cleavage by PEDV 3CL^{pro}.

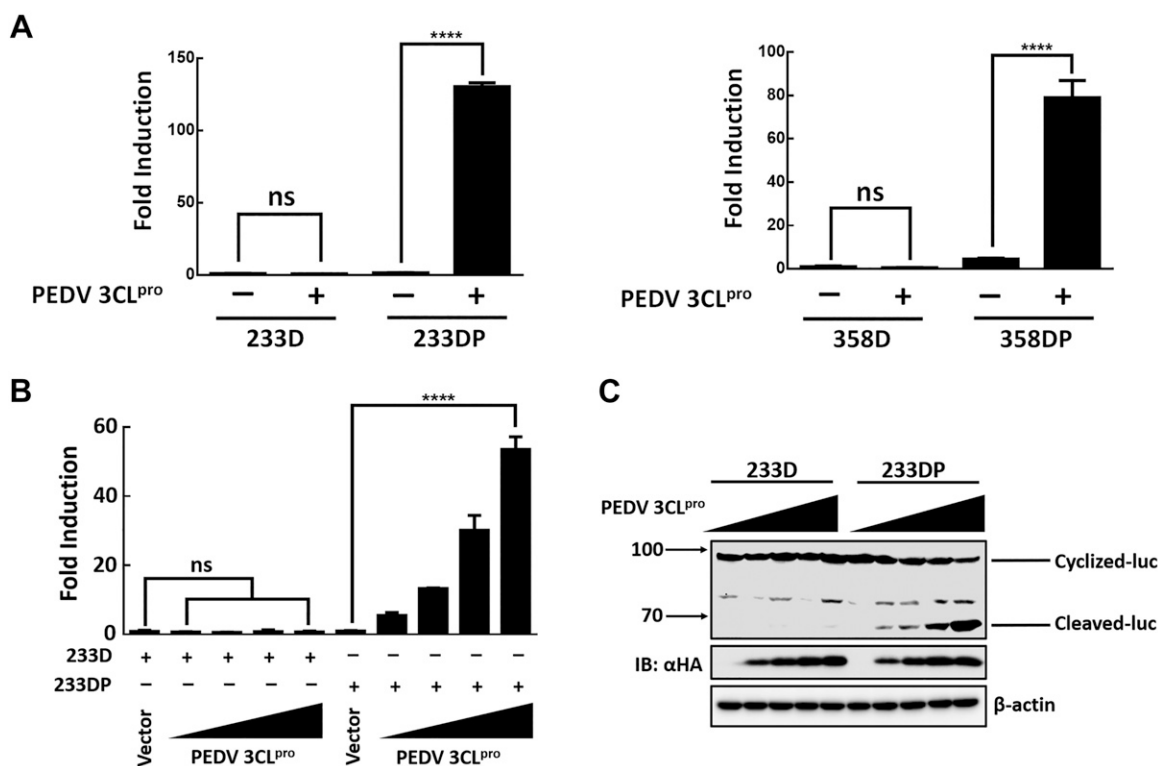


Figure 2. Reporter 233DP reliably detects PEDV 3CL^{pro} activity in cells. **A)** HEK-293T cells in 24-well plates were transfected with each of the 2 reporters or their corresponding controls, the pRL-TK plasmid, and the PEDV 3CL^{pro} expression plasmid. Luciferase assays were performed 36 h post-transfection. ns, not significant. *****P* < 0.0001. **B)** HEK-293T cells were transfected with 233DP, pRL-TK, and various concentrations of the PEDV 3CL^{pro} expression plasmid. The transfected cells were lysed for a dual-luciferase assay at 36 h post-transfection. **C)** HEK-293T cells cotransfected with PEDV 3CL^{pro} and the 233DP expression plasmid. Cell lysates were prepared 30 h post-transfection and then subjected to Western blotting. αHA, antihemagglutinin; IB, immunoblotting; luc, luciferase.

Identifying the novel amino acid residues involved in PEDV 3CL^{pro} activity

CoV 3CL^{pro} employs conserved cysteine and histidine residues (Cys144 and His41 in the case of PEDV 3CL^{pro}) as the principal nucleophile and general acid-base catalyst, respectively, at its catalytic site (49–51). To screen for additional amino acids impinging on the activity of PEDV 3CL^{pro}, 7 aa sites (Trp31, Cys38, Phe139, Gly142, Gly145, Tyr160, and His162) were chosen because they are highly conserved in CoV 3CL^{pro} (Fig. 3A) and have a strong interaction network with other amino acids in 3CL^{pro} (Fig. 3B) (52). The 233DP reporter system was used to access the protease activities of the single-substitution variants (W31A, C38A, F139A, G142A, G145A, Y160A, and H162A), with C41A and H144A variants used as the positive controls. As shown in Fig. 3C, WT PEDV 3CL^{pro}

and 3CL^{pro}-C38A successfully induced luciferase activity, whereas the other mutants failed to activate the 233DP reporter. To explore the mechanism underlying the failure of the overexpressed 3CL^{pro} mutants to induce reporter luciferase activity, WT PEDV 3CL^{pro} or the 3CL^{pro} mutant was overexpressed in the presence of the recombinant firefly luciferase 233DP reporter. The cyclized form of the recombinant firefly luciferase was cleaved normally by WT PEDV 3CL^{pro} and 3CL^{pro}-C38A, generating faster migrating protein bands on Western blots. However, no obvious cleavage products were observed when the other mutants were overexpressed. Interestingly, the protein abundance from the G145A and Y160A mutants was significantly reduced when compared with that of WT PEDV 3CL^{pro} (Fig. 3D). Unfortunately, it is difficult to determine whether a decrease in protein expression or protease activity leads to the occurrence of this phenomenon, because

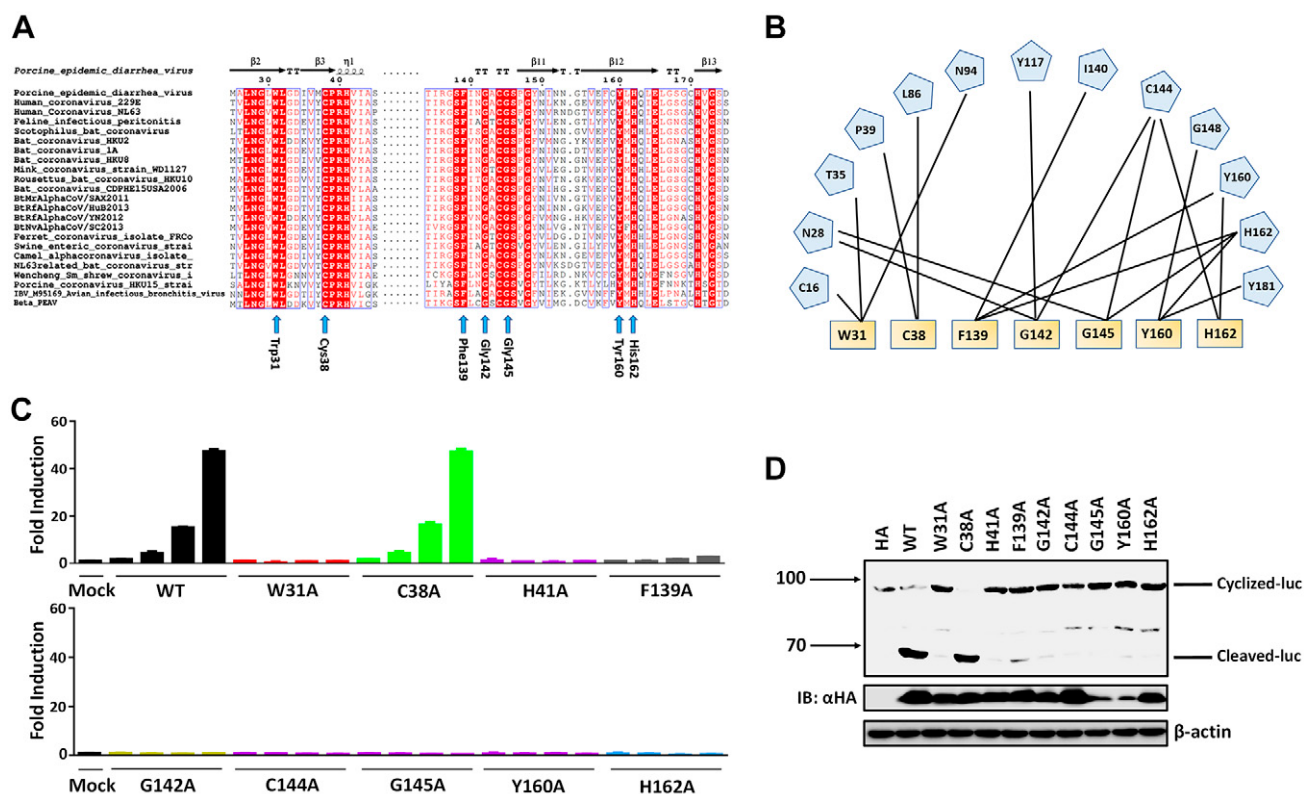


Figure 3. Identifying the novel amino acid residues involved in PEDV 3CL^{pro} activity. **A**) Amino acid alignment of the conserved region of the CoV 3CL^{pro} (numbering is based on PEDV 3CL^{pro}). Secondary structural elements in PEDV 3CL^{pro} are represented as η for 310 helix, arrows for β-strands, and T for β-turns. Residues conserved in all CoV 3CL^{pro}s are presented in white on a red background. Conserved residues in most of the CoV 3CL^{pro}s are presented in red and boxed with a white background. Blue arrows indicate the residues we selected. The sequences were derived from GenBank entries with the following accession numbers: PEDV, AF353511; Human_coronavirus_229E, AF304460; Human_Coronavirus_NL63, AY567487; Feline_infectious_peritonitis, AY994055; Scotophilus_bat_coronavirus, DQ648858; Bat_coronavirus_HKU2, EF203064; Bat_coronavirus_1A, EU420138; Bat_coronavirus_HKU8, EU420139; Mink_coronavirus_strain_WD1127, HM245925; Roussetus_bat_coronavirus_HKU10, JQ989270; Bat_coronavirus_CDPHE15USA2006, KF430219; BtMrAlphaCoV/SAX2011, KJ473806; BtRfAlphaCoV/HuB2013, KJ473807; BtRfAlphaCoV/YN2012, KJ473808; BtNvAlphaCoV/SC2013, KJ473809; Ferret_coronavirus_isolate_FRCO, KM347965; Swine_enteric_coronavirus_strain, KR061459; Camel_alphacoronavirus_isolate, KT368907; NL63related_bat_coronavirus_strain, KY073744; Wencheng_Sm_shrew_coronavirus, KY967717; Porcine_coronavirus_HKU15_strain, JQ065042; Avian_infectious_bronchitis_virus, M95169; and Beta_PEA, MG742313. **B**) The predicted related amino acids are listed and connected with black solid lines. **C**) HEK-293 cells were transfected with 233DP and pRL-TK and with increasing quantities of plasmid encoding the WT or mutant PEDV 3CL^{pro}. The cells were harvested after 36 h and subjected to a dual-luciferase assay. **D**) HEK-293T cells were cotransfected with the WT or mutant PEDV 3CL^{pro} and the 233DP expression plasmid. Cell lysates were prepared at 30 h post-transfection and analyzed by Western blotting. αHA, antihemagglutinin; IB, immunoblotting; luc, luciferase.

the mutations (G145A and Y160A) abrogated not only the catalytic activity but also the protein expression of PEDV 3CL^{Pro}. Thus, the G145A and Y160A mutations, which both mediated a reduction in the catalytic activity of PEDV 3CL^{Pro}, were not investigated further in this study.

Hydrogen bond interactions of substrates vs. protease

To further investigate the potential mechanism involving the single-substitution variant and in support of our experimental data, MD simulations were performed to investigate the dynamic mechanism used by the proteolytically inactive PEDV 3CL^{Pro} mutants. Based on the crystal structures of the PEDV-3CL^{Pro} complexes (Protein Data Bank identifier: 4ZUH) (53), 3 replicates of the 50 ns MD simulations were performed for each PEDV-3CL^{Pro} complex. In each simulation, the RMSD values of the C α atoms during the simulation calculations converged and remained stable. In the simulations over the last 40 ns, the overall binding modes of the WT and mutant complexes were conserved when bound to the substrate (Supplemental Fig. S1). As shown in the structure, only P4–P2' in the substrate fit comfortably in the active site, whereas the other residues floated out of the protease pockets (Fig. 4A, B). Then, we calculated the mean times for the hydrogen bonds during the MD simulations to better capture the intermolecular polar interactions. Overall, the hydrogen bonding network in the substrate packing was stably retained within the WT MD simulation. The most prevalent hydrogen bonds in the active site are formed by P1-Gln and residues Gly142, Cys144, His162, or Gln163 in the S1 pocket, which may help to stabilize the substrate in the active site during the cleavage reactions (Fig. 4C and Supplemental Fig. S2). The N ϵ 2 atom of His162 and the O ϵ 1 atom of P1-Gln form a hydrogen bond, which is stable by the π - π stacking interactions between Phe139 and His162. There is an oxyanion hole constituted by the main chain amides of Gly142 and Cys144 to stabilize the carbonyl oxygen of P1-Gln, which is reported to be critical for cleavage (Supplemental Fig. S3) (53). In addition, residue with stronger hydrogen bonds to the substrate is Glu165 in the S4 pocket, whose backbone links tightly to P4-Ser and P3-Thr (Fig. 4C and Supplemental Fig. S2). The stable hydrogen bond network in the WT complex between the S1 and S4 pockets is consistent with the crystal structure elucidated in previous studies, further verifying the importance of the S1 and S4 pockets in substrate binding (14, 54).

The Phe139, Gly142, and His162 residues in the protease active site make direct hydrogen bonds with the substrate. Compared with the WT complex, the residue 139 mutation resulted in no considerable difference in the hydrogen bonding between the main side of Phe139 and P1-Gln. However, the loss of π - π stacking interactions between Phe139 and His162 caused subtle rearrangements in the structure that resulted in decreased interactions in His162–P1-Gln and Gln163–P1-Gln (–35 and –25%, respectively) (Supplemental Fig. S4). Interestingly, the G142A substitution decreased the interactions not only at

this position with P1-Gln but also at the other active sites, especially Q163–P1-Gln (–72%) (Fig. 4D and Supplemental Figs. S4 and S5). Notably, the H162A variant was the most disrupted with 10 hydrogen bonds changing by >15% relative to the WT complex throughout the dimer, with 9 being weakened including most dramatically the interactions of the side chain of His162 with P1-Gln (a 96% reduction) (Fig. 4E). In addition to the mutants within the protease active site, a remote mutant site, W31A, resulted in the loss of at least 2 intermolecular hydrogen bonds at the S4 pocket. As expected, the C38A substitution did not cause any further considerable changes in the active site relative to the WT complexes (Fig. 4D and Supplemental Fig. S5). Overall, the active site polymorphisms in the S1 and S4 pockets severely disrupted the intermolecular hydrogen bonding network in the active site as well as affecting the substrate binding.

Differences in the activity of PEDV 3CL^{Pro} alter substrate packing

In addition to the hydrogen bond interactions shown for the packed substrate at the active site, we calculated the vdW contact energies for the active site and substrate in each complex for more detail. The total vdW contact energies were conserved between the WT and C38A complexes (–98.9 and –98.1 kcal/mol, respectively), but striking energy losses were evident for the W31A, F139A, G142A, and H162A complexes when compared with the WT value (–90.5, –91.5, –89.2, and –86.7 kcal/mol, respectively) (Fig. 5A and Table 1), a result consistent with the experimental loss in protease activity and the severe disruption of hydrogen bonding network (Fig. 4 and Supplemental Fig. S5).

To quantify the interactions of the substrates with the individual active site residues, intermolecular vdW interactions over the MD trajectories for each residue at the active site were calculated. In line with the conserved overall binding mode, the strongest substrate–protease interaction occurred with the Met25 and Asn141 residues in the WT complex (–3.72 and –3.25 kcal/mol, respectively). Compared with the WT protease, the contact energy landscape in the C38A complex was highly conserved, but disrupted in the W31A, F139A, G142A, and H162A complexes with a conspicuous loss of interactions over 1.2 kcal/mol in the Met25 residue (Fig. 5B and Supplemental Fig. S6). Moreover, a considerable loss of contacts in the S1' pocket, in particular, Asn24 and Ala26, was evident. The decline of the vdW between S1' pocket and the substrate indicates that the C terminus of substrate in single mutant complexes became more disordered. According to the vdW equation, we calculated the distance between Met25 and the substrate and found that the P4'-Arg had markedly moved away from the Met25 residue in the W31A, F139A, G142A, and H162A mutants (Supplemental Fig. S7). Interestingly, previous research has shown that the M25T single-substitution mutant failed to cleave an NF- κ B essential modulator signaling–derived substrate (53), indicating that the Met25 residue plays a crucial role in substrate binding.

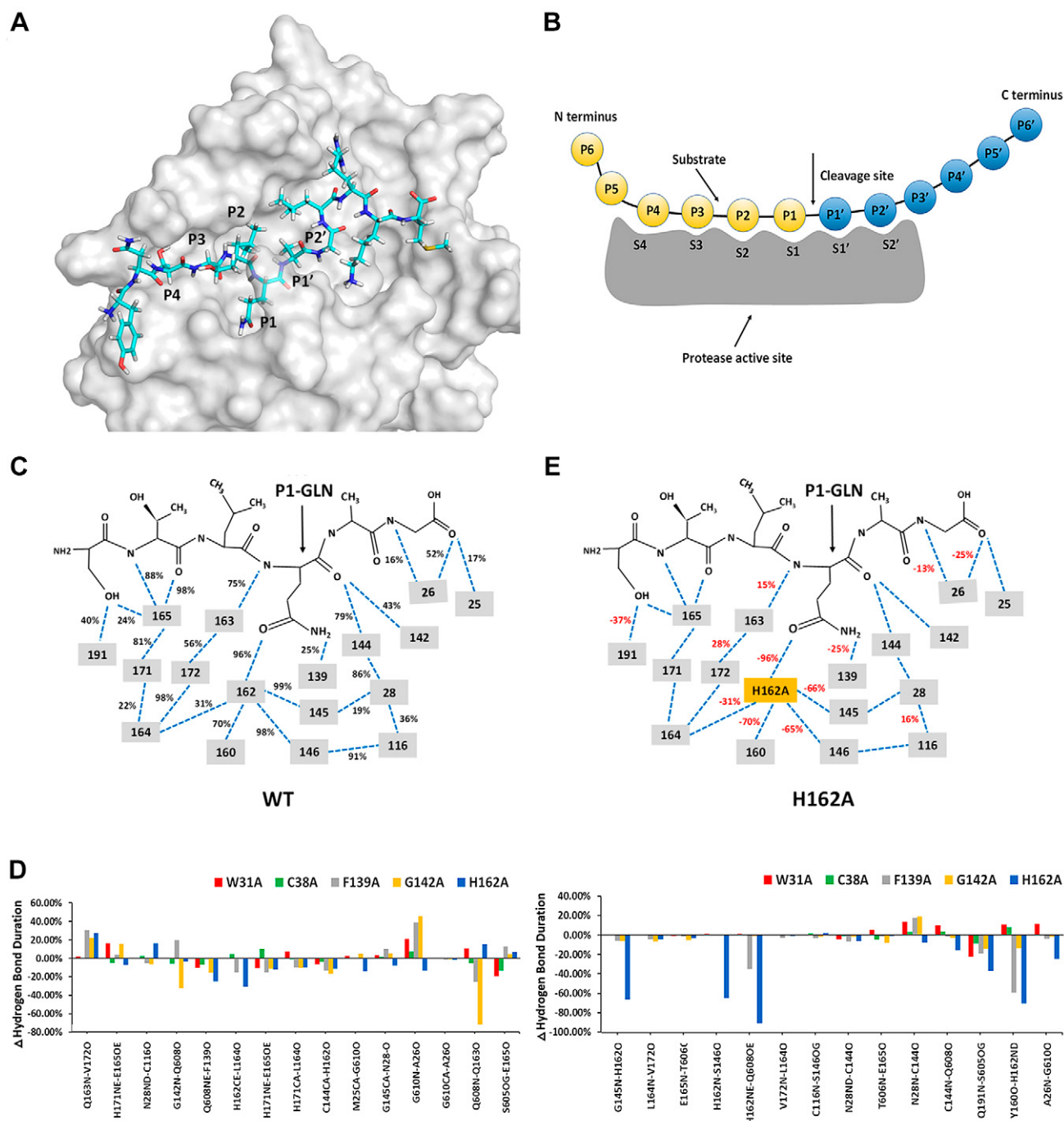


Figure 4. Hydrogen bond interactions of substrates *vs.* protease. *A*) Crystal structure of the substrate bound to the active site with only one protease monomer shown for clarity. The substrate is shown in blue. *B*) Diagram of the substrate envelope model. Only P4–P2' in the substrate fit nicely in the pockets of the active site. *C*) Histograms of the changes in the percentage of times that hydrogen bonds are formed relative to the WT simulation for each of the complexes. *D*) Schematic hydrogen bond network in the active site with the percentage of times hydrogen bonds are formed during the WT simulation. The dashed blue lines represent the hydrogen bond interactions. The substrate is shown from P4-Ser to P2'-Gly considering the substrate envelope model and previous research. *E*) Schematic representation of the H162A complex simulation with changes in hydrogen bonding relative to the WT simulations. The schematic for the remaining variants is shown in Supplemental Fig. S5.

Loss of dynamic correlations during protease-peptide atomic fluctuations in the proteolytically inactive mutants

In principle, tight binding substrates are characterized by strong intermolecular interactions with their cognate

proteins, which persist over the dynamics of individual enzymes (25). Conservation of protease-inhibitor dynamic cross-correlations is often incorporated into the rational design and computational evaluation of protein inhibitors in structure-based drug design (26, 27). To further investigate the coupling of atomic fluctuations between the

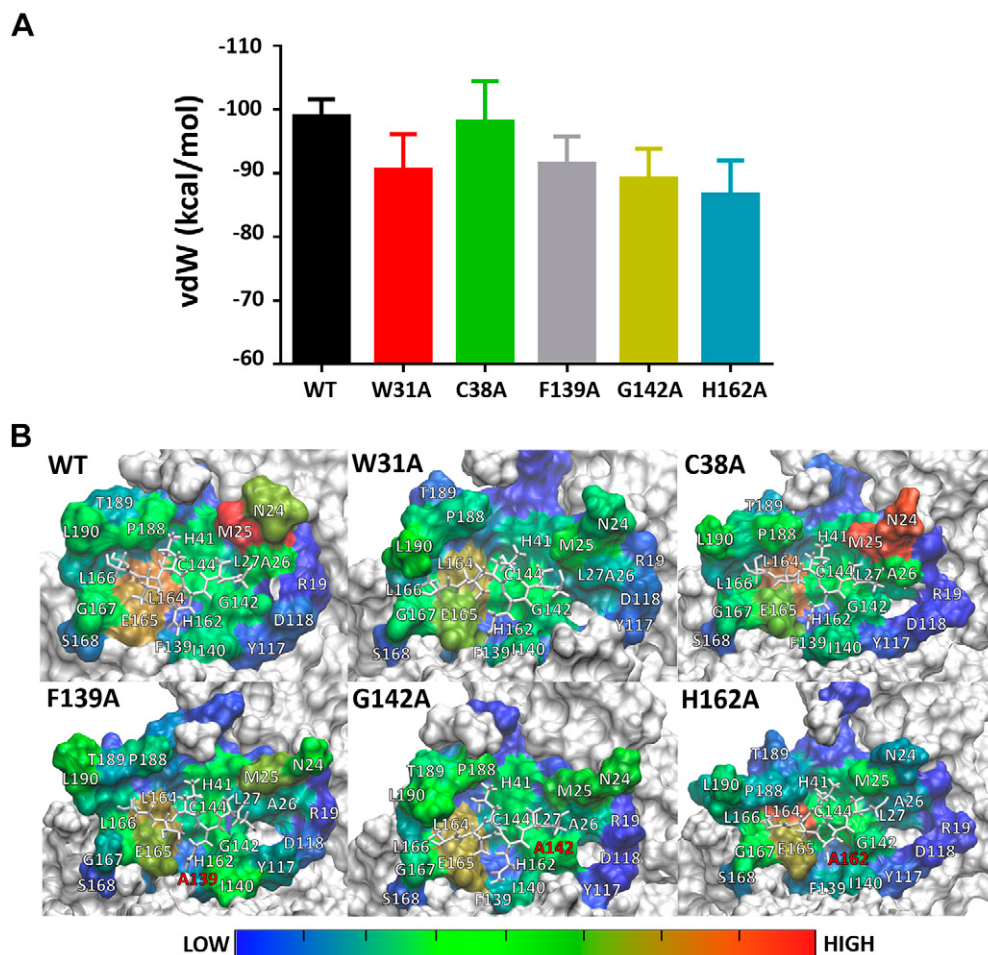


Figure 5. Differences in the activity of PEDV 3CL^{pro} alter substrate packing. **A)** Histograms of the total vdW contact energies of the 6 complexes. **B)** The vdW contact potentials averaged from MD simulations of protease active site residues for the substrate bound to WT, W31A, C38A, F139A, G142A, and H162A proteases, respectively. The warmer (red) and cooler (blue) colors indicate more and less contact with the substrate, respectively. Single-substitution residues are highlighted in red. Trp31 and Cys38, outside the active site, are not marked. See also Supplemental Fig. S6.

active site surface of the protease and the substrate, we calculated the cross-correlation coefficients between the atomic fluctuations of the protease's backbone and the peptide substrate's atoms. On the basis of our binding model and previous research, only P4–P2' in the substrate are shown in the results (14, 53). As shown in Fig. 6A, the protease active site is mainly composed of 4 pockets, S4, S2, S1, S1', which correspond to P4, P2, P1, P1' sites in the substrate, respectively. In the WT complex, the dynamics of the substrate were highly correlated with the motions of the residues in the protease's active site. In addition to the conserved S1-P1 and S4-P4 interaction revealed by the hydrogen bonding interactions, this coupling was the most pronounced for the 162–165 active site residues, displaying correlations

with most of the substrate's moieties. Additionally, the dynamics of the P1-Gln moiety of the substrate were highly coupled with the dynamics of residue Gly142 and the Cys144 catalytic site. Neither of these correlations changed when Cys38 was mutated to an alanine, consistent with the stabilization of the intramolecular dynamics in the C38A complexes (Fig. 6B).

In contrast, there was a considerable loss of correlation between P2-Leu, P3-Thr, or P4-Ser and residues 162–167 in both the W31A and F139A complexes. However, the reduction in the F139A complex was much more serious than that in the W31A complex. The H162A substitution severely reduced the dynamic coupling of the substrate to the protease's active site. In addition to the striking loss of P2-Leu, P3-Thr, or P4-Ser with residues 162–167, the correlation between P1-Gln and the catalytic residue Cys144 was severely reduced (Fig. 6B). The disrupted correlations for the pocket residues for S1 and S4 agree with the observed loss of intermolecular interactions for P1-Gln and P4-Ser in this substrate. Interestingly, in the G142A protease, the loss of hydrogen bond between Cys142 and the carbonyl oxygen of P1-Gln undermined the stability of the oxyanion hole, leading to complete disruption of substrate–active site correlations (Figs. 4 and 6B). Overall, the loss of coupling between substrate and protease dynamics might be found to be correlated with reduced protease activity against the single mutant variant.

TABLE 1. *vdW interactions between protease and substrate in WT and mutant complexes*

System	ΔE_{vdW} (kcal/mol)	$\Delta\Delta E_{\text{vdW}}$ (kcal/mol)
WT	−98.9	—
W31A	−90.5	+8.4
C38A	−98.1	+0.8
F139A	−91.5	+7.4
G142A	−89.2	+9.7
H162A	−86.7	+12.2

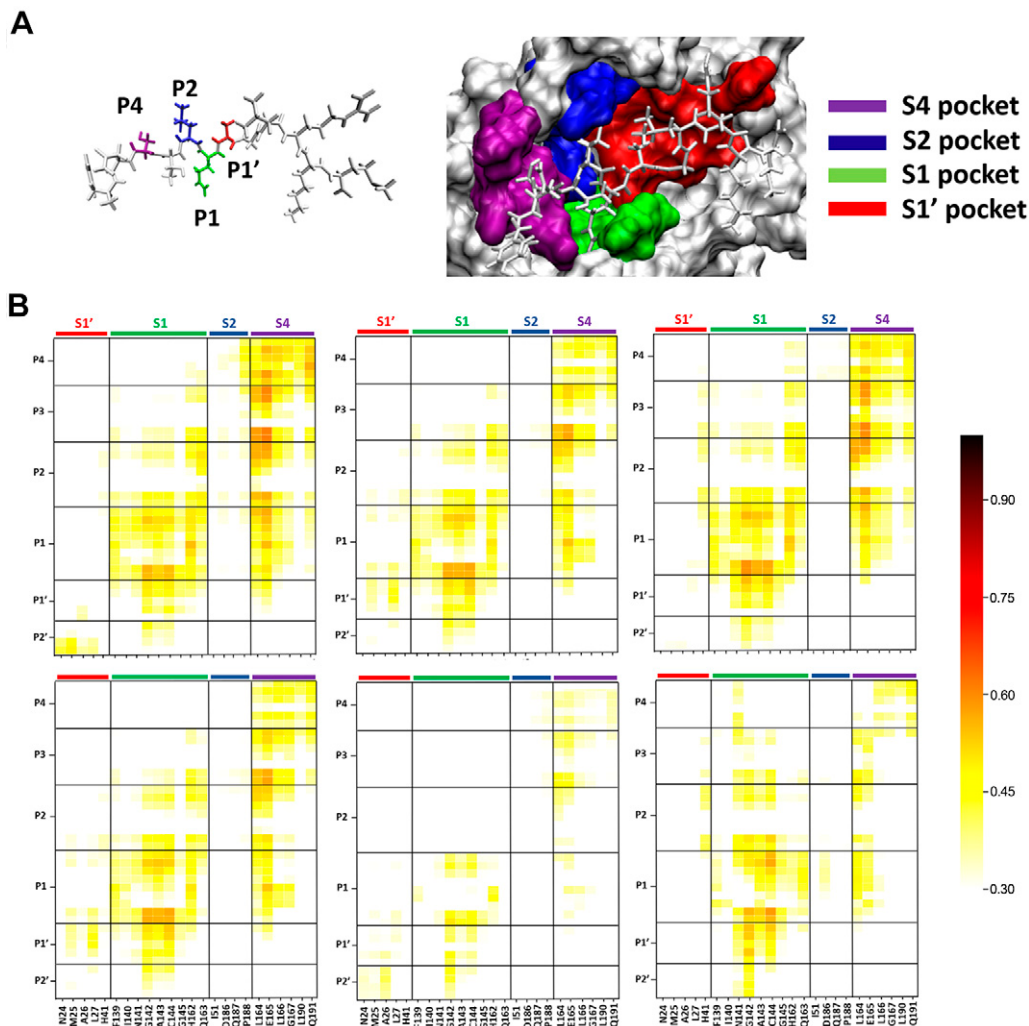


Figure 6. Protease-substrate dynamic coupling of an oligopeptide bound to WT, W31A, C38A, F139A, G142A, and H162A proteases. *A*) Diagram of the binding pocket in the active site. S1': Asn24, Met25, Ala26, Leu27, His41; S1: Phe139, Ile140, Asn141, Gly142, Ala143, Cys144, Gly145, His162, Gln163; S2: Ile51, Asp186, Gln187, Pro188; S4: Leu164, Glu165, Leu166, Gly167, Leu190, Gln191. *B*) Cross-correlations between atomic fluctuations of protease active site residues and substrate in different complexes. Warm colors in the matrices indicate increased correlations. Residues are colored on the surface to indicate their locations in the active site.

DISCUSSION

With their pivotal roles in the multiplication and proliferation of CoVs, 3CL^{pro} are recognized as the major targets of protein inhibitors in anti-CoV therapies (14, 21, 35). Developing assays that efficiently detect the activity of proteases in CoVs is a key step toward the goal of screening for specific PEDV 3CL^{pro} inhibitors or broad-spectrum inhibitors of CoV proteases. Thus, analyzing 3CL^{pro} activity in live cells with an efficient, high-throughput strategy is critical to moving this field forward. In the present study, we developed a luciferase-based protease activity biosensor, which contained DnaE and the N-terminal auto-cleavage sequences of PEDV 3CL^{pro}. DnaE was used to cyclize the 2 domains of firefly luciferase to generate a cyclized recombinant firefly luciferase without affecting the activities (31, 47, 48). With a circular permuted luciferase, the movement of the 2 firefly luciferase domains is restricted, locking the enzyme into its less-active open form (32). This

underpins the high sensitivity and low background of the 233DP reporter system (Fig. 2B), confirming the outstanding prospects of the 233DP reporter in the analysis of PEDV 3CL^{pro} activity.

Thorough elucidation of substrate-active site interactions is crucial for rational drug design, and further improvements in this area are needed if broader potencies are to be achieved (26, 27, 55–57). Following the SARS outbreak, a series of inhibitors that prevent viral replication were reported to act against the 3CL^{pro} of SARS-CoV (17–19). However, the binding mode of the substrate and active site was merely described for several crystal structures. Previous studies have shown that therapeutic efforts against HCV and HIV have greatly benefited from MD-based drug design, specifically in developing viral protease inhibitors (23–27). Our MD analysis has revealed the potential structural mechanism for substrate binding in more detail. The side chain of the conserved P1-Gln fits comfortably in the S1 pocket, stabilized by a hydrogen

bonding network, whereas the prevalent hydrogen bonds of P3-Thr and P4-Ser are with residues Glu165 and Gln191 in the S4 pocket. Notably, we discovered the dynamic cross-correlations with the target active site for substrate binding, which revealed the strong positive interdependency on P1-S1 and P4-S4. We further explored the protein surface of PEDV 3CL^{pro} using a variety of small “probe” molecules by FTMap (58), which showed that the druggable sites in PEDV 3CL^{pro} comprise a cluster of binding hot spots in the S1 and S4 pockets (Supplemental Fig. S8). Taken together, the intermolecular hydrogen bonding network and intermolecular dynamic correlations further confirm that the S1 and S4 pockets are key to substrate recognition and comprise the ideal target for drug design.

More importantly, the high replication rate and high mutation frequency that occur during the replication of viruses such as SARS-CoV and PEDV (9, 10, 59, 60) mean that every possible nonsense mutation is likely introduced into the viral genome on a daily basis, which may lead to resistance-associated substitutions in the target proteins. CoV 3CL^{pro} employs conserved cysteine and histidine residues in S1 pocket as the principal nucleophile and general acid-base catalyst, respectively, at its catalytic site (49–51). One strategy used to avoid drug resistance in the rational drug design of protein inhibitors tends to stack on the catalytic residues and have direct physical interactions with them. The catalytic dimer residues are critical for the biologic function of the protease and thus are almost always invariant (27). We screened 7 aa through interaction networks and the conservation of PEDV 3CL^{pro}. Four of them, the W31A, F139A, G142A, and H162A mutants, were incapable of activating the luciferase-based biosensor. The irreplaceability of these 4 sites in PEDV 3CL^{pro} for the N-terminal cleavage of PEDV 3CL^{pro} indicated that they are critical for the maturation of this protease. Specifically, Phe139, Gly142, and His162 are within the protease active site. These 3 mutations (F139A, G142A, H162A) unfavorably alter or completely disrupt the active site’s intermolecular network of hydrogen bonds and the intermolecular dynamic correlations in S1 and S4 pockets required for tight substrate binding, further leading to the loss of protease activity. Interestingly, although the Trp31 is out of the active site and away from binding surface for dimerization, the distance between Trp31 and S1’ pocket is close (Supplemental Fig. S9). The single substitution might utilize a common mechanism or pathway for altering the protease-substrate interactions (23). This mutation (W31A) might cause subtle but significant rearrangements in the structure resulting in altered interactions with the bound substrate, as well as impacting the dynamic assemblies of the complexes; these changes were particularly obvious in the MD analysis at the S1 and S4 pockets.

Here, Phe139, Gly142, and His162 are crucial for protease activity and further irreplaceable for viral replication. More importantly, these 3 aa locate on the surface of S1 pocket. Their side chains or main chains are potential sites

for hydrogen bond interactions formed during the binding of small molecular drugs. These characteristics give them the potential to be target sites for the newer-generation inhibitors to work with and conquer drug resistance. Furthermore, we incorporated conformational dynamics into a dynamic substrate envelope model using MD simulations and compared the WT complex with various mutant complexes to effectively explain the mechanism involved in the substrate–active site interactions, something that will be necessary for the design of protein inhibitors and for improving the potency of these inhibitors against any emerging inhibitor-resistant variants. The 3C^{pro} or 3CL^{pro} encoded by viruses exist in plus-stranded RNA viruses and double-stranded RNA viruses (61). Therefore, the method established in the present study can be used as a reference for other viruses.

In summary, the 233DP reporter system we developed in this study is an improvement over other biosensors in that it is sensitive and yields reproducible data. We screened 4 proteolytically inactive mutants using this method and further elucidated the potential molecular mechanism by MD analysis. The proteolytically crucial sites in the active site might enrich the target residue for drug design. We also confirmed the importance of the S1 and S4 pockets in substrate binding. The MD simulations of the PEDV WT 3CL^{pro} and its mutant complexes enabled a detailed analysis of the protease-substrate binding model to be undertaken, which is critical for the rational design of protease inhibitors. Overall, our successful development of a luciferase-based biosensor for measuring protease activity will facilitate the screening and identification of effective protease inhibitors against PEDV 3CL^{pro} and future emerging CoVs. FJ

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (2016YFD0500103 and 2017YFB0203405), the National Natural Science Foundation of China (21873034, 31672569, and 31672566), the Natural Science Foundation of Hubei Province (2017CFA059), the Special Project for Technology Innovation of Hubei Province (2017ABA138), and the Fundamental Research Funds for the Central Universities (2662016PY070, 2662018JC027, and 2015RC008). The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

J. Zhou, D. Wang, J. Gao, and S. Xiao conceived and designed the study; J. Zhou wrote the manuscript, with critical input from L. Fang, D. Wang, J. Gao, and S. Xiao; J. Zhou performed the molecular dynamics simulations; J. Zhou, Z. Yang, S. Xu, M. Lv, Z. Sun, and J. Chen performed cellular experiments; and all authors discussed results and contributed to manuscript preparation.

REFERENCES

1. Pereira, H. (1989) *Coronaviridae. Andrewes' Viruses of Vertebrates*, 5th ed. pp. 42–57, Bailliere Tindall, London

2. Holmes, K. V., and Lai, M. M. C. (1996) Coronaviridae: the viruses and their replication. *Virology* **1**, 1075–1093
3. Peiris, J. S., Yuen, K. Y., Osterhaus, A. D., and Stöhr, K. (2003) The severe acute respiratory syndrome. *N. Engl. J. Med.* **349**, 2431–2441
4. Memish, Z. A., Mishra, N., Olival, K. J., Fagbo, S. F., Kapoor, V., Epstein, J. H., Alhakeem, R., Durosounlou, A., Al Asmari, M., Islam, A., Kapoor, A., Briese, T., Daszak, P., Al Rabecah, A. A., and Lipkin, W. I. (2013) Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. *Emerg. Infect. Dis.* **19**, 1819–1823
5. De Groot, R. J., Baker, S. C., Baric, R. S., Brown, C. S., Drosten, C., Enjuanes, L., Fouchier, R. A., Galiano, M., Gorbelenya, A. E., Memish, Z. A., Perlman, S., Poon, L. L., Snijder, E. J., Stephens, G. M., Woo, P. C., Zaki, A. M., Zambon, M., and Ziebuhr, J. (2013) Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J. Virol.* **87**, 7790–7792
6. Wood, E. N. (1977) An apparently new syndrome of porcine epidemic diarrhoea. *Vet. Rec.* **100**, 243–244
7. Sun, R.-Q., Cai, R.-J., Chen, Y.-Q., Liang, P.-S., Chen, D.-K., and Song, C.-X. (2012) Outbreak of porcine epidemic diarrhoea in suckling piglets, China. *Emerg. Infect. Dis.* **18**, 161–163
8. Liu, C., Tang, J., Ma, Y., Liang, X., Yang, Y., Peng, G., Qi, Q., Jiang, S., Li, J., Du, L., and Li, F. (2015) Receptor usage and cell entry of porcine epidemic diarrhoea coronavirus. *J. Virol.* **89**, 6121–6125
9. Song, D., and Park, B. (2012) Porcine epidemic diarrhoea virus: a comprehensive review of molecular epidemiology, diagnosis, and vaccines. *Virus Genes* **44**, 167–175
10. Kweon, C.-H., Kwon, B.-J., Lee, J.-G., Kwon, G.-O., and Kang, Y.-B. (1999) Derivation of attenuated porcine epidemic diarrhoea virus (PEDV) as vaccine candidate. *Vaccine* **17**, 2546–2553
11. Perlman, S., and Netland, J. (2009) Coronaviruses post-SARS: update on replication and pathogenesis. *Nat. Rev. Microbiol.* **7**, 439–450
12. Zhu, X., Wang, D., Zhou, J., Pan, T., Chen, J., Yang, Y., Lv, M., Ye, X., Peng, G., Fang, L., and Xiao, S. (2017) Porcine deltacoronavirus nsp5 antagonizes type I interferon signaling by cleaving STAT2. *J. Virol.* **91**, e00003-17
13. Wang, D., Fang, L., Shi, Y., Zhang, H., Gao, L., Peng, G., Chen, H., Li, K., and Xiao, S. (2015) Porcine epidemic diarrhoea virus 3C-like protease regulates its interferon antagonism by cleaving NEMO. *J. Virol.* **90**, 2090–2101
14. Yang, H., Xie, W., Xue, X., Yang, K., Ma, J., Liang, W., Zhao, Q., Zhou, Z., Pei, D., Ziebuhr, J., Hilgenfeld, R., Yuen, K. Y., Wong, L., Gao, G., Chen, S., Chen, Z., Ma, D., Bartlam, M., and Rao, Z. (2005) Design of wide-spectrum inhibitors targeting coronavirus main proteases. *PLoS Biol.* **3**, e324; erratum: e428
15. Deeks, S. G., Smith, M., Holodniy, M., and Kahn, J. O. (1997) HIV-1 protease inhibitors. A review for clinicians. *JAMA* **277**, 145–153
16. Lamarre, D., Anderson, P. C., Bailey, M., Beaulieu, P., Bolger, G., Bonneau, P., Bös, M., Cameron, D. R., Cartier, M., Cordingley, M. G., Faucher, A. M., Goudreau, N., Kawai, S. H., Kukolj, G., Lagacé, L., LaPlante, S. R., Narjes, H., Poupard, M. A., Rancourt, J., Sentjens, R. E., St George, R., Simoneau, B., Steinmann, G., Thibeault, D., Tsantrizos, Y. S., Weldon, S. M., Yong, C. L., and Llinàs-Brunet, M. (2003) An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus. *Nature* **426**, 186–189; erratum: 246
17. Wu, C.-Y., Jan, J.-T., Ma, S.-H., Kuo, C.-J., Juan, H.-F., Cheng, Y.-S. E., Hsu, H.-H., Huang, H.-C., Wu, D., Brik, A., Liang, F. S., Liu, R. S., Fang, J. M., Chen, S. T., Liang, P. H., and Wong, C. H. (2004) Small molecules targeting severe acute respiratory syndrome human coronavirus. *Proc. Natl. Acad. Sci. USA* **101**, 10012–10017
18. Bacha, U., Barrila, J., Velazquez-Campoy, A., Leavitt, S. A., and Freire, E. (2004) Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro. *Biochemistry* **43**, 4906–4912
19. Blanchard, J. E., Elowe, N. H., Huitema, C., Fortin, P. D., Cechetto, J. D., Eltis, L. D., and Brown, E. D. (2004) High-throughput screening identifies inhibitors of the SARS coronavirus main proteinase. *Chem. Biol.* **11**, 1445–1453
20. Jain, R. P., Pettersson, H. I., Zhang, J., Aull, K. D., Fortin, P. D., Huitema, C., Eltis, L. D., Parrish, J. C., James, M. N., Wishart, D. S., and Vederas, J. C. (2004) Synthesis and evaluation of keto-glutamine analogues as potent inhibitors of severe acute respiratory syndrome 3CLpro. *J. Med. Chem.* **47**, 6113–6116
21. Kao, R. Y., Tsui, W. H., Lee, T. S., Tanner, J. A., Watt, R. M., Huang, J.-D., Hu, L., Chen, G., Chen, Z., Zhang, L., He, T., Chan, K. H., Tse, H., To, A. P., Ng, L. W., Wong, B. C., Tsoi, H. W., Yang, D., Ho, D. D., and Yuen, K. Y. (2004) Identification of novel small-molecule inhibitors of severe acute respiratory syndrome-associated coronavirus by chemical genetics. *Chem. Biol.* **11**, 1293–1299
22. Tanner, J. A., Zheng, B.-J., Zhou, J., Watt, R. M., Jiang, J.-Q., Wong, K.-L., Lin, Y.-P., Lu, L.-Y., He, M.-L., Kung, H.-F., Kesel, A. J., and Huang, J. D. (2005) The adamantane-derived bananins are potent inhibitors of the helicase activities and replication of SARS coronavirus. *Chem. Biol.* **12**, 303–311
23. Ragland, D. A., Nalivaika, E. A., Nalam, M. N., Prachanronarong, K. L., Cao, H., Bandaranayake, R. M., Cai, Y., Kurt-Yilmaz, N., and Schiffer, C. A. (2014) Drug resistance conferred by mutations outside the active site through alterations in the dynamic and structural ensemble of HIV-1 protease. *J. Am. Chem. Soc.* **136**, 11956–11963
24. Matthew, A. N., Zephyr, J., Hill, C. J., Jahangir, M., Newton, A., Petropoulos, C. J., Huang, W., Kurt-Yilmaz, N., Schiffer, C. A., and Ali, A. (2017) Hepatitis C virus NS3/4A protease inhibitors incorporating flexible P2 quinoxalines target drug resistant viral variants. *J. Med. Chem.* **60**, 5699–5716
25. Ozen, A., Sherman, W., and Schiffer, C. A. (2013) Improving the resistance profile of hepatitis C NS3/4A inhibitors: dynamic substrate envelope guided design. *J. Chem. Theory Comput.* **9**, 5693–5705
26. Soumana, D. I., Kurt Yilmaz, N., Ali, A., Prachanronarong, K. L., and Schiffer, C. A. (2016) Molecular and dynamic mechanism underlying drug resistance in genotype 3 hepatitis C NS3/4A protease. *J. Am. Chem. Soc.* **138**, 11850–11859
27. Matthew, A. N., Leidner, F., Newton, A., Petropoulos, C. J., Huang, W., Ali, A., Kurt-Yilmaz, N., and Schiffer, C. A. (2018) Molecular mechanism of resistance in a clinically significant double-mutant variant of HCV NS3/4A protease. *Structure* **26**, 1360–1372.e5
28. Soumana, D. I., Kurt Yilmaz, N., Prachanronarong, K. L., Aydin, C., Ali, A., and Schiffer, C. A. (2016) Structural and thermodynamic effects of macrocyclization in HCV NS3/4A inhibitor MK-5172. *ACS Chem. Biol.* **11**, 900–909
29. Galbán, S., Jeon, Y. H., Bowman, B. M., Stevenson, J., Sebolt, K. A., Sharkey, L. M., Lafferty, M., Hoff, B. A., Butler, B. L., Wigdal, S. S., Binkowski, B. F., Otto, P., Zimmerman, K., Vidugiris, G., Encell, L. P., Fan, F., Wood, K. V., Galbán, C. J., Ross, B. D., and Rehemtulla, A. (2013) Imaging proteolytic activity in live cells and animal models. *PLoS One* **8**, e66248
30. Kilianski, A., Mielech, A. M., Deng, X., and Baker, S. C. (2013) Assessing activity and inhibition of Middle East respiratory syndrome coronavirus papain-like and 3C-like proteases using luciferase-based biosensors. *J. Virol.* **87**, 11955–11962
31. Shyu, Y. J., Liu, H., Deng, X., and Hu, C. D. (2006) Identification of new fluorescent protein fragments for bimolecular fluorescence complementation analysis under physiological conditions. *Biotechniques* **40**, 61–66
32. Wigdal, S. S., Anderson, J. L., Vidugiris, G. J., Shultz, J., Wood, K. V., and Fan, F. (2008) A novel bioluminescent protease assay using engineered firefly luciferase. *Curr. Chem. Genomics* **2**, 16–28
33. Zhou, J., Wang, D., Xi, Y., Zhu, X., Yang, Y., Lv, M., Luo, C., Chen, J., Ye, X., Fang, L., and Xiao, S. (2017) Assessing activity of Hepatitis A virus 3C protease using a cyclized luciferase-based biosensor. *Biochem. Biophys. Res. Commun.* **488**, 621–627
34. Cheng, S.-C., Chang, G.-G., and Chou, C.-Y. (2010) Mutation of Glu-166 blocks the substrate-induced dimerization of SARS coronavirus main protease. *Biophys. J.* **98**, 1327–1336
35. Anand, K., Yang, H., Bartlam, M., Rao, Z., and Hilgenfeld, R. (2005) Coronavirus main proteinase: target for antiviral drug therapy. In *Coronaviruses With Special Emphasis on First Insights Concerning SARS*, pp. 173–199, Springer, Berlin, Germany.
36. Shi, J., Wei, Z., and Song, J. (2004) Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme: defining the extra domain as a new target for design of highly specific protease inhibitors. *J. Biol. Chem.* **279**, 24765–24773
37. Case, D. A., Betz, R. M., Cerutti, D. S., Cheatham, T. E. III, Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Luchko, T., Luo, R., Madej, B., Merz, K. M., Monard, G., Needham, P., Nguyen, H., Nguyen, H. T., Omelyan, I., Onufriev, A., Roe, D. R., Roitberg, A., Salomon-Ferrer, R., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., York, D. M., and Kollman, P. A. (2015) *AMBER 2015*. University of California, San Francisco
38. Case, D. A., Cheatham III, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688

39. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935
40. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802
41. Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341
42. Izaguirre, J. A., Catarello, D. P., Wozniak, J. M., and Skeel, R. D. (2001) Langevin stabilization of molecular dynamics. *J. Chem. Phys.* **114**, 2090–2098
43. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092
44. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593
45. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28
46. Ichiye, T., and Karplus, M. (1991) Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* **11**, 205–217
47. Iwai, H., and Plückthun, A. (1999) Circular β -lactamase: stability enhancement by cyclizing the backbone. *FEBS Lett.* **459**, 166–172
48. Zettler, J., Schütz, V., and Mootz, H. D. (2009) The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett.* **583**, 909–914
49. Hegyi, A., and Ziebuhr, J. (2002) Conservation of substrate specificities among coronavirus main proteases. *J. Gen. Virol.* **83**, 595–599
50. Anand, K., Palm, G. J., Mesters, J. R., Siddell, S. G., Ziebuhr, J., and Hilgenfeld, R. (2002) Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α -helical domain. *EMBO J.* **21**, 3213–3224
51. Thiel, V., Ivanov, K. A., Putics, A., Hertzog, T., Schelle, B., Bayer, S., Weissbrich, B., Snijder, E. J., Rabenau, H., Doerr, H. W., Gorbalenya, A. E., and Ziebuhr, J. (2003) Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **84**, 2305–2315
52. Colell, E. A., Iserte, J. A., Simonetti, F. L., and Marino-Buslje, C. (2018) MISTIC2: comprehensive server to study coevolution in protein families. *Nucleic Acids Res.* **46**, W323–W328
53. Ye, G., Deng, F., Shen, Z., Luo, R., Zhao, L., Xiao, S., Fu, Z. F., and Peng, G. (2016) Structural basis for the dimerization and substrate recognition specificity of porcine epidemic diarrhea virus 3C-like protease. *Virology* **494**, 225–235
54. Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., and Hilgenfeld, R. (2003) Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* **300**, 1763–1767
55. Ali, A., Aydin, C., Gildemeister, R., Romano, K. P., Cao, H., Özen, A., Soumana, D., Newton, A., Petropoulos, C. J., Huang, W., and Schiffer, C. A. (2013) Evaluating the role of macrocycles in the susceptibility of hepatitis C virus NS3/4A protease inhibitors to drug resistance. *ACS Chem. Biol.* **8**, 1469–1478
56. Romano, K. P., Ali, A., Royer, W. E., and Schiffer, C. A. (2010) Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding. *Proc. Natl. Acad. Sci. USA* **107**, 20986–20991
57. Romano, K. P., Laine, J. M., Deveau, L. M., Cao, H., Massi, F., and Schiffer, C. A. (2011) Molecular mechanisms of viral and host cell substrate recognition by hepatitis C virus NS3/4A protease. *J. Virol.* **85**, 6106–6116
58. Kozakov, D., Grove, L. E., Hall, D. R., Bohnuud, T., Mottarella, S. E., Luo, L., Xia, B., Beglov, D., and Vajda, S. (2015) The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.* **10**, 733–755
59. Brown, E. G., Liu, H., Kit, L. C., Baird, S., and Nesrallah, M. (2001) Pattern of mutation in the genome of influenza A virus on adaptation to increased virulence in the mouse lung: identification of functional themes. *Proc. Natl. Acad. Sci. USA* **98**, 6883–6888
60. Lai, M., and Holmes, K. (2001) Coronaviridae: The viruses and their replication. In *Fields Virology* (Knipe, D. M., and Howley, P. M., eds.), Lippincott, Philadelphia
61. Kim, Y., Lovell, S., Tiew, K.-C., Mandadapu, S. R., Alliston, K. R., Battaile, K. P., Groutas, W. C., and Chang, K.-O. (2012) Broad-spectrum antivirals against 3C or 3C-like proteases of picornaviruses, noroviruses, and coronaviruses. *J. Virol.* **86**, 11754–11762

Received for publication July 1, 2019.

Accepted for publication September 23, 2019.